Markov Chains

A random process X is a family $\{X_t : t \in T\}$ of random variables indexed by some set T. When $T = \{0, 1, 2, \dots\}$ one speaks about a '**discrete-time**' process, for $T = \mathbb{R}$ or $T = [0, \infty)$ one has a '**continuous-time**' process.

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and let $\{X_0, X_1, \dots\}$ be a sequence of random variables which take values in some **countable** set S, called the **state space**. We assume that each X_n is a discrete random variable which takes one of N possible values, where $N = |S|$ (N may equal $+\infty$).

Markov property

Def.5.1: The process X is a **Markov chain** if it has the **Markov property**:

$$\begin{aligned} \mathsf{P}(X_{n+1} = x_{n+1} \,|\, X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) \\ = \mathsf{P}(X_{n+1} = x_{n+1} \,|\, X_n = x_n) \end{aligned} \tag{5.1}$$

for all $n \geq 1$ and all $x_0, x_1, \ldots, x_{n+1} \in S$.

With n being the 'present' and $n+1$ a 'future' moment of time, the Markov property (5.1) says :

"**given** *the* **present** *value of a Markov chain,*
its **future** *behaviour* **does not depend** *on the* **past***".*

Remark : It is straightforward to check that the Markov property (5.1) is equivalent to the following statement:

for each $s \in S$ *and every sequence* $\{x_k : k \geq 0\}$ *in* S*,*

$$\begin{aligned} \mathrm{P}(X_{n+m} = s \,|\, X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) \\ = \mathrm{P}(X_{n+m} = s \,|\, X_n = x_n) \end{aligned}$$

for any $m, n \geq 0$*.*

The evolution of a Markov chain is described by its '**initial distribution**'

$$\mu_k^0 \stackrel{\text{def}}{=} \mathsf{P}(X_0 = k)$$

and its '**transition probabilities**'

$$\mathsf{P}(X_{n+1} = j \,|\, X_n = i)\,;$$

it can be quite complicated in general since these probabilities depend upon the **three** quantities n, i, and j.

We shall restrict our attention to the case when they **do not depend on** n but **only upon** i and j.

Def.5.2: A Markov chain X is called **homogeneous** if

$$\mathsf{P}(X_{n+1} = j \,|\, X_n = i) \equiv \mathsf{P}(X_1 = j \,|\, X_0 = i)$$

for all n, i, j. The **transition matrix** $\mathbf{P} = (p_{ij})$ is the $|S| \times |S|$ matrix of **transition probabilities**

$$p_{ij} = \mathsf{P}(X_{n+1} = j \,|\, X_n = i)\,.$$

In what follows we shall **only** consider **homogeneous** Markov chains.

The next claim **characterizes** transition matrices.

Theorem 5.3: *The transition matrix* **P** *is a* **stochastic matrix**, *which is to say that*

a) **P** *has non-negative entries,* $p_{ij} \geq 0$*;*

b) **P** *has row sums equal to one,* $\sum_j p_{ij} = 1$*.*

Bernoulli process

Example 5.4: Let $S = \{0, 1, 2, \dots\}$ and define the Markov chain Y by $Y_0 = 0$ and

$$\mathrm{P}(Y_{n+1} = s + 1 \mid Y_n = s) = p\,, \qquad \mathrm{P}(Y_{n+1} = s \mid Y_n = s) = 1 - p\,,$$

for all $n \geq 0$, where $0 < p < 1$.

You may think of Y_n as the number of heads thrown in n tosses of a coin.

Simple random walk

Example 5.5: Let $S = \{0, \pm 1, \pm 2, \dots\}$ and define the Markov chain X by $X_0 = 0$ and

$$p_{ij} = \begin{cases} p, & \text{if } j = i+1, \\ q = 1-p, & \text{if } j = i-1, \\ 0, & \text{otherwise.} \end{cases}$$

Ehrenfest chain

Example 5.6: Let $S = \{0, 1, \ldots, r\}$ and put

$$p_{k,k+1} = \frac{r-k}{r}\,, \qquad p_{k,k-1} = \frac{k}{r}\,, \qquad p_{ij} = 0 \quad \text{otherwise}\,.$$

In words, there is a total of r balls in two urns; k in the first and $r-k$ in the second. We pick one of the r balls at random and move it to the other urn.

Ehrenfest used this to model the division of air molecules between two chambers (of equal size and shape) which are connected by a small hole.

Birth and death chains

Example 5.7: Let $S = \{0, 1, 2, \ldots, \}$. These chains are defined by the restriction $p_{ij} = 0$ when $|i - j| > 1$ and, say,

$$p_{k,k+1} = p_k \, , \qquad p_{k,k-1} = q_k \, , \qquad p_{kk} = r_k$$

with $q_0 = 0$.

The fact that these processes cannot jump over any integers makes their analysis particularly simple.

Def.5.8: The *n*-**step transition matrix** $\mathbf{P}_n = \big(p_{ij}(n)\big)$ is the matrix of *n*-**step transition probabilities**

$$p_{ij}(n) \equiv p_{ij}^{(n)} \stackrel{\text{def}}{=} \mathsf{P}(X_{m+n} = j \,|\, X_m = i)\,.$$

Of course, $\mathbf{P}_1 = \mathbf{P}$.

Chapman-Kolmogorov equations

Theorem 5.9: *We have*

$$p_{ij}(m+n) = \sum_k p_{ik}(m)\, p_{kj}(n)\,.$$

Hence $\mathbf{P}_{m+n} = \mathbf{P}_m \mathbf{P}_n$, *and so* $\mathbf{P}_n = \mathbf{P}^n \equiv (\mathbf{P})^n$, *the n-th power of the transition matrix* $\mathbf{P}$.

Proof

Using the identity

$$\mathsf{P}(A \cap B \mid C) = \mathsf{P}(A \mid B \cap C)\,\mathsf{P}(B \mid C)$$

and the Markov property, we get

$$\begin{aligned}
p_{ij}(m+n) &= \mathsf{P}(X_{m+n} = j \mid X_0 = i) \\
&= \sum_k \mathsf{P}(X_{m+n} = j, X_m = k \mid X_0 = i) \\
&= \sum_k \mathsf{P}(X_{m+n} = j \mid X_m = k, X_0 = i)\,\mathsf{P}(X_m = k \mid X_0 = i) \\
&= \sum_k \mathsf{P}(X_{m+n} = j \mid X_m = k)\,\mathsf{P}(X_m = k \mid X_0 = i) \\
&= \sum_k p_{kj}(n)\,p_{ik}(m) = \sum_k p_{ik}(m)\,p_{kj}(n)\,.
\end{aligned}$$

Let $\mu_i^{(n)} \stackrel{\text{def}}{=} \mathsf{P}(X_n = i)$, $i \in S$, be the mass function of X_n; we write $\boldsymbol{\mu}^{(n)}$ for the row vector with entries $(\mu_i^{(n)} : i \in S)$.

Lemma 5.10: *We have*

$$\boldsymbol{\mu}^{(m+n)} = \boldsymbol{\mu}^{(m)} \mathbf{P}_n \,,$$

and hence

$$\boldsymbol{\mu}^{(n)} = \boldsymbol{\mu}^{(0)} \mathbf{P}^n \,.$$

Example 5.11: Consider the three-state Markov chain with the transition matrix

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 2/3 & 1/3 \\ 1/16 & 15/16 & 0 \end{pmatrix} .$$

Find a general formula for $p_{11}^{(n)}$.

General method

To find a formula for $p_{ij}^{(n)}$ for any M-state chain and any states i, j:

a) Compute the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_M$ of $\mathbf{P}$ by solving the characteristic equation;

b1) If the eigenvalues are **distinct** then $p_{ij}^{(n)}$ has the form

$$p_{ij}^{(n)} = a_1 \big(\lambda_1\big)^n + \cdots + a_M \big(\lambda_M\big)^n$$

for some constants $a_1, \dots, a_M$ (**depending on** i **and** j).

b2) If an eigenvalue λ is **repeated** (once, say) then the general form includes the term $(a_1 + a_2 n)\lambda^n$.

b3) As roots of a polynomial with real coefficients, complex eigenvalues will come in **conjugate pairs** and these are best written using $\cos\varphi$ and $\sin\varphi$, ie., for the eigenvalues $\lambda_{1,2} = r\, e^{\pm i\varphi}$ use $r^n\big(a_1 \cos(n\varphi) + a_2 \sin(n\varphi)\big)$.

Class structure

Def.5.12: We say that state i *leads to* state j and write $i \to j$ if

$$\mathsf{P}_i(X_n = j \text{ for some } n \geq 0) \equiv \mathsf{P}\big(X_n = j \text{ for some } n \geq 0 \,|\, X_0 = i\big) > 0\,.$$

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

Theorem 5.13: *For distinct states i and j the following are equivalent:*

a) $i \to j$;

b) $p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n} > 0$ *for some states* $i_0 \equiv i$, i_1, i_2, ..., i_{n-1}, $i_n \equiv j$;

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

Remark : It is clear from b) that $i \to j$ and $j \to k$ imply $i \to k$. Also, $i \to i$ for any state i.

So the communication relation $\leftrightarrow$ satisfies the conditions for an **equivalence relation** on S and thus **partitions** S into **communicating classes**.

Def.5.14: We say that a class C is *closed* if

$$i \in C\,, \quad i \to j \qquad \Longrightarrow \qquad j \in C\,.$$

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

Exercise 5.15: Show that every transition matrix on a **finite** state space has at least one **closed** communicating class.

Find an example of a transition matrix with no closed communicating classes.

Def.5.16: A Markov chain or its transition matrix P is called **irreducible** if its state space S forms a single communicating class.

Example 5.17: Find the communicating classes associated to the stochastic matrix

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

Def.5.18: The **period** $d(i)$ of a state i is defined by

$$d(i) = \text{gcd}\{n > 0 : p_{ii}^{(n)} > 0\},$$

the greatest common divisor of the epochs at which return is possible (ie., $p_{ii}^{(n)} = 0$ unless n is a multiple of $d(i)$).

We call i **periodic** if $d(i) > 1$ and **aperiodic** if $d(i) = 1$.

Lemma 5.19: *If states i and j are communicating, then i and j have the same period.*

Example 5.20: It is easy to see that both the simple random walk (Example 5.6) and the Ehrenfest chain (Example 5.7) have period 2.

On the other hand, the birth and death process (Example 5.8) with all $p_k \equiv p_{k,k+1} > 0$, all $q_k \equiv p_{k,k-1} > 0$ and at least one $r_k \equiv p_{kk}$ positive is aperiodic (however, if all r_{kk} vanish, the birth and death chain has period 2).

Hitting times and absorption probabilities

Example 5.21: A man is saving up to buy a new car at a cost of N units of money. He starts with k $(0 < k < N)$ units and tries to win the remainder by the following gamble with his bank manager. He tosses a coin repeatedly; if the coin comes up **heads** then the manager pays him one unit, but if it comes up **tails** then he pays the manager one unit. The man plays this game repeatedly until one of two events occurs: either he runs out of money and is bankrupt or he wins enough to buy the car.

What is the probability that he is ultimately bankrupt?

Def.5.22: Let $(X_n)_{n\geq 0}$ be a Markov chain with transition matrix $\mathbf{P}$. The **hitting time** of a subset $A \subset S$ is the random variable $H^A : \Omega \to \{0, 1, 2, \dots\} \cup \{\infty\}$ given by

$$H^A(\omega) \stackrel{\text{def}}{=} \inf\Big\{ n \geq 0 : X_n(\omega) \in A \Big\} \tag{5.3}$$

where we agree that the infimum over the empty set $\varnothing$ is ∞.

The probability starting from i that $(X_n)_{n\geq 0}$ ever hits A is

$$h_i^A \stackrel{\text{def}}{=} \mathsf{P}_i(H^A < \infty) \equiv \mathsf{P}\big(H^A < \infty \,|\, X_0 = i\big)\,.$$

When A is a closed class, h_i^A is called the **absorption probability**.

Example 5.23: Consider the chain on $\{1,2,3,4\}$ with the following transition matrix:

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

Starting from 2, what is the probability of absorption in 4?

Theorem 5.24: *Fix $A \subset S$. The vector of hitting probabilities $h^A \equiv (h_i^A : i \in S)$ solves the following system of linear equations:*

$$\begin{cases} h_i^A = 1\,, & \text{for } i \in A, \\ h_i^A = \sum_{j\in S} p_{ij}\, h_j^A\,, & \text{for } i \in A^c. \end{cases} \tag{5.5}$$

One can show that $h^A = (h_i^A : i \in S)$ is the **smallest non-negative** solution to (5.5),

$$\begin{cases} h_i^A = 1\,, & \text{for } i \in A, \\ h_i^A = \sum_{j \in S} p_{ij}\, h_j^A\,, & \text{for } i \in A^c. \end{cases}$$

in that if $x = (x_i : i \in S)$ is another solution to (5.5) with $x_i \geq 0$ for **all** $i \in S$, then $x_i \geq h_i^A$ for **all** $i \in S$.

This property is especially useful if the state space S is **infinite**.

Gambler's ruin

Example 5.25: Imagine the you enter a casino with a fortune of £i and gamble, £1 at a time, with probability p of doubling your stake and probability q of losing it. The resources of the casino are regarded as infinite, so there is no upper limit to your fortune. What is the probability that you leave broke?

In other words, consider the Markov chain on $\{0, 1, 2, \dots\}$ with the transition probabilities

$$p_{00} = 1\,, \qquad p_{k,k+1} = p\,, \qquad p_{k,k-1} = q \qquad (k \geq 1)$$

where $0 < p = 1 - q < 1$. Find $h_i = \mathrm{P}_i(H^{\{0\}} < \infty)$.

It is often useful to know the **expected time** before absorption,

$$k_i^A \stackrel{\text{def}}{=} \mathsf{E}_i\big(H^A\big) \equiv \mathsf{E}\big(H^A \,|\, X_0 = i\big) = \sum_{n<\infty} n\,\mathsf{P}_i(H^A = n) + \infty \cdot \mathsf{P}_i(H^A = \infty)\,. \tag{5.6}$$

Example 5.23 [cont'd]: Assuming that $X_0 = 2$, find the mean time until the chain is absorbed in states 1 or 4.

Theorem 5.26: *Fix $A \subset S$. The vector of* **mean hitting times** $k^A \equiv (k_i^A, i \in S)$ *is the* **minimal non-negative** *solution to the following system of linear equations:*

$$\begin{cases} k_i^A = 0\,, & \text{for } i \in A, \\ k_i^A = 1 + \sum_{j \in S} p_{ij}\, k_j^A\,, & \text{for } i \in A^c. \end{cases}$$

Recurrence and transience

Let X_n, $n \geq 0$, be a Markov chain with a discrete state space S.

Def.5.27: State i is called **recurrent** if, starting from i the chain eventually returns to i with probability 1, ie.,

$$\mathsf{P}\big(X_n = i \textit{ for some } n \geq 1 \mid X_0 = i\big) = 1\,.$$

State i is called **transient** if this probability is smaller than 1.

If $j \in S$, the **first passage time** to state j for X_n is

$$T_j = \inf\Big\{ n \geq 1 : X_n = j \Big\} . \tag{5.8}$$

Consider

$$f_{ij}^{(n)} \stackrel{\text{def}}{=} \mathsf{P}\big(X_1 \neq j, X_2 \neq j, \ldots, X_{n-1} \neq j, X_n = j \,|\, X_0 = i \big) \,,$$

the probability of the event "the first visit to state j, starting from i, takes place at nth step", ie., $\mathsf{P}_i\big(T_j = n\big)$. Then

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)} \equiv \mathsf{P}_i\big(T_j < \infty\big)$$

is the probability that the **chain ever visits j, starting from i**.

Of course, state j is recurrent iff

$$f_{jj} = \sum_{n=1}^{\infty} f_{jj}^{(n)} = \mathsf{P}_j\big(T_j < \infty\big) = 1 \,.$$

If $f_{jj} = 1$, we have $\mathsf{P}_j(X_n \text{ returns to state } j \text{ at least once}) = 1$.

By induction,

$$\mathsf{P}_j(X_n \text{ returns to state } j \text{ at least } m \text{ times}) = 1,$$

and therefore

$$\mathsf{P}_j(X_n \text{ returns to state } j \text{ infinitely many times}) = 1.$$

If $f_{jj} < 1$, the number of returns R_j to state j is geometrically distributed with parameter $1 - f_{jj} > 0$, and thus **with probability one** R_j is finite (and has finite expectation).

In other words, for **every state** j

$$\mathsf{P}_j(X_n \text{ returns to state } j \text{ infinitely many times}) \in \{0, 1\}.$$

Remark 5.27.1: Clearly, for $i \neq j$ we obviously have

$$f_{ij} = \mathsf{P}_i\big(T_j < \infty\big) = \mathsf{P}_i(H^{\{j\}} < \infty) = h_i^{\{j\}},$$

the probability that starting from i the chain ever hits j.

Remark 5.27.2: A recurrent state j is **positive recurrent** if

$$\mathsf{E}_j\big(T_j\big) \equiv \mathsf{E}\big(T_j \mid X_0 = j\big) = \sum_{n=1}^{\infty} n\, f_{jj}^{(n)} + \infty \cdot \mathsf{P}_j\big(T_j = \infty\big) < \infty\,.$$

Otherwise j is **null recurrent**.

Lemma 5.28: *Let i, j be two states. Then for all $n \geq 1$,*

$$p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)}\, p_{jj}^{(n-k)}\,. \tag{5.10}$$

Eqn. (5.10) is often called the **first passage** decomposition.

Let $\mathcal{P}_{ij}(s)$ and $\mathcal{F}_{ij}(s)$ be the generating functions of the sequences $p_{ij}^{(n)}$ and $f_{ij}^{(n)}$ respectively,

$$\mathcal{P}_{ij}(s) \stackrel{\text{def}}{=} \sum_{n} p_{ij}^{(n)}\, s^n\,, \qquad \mathcal{F}_{ij}(s) \stackrel{\text{def}}{=} \sum_{n} f_{ij}^{(n)}\, s^n\,.$$

Then (5.10) reads

$$\mathcal{P}_{ij}(s) = \delta_{ij} + \mathcal{F}_{ij}(s)\,\mathcal{P}_{jj}(s)\,, \tag{5.11}$$

where δ_{ij} is the Kronecker delta-function.

Corollary 5.29: *The following dichotomy holds:*

a) *if* $\sum_n p_{jj}^{(n)} = \infty$, *then the state* j *is* **recurrent**;
in this case $\sum_n p_{ij}^{(n)} = \infty$ *for* **all** i **such that** $f_{ij} > 0$.

b) *if* $\sum_n p_{jj}^{(n)} < \infty$, *then the state* j *is* **transient**;
in this case $\sum_n p_{ij}^{(n)} < \infty$ *for* **all** i.

Example 5.30: If $j \in S$ is **transient**, then $\mathsf{P}\big(X_n = j \text{ i.o.}\big) = 0$, ie., *with probability one* there are only finitely many visits to state j.

Corollary 5.31: *If* $j \in S$ *is* **transient**, *then* $p_{ij}^{(n)} \to 0$ *as* $n \to \infty$ *for all* $i \in S$.

Example 5.32: Determine recurrent and transient states for the Markov chain on $\{1, 2, 3\}$ with the following transition matrix:

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

Recurrence and transience are class properties:

Lemma 5.33: *Let C be a communicating class. Then* **either all** *states of C are transient* **or all** *are recurrent.*

Lemma 5.34: Every recurrent *class is* **closed**.

Remark 5.34.1: Every finite closed class is **recurrent**.

Strong Markov property

The *usual* Markov property states:

> *Let* $(X_n)_{n\geq 0}$ *be a Markov chain with state space* S*.*
> *Given* $X_n = x \in S$*, the Markov chain* $(X_m)_{m\geq 0}$ *has the same distribution as* $(X_m)_{m\geq 0}$ *started at* $X_0 = x$*.*

In other words,

> *Markov chain* $(X_m)_{m\geq n}$ *starts afresh from the state* $X_n = x$ *at* **deteministic** *time* n*.*

It is often desirable to extend the validity of this property from

☞ **deterministic** times n to **random** times T.

Stopping times

✍ A random time T is called a *stopping time* for the Markov chain $(X_n)_{n\geq 0}$, if for any $n \geq 0$ the event $\{T \leq n\}$ is only determined by $(X_k)_{k\leq n}$ (and thus *does not depend on the future* evolution of the chain).

Typical examples of stopping times include the hitting times H^A from (5.3) and the first passage times T_j from (5.8).
Notice that the example $T^* = T_j - 1$ above, predetermines the first
☞ *future jump* of the Markov chain and thus is not a *stopping time*.

Strong Markov property

Lemma 5.35: *Let T be a stopping time for a Markov chain $(X_n)_{n\geq 0}$ with state space S. Then, given $\{T < \infty\}$ and $X_T = i \in S$, the process $(Y_n)_{n\geq 0}$ defined via $Y_n = X_{T+n}$ has the same distribution as $(X_n)_{n\geq 0}$ started from $X_0 = i$.*

In other words,

> *Markov chain $(X_m)_{m\geq T}$ starts afresh from the state $X_T = i$ at* **random** *time T.*

Long time properties

Consider a Markov chain on $S = \{1, 2\}$ with transition matrix

$$\mathbf{P} = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}, \quad 0 < a < 1, \quad 0 < b < 1,$$

and initial distribution $\mu^{(0)} = (\mu_1^{(0)}, \mu_2^{(0)})$. Since

$$\mathbf{P}^n = \frac{1}{a+b}\begin{pmatrix} b & a \\ b & a \end{pmatrix} + \frac{(1-a-b)^n}{a+b}\begin{pmatrix} a & -a \\ -b & b \end{pmatrix}, \qquad n \geq 0,$$

the distribution $\mu^{(n)}$ of X_n satisfies, as $n \to \infty$,

$$\mu^{(n)} = \mu^{(0)}\,\mathbf{P}^n \to \boldsymbol{\pi}, \qquad \text{where} \qquad \boldsymbol{\pi} \stackrel{\text{def}}{=} \Big(\frac{b}{a+b}, \frac{a}{a+b}\Big).$$

We see that $\boldsymbol{\pi}\,\mathbf{P} = \boldsymbol{\pi}$, ie., the distribution $\boldsymbol{\pi}$ is 'invariant' for $\mathbf{P}$.

Stationary distributions

Def.5.36: A vector $\boldsymbol{\pi} = (\pi_j : j \in S)$ is a **stationary distribution** of a Markov chain on S with the transition matrix $\mathbf{P}$, if:

a) $\boldsymbol{\pi}$ is a distribution, ie., $\pi_j \geq 0$ for all $j \in S$, and $\sum_j \pi_j = 1$;

b) $\boldsymbol{\pi}$ is stationary, ie., $\boldsymbol{\pi} = \boldsymbol{\pi}\,\mathbf{P}$, which is to say that $\pi_j = \sum_i \pi_i p_{ij}$ for all $j \in S$.

Remark 5.36.1: Property b) implies that $\boldsymbol{\pi}\,\mathbf{P}^n = \boldsymbol{\pi}$ for all $n \geq 0$, that is if X_0 has distribution $\boldsymbol{\pi}$ then X_n has distribution $\boldsymbol{\pi}$ for **all** $n \geq 0$, showing that the distribution of X_n is **stationary in time**.

Example 5.37: Find a stationary distribution for the Markov chain with the transition matrix

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

Lemma 5.38: *Let S be finite and $i \in S$ be fixed. If for all $j \in S$,*

$$p_{ij}^{(n)} \to \pi_j \qquad \text{as } n \to \infty,$$

then the vector $\boldsymbol{\pi} = (\pi_j : j \in S)$ is a stationary distribution.

Theorem 5.39 [Convergence to equilibrium]: *Let $(X_n)_{n \geq 0}$, be an irreducible aperiodic Markov chain on a finite state space S with transition matrix $\mathbf{P}$. Then there exists a* **unique** *probability distribution $\boldsymbol{\pi}$ such that for* **all** *$i, j \in S$*

$$\lim_{n \to \infty} p_{ij}^{(n)} = \pi_j \, .$$

In particular, $\boldsymbol{\pi}$ is **stationary** *for $\mathbf{P}$, and for every initial distribution*

$$\mathsf{P}(X_n = j) \to \pi_j \qquad \text{as } n \to \infty.$$

Example 5.40: For a *periodic* Markov chain on $S = \{1, 2\}$ with transition probabilities $p_{12} = p_{21} = 1$ and $p_{11} = p_{22} = 0$ we have

$$\mathbf{P}^{2k+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \mathbf{P}^{2k} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

for all $k \geq 0$, so that there is **no convergence**.

Example 5.41: For a *reducible* Markov chain on $S = \{1, 2, 3\}$ with transition matrix (with $a > 0$, $b > 0$, $c > 0$ s.t. $a + b + c = 1$)

$$\mathbf{P} = \begin{pmatrix} a & b & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

every stationary distribution $\boldsymbol{\pi}$ solves the equations

$$\pi_1 = a\,\pi_1 \qquad \pi_2 = b\,\pi_1 + \pi_2\,, \qquad \pi_3 = c\,\pi_1 + \pi_3\,,$$

so every $\mu_\rho \stackrel{\text{def}}{=} (0, \rho, 1 - \rho)$ with $0 \leq \rho \leq 1$ is stationary for $\mathbf{P}$.

Random Walk on a cycle

(cf. MC-14,15)

A flea hops randomly on vertices of a regular k-gon, hopping to the neighbouring vertex on the right with probability p and to the neighbouring vertex on the left with probability $1-p$.

- Describe the probability distribution of the flea position after n jumps.
- Find the corresponding stationary distribution.

▶R

Let T_j be the first passage time of a Markov chain X_n,

$$T_j \stackrel{\text{def}}{=} \min\{n \geq 1 : X_n = j\}.$$

Lemma 5.42: *If the transition matrix* $\mathbf{P}$ *of a Markov chain* X_n *is* γ*-positive,*

$$\min_{j,k} p_{jk} \geq \gamma > 0,$$

then there exist positive constants C *and* α *such that for all* $n \geq 0$

$$\mathsf{P}(T_j > n) \leq C \exp\{-\alpha n\}.$$

In other words, the **exponential** moments $\mathsf{E}\big(e^{cT_j}\big)$ exist for all $c > 0$ small enough, in particular, **all polynomial** moments $\mathsf{E}\big((T_j)^p\big)$ with $p > 0$ are finite.

Lemma 5.43: *Let X_n be an irreducible recurrent Markov chain with stationary distribution $\boldsymbol{\pi}$. Then for every state j, the expected return time $\mathsf{E}_j(T_j)$ satisfies the identity*

$$\pi_j \, \mathsf{E}_j(T_j) = 1 \, .$$

The argument applies even to the **countable state space** and shows that the stationary distribution is **unique**, **provided** it exists.

Let $V_k(n)$ denote the number of visits to k before time n,

$$V_k(n) = \sum_{l=1}^{n-1} \mathbb{1}_{\{X_l = k\}}\,.$$

Theorem 5.44 [Ergodic theorem]: *If $(X_n)_{n\geq 0}$ is an irreducible Markov chain, then*

$$\mathsf{P}\Big(\frac{V_k(n)}{n} \to \frac{1}{\mathsf{E}_k(T_k)} \text{ as } n \to \infty\Big) = 1\,,$$

where $\mathsf{E}_k(T_k)$ is the expected return time to state k.

Moreover, if $\mathsf{E}_k(T_k) < \infty$, then for any **bounded** *fn $f : S \to \mathbb{R}$*

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

where $\bar{f} \stackrel{\text{def}}{=} \sum_{k\in S} \pi_k\, f(k)$ and $\boldsymbol{\pi} = (\pi_k)_{k\in S}$ is the unique stationary distribution.

Example [2008, Q.3]

Like a good boy, Harry visits the dentist every six months. Because of his sweet tooth, the condition of his teeth varies according to a Markov chain on the states $\{0, 1, 2, 3\}$, where 0 means no work is required, 1 means a cleaning is required, 2 means a filling is required and 3 means root canal work is needed. Charges for each visit to the dentist depend on the work done. State 0 has a charge of £10, state 1 has a charge of £20, state 2 has a charge of £30 and state 3 has the disastrous charge of £150. Transitions from state to state are governed by the matrix

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

What is the percentage of visits that are disastrous?

What is Harry's long run cost rate for maintaining his teeth?

In your answer you should give a clear statement of any result you use.

Detailed balance equations

Def.5.49: A stochastic matrix **P** and a measure λ are said to be in **detailed balance** if

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

Lemma 5.50: *If* **P** *and* λ *are in detailed balance, then* λ *is invariant for* **P***, ie.,* $\lambda\mathbf{P} = \lambda$.

Reversibility

Def.5.51: Let $(X_n)_{n\geq 0}$ be a Markov chain with state space S and transition matrix $\mathbf{P}$. A probability measure $\boldsymbol{\pi}$ on S is said to be *reversible* for the chain (or for the matrix $\mathbf{P}$) if $\boldsymbol{\pi}$ and $\mathbf{P}$ are in detailed balance, ie.,

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

A Markov chain is said to be *reversible* if it has a reversible distribution.

Example 5.52: A Markov chain with the transition matrix

$$\begin{pmatrix} 0 & p & q \\ q & 0 & p \\ p & q & 0 \end{pmatrix}, \qquad 0 < p = 1 - q < 1\,,$$

has stationary distribution $\boldsymbol{\pi} = (1/3, 1/3, 1/3)$. For the latter to be reversible for this Markov chain we need

$$\pi_1\, p_{12} = \pi_2\, p_{21}\,, \qquad \text{i.e.} \qquad p = q\,.$$

If $p = q = \frac{1}{2}$, then DBE hold for all pairs of states, ie., the chain is reversible. Otherwise, the chain is not reversible.

Time-reversal

Exercise 5.53: Let $\mathbf{P}$ be irreducible and have an invariant distribution π. Suppose that $(X_n)_{0\le n\le N}$ is a Markov chain with transition matrix $\mathbf{P}$ and the initial distribution π, and set $Y_n \stackrel{\text{def}}{=} X_{N-n}$. Show that $(Y_n)_{0\le n\le N}$ is a Markov chain with initial distribution π and the transition matrix $\widehat{\mathbf{P}} = (\hat{p}_{ij})$ given by

$$\pi_j \hat{p}_{ji} = \pi_i p_{ij} \qquad \text{for all } i,\ j.$$

The chain $(Y_n)_{0\le n\le N}$ is called the *time reversal* of $(X_n)_{0\le n\le N}$.

Clearly, a Markov chain is *reversible* if its distribution coincides with that of its time reversal.

Random walk on a graph

A graph G is a countable collection of states, usually called vertices, some of which are joined by edges. The valency v_j of vertex j is the number of edges at j, and we assume that every vertex in G has finite valency. The random walk on G is a Markov chain with transition probabilities

$$p_{jk} = \begin{cases} 1/v_j\,, & \text{if } (j,k) \text{ is an edge} \\ 0\,, & \text{otherwise.} \end{cases}$$

We assume that G is connected, so that $\mathbf{P}$ is irreducible. It is easy to see that $\mathbf{P}$ is in detailed balance with $\boldsymbol{v} = (v_j : j \in G)$. As a result, if the total valency $V = \sum_j v_j$ is finite, then $\boldsymbol{\pi} \stackrel{\text{def}}{=} \boldsymbol{v}/V$ is a stationary distribution and $\mathbf{P}$ is reversible.

Problem MC-57

A random walker on the standard chessboard makes each permissible move with equal probability. If it starts in a corner, how long on average will it take to return, if:

a) only horizontal and vertical moves are allowed (ie, in the middle of the chessboard there are four permissible moves)?

b) the diagonal moves are also allowed (ie, in the middle of the chessboard there are eight permissible moves)?

Markov chain Monte Carlo (MCMC)

Example Let $G = (V, E)$ be a graph with $V = \{v_1, \dots, v_k\}$ and $E = \{e_1, \dots, e_\ell\}$. A set $A \subset V$ is an *independent set*, if no two vertices in A are adjacent in G. The following Markov chain generates a random set in $S_G \stackrel{\text{def}}{=} \{ \textit{ all independent sets in } G \}$:

Let $A \in S_G$ be given (eg., $A = \varnothing$).

1. Pick $v \in V$ uniformly at random;
2. Flip a fair coin;
3. **If** 'heads' and no neighbour of v is in A, add v to A:
$$A \mapsto A \cup \{v\};$$
otherwise, remove v from A:
$$A \mapsto A \setminus \{v\}.$$

Exercise: Check that this chain is irreducible, aperiodic, and has the correct stationary distribution.
Estimate $|S_G|$, if G is a 10×10 subset in $\mathbb{Z}^2$.

MCMC-2: random q-colourings

Let $G = (V, E)$ be a graph and $q \geq 2$ be an integer. A *q-colouring* of G is an assignment

$$V \ni v \mapsto \xi_v \in S \stackrel{\text{def}}{=} \{1, 2, \ldots, q\}$$

such that *if* v_1 *and* v_2 *are adjacent, then* $\xi_{v_1} \neq \xi_{v_2}$. The following Markov chain in S^V generates a random q-colouring in G:

Let a colouring $C \in S^V$ be given.

1. Pick $v \in V$ uniformly at random;
2. Re-colour v in an admissible colour (ie., not used by any of the neighbours of v) uniformly at random.

Exercise: Check that this chain is irreducible, aperiodic, and has the correct stationary distribution (provided q is large enough!).

Question:

How would you generate a general distribution π on a finite set S?

$\rightsquigarrow$ Consruct a Markov chain on S having π as its only stationary distribution!

$\rightsquigarrow$ Gibbs sampler.

Markov chains

By the end of this section you should be able to:

- define a Markov chain, verify whether a given process is a Markov chain;
- compute the n-step transition probabilities $p_{ij}^{(n)}$ for a given Markov chain;
- identify classes of communicating states, explain which classes are closed and which are not; identify absorbing states;
- check whether a given Markov chain is irreducible;
- determine the period of every state for a given Markov chain;
- compute absorption probabilities and expected hitting times;
- define transient and (positive/null) recurrent states;
- compute the first passage probabilities f_{ij} and use them to classify states into transient and recurrent;
- use the transition probabilities $p_{ij}^{(n)}$ to classify states into transient and recurrent;
- find all stationary distributions for a given Markov chain;
- state and apply the convergence to equilibrium theorem for Markov chains;
- state and apply the Ergodic theorem;
- state and use the relation between the stationary distribution and the mean return times;
- define reversible measures and state their main properties;
- check whether a given Markov chain is reversible;
- find stationary distributions for random walks on graphs.