Markov chains and applications

M. Simon

M1 Data Science 2019-2020, Université de Lille

Contents

Bibliography

[1] S. Asmussen. *Applied probability and queues*, volume 51 of Applications of Mathematics. Springer-Verlag, New York, 2nd edition, 2003. ISBN 0-387-00211-1. Stochastic Modelling and Applied Probability.

[2] P. Brémaud. *Markov chains: Gibbs Fields, Monte Carlo Simulations and Queues*, volume 31 of Texts in Applied Mathematics. Springer-Verlag, New York, 1999. ISBN 978-0-387-98509-1. Probability Theory and Stochastic Processes.

[3] J.-F. Delmas and B. Jourdain. *Modèles Aléatoires*, volume 57 of Mathématiques et Applications. Springer-Verlag, Berlin Heidelberg, 2006. ISBN 978-3-540-33282-4. Probability Theory and Stochastic Processes.

Chapter 1

Markov chains in discrete time

Let $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and let* S *be a finite or countable set. A random variable with possible values in a state space* S *is a measurable map* $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (S, \mathcal{P}(S))$ *where* $\mathcal{P}(S)$ *is the family of all subsets of* S.

1.1 Definition

Definition 1.1 (Discrete-time stochastic process)**.** *A* stochastic process in discrete time *is a family* $(X(n))_{n\in\mathbb{N}_0}$ *of random variables indexed by the numbers* $\mathbb{N}_0 = \{0, 1, 2, \dots\}$*. The set* S *of possibles values of* $X(n)$ *is the* state space *of the process. In this course* S *will always be finite or countable.*

The distribution of a discrete-time stochastic process with countable state space S *is characterized by the* point probabilities

$$\mathbb{P}\big(X(n) = i_n, \dots, X(0) = i_0\big), \qquad i_n, i_{n-1}, \dots, i_0 \in S.$$

Example 1.2 (Queues). For example, in a queue, $X(n)$ represents the time (in number of minutes) that the n–th customer waits after arrival before receiving service.

Definition 1.3 (Markov chain)**.** *A* discrete-time Markov chain on a countable state space S *is a stochastic process satisfying the* Markov property

$$\mathbb{P}\big(X(n) = i_n \mid X(n-1) = i_{n-1}, \dots, X(0) = i_0\big) = \mathbb{P}\big(X(n) = i_n \mid X(n-1) = i_{n-1}\big),$$

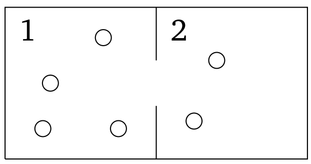
for any $i_n, \dots, i_0 \in S$ *and any* $n \in \mathbb{N}_0$.

Example 1.4 (Gambling banker). You take part in a roulette game and you start with a capital of m euros. At each round you gamble 1€ :

- if the roulette gives an even number, you lose it;
- if the roulette gives an odd number, you double it (*i.e.* you receive 2€).

Let $X(n)$ be your fortune (in euros) after the n–th round. This defines a Markov process. An interesting question is to know the probability that you will leave the casino broke.

EXAMPLE 1.5 (Ehrenfest model). We start with N particles in a closed box, divided into two compartments which are in contact with each other. Particles may move between compartments: at each time, one particle is chosen uniformly at random and moved from its current compartment to the other compartment. Let $X(n)$ be the number of particles in compartment 1 at step n. This stochastic process is a Markov chain.



$X(n) = 4$

@home: Compute the probabilities

$$\mathbb{P}\big[X(n+1) = j \mid X(n) = k\big]$$

for any $j, k \in \mathbb{N}_0$.

EXAMPLE 1.6 (GOOGLE's algorithm to rank pages). Let N be the number of webpages. For any $i \in \{1, \ldots, N\}$, let L_i be the number of links from page i.

A random surfer goes from one webpage to another, using the following algorithm: if the surfer is on page i,

- with probability $q \in (0,1)$, pick one page from $\{1, \ldots, N\}$ uniformly at random (page i is not interesting any more)
- with probability $1 - q$, pick one of the links among the L_i links uniformly at random.

If $X(n)$ denotes the page visited by the surfer at time n, then $(X(n))_{n\in\mathbb{N}}$ is a Markov chain. The long time behavior of this chain gives the final rank of each page, used by Google!

PROPOSITION 1.7. *Let* $(X(n))_{n\in\mathbb{N}_0}$ *be a Markov chain and let us introduce the notation*

$$p_{i,j}(k) = \mathbb{P}\big(X(k+1) = j \mid X(k) = i\big), \qquad \text{for any } k \in \mathbb{N}_0.$$

Then the following formula holds: for any $i_0, \dots, i_n \in S$,

$$\begin{aligned}&\mathbb{P}\big(X(n) = i_n, X(n-1) = i_{n-1}, \dots, X(0) = i_0\big)\\ &\qquad = p_{i_{n-1},i_n}(n-1) \times p_{i_{n-2},i_{n-1}}(n-2) \times \cdots \times p_{i_0,i_1}(0) \times \mathbb{P}\big(X(0) = i_0\big).\end{aligned}$$

Proof. The proof goes by induction. □

DEFINITION 1.8. *We say that the Markov chain* $(X(n))_{n\in\mathbb{N}_0}$ *is* time homogeneous *if the* transition probabilities $p_{i,j}(k) =: p_{i,j}$ *do not depend on the time index* $k \in \mathbb{N}$.

> TO SUM UP:
> A *time-homogeneous Markov chain on a finite or countable state space* S is a family of random variables $(X(n))_{n\in\mathbb{N}_0}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that
>
> $$\mathbb{P}\big(X(n) = j \mid X(n-1) = i, X(n-1) = i_{n-2}, \dots, X(0) = i_0\big) = p_{i,j}$$
>
> for any $i, j, i_0, \dots, i_{n-2} \in S$ and $n \in \mathbb{N}_0$. The distribution of the Markov chain is uniquely determined by the *initial distribution*
>
> $$\Phi(i) = \mathbb{P}\big(X(0) = i\big)$$
>
> and the *transition probabilities*
>
> $$p_{i,j} = \mathbb{P}\big(X(n+1) = j \mid X(n) = i\big)$$

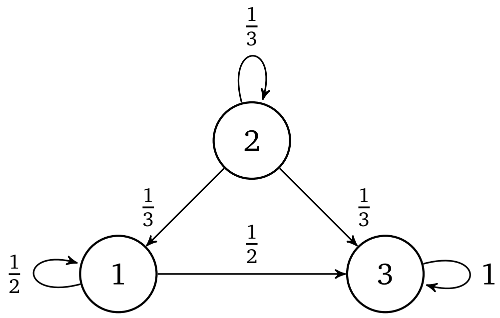
If the state space S if *finite*, the initial distribution $\overline{\Phi} = (\Phi(i))_{i\in S}$ is a *probability vector* (the sum of its components equals 1), and the transition probabilities $P = (p_{i,j})_{i,j\in S}$ define a square *stochastic matrix* which satisfies:

$$p_{i,j} \in [0,1] \qquad \text{and} \qquad \sum_{j\in S} p_{i,j} = 1, \qquad \text{for any } i \in S.$$

An alternative representation of the transition probability matrix P is given by the *transition diagram* with nodes representing the individual states of the chain and directed edges labeled by the probability of possible transitions.

EXAMPLE 1.9. *@home: Write the transition matrix and diagram of first two example (the gambling banker and the Ehrenfest model).*

Write the transition matrix associated with the following transition diagram:



THEOREM 1.10 (n-step transition probabilities)**.** *Let $(X(n))_{n\in\mathbb{N}_0}$ be a Markov chain on a finite state space $S = \{1,\ldots,N\}$ with transition probability matrix $P \in \mathfrak{M}_{N,N}(\mathbb{R})$ and initial distribution $\overline{\Phi} = (\Phi(1),\ldots,\Phi(N))$ (row vector).*

Then, the distribution of $X(n)$ is given by

$$\mathbb{P}\big(X(n) = j\big) = \big(\overline{\Phi}P^n\big)_j = \sum_{i=1}^{N} \Phi(i)(P^n)_{i,j}, \tag{1.1}$$

i.e. the j–th component of the (row) vector $\overline{\Phi}P^n$. Moreover, for any $k \in \mathbb{N}$,

$$\mathbb{P}\big(X(n+k) = j \mid X(k) = i\big) = (P^n)_{i,j}. \tag{1.2}$$

*The matrix P^n is called the n–*step transition matrix.

Proof. The proof goes by induction. For $n = 1$ we have

$$\begin{aligned}\mathbb{P}\big(X(1) = j\big) &= \sum_{i\in S} \mathbb{P}\big(X(0) = i, X(1) = j\big)\\ &= \sum_{i\in S} \mathbb{P}\big(X(0) = i\big)\mathbb{P}\big(X(1) = j \mid X(0) = i\big)\\ &= \sum_{i\in S} \Phi(i)P_{i,j}.\end{aligned}$$

We get for $n + 1$ that

$$\begin{aligned}\mathbb{P}\big(X(n+1) = j\big) &= \sum_{i\in S} \mathbb{P}\big(X(n) = i, X(n+1) = j\big)\\ &= \sum_{i\in S} \mathbb{P}\big(X(n) = i\big)\mathbb{P}\big(X(n+1) = j \mid X(n) = i\big)\\ &= \sum_{i\in S} \mathbb{P}\big(X(n) = i\big)P_{i,j} = \sum_{i\in S} \big(\Phi P^n\big)_i P_{i,j} = \big(\Phi P^{n+1}\big)_j.\end{aligned}$$

since we assume (1.1) at rank n. The second identity (1.2) is proved in the exact same way, by induction on n. □

1.2 Classification of states

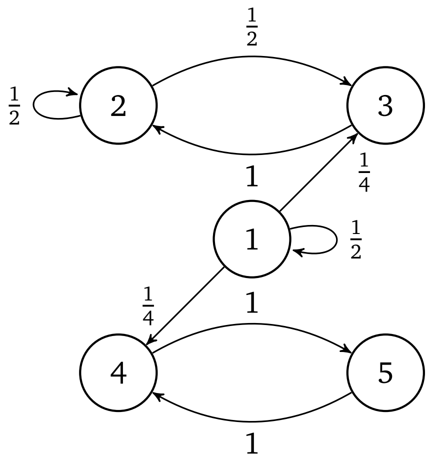
DEFINITION 1.11. *For a discrete-time Markov chain with state space* S *and transition probabilities* $P = (p_{i,j})_{i,j \in S}$ *we say that:*

1. *there is a* possible path *from state* i *to state* j *if there exists* $m \in \mathbb{N}_0$ *and a sequence of states*
$$i = i_0 \to i_1 \to \cdots \to i_m = j$$
such that the transitions satisfy $p_{i_{\ell-1},i_\ell} > 0$ *for any* $\ell \in \{1, \ldots, m\}$. *By convention, there always exists a possible path from* i *to* i, *taking* $m = 0$.
Equivalently[1], *there exists* $m \in \mathbb{N}_0$ *such that* $(P^m)_{i,j} > 0$.

2. *two states* $i, j \in S$ communicate *if there is a possible path from* i *to* j *and from* j *to* i. *In that case we write* $i \leftrightarrow j$. *The relation* $\leftrightarrow$ *is an equivalence relation, which partitions the state space* S *into disjoint* communication classes *;*

3. *the Markov chain is* irreducible *if there is only one communication class.*

EXAMPLE 1.12. We consider the following transition diagram:



[1] *Proof*: note that
$$(P^m)_{i,j} = \sum_{i_1,\ldots,i_{m-1} \in S} p_{i,i_1} p_{i_1,i_2} \cdots p_{i_{m-1},j}.$$

This Markov chain has 3 communication classes: $\{1\}, \{2,3\}$ and $\{4,5\}$.

Definition 1.13. *We say that a communication class* C *is closed if, for any* $i \in C$, *we have*

$$\sum_{j \in C} p_{i,j} = 1.$$

If C *is finite, then* C *is closed if the submatrix of transition probabilities restricted to* C *has all row sums equal to 1.*

Example 1.14. In Example 1.12, $\{1\}$ is not closed, but $\{2,3\}$ and $\{4,5\}$ are closed.

Remark 1.15. The restriction of a Markov chain to a closed communication class is an irreducible Markov chain.

Definition 1.16. *For any state* $i \in S$ *we define the* hitting time *of* i *by*

$$T_i = \inf\{n > 0 \,;\, X(n) = i\} \in \mathbb{N} \cup \{\infty\}.$$

If two states i and j communicate, we know that

$$\mathbb{P}\big(T_i < +\infty \mid X(0) = j\big) > 0, \qquad \text{and} \qquad \mathbb{P}\big(T_j < +\infty \mid X(0) = i\big) > 0.$$

In other words: it is possible (with positive probability) to get from i to j and from j to i. A much more relevant question is whether these probabilities are equal to 1. This leads us to the following definition:

Definition 1.17 (Recurrence and transience)**.** *For a discrete-time Markov chain on* S *we say that a state* $i \in S$ *is* recurrent *if and only if*

$$\mathbb{P}\big(T_i < +\infty \mid X(0) = i\big) = 1.$$

If $\mathbb{P}\big(T_i < +\infty \mid X(0) = i\big) < 1$ *then* i *is said to be a* transient *state.*

Remark 1.18. Note that the probability $\mathbb{P}\big(T_i < +\infty \mid X(0) = i\big)$ is also equal to

$$\sum_{n=1}^{\infty} \mathbb{P}\big(T_i = n \mid X(0) = i\big).$$

Example 1.19. Going back to Example 1.12 we have

1. $\mathbb{P}\big(T_1 < +\infty \mid X(0) = 1\big) = \frac{1}{2}$ therefore 1 is transient.
2. $\mathbb{P}\big(T_4 < +\infty \mid X(0) = 4\big) = \mathbb{P}\big(T_4 = 2 \mid X(0) = 4\big) = 1$ and 4 is recurrent.

   The same holds for 5, which is also recurrent.

3. $\mathbb{P}\big(T_2 = 1 \mid X(0) = 2\big) = \frac{1}{2}$ and $\mathbb{P}\big(T_2 = 2 \mid X(0) = 2\big) = \frac{1}{2}$ and then 2 is recurrent.

4. Finally, state 3 is recurrent, but this requires more work and will be seen in exercise.

We now give one useful criterion.

THEOREM 1.20 (Recurrence criterion). *For a discrete-time Markov chain with transition probability matrix* P*, the state i is recurrent if and only if*

$$\sum_{n=1}^{\infty} (P^n)_{i,i} = +\infty.$$

Proof. Admitted. See for instance [1, Chapter 1]. □

COROLLARY 1.21. *All states in a communication class are either all recurrent or all transient.*

Therefore, a communication class C *is either* recurrent *(if all the states in* C *are recurrent) or* transient *(if all the states in* C *are transient).*

Proof. Assume that states i and j communicate. Then there exists $\ell, m \in \mathbb{N}$ such that $(P^\ell)_{i,j} > 0$ and $(P^m)_{j,i} > 0$. By the Markov property, the quantity

$$(P^\ell)_{i,j}(P^k)_{j,j}(P^m)_{j,i}$$

describes the probability of a loop of length $\ell + k + m$ from state i to state i, which visits j after ℓ steps and after $\ell + k$ steps. This is smaller than or equal to the probability of having a loop of length $\ell + k + m$ from state i to state i, which equals $(P^{\ell+k+m})_{i,i}$. Therefore, we can write the inequality

$$\sum_{n=1}^{\infty} (P^n)_{i,i} \geqslant \sum_{k=1}^{\infty} (P^{\ell+k+m})_{i,i} \geqslant \sum_{k=1}^{\infty} (P^\ell)_{i,j}(P^k)_{j,j}(P^m)_{j,i} = (P^\ell)_{i,j}\left(\sum_{k=1}^{\infty} (P^k)_{j,j}\right)(P^m)_{j,i}.$$

Either i and j are transient, or we assume for instance that j is recurrent, which by Theorem 1.20 implies $\sum_{n=1}^{\infty} (P^n)_{j,j} = +\infty$. From the inequality above this shows that $\sum_{n=1}^{\infty} (P^n)_{i,i} = +\infty$ and i is also recurrent. □

COROLLARY 1.22. *A finite communication class is recurrent if and only if it is closed.*

Proof. Assume that C is a finite closed communication class. Then, for any $i \in \mathrm{C}$, and $k \in \mathbb{N}$,

$$\sum_{j \in \mathrm{C}} (\mathrm{P}^k)_{i,j} = 1.$$

By interchanging the order of summation (by positivity) we get

$$\sum_{j \in \mathrm{C}} \sum_{k=1}^{\infty} (\mathrm{P}^k)_{i,j} = \sum_{k=1}^{\infty} \sum_{j \in \mathrm{C}} (\mathrm{P}^k)_{i,j} = +\infty.$$

Since C is finite, there exists $j \in \mathrm{C}$ such that

$$\sum_{k=1}^{\infty} (\mathrm{P}^k)_{i,j} = +\infty.$$

Since i and j communicate, there exists $m \in \mathbb{N}$ such that $(\mathrm{P}^m)_{i,j} > 0$. We conclude:

$$\sum_{n=1}^{\infty} (\mathrm{P}^n)_{i,i} \geqslant \sum_{k=1}^{\infty} (\mathrm{P}^{k+m})_{i,i} \geqslant \sum_{k=1}^{\infty} (\mathrm{P}^k)_{i,j} (\mathrm{P}^m)_{j,i} = \left(\sum_{k=1}^{\infty} (\mathrm{P}^k)_{i,j} \right) (\mathrm{P}^m)_{j,i} = +\infty.$$

Assume now that C is not closed. Then, there exists $i \in \mathrm{C}$ such that

$$\sum_{j \in \mathrm{C}} p_{i,j} < 1.$$

Let $k \notin \mathrm{C}$ be any state of the Markov chain with $p_{i,k} > 0$. Then we have

$$\mathbb{P}\big(\mathrm{T}_i = \infty \mid \mathrm{X}(0) = i\big) \geqslant \mathbb{P}\big(\mathrm{X}(1) = k \mid \mathrm{X}(0) = i\big).$$

This implies

$$\mathbb{P}\big(\mathrm{T}_i < \infty \mid \mathrm{X}(0) = i\big) \leqslant 1 - p_{i,k} < 1.$$

Therefore i is transient. From Corollary 1.21, this means that C is transient. □

BE CAREFUL:

- For an *irreducible* Markov chain there is a path with positive probability between any two states $i \neq j$.
- An irreducible Markov chain is said to be *recurrent* (resp. *transient*) if its unique communication class is recurrent (resp. *transient*). For a *recurrent* Markov chain there will eventually be a transition between any two states $i \neq j$ with probability one[a].

[a] *Proof of the last statement*: Assume that $\mathbb{P}(T_i = +\infty \mid X(0) = i) = 0$ and that i communicates with $j \neq i$. Then, there exists $m \in \mathbb{N}$ and $i_1, \dots, i_{m-1} \in S \backslash \{j\}$ such that

$$\mathbb{P}(X(m) = j, X(m-1) = i_{m-1}, \dots, X(1) = i_1 | X(0) = i) > 0. \tag{1.3}$$

By contradiction, assume that $\mathbb{P}(T_i = +\infty \mid X(0) = j) > 0$. Then we have

$$\begin{aligned} \mathbb{P}(T_i = +\infty \mid X(0) = i) &\geqslant \mathbb{P}(T_i = \infty, X(m) = j, X(m-1) = i_{m-1}, \dots, X(1) = i_1 \mid X(0) = i) \\ &= \mathbb{P}(T_i = \infty \mid X(m) = j, X(m-1) = i_{m-1}, \dots, X(1) = i_1, X(0) = i) \\ &\quad \times \mathbb{P}(X(m) = j, X(m-1) = i_{m-1}, \dots, X(1) = i_1 \mid X(0) = i). \end{aligned} \tag{1.4}$$

One then needs to use the *strong Markov property* (admitted here), which tells that

$$\mathbb{P}(T_i = +\infty \mid X(m) = j, X(m-1) = i_{m-1}, \dots, X(1) = i_1, X(0) = i) = \mathbb{P}(T_i = +\infty \mid X(0) = j),$$

and by assumption this last quantitiy is > 0. Using (1.3) and (1.4), one can see the contradiction.

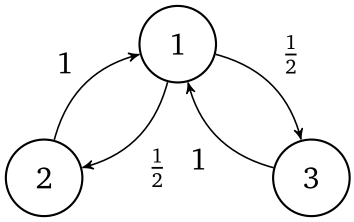
1.3 Limit results and invariant probabilities

The aim of this section is to study the long time behavior of a Markov chain. More precisely we want to study $\mathbb{P}(X(n) = j)$ when $n \to \infty$, which boils down to understanding the behaviour of the n–step transition probabilities $(P^n)_{i,j}$, due to the identity

$$\mathbb{P}(X(n) = j) = \sum_{i \in S} \Phi(i)(P^n)_{i,j}.$$

We first need to introduce the *period* of the chain. Let us start with an example:

Starting from state 1 at time $n = 0$, the chain will be in state 1 at any even time $n = 2p$ with probability 1. At odd times, the chain will belong to the set $\{2, 3\}$ with probability 1. In other words, the chain passes alternately from $\{1\}$ to $\{2, 3\}$, and the state space $S = \{1, 2, 3\}$ can be decomposed as $S = E_0 \cup E_1$ with the following property: if the chain starts from $i \in E_0$ then in one step it can only go to a state $j \in E_1$, and the converse is true. In this sense, the chain



has a periodic behavior.

The general definition is the following:

THEOREM 1.23. *For any irreducible Markov chain, one can find a unique partition of* S *into* d *sets* $\mathrm{E}_0, \mathrm{E}_1, \dots, \mathrm{E}_{d-1}$ *such that, for all* $k \in \{0, 1, \dots, d-1\}$ *and for any* $i \in \mathrm{E}_k$,

$$\sum_{j \in \mathrm{E}_{k+1}} p_{i,j} = 1,$$

where by convention $\mathrm{E}_d = \mathrm{E}_0$, *and where* d *is maximal (that is, there is no other such partition* $\mathrm{E}_0, \mathrm{E}_1, \dots, \mathrm{E}_{d'-1}$ *with* $d' > d$.

The number d *is called the* period *of the chain. The chain therefore moves from one class to the other at each transition, and this cyclically.*

We say that the chain is aperiodic *if* $d = 1$.

Proof. Admitted. See [2, Theorem 4.1]. □

REMARK 1.24. In other words, the transition matrix can we written with blocks as follows:

$$\mathrm{P} = \begin{array}{c} \\ \mathrm{E}_0 \\ \mathrm{E}_1 \\ \mathrm{E}_2 \\ \vdots \\ \mathrm{E}_{d-1} \end{array} \begin{array}{c} \begin{array}{ccccc} \mathrm{E}_0 & \mathrm{E}_1 & \mathrm{E}_2 & \dots & \mathrm{E}_{d-1} \end{array} \\ \begin{pmatrix} 0 & \mathbf{A}_0 & 0 & \dots & 0 \\ 0 & \ddots & \mathbf{A}_1 & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & & 0 & 0 & \mathbf{A}_{d-2} \\ \mathbf{A}_{d-1} & 0 & \cdots & \cdots & 0 \end{pmatrix} \end{array}$$

and it can be checked that P^d is block-diagonal: this means that the cyclic classes $\mathrm{E}_0, \mathrm{E}_1, \dots, \mathrm{E}_{d-1}$ are exactly the communication classes of P^d.

We now have the first result concerning the long run behavior of Markov chains:

THEOREM 1.25. *Let* $(\mathrm{X}(n))_{n \in \mathbb{N}}$ *be an irreducible, recurrent, and aperiodic Markov chain.*

Then, for any state i and any initial distribution, it holds

$$\lim_{n\to\infty} \mathbb{P}\big(X(n) = i\big) = \frac{1}{\mathbb{E}\big[T_i \mid X(0) = i\big]}, \qquad \textit{where } T_i = \inf\big\{n > 0\,;\, X(n) = i\big\}.$$

Proof. Admitted. See [1, Chapter 1.2]. □

REMARK 1.26. If state i is *recurrent*, which means $\mathbb{P}(T_i = +\infty \mid X(0) = i) = 0$ then the expectation in Theorem 1.25 can be computed as

$$\mathbb{E}\big[T_i \mid X(0) = i\big] = \sum_{n=1}^{\infty} n\mathbb{P}\big(T_i = n \mid X(0) = i\big).$$

Note that the expectation may or may not be finite.

DEFINITION 1.27. *A recurrent state i is said to be* positive recurrent *if and only if the mean return time to state i is finite :*

$$\mathbb{E}\big[T_i \mid X(0) = i\big] < +\infty.$$

Otherwise the recurrent state is said to be null recurrent. *It can be shown that all states belonging to the same recurrent class are either positive recurrent or null recurrent.*

Therefore, from Theorem 1.25, an irreducible, *positive recurrent*, aperiodic Markov chain has a nontrivial limit. However, its formulation is not very useful since we are rarely able to compute $\mathbb{E}\big[T_i \mid X(0) = i\big]$. Fortunately, there is another way to characterize the limit. Let us first give a formal explanation: we have the following identity

$$\begin{aligned}
&\mathbb{P}\big(X(n+1) = j \mid X(0) = i\big) \\
&\qquad = \sum_{k\in S} \mathbb{P}\big(X(n+1) = j, X(n) = k \mid X(0) = i\big) \\
&\qquad = \sum_{k\in S} \mathbb{P}\big(X(n+1) = j \mid X(n) = k, X(0) = i\big)\mathbb{P}\big(X(n) = k \mid X(0) = i\big) \\
&\qquad = \sum_{k\in S} \mathbb{P}\big(X(n+1) = j \mid X(n) = k\big)\mathbb{P}\big(X(n) = k \mid X(0) = i\big) \\
&\qquad = \sum_{k\in S} p_{k,j}\, \mathbb{P}\big(X(n) = k \mid X(0) = i\big).
\end{aligned}$$

Assuming that the limits $\lim_{n\to\infty} \mathbb{P}\big(X(n) = k \mid X(0) = i\big)$ exist and do not depend on i, and assuming that we can interchange the summation, we obtain that

$$\pi(j) = \lim_{n\to\infty} \mathbb{P}\big(X(n) = k \mid X(0) = i\big)$$

solves the system of equations

$$\pi(j) = \sum_{k \in S} \pi(k) p_{k,j}. \tag{1.5}$$

Definition 1.28. *A non-negative vector* $\overline{\pi} = (\pi(j))_{j \in S}$ *solving the system of equations* (1.5) *is called an* invariant measure *for the transition probabilities* $P = (p_{i,j})$. *If* S *is finite, then* (1.5) *is equivalent to* $\overline{\pi} = \overline{\pi} P$.

If $\overline{\pi}$ *is a probability (i.e.* $\sum_{j \in S} \pi(j) = 1$*) then* $\overline{\pi}$ *is called an* invariant probability distribution.

Theorem 1.29. *Let* $(X(n))_{n \in \mathbb{N}}$ *be an irreducible, recurrent, Markov chain.*

There exists a unique (up to multiplication) invariant measure $\overline{\pi}$ *solving* (1.5).

The solution can be normalized into a unique invariant probability if and only if the Markov chain is positive recurrent, and in that case

$$\overline{\pi}(i) = \frac{1}{\mathbb{E}\big[T_i \mid X(0) = i\big]}.$$

Sketch of the proof. Show, using the Markov property, that the measure defined by

$$\nu(j) = \mathbb{E}\left[\sum_{n=0}^{T_i - 1} \mathbf{1}_{\{X(n) = j\}} \,\middle|\, X(0) = i\right]$$

solves (1.5), and that its total mass is

$$\sum_{j \in S} \nu(j) = \mathbb{E}\big[T_i \mid X(0) = i\big],$$

therefore ν can be normalized into a probability measure if and only if one has $\mathbb{E}\big[T_i \mid X(0) = i\big] < \infty$ which means that the Markov chain is positive recurrent. It remains to show uniqueness, which is done for instance in [2, Chapter 3]. □

The final (and main) result concerning the limiting behavior for discrete Markov chains is the following:

Theorem 1.30. *Let* $(X(n))_{n \in \mathbb{N}}$ *be an irreducible, positive recurrent, aperiodic, Markov chain.*

Then, for any state i *and any initial distribution, it holds*

$$\lim_{n \to \infty} \mathbb{P}\big(X(n) = i\big) = \pi(i) = \frac{1}{\mathbb{E}\big[T_i \mid X(0) = i\big]},$$

where $\overline{\pi} = (\pi(j))_{j\in S}$ *is the unique invariant probability vector solving the system of equations*

$$\pi(j) = \sum_{i\in S} \pi(i) p_{i,j}.$$

Finally, we can also say something about the limit for null-recurrent states and transient states:

THEOREM 1.31. *Let* $(X(n))_{n\in\mathbb{N}}$ *be an irreducible, aperiodic, Markov chain.*

For a null-recurrent, or transient, state j, it holds

$$\lim_{n\to\infty} \mathbb{P}\big(X(n) = j\big) = 0,$$

for any choice of initial distribution.

Proof. Admitted. See [1, Chapter 1]. □

TO SUM UP:

- If the chain is **irreducible**, **recurrent** and **aperiodic**, then the limits

$$\lim_{n\to\infty} \mathbb{P}\big(X(n) = j\big)$$

  exist, are independent of the initial distribution, and belong to $[0, 1]$.

- If the chain is **irreducible** and **recurrent**, then
  - there exists a unique invariant measure $\overline{\pi}$ solving $\overline{\pi} = \overline{\pi}P$, up to multiplication ;
  - this invariant measure $\overline{\pi}$ can be normalized to 1 (which is equivalent to $\sum \pi(j) < \infty$) **if and only if** the chain is positive recurrent.
    1. In **that case**, if the chain is **moreover aperiodic**, then

$$\lim_{n\to\infty} \mathbb{P}\big(X(n) = j\big) = \overline{\pi}(j)$$

    2. If it is **not the case**, then the chain is **null recurrent**. If the chain is **moreover aperiodic**, then

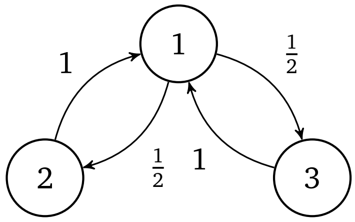
$$\lim_{n\to\infty} \mathbb{P}\big(X(n) = j\big) = 0, \qquad \text{for any } j.$$

- If the chain is **irreducible**, **aperiodic** and **transient**, then

$$\lim_{n\to\infty} \mathbb{P}\big(X(n) = j\big) = 0, \qquad \text{for any } j.$$

1. An **irreducible positive-recurrent** Markov chain which is **not aperiodic**:

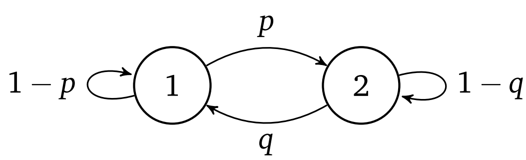
   ⊳ the Markov chain associated with the diagram:

   Its unique probability distribution is $\overline{\pi} = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.

2. An **irreducible positive-recurrent aperiodic** Markov chain:

   ⊳ any Markov chain on a finite state space, which is irreducible and aperiodic (for which there exists i such that $p_{i,i} > 0$, for instance), like

3. An **irreducible null-recurrent** Markov chain which is **not aperiodic**:

   ⊳ the simple symmetric random walk on $\mathbb{Z}$, which jumps to $k+1$ with probability $\frac{1}{2}$ and $k-1$ with probability $\frac{1}{2}$.

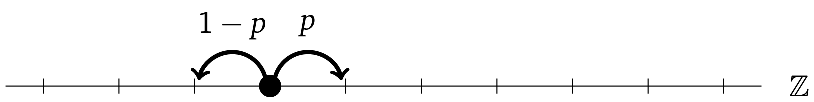
   

4. An **irreducible null-recurrent aperiodic** Markov chain:

   ⊳ let $X(n) = S(2n)$ where S is the simple symmetric random walk on $\mathbb{Z}$ as above, then $(X(n))$ is aperiodic on $2k\mathbb{Z}$.

5. An **irreducible transient** Markov chain which is **not aperiodic**:

   ⊳ the simple asymmetric random walk on $\mathbb{Z}$, which jumps to $k+1$ with probability $p \neq \frac{1}{2}$ and $k-1$ with probability $1-p$.

6. An **irreducible transient aperiodic** Markov chain:

   ⊳ let $X(n) = S(2n)$ where S is the simple asymmetric randow walk on Z as above, then $(X(n))$ is aperiodic on $2k\mathbb{Z}$.

1.4 Absorption probabilities

Recall that recurrent classes are closed: once the Markov chain enters a recurrent class then it stays there forever. Transient classes may or may not be closed, but in either case we know that $\lim_{n\to\infty} \mathbb{P}\big(X(n) = j\big) = 0$ for any transient state j. This naturally raises the following questions: if a Markov chain with multiple communication classes is started in a transient state i, how many times will it visit state i before it leaves the state forever? And what is the probability that it will be *absorbed* by a closed (recurrent) class?

When the state space S is *finite*, these quantities can be evaluated by a technique called *first-step analysis*, which is the motor of most computations in Markov chain theory, and is best illustrated by examples.

1.4.1 Probabilities

EXAMPLE 1.32 (Gambler's ruin, part I). Two players A and B play *heads or tails*, where heads occur with probability $p \in (0,1)$. The successive outcomes form an i.i.d. sequence indexed by $n \in \mathbb{N}$.

We call $X(n)$ the fortune in euros of player A at time n. Then

$$X(n+1) = X(n) + Z(n+1),$$

where

$$Z(n+1) = \begin{cases} +1 & \text{if the result of the toss is heads} \\ -1 & \text{if the result of the toss is tails} \end{cases}$$

In other words, A bets 1€ on heads at each toss, and B bets 1€ on tails. The respective initial fortunes of A and B are a and b. The gain ends when a player is ruined. The process $(X(n))_{n\in\mathbb{N}_0}$ is a random walk on the state space $S = \{0, 1, \ldots, a, a+1, \ldots, a+b\}$. The duration of the game is denoted by T, it is the hitting time of $\{0, a+b\}$, namely

$$T = \inf\big\{n > 0 \,;\, X(n) = 0 \quad \text{or} \quad X(n) = a+b\big\}.$$

We denote the probability of winning for A by

$$u(a) = \mathbb{P}\big(X(T) = a+b \mid X(0) = a\big).$$

@home: Draw an example of trajectory.

Instead of computing $u(a)$ alone, the *first-step analysis* consists in computing

$$u(i) = \mathbb{P}\big(X(T) = a+b \mid X(0) = i\big), \qquad \text{for any } i \in \{0, 1, \ldots, a+b\}.$$

We obtain a *recurrence relation* as follows:

- first, note that $u(a+b) = 1$ and $u(0) = 0$;
- if $X(0) = i \in \{1, \dots, a+b-1\}$ then
  - with probability p, $X(1) = i+1$ and the probability of winning for A with initial fortune $i+1$ is $u(i+1)$
  - with probability $1-p$, $X(1) = i-1$ and the probability of winning for A with initial fortune $i-1$ is $u(i-1)$.

  Therefore, for any $i \in \{1, \dots, a+b-1\}$

$$u(i) = pu(i+1) + (1-p)u(i-1). \tag{1.6}$$

The solution of this linear recurrence equation is

$$u(i) = \begin{cases} \dfrac{1 - \left(\frac{1-p}{p}\right)^i}{1 - \left(\frac{1-p}{p}\right)^{a+b}} & \text{if } p \neq \frac{1}{2}, \\ \dfrac{i}{a+b} & \text{if } p = \frac{1}{2}. \end{cases}$$

Proof of (1.6). A rigorous proof can be given: let $(Y(n))_{n\in\mathbb{N}_0}$ denote the Markov chain obtained by shifting $(X(n))$ by one time unit, namely:

$$Y(n) = X(n+1), \qquad n \in \mathbb{N}_0.$$

If $X(0) \in \{1, \dots, a+b-1\}$ then the events "X is absorbed at 0" and "Y is absorbed at 0" are equal. Therefore,

$$\begin{aligned} &\mathbb{P}\big(X \text{ is absorbed at } 0, X(1) = i \pm 1, X(0) = i\big) \\ &\quad = \mathbb{P}\big(Y \text{ is absorbed at } 0, X(1) = i \pm 1, X(0) = i\big) \\ &\quad = \mathbb{P}\big(Y \text{ is absorbed at } 0 \mid Y(0) = i \pm 1, X(0) = i\big)\mathbb{P}\big(X(1) = i \pm 1, X(0) = i\big) \\ &\quad = \mathbb{P}\big(Y \text{ is absorbed at } 0 \mid Y(0) = i \pm 1\big)\mathbb{P}\big(X(1) = i \pm 1, X(0) = i\big), \end{aligned}$$

since $X(0)$ and $(Y(n))_{n\in\mathbb{N}_0}$ are independent given $Y(0)$. The two chains have the same transition matrix, and therefore when they have the same initial state they have the same distribution. Hence,

$$\mathbb{P}\big(Y \text{ is absorbed at } 0 \mid Y(0) = i \pm 1\big) = u(i \pm 1).$$

The rest of the proof is straightforward. □

1.4.2 Average times

The same analysis can be used to compute average times before absorption.

EXAMPLE 1.33 (Gambler's ruin, part II). We continue Example 1.32. We denote the *average duration* of the game by

$$m(i) = \mathbb{E}\big[\mathrm{T} \mid \mathrm{X}(0) = i\big].$$

We obtain in the same way a recurrence relation as follows:

- first, note that $m(0) = 0$ and $m(a+b) = 0$;
- if $\mathrm{X}(0) = i \in \{1, \ldots, a+b-1\}$, then the game will last at least one unit time. Then,
  - with probability p, the fortune of A will be $i+1$ and therefore $m(i+1)$ more tosses will be needed in average to end the game,
  - with probability $1-p$, the fortune of A will be $i-1$ and therefore $m(i+1)$ more tosses will be needed in average to end the game.

  Therefore, for any $i \in \{1, \ldots, a+b-1\}$,

$$m(i) = 1 + pm(i+1) + (1-p)m(i-1).$$

The solution of this linear recurrence equation, for $p = \frac{1}{2}$, is

$$m(i) = i(a+b-i).$$

1.4.3 General case: countable state space

For Markov chains on a countable state space, absorption probabilities may be found by solving a countably infinite system of equations:

THEOREM 1.34. *For a Markov chain on* S*, let* C *be a* recurrent class*, and let* T *be the set of all transient states. Then, the probabilities*

$$u(j) = \mathbb{P}\big(\mathrm{X} \textit{ is absorbed in } \mathrm{C} \mid \mathrm{X}(0) = j\big), \qquad j \in \mathrm{T}$$

(that the chain will ever visit class C *and stay there forever) solve the system of equations*

$$u(j) = \sum_{k \in \mathrm{T}} p_{j,k}\, u(k) + \sum_{k \in \mathrm{C}} p_{j,k}. \tag{1.7}$$

The absorption probabilities $(u(j))_{j \in \mathrm{T}}$ is the smallest non-negative solution to (1.7)*. There is a unique bounded solution to* (1.7) *if and only if there is zero probability that the Markov chain stays in the transient states forever.*

Chapter 2

Poisson processes and Queues

A random *point process* is, roughly speaking, a countable random set of points of the real line. In most applications, these real points are *times of occurence* of some event, for instance: the arrival times of customers at the desk of a post office (remember the first example of these lecture notes), or times of birth of an biologic organism, etc.

REMINDER:

1. The *exponential distribution* is a probability distribution on $\mathbb{R}_+$, parametrized by some $\lambda > 0$, and whose density function is

$$f_\lambda(x) = \lambda e^{-\lambda x} \mathbf{1}_{x \geqslant 0}.$$

   The parameter λ is often called the *rate parameter*. If a random variable X has this distribution, we write $X \sim \text{Exp}(\lambda)$. One can easily check that its expectation and variance are given by

$$\mathbb{E}[X] = \frac{1}{\lambda}, \qquad \text{Var}[X] = \frac{1}{\lambda^2}.$$

   The most important property of an exponential random variable X is the following *memorylessness* property: for any $s, t \geqslant 0$,

$$\mathbb{P}\big(X > s + t \mid X > s\big) = \mathbb{P}\big(X > t\big).$$

   @home: Prove it, using the cumulative distribution function.

2. The *Poisson distribution* is a probability distribution on $\mathbb{N}_0$, parametrized by some $\lambda > 0$, whose probability mass function is given by

$$f_\lambda(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \qquad k = 0, 1, 2, \ldots$$

One can easily check that its expectation and variance are given by

$$\mathbb{E}[X] = \mathrm{Var}[X] = \lambda.$$

2.1 Poisson processes

DEFINITION 2.1 (Random point process). *A* random point process *on* $[0,+\infty)$ *is a sequence* $(T_n)_{n\in\mathbb{N}_0}$ *of nonnegative random variables such that*

1. $T_0 \equiv 0$
2. $0 < T_1 < T_2 < \cdots$
3. $\lim_{n\to\infty} T_n = +\infty$ *a.s.*

The sequence $\{S_n = T_n - T_{n-1}\}_{n\geqslant 0}$ is the sequence of *inter-arrival times*.

DEFINITION 2.2 (Homogeneous Poisson point process). *Let* $(S_n)_{n\in\mathbb{N}_0}$ *be a sequence of i.i.d. exponential random variables:*

$$S_n \sim \mathrm{Exp}(\lambda)$$

for some parameter $\lambda \in (0,+\infty)$. *Let*

$$T_0 = 0, \qquad T_n = \sum_{k=1}^{n} S_k, \quad n \geqslant 1.$$

Then the process defined for any $t \geqslant 0$ *by*

$$N(t) = \sum_{n=1}^{\infty} \mathbf{1}_{\{T_n\in(0,t]\}} = \#\{n \geqslant 1 \,;\, T_n \leqslant t\}$$

is a random point process which is called homogeneous Poisson process with intensity $\lambda > 0$.

REMARK 2.3. Note that $N(0) = 0$. Since the interval $(0,t]$ is closed on the right, the trajectories $t \mapsto N(t,\omega)$ are right-continuous for almost every $\omega \in \Omega$. The trajectories are nondecreasing, have limits on the left at every time t, and jump one unit upwards at each random event T_k.

REMARK 2.4. By induction, one can compute the *probability density function* of T_n for any n: it is given by

$$f_{T_n}(t) = \frac{\lambda^n}{(n-1)!} y^{n-1} e^{-\lambda y}.$$

@home: Prove it.

PROPOSITION 2.5. *A Poisson process of intensity $\lambda > 0$ has the following properties:*

1. *For any $t \geqslant 0$, the random variable $N(t)$ is distributed according to a Poisson distribution with mean λt.*
2. *The process has* independent increments, *namely: for any sequence of times $0 = t_0 \leqslant t_1 \leqslant \cdots \leqslant t_n$, the variables $N(t_j) - N(t_{j-1}), j \in \{1, \ldots, n\}$ are independent.*
3. *The process has* stationary increments, *namely: for any $s \leqslant t$, the random variables $N(t+s) - N(t)$ and $N(s)$ have the same distribution.*

Proof. To prove the first claim, it is enough to compute the *generating function* $\mathbb{E}[q^{N(t)}]$ for any $q \in (0,1)$ (since it characterizes the distribution). We get

$$\mathbb{E}\big[q^{N(t)}\big] = \sum_{n=0}^{\infty} q^n \mathbb{P}\big(N(t) = n\big) = \sum_{n=0}^{\infty} q^n \mathbb{P}\big(T_n \leqslant t, T_{n+1} > t\big).$$

Note that

$$\begin{aligned} \mathbb{P}\big(T_n \leqslant t\big) &= \mathbb{P}\big(T_n \leqslant t, T_{n+1} \leqslant t\big) + \mathbb{P}\big(T_n \leqslant t, T_{n+1} > t\big) \\ &= \mathbb{P}\big(T_{n+1} \leqslant t\big) + \mathbb{P}\big(T_n \leqslant t, T_{n+1} > t\big). \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{E}\big[q^{N(t)}\big] &= \sum_{n=0}^{\infty} q^n \Big(\mathbb{P}\big(T_n \leqslant t\big) - \mathbb{P}\big(T_{n+1} \leqslant t\big)\Big) \\ &= 1 - \sum_{j=1}^{\infty} q^{j-1}(1-q)\mathbb{P}\big(T_j \leqslant t\big) \\ &= 1 - \sum_{j=1}^{\infty} q^{j-1}(1-q) \int_0^t \frac{\lambda^j}{(j-1)!} y^{j-1} e^{-\lambda y} dy \\ &= 1 - \int_0^t (1-q)\lambda e^{-\lambda y + \lambda q y} dy = e^{-\lambda t(1-q)}, \end{aligned}$$

which is the probability generating function of the Poisson distribution with parameter λt.

The claims *2.* and *3.* are consequences of the *Markov property* for continuous-time processes, which we are going to define now. We will end the proof in the following section (see also [3, Chapter 8]). □

2.2 Markov processes in continuous time

2.2.1 Minimal construction

Definition 2.6 (Homogeneous Markov process)**.** *A* continuous-time Markov process *on a finite or countable set* S *is a family of random variables* $(X(t))_{t\geqslant 0}$ *such that the following* Markov property *is satisfied:*

for any $j, i, i_{n-1}, \dots, i_0$*, for any* $0 \leqslant t_0 < t_1 < \dots < t_n < t_{n+1}$*,*

$$\begin{aligned}\mathbb{P}\big(X(t_{n+1}) = j \mid X(t_n) = i, X(t_{n-1}) = i_{n-1}, \dots, X(t_0) = i_0\big) \\ = \mathbb{P}\big(X(t_{n+1}) = j \mid X(t_n) = i\big) \\ = \mathbb{P}\big(X(t_{n+1} - t_n) = j \mid X(0) = i\big) \\ =: p_{i,j}(t_{n+1} - t_n). \end{aligned} \tag{2.1}$$

The distribution of the Markov process is determined by the initial distribution

$$\Phi(i) = \mathbb{P}\big(X(0) = i\big)$$

and the transition probabilities

$$p_{i,j}(t) = \mathbb{P}\big(X(t+s) = j \mid X(s) = i\big)$$

through the identity

$$\begin{aligned}\mathbb{P}\big(X(t_{n+1}) = j, X(t_n) = i, X(t_{n-1}) = i_{n-1}, \dots, X(t_0) = i_0\big) \\ = p_{i,j}(t_{n+1} - t_n) p_{i_{n-1},i}(t_n - t_{n-1}) \cdots p_{i_0,i_1}(t_1 - t_0)\Phi(i_0).\end{aligned}$$

Note that this definition does not say what conditions on the transition probabilities $(p_{i,j}(t))_{i,j\in S, t\geqslant 0}$ must satisfy in order to define a Markov process. This is one of the main problems with this construction. We now present a dynamical construction of a class of stochastic processes which satisfy the Markov property (2.1), and which we call the *minimal construction*. All Markov processes in this course will be constructed in this way.

DEFINITION 2.7 (Minimal construction). *Let* $\overline{\Phi} = (\Phi(i))_{i\in S}$ *be a probability vector, and let* $Q = (q_{i,j})_{i,j\in S}$ *be real numbers with the following properties:*

$$q_{i,j} \geqslant 0, \quad \text{for any } i \neq j,$$
$$q_{i,i} = -\sum_{j\neq i} q_{i,j}.$$

The homogeneous Markov process *with initial distribution* $\overline{\Phi}$ *and* transition intensity Q *is the stochastic process* $(X(t))_{t\geqslant 0}$ *given by the following construction:*

1. *Let* $Y(0)$ *be distributed according to* $\overline{\Phi}$*, namely:*

$$\mathbb{P}\big(Y(0) = i\big) = \Phi(i).$$

2. *Given* $Y(0)$*, choose* S_1 *according to an exponential distribution with rate parameter* $-q_{Y(0),Y(0)} \geqslant 0$*, let* $T_1 = S_1$ *and define*

$$X(t) = Y(0), \quad \text{for any } t \in [0, T_1).$$

3. *Given* $Y(0)$ *and* T_1*, chose* $Y(1)$ *such that*

$$\mathbb{P}\big(Y(1) = j \mid Y(0)\big) = \frac{q_{Y(0),j}}{-q_{Y(0),Y(0)}}, \quad j \neq Y(0).$$

4. *Recursively, given* $Y(0), Y(1), \dots, Y(n), S_1, \dots, S_n$*,*

   (a) *choose* S_{n+1} *according to an exponential distribution with parameter* $-q_{Y(n),Y(n)}$*, let* $T_{n+1} = T_n + S_{n+1}$ *and define*

$$X(t) = Y(n), \quad \text{for any } t \in [T_n, T_{n+1})$$

REMARK 2.8. If the Markov chain at time T_n jumps to a state $Y(n) = i$ such that $-q_{i,i} = 0$, then we let

$$X(t) = Y(n), \qquad \text{for any } t \geqslant T_n,$$

and we say that the Markov chain is *absorbed* at state i. In fact, the construction still makes sense if we say that an exponential distribution with rate parameter 0 is a random variable with probability mass 1 at $+\infty$.

THEOREM 2.9. *Take a continuous Markov chain with transition intensity* $Q = (q_{i,j})_{i,j\in S}$ *given by the minimal construction above. Then, it satisfies the Markov property* (2.1).

THEOREM 2.10 (Embedded Markov chain). *Take a continuous Markov chain with transition intensity* $Q = (q_{i,j})_{i,j\in S}$ *given by the minimal construction above, and assume that* $\lim_{n\to\infty} T_n = \infty$ *a.s.*

Then, the sequence $(Y(n))_{n\in\mathbb{N}_0}$ *of visited states is a* discrete-time Markov chain *with transition probabilities*

$$P_{i,j} = \begin{cases} \dfrac{-q_{i,j}}{q_{i,i}} & \text{if } j \neq i \text{ and } q_{i,i} \neq 0 \\ 0 & \text{if } j \neq i \text{ and } q_{i,i} = 0 \\ 0 & \text{if } j = i \text{ and } q_{i,i} \neq 0 \\ 1 & \text{if } j = i \text{ and } q_{i,i} = 0. \end{cases}$$

Theorems 2.9 and 2.10 are admitted in this lecture.

2.2.2 Birth-and-death processes

The *birth-and-death process* is a special case of continuous time Markov process on $\mathbb{N}_0$, where the states (for example) represent a current size of a population and the transitions are limited to birth and death. When a birth occurs, the process goes from state i to state $i+1$. Similarly, when death occurs, the process goes from state i to state $i-1$. It is assumed that the birth and death events are independent of each other. The birth-and-death process is characterized by the *birth rates* $(\beta_i)_{i\in\mathbb{N}_0}$ and *death rates* $(\delta_i)_{i\in\mathbb{N}}$ which vary according to state i of the system.

In terms of transition intensities, we have

$$\begin{aligned} q_{i,j} &= 0 && \text{if} \quad |i-j| \geqslant 2, \\ q_{i,i+1} &= \beta_i > 0, && i \in \mathbb{N}_0, \\ q_{i,i-1} &= \delta_i > 0, && i \in \mathbb{N}. \end{aligned}$$

In other words

$$Q = \begin{pmatrix} -\beta_0 & \beta_0 & 0 & 0 & \cdots \\ \delta_1 & -(\delta_1+\beta_1) & \beta_1 & 0 & \cdots \\ 0 & \delta_2 & -(\delta_2+\beta_2) & \beta_2 & \cdots \\ \vdots & 0 & \delta_3 & \ddots & \ddots \end{pmatrix}$$

The dynamics is very simple: if the process is in state i, then the waiting time to the next jump follows an exponential distribution with rate $\beta_i + \delta_i$. At the (random) time of the jump the process moves one step $\nearrow$ with probability $\frac{\beta_i}{\beta_i+\delta_i}$ and one step $\searrow$ with probability $\frac{\delta_i}{\beta_i+\delta_i}$.

REMARK 2.11. A Poisson process of intensity $\lambda > 0$ is a pure birth process on $\mathbb{N}_0$ with intensities $q_{i,i+1} = \lambda$ for any $i \in \mathbb{N}_0$. Therefore, it satisfies the Markov property (2.1) and we have all in hands to end the proof of Proposition 2.5. Let us compute

$$\begin{aligned}
&\mathbb{P}\big(\mathrm{N}(t_1) - \mathrm{N}(t_0) = n_1, \dots, \mathrm{N}(t_p) - \mathrm{N}(t_{p-1}) = n_p\big) \\
&\quad = \mathbb{P}\Big(\mathrm{N}(t_1) - \mathrm{N}(0) = n_1, \dots, \mathrm{N}(t_p) = \sum_{i=1}^{p} n_i\Big) \\
&\quad = \mathbb{P}\big(\mathrm{N}(t_1) - \mathrm{N}(0) = n_1\big)\mathbb{P}\Big(\mathrm{N}(t_2) = n_1 + n_2, \dots, \mathrm{N}(t_p) = \sum_{i=1}^{p} n_i \mid \mathrm{N}(t_1) = n_1\Big) \\
&\quad = \cdots \\
&\quad = \mathbb{P}\big(\mathrm{N}(t_1) - \mathrm{N}(0) = n_1\big) \prod_{k=2}^{p} \mathbb{P}\Bigg(\mathrm{N}(t_k) = \sum_{i=1}^{k} n_i \mid \mathrm{N}(t_1) = n_1, \dots, \mathrm{N}(t_{k-1}) = \sum_{i=1}^{k-1} n_i\Bigg) \\
&\quad = \mathbb{P}\big(\mathrm{N}(t_1) - \mathrm{N}(0) = n_1\big) \prod_{k=2}^{p} \mathbb{P}\big(\mathrm{N}(t_k - t_{k-1}) = n_k \mid \mathrm{N}(0) = 0\big) \\
&\quad = \prod_{k=1}^{p} \mathbb{P}\big(\mathrm{N}(t_k - t_{k-1}) = n_k\big).
\end{aligned}$$

This proves *(ii)* and *(iii)* of Proposition 2.5.

PROPOSITION 2.12 (Kolmogorov equations). *Let $(\mathrm{X}(t))_{t\in\mathbb{R}}$ be a birth-and-death process on $\mathbb{N}_0$ with positive birth rates $(\beta_i)_{i\in\mathbb{N}_0}$ and death rates $(\delta_i)_{i\in\mathbb{N}}$. Recall the definition*

$$p_{i,j}(t) = \mathbb{P}\big(\mathrm{X}(t) = j \mid \mathrm{X}(0) = i\big) = \mathbb{P}\big(\mathrm{X}(s+t) = j \mid \mathrm{X}(s) = i\big),$$

for any $s, t \geqslant 0$.

The (infinite) matrix $\mathrm{P}(t) = (p_{i,j}(t))_{i,j\in\mathbb{N}_0}$ satisfies the backward Kolmogorov differential equation

$$\mathrm{P}'(t) = \mathrm{QP}(t), \qquad \mathrm{P}(0) = \mathrm{Id}, \tag{2.2}$$

where Id *is the infinite identity matrix.*

It also satisfies the forward Kolmogorov differential equation

$$\mathrm{P}'(t) = \mathrm{P}(t)\mathrm{Q}, \qquad \mathrm{P}(0) = \mathrm{Id}. \tag{2.3}$$

Proof. Admitted. □

@home: Find the differential equations satisfied by the real quantities $p_{i,j}(t)$ for any $i, j \in \mathbb{N}_0$.

The *backward equations* can be rewritten as

$$\begin{aligned} p'_{0,j}(t) &= -\beta_0 p_{0,j}(t) + \beta_0 p_{1,j}(t) \\ p'_{i,j}(t) &= \delta_1 p_{i-1,j}(t) - (\beta_i + \delta_i) p_{i,j}(t) + \beta_{i+1} p_{i+1,j}(t), \qquad i \geqslant 1. \end{aligned}$$

The *forward equations* can be rewritten as

$$\begin{aligned} p'_{i,0}(t) &= -\beta_0 p_{i,0}(t) + \delta_1 p_{i,1}(t) \\ p'_{i,j}(t) &= \beta_{j-1} p_{i,j-1}(t) - (\beta_j + \delta_j) p_{i,j}(t) + \delta_{j+1} p_{i,j+1}(t), \qquad j \geqslant 1. \end{aligned}$$

REMARK 2.13. In general the rates (β_i, δ_i) may not determine a *unique* stochastic process. There exist sufficient conditions to ensure this is the case, but we will not pursue in that direction, since in our examples the process will always be uniquely defined.

Let us now briefly describe the behavior of $p_{i,j}(t)$ as t becomes large. It can be proved the following (admitted):

THEOREM 2.14. *The limits*

$$\lim_{t \to \infty} p_{i,j}(t) = \pi(j)$$

exist and are independent of the initial state i, and they satisfy the equations

$$0 = -\beta_0 \pi(0) + \delta_1 \pi(1) \tag{2.4}$$
$$0 = \beta_{n-1} \pi(n-1) - (\beta_n + \delta_n) \pi(n) + \delta_{n+1} \pi(n+1), \quad n \in \mathbb{N}, \tag{2.5}$$

which is obtained by solving $\pi Q = 0$.

If moreover $\sum_j \pi(j) = 1$ then the sequence $(\pi(i))$ is called a stationary distribution.

2.3 Queuing theory

Queues are common in computer systems: there are queues of inquiries waiting to be processed by an interactive computer system, queues of data base requests, etc.

Typically, a queue has one *service facility*, although there may be more than one server in the service facility, and a *waiting room* of finite or infinite *capacity*.

Customers from a population enter a queuing system to receive some service. Upon arrival a customer joins the waiting room if all servers in the service center are busy. When a customer has been served, it leaves the system.

There is a specific way, called *Kendall's notation*, to describe a queuing system: the notation has the form

$$A/B/c/K$$

where

- A describes the *interarrival time* distribution;
- B describes the *service time* distribution;
- c is the number of servers;
- K is the size of the service capacity (= waiting room + servers).

The symbols used for A and B are:

- M for exponential distribution (M stands for Markov);
- D for deterministic distribution;
- G for general distribution.

The most common systems are: $M/M/1/\infty$, $M/M/c/\infty$, $M/G/1/\infty$ and $G/M/1/\infty$.

2.3.1 The $M/M/1/\infty$ queue

In that case, there is one server, and there is no bound imposed on the number of customers waiting for service. Moreover, the interarrival times are exponential random variables with parameter $\lambda > 0$ and the service times are exponential random variables with parameter $\mu > 0$. Let $X(t)$ be the number of customers in the system at time t. This queue process is therefore a birth-and-death process on $\mathbb{N}_0$ with constant birth rates λ and constant death rates μ.

PROPOSITION 2.15. *If $\lambda < \mu$ (meaning that the mean arrival time is bigger than the mean service time), then there exists a unique stationary distribution given by*

$$\pi(i) = \left(1 - \tfrac{\lambda}{\mu}\right)\left(\tfrac{\lambda}{\mu}\right)^i.$$

Proof. Solve equations (2.5). □

Therefore, from Proposition 2.15, one can compute the *mean number of customers* (in the long-time run, *i.e.* according to the stationary state) which is equal to

$$\overline{m} = \sum_{i=0}^{+\infty} i\pi(i) = \frac{\frac{\lambda}{\mu}}{1-\frac{\lambda}{\mu}}.$$

Note that the queue will be empty infinitely many times, since $\pi(0) > 0$.

From now on we define the *traffic intensity*

$$\rho = \frac{\lambda}{\mu}.$$

The condition $\rho < 1$ means that the system is *stable*: the work which is brought to the system is strictly smaller than the processing rate. Observe that

$$\overline{m} \xrightarrow[\rho\to 1]{} +\infty,$$

therefore, in practice, if the system is not stable, then the queue will explode. Finally, one may also be interested in the probability that the queue exceeds K customers (in the stationary state). This is equal, from Proposition 2.15:

$$\mathbb{P}_\pi\left(X \geqslant K\right) = \rho^K \xrightarrow[K\to\infty]{} 0.$$

2.3.2 The $M/M/c/\infty$ queue

This is the $M/M/1/\infty$ queue except that there are $c \in \mathbb{N}$ servers. Therefore, the birth and death rates are now given by

$$\begin{aligned} \beta_i &= \lambda > 0, & i \in \mathbb{N}_0, \\ \delta_i &= \inf\{i, c\}\mu > 0, & i \in \mathbb{N}. \end{aligned}$$

Indeed: if there are $X(t) = i \leqslant c$ customers, there are i independent exponentials of mean $1/\mu$ which can provoke a transition $i \to i-1$. If there are $X(t) = i > c$ customers, then only c exponentials are active since there are only c servers.

In this case, the system of equations satisfied by the stationary state can be written as

$$\begin{aligned} -\lambda\pi(0) + \mu\pi(1) &= 0 \\ \lambda\pi(0) - (\mu+\lambda)\pi(1) + 2\mu\pi(2) &= 0 \\ &\vdots \\ \lambda\pi(i-1) - (i\mu+\lambda)\pi(i) + (i+1)\mu\pi(i+1) &= 0, \qquad i \leqslant c-1, \\ \lambda\pi(i-1) - (c\mu+\lambda)\pi(i) + c\mu\pi(i+1) &= 0, \qquad i \geqslant c. \end{aligned}$$

Solving this system gives:

PROPOSITION 2.16. *If $\rho < c$ (meaning that the mean incoming work is smaller than the maximum service time), then there exists a unique stationary distribution given by*

$$\pi(i) = \begin{cases} \pi(0)\,\dfrac{\rho^i}{i!} & \text{if } i \in \{0,1,\dots,c\} \\[2ex] \pi(0)\,\dfrac{\rho^i c^{c-i}}{c!} & \text{if } i \geqslant c, \end{cases}$$

where

$$\pi(0) = \left(\sum_{i=0}^{c-1}\frac{\rho^i}{i!} + \left(\frac{\rho^c}{c!}\frac{1}{1-\frac{\rho}{c}}\right)\right)^{-1}.$$

Therefore, the probability that an arriving customer cannot find any available server is (in the stationary state)

$$\mathbb{P}_\pi\big(X \geqslant c\big) = \sum_{i=c}^{\infty}\pi(0)\frac{\rho^i c^{c-i}}{c!}.$$

This probability is widely used in telephony to help understand the functioning of a group of agents taking incoming calls in a call center. It is referred to as *Erlang's waiting formula*.

@home: show that the average number of working servers is

$$\mathbb{E}_\pi\big[\inf\{X,c\}\big] = \rho.$$

Show that the average number of customers inside the system is

$$\mathbb{E}_\pi\big[X\big] = \rho + \pi(0)\,\frac{\frac{\rho}{c}}{(1-\frac{\rho}{c})^2}\frac{\rho^c}{c!}.$$

Chapter 3

Monte Carlo methods

Monte Carlo methods are computational algorithms that rely on repeated random sampling to obtain numerical results. In other words, one uses randomness in order to solve problems which might be deterministic in principle. The most famous examples are:

- the computation of number π using the experiment of a "uniform rain" on the square $[-1,1] \times [-1,1]$. Indeed, the probability that a raindrop falls into the unit circle is

  $$\mathbb{P}\big(\text{drop within circle}\big) = \frac{\text{area of the unit circle}}{\text{area of the square}} = \frac{\pi}{4}.$$

  The probability can we estimated by the empirical average of the number of points inside the circle (and is better estimated is the number of raindrops is very large) ;

- the evaluation of any integral $\int_0^1 f(x)\mathrm{d}x$.

The Monte Carlo methods have the advantage of being relatively simple and easy to implement on a computer.

3.1 Metropolis-Hastings algorithm

Let π be a probability measure on a finite or countable state space S. The goal is to simulate a draw from the distribution π (our *target* distribution), which we know up to a constant. In other words, we know how to compute a function proportional to π, which is $f(i) = K\pi(i)$ (but we don't know how to compute nor π neither $K \in \mathbb{R}$). We will generate a Markov chain with invariant probability distribution π. This means that, in the long run, the samples from the Markov chain look like the samples from π. As we will see, the algorithm is incredibly simple and flexible. Its main limitation is that, for difficult problems, "in the long run" may mean after a very long time. However, for simple problems the algorithm can work well.

3.1.1 The transition matrix

First of all, the user must provide a *transition matrix* P. The key assumption we will make is that the Markov chain is *reversible*:

DEFINITION 3.1. *A Markov chain with transition matrix* P *is called* reversible *if there exists a probability distribution* $\pi^\star$ *on* S *such that, for any* $i, j \in S$,

$$\pi^\star(i)P_{i,j} = \pi^\star(j)P_{j,i}. \tag{3.1}$$

PROPOSITION 3.2. *If* $\pi^\star$ *satisfies the* reversibility condition (3.1)*, then* $\pi^\star$ *is an invariant probability distribution.*

Proof. For any state j, we have

$$\sum_{i\in S} \pi^\star(i)P_{i,j} = \sum_{i\in S} \pi^\star(j)P_{j,i} = \pi^\star(j).$$

□

The Metropolis-Hastings algorithm designs a transition matrix P so that the Markov chain is reversible, and so that the target distribution π satisfies (3.1). Let $Q = (Q_{i,j})_{i,j\in S}$ be any *proposal transition matrix*. We construct the transition matrix P as:

$$P_{i,j} = \begin{cases} Q_{i,j}\alpha(i,j) & \text{if } j \neq i, \\ Q_{i,i} + \sum_{k\neq i} Q_{i,k}\big(1-\alpha(i,k)\big) & \text{otherwise,} \end{cases}$$

where

$$\alpha(i,j) = \min\left\{1, \frac{\pi(j)Q_{j,i}}{\pi(i)Q_{i,j}}\right\}.$$

Note that $\alpha(i,j)$ can be rewritten in terms of f as

$$\alpha(i,j) = \min\left\{1, \frac{f(j)Q_{j,i}}{f(i)Q_{i,j}}\right\}, \tag{3.2}$$

which is better for our simulation purposes.

PROPOSITION 3.3. *The target measure π satisfies the reversibility condition (3.1) with respect to the matrix P constructed above.*

Proof. By definition of α, we can assume without loss of generality that $\alpha(j,i) = 1$ and $\alpha(i,j) = \frac{f(j)Q_{j,i}}{f(i)Q_{i,j}}$ (the other case being symmetric). Then,

$$\pi(i)P_{i,j} = Kf(i)Q_{i,j}\frac{f(j)Q_{j,i}}{f(i)Q_{i,j}} = Kf(j)Q_{j,i} = \pi(j)P_{j,i}.$$

□

The reversibility condition does not provide uniqueness of the invariant probability distribution: one needs to ensure that the Markov chain is irreducible and aperiodic. This depends on the proposal transition matrix Q and target probability distribution π.

3.1.2 The algorithm

According to the definition of P, one can now implement:

METROPOLIS-HASTINGS ALGORITHM:

1. Initialize $X(0) \in S$.
2. Given $X(n) = i \in S$, move to $X(n+1)$ as follows:
   (a) Draw a sample j from the transition matrix Q, *i.e.* choose $j \in S$ with probability $Q_{i,j}$.
   (b) Accept the move $i \mapsto j$ with probability $\alpha(i,j)$ where α is given in (3.2). In other words: if $\alpha(i,j) = 1$ then $X(n+1) = j$. If $\alpha(i,j) < 1$, then draw $u \sim \mathcal{U}([0,1])$ (distributed according to a uniform random variable on $[0,1]$, then:
      - if $u < \alpha(i,j)$ then $X(n+1) = j$,
      - if $u \geqslant \alpha(i,j)$ then $X(n+1) = i$.

In practice, we start the chain with an arbitrary $X(0)$, run the algorithm many times (say M times), and then use the last $N \ll M$ draws as samples from π.

The algorithm is often used with *symmetric* candidate probabilities, satisfying $Q_{i,j} = Q_{j,i}$ for any $i, j \in S$. In this case, the formula for α reduces to a simpler form and the behavior of the algorithm is easier to understand: it will always make a transition to a state j whenever $\pi(j) > \pi(i)$, but it can also make a transition to a state j if $\pi(j) \leqslant \pi(i)$; the probability of such a transition is equal to the ratio $\pi(j)/\pi(i)$.

3.1.3 An example: the Ising model

The *Ising model* is a model of ferromagnetism in statistical physics. Consider the state space $S = \{-1, +1\}^{N \times N}$ constituted by all square matrices of size $N \times N$ whose coefficients are either -1 or $+1$. An element $B_{i,j}$ of $B \in S$ is called *spin* at *site* (i, j). We say that two sites (i_1, j_1) and (i_2, j_2) are *adjacent* if both their coordinates differ at most by 1, and we write $(i_1, j_1) \sim (i_2, j_2)$.

For any $B \in S$, its probability $\pi(B)$ is a function of its *energy*, which is given by the formula

$$\mathcal{H}(B) = -\sum_{(i_1,j_1)\sim(i_2,j_2)} B_{i_1,j_1} B_{i_2,j_2}.$$

Note that two adjacent sites with the same spin reduce the value of $\mathcal{H}$ by 1, while two adjacent sites with the opposite spin contribute by 1. The probability of a state B (using Boltzmann's theory) favors the states with *low energy*, and is given by

$$\pi(B) = \frac{1}{Z_\beta} e^{-\beta \mathcal{H}(B)}, \tag{3.3}$$

where $\beta > 0$ is a constant which can be chosen, and Z_β is the normalizing constant, which is very hard to calculate, even numerically.

One can see the advantage of the Metropolis-Hastings algorithm: it is hopeless to generate a random state distributed according to π using the formula (3.3), because of the size of S (which is 2^{N^2}). In the algorithm, we need to

1. choose a symmetric candidate probability Q, for instance: we move from $B \mapsto B'$ according to the uniform distribution on all states which differ from B at exactly one site, namely:

$$Q_{B,B'} = \begin{cases} \frac{1}{N^2} & \text{if B and B}' \text{ differ at exactly one site } (i,j) \\ 0 & \text{otherwise;} \end{cases}$$

2. compute the *acceptance probability* α, which here is given by

$$\frac{\pi(B')}{\pi(B)} = e^{-\beta(\mathcal{H}(B')-\mathcal{H}(B))}.$$

Since B and B′ differ by only one site, say (i, j), one can easily check that $\mathcal{H}(B') - \mathcal{H}(B)$ actually contains a finite number of terms, namely:

$$\mathcal{H}(B') - \mathcal{H}(B) = -\sum_{(i_1,j_1)\sim(i,j)} \left(B'_{i_1,j_1}B'_{i,j} - B_{i_1,j_1}B_{i,j}\right) = 2B_{i,j}\sum_{(i_1,j_1)\sim(i,j)} B_{i_1,j_1},$$

since $B_{i,j} = -B'_{i,j}$ and $B_{i_1,j_1} = B'_{i_1,j_1}$ by definition.

3.2 Simulated Annealing Algorithm

The Metropolis-Hastings algorithm has other applications. Let us assume that we are given a function $\mathcal{H} : S \to \mathbb{R}$ where S is *finite*, and we want to compute the points where $\mathcal{H}$ achieves its minimum. The function $\mathcal{H}$ is often called *energy*, by analogy with statistical physics theory (as in the Ising model).

DEFINITION 3.4. *The* Gibbs measure *associated to the energy function* H *at temperature* $T > 0$ *is the probability measure* μ_T *defined on* S *by*

$$\mu_T(x) = \frac{1}{Z_T}e^{-\mathcal{H}(x)/T}, \qquad x \in S,$$

where Z_T *is the normalization constant also called partition function.*

When $T \to 0$, the Gibbs measures concentrate on the points where $\mathcal{H}$ achieves its minimum: let $\mathcal{M} \subset S$ be the set of all minimas of $\mathcal{H}$, then

$$\mu_T(x) \xrightarrow[T\to 0]{} \begin{cases} 0 & \text{if } x \notin \mathcal{M} \\ \dfrac{1}{|\mathcal{M}|} & \text{if } x \in \mathcal{M}. \end{cases}$$

The general idea of the *Simulated Annealing Algorithm* is:

- generate a Markov chain which is irreducible, aperiodic, with invariant probability μ_T, like in the Metropolis-Hastings algorithm;
- let this chain evolve until it becomes close to the invariant measure:

- at each step, decrease the temperature.

At the end, the chain should always visit states in $\mathcal{M}$. We have the following proposition:

PROPOSITION 3.5. *Let $(X(n))_{n\in\mathbb{N}_0}$ be a Markov chain which is irreducible, aperiodic, with invariant probability measure μ_T. Then, for any initial distribution,*

$$\lim_{T\to 0}\lim_{n\to\infty} \mathbb{P}\big[X(n) = x\big] = \begin{cases} 0 & \text{if } x \notin \mathcal{M} \\ \dfrac{1}{|\mathcal{M}|} & \text{if } x \in \mathcal{M}. \end{cases}$$

For practical reasons, it is much more convenient to undertake both limits simultaneously, *i.e.* to choose $T(n) \to 0$ as $n \to \infty$. We obtain an *inhomogeneous* Markov chain, with transition probabilities at step n which depend on $T(n)$. Note that $T(n)$ should not decrease *too fast* towards 0, so that the chain has enough time to reach its invariant measure. In practice, we use the following theorem:

THEOREM 3.6. *There exists $h^\star > 0$ such that, for any $h \geqslant h^\star$, the simulated annealing algorithm associated to the Metropolis-Hastings scheme at temperature $T(n) = h/\log(n)$, whose transition matrix is irreducible and aperiodic, converges as $n \to \infty$ to the uniform measure on $\mathcal{M}$.*

Proof. Admitted, from Hajek (1988). □

3.2.1 An example: the traveling salesman problem

The concrete problem of a traveling salesman is the following: there are m cities which are situated at distinct locations $V_1, \dots, V_m$. The salesman needs to find the order of cities to visit which minimizes the total distance: in other words, find the permutation $\sigma \in \Sigma_m$ which minimizes the function

$$\sigma \in \Sigma_m \mapsto \mathcal{H}(\sigma) := \sum_{i=1}^{m} \mathrm{dist}\big(V_{\sigma(i)}, V_{\sigma(i+1)}\big), \qquad \sigma(m+1) = \sigma(1).$$

It is hopeless to compute $\mathcal{H}$ for any permutation σ (even for $m \geqslant 10$ it is not feasible).

We are therefore going to apply the Simulated Annealing Algorithm by constructing a Markov chain on the set of permutations: one needs to define the possible transitions. The most efficient way is the following: we say that $\sigma = (x_1, \ldots, x_m)$ and $\sigma' = (y_1, \ldots, y_m)$ are *neighbours* if there exists $1 \leqslant i < k \leqslant m$ such that

$$(y_1, \ldots, y_m) = (x_1, \ldots, x_{i-1}, x_k, x_{k-1}, \ldots, x_{i+1}, x_i, x_{k+1}, \ldots, x_m).$$

In other words, the cities order between rank i and rank k is reversed.

SIMULATED ANNEALING ALGORITHM (TRAVELING SALESMAN PROBLEM):

1. Choose an irreducible transition probability matrix P, aperiodic, on Σ_m, such that $p_{\sigma,\sigma'} > 0$ if and only if σ and σ' are neighbours. *An exemple will be given during the practical session.*
2. Initialize $X(0) = \sigma_0 \in \Sigma_m$.
3. Repeat the Metropolis-Hastings Algorithm scheme to construct $X(n)$, changing $T(n) = c/\log(n)$ at any step. *The choice of the constant c will be investigated in the practical session.*