Chapter 1

Basic theory

Markov chains are the principal tool for modeling random dynamic phenomena and computing related probabilities.

In this chapter, the principal basic results about discrete-space Markov chains are reviewed.

The power and effectiveness of Markov chain modeling is due in a large part to the fact that many probability computations can be reduced to linear algebra computations. Appendix A provides the necessary notions.

1.1 Discrete-time Markov chains

1.1.1 Definitions

Throughout this document, $\mathcal{E}$ will denote a set at most countable, that is, either finite or countably infinite.

Definition 1.1 (Discrete-time Markov chain)**.** *Let $\mathcal{E}$ be a set at most countable. A process $\{X(n), n \in \mathbb{N}\}$ is a discrete-time Markov chain (DTMC for short) over $\mathcal{E}$ if and only if: $X(n) \in \mathcal{E}$ for all n, and:*

i/ (Markov property) for all $t \in \mathbb{N}$, and all $t+2-$ uple $(j_0, j_1, \ldots, j_t, j_{t+1})$ of elements in $\mathcal{E}$:

$$\mathbb{P}(X(t+1) = j_{t+1} | X(t) = j_t, \ldots, X(0) = j_0) \;=\; \mathbb{P}(X(t+1) = j_{t+1} | X(t) = j_t) \;; \qquad (1.1)$$

This chain is homogeneous *if in addition:*

ii/ (homogeneity) for all pair (i, j) in $\mathcal{E}$, there exists a number $P_{i,j}$ such that:

$$\mathbb{P}(X(t+1) = j | X(t) = i) \;=\; P_{i,j} \;, \qquad (1.2)$$

for any value of t.

We shall often use "DTMC" for Discrete-time Markov chain.

1.1.2 Finite-state Markov chains

In this section and in the following, the state space $\mathcal{E}$ will be assumed to be *finite*. In that case, the Markov chain can be "represented" by a matrix, and many probabilistic properties can be deduced from the properties of this matrix.

1.1.2.1 Markov chains and matrices

The numbers $P_{i,j}$, $(i,j) \in \mathcal{E} \times \mathcal{E}$ are the *transition probabilities*, and the matrix $\mathbf{P}$, which components are the $P_{i,j}$, is the *transition matrix*. This matrix is stochastic:

Definition 1.2 (Stochastic Matrix)**.** *A stochastic matrix is a square matrix which entries are all real and positive, and such that, the sum over each row is equal to 1.*

This property can be expressed in the notation of linear algebra. Let $\mathbf{1}$ be the column vector made of "1": $\mathbf{1} = (1, \ldots, 1)^T$. Then

$$\mathbf{P}\ \mathbf{1} \ = \ \mathbf{1}\ . \tag{1.3}$$

Conversely, given a stochastic matrix $\mathbf{P}$, it is easy to define a process $\{X(n), n \in \mathbb{N}\}$ which is an homogeneous Markov chain with $\mathbf{P}$ as transition matrix. See Section 3.1 for the illustration of this principle in simulation.

A Markov chain is therefore *characterized* by its transition probability matrix.

Since it is possible to associate a weighted graph to every matrix (see Section A.2), we associate a graph $\mathcal{G}$ to every Markov chain: this graph is called the **transition diagram**. An example is shown in Section 1.2.5.

1.1.2.2 Transient Probabilities

The first problem which arises is to compute the probability that the Markov chain is in some given state after some given number of jumps. Those are called the "transient probabilities".

Actually, two types of probabilities are interesting. The **n-step transition probabilities**:

$$p(i,j;n) \ = \ \mathbb{P}(X(n) = j \mid X(0) = i)\ , \tag{1.4}$$

and the n-th step **state probabilities**:

$$\pi_n(j) \ = \ \mathbb{P}(X(n) = j)\ . \tag{1.5}$$

The first one is the probability of going from state i to state j in n steps. The second one is the probability of being in state i after n steps.

Denote as $P(n)$ the matrix constructed with the numbers $p(i,j;n), (i,j) \in \mathcal{E} \times \mathcal{E}$.

These probabilities $p(i,j;n)$ are given by the theorem:

Theorem 1.1 (Transient probabilities)**.** *For all $n \in \mathbb{N}$:*

$$P(n) \ = \ \mathbf{P}^n\ .$$

Given Theorem A.1 which relates powers of matrices and sequences of transitions, we also have the following expression. The definition of $\Gamma^n(i,j)$ is in the theorem.

Corollary 1.2. *For all $n \in \mathbb{N}$, $(i,j) \in \mathcal{E} \times \mathcal{E}$, we have:*

$$p(i,j;n) \;=\; \sum_{(i,i_1,\ldots,i_{n-1},j)\,\in\,\Gamma^n(i,j)} p_{i,i_1}\; p_{i_1,i_2}\;\cdots\; p_{i_{n-1},j}\;. \tag{1.6}$$

The interpretation of this formula is the following: consider a path of length n (measured as the number of arcs) in graph $\mathcal{G}$: $i = i_0, i_1, \ldots, i_n = j$. To this path is associated the weight: $p_{i_0,i_1} \ldots p_{i_{n-1},i_n}$. The probability $p(i,j;n)$ is the sum of the weights of all paths of length n going from i to j.

This probability is therefore strictly positive *if and only if* there exists a path of length n going from i to j in the transition diagram.

Consider now the probabilities $\pi_n(j)$, $n \in \mathbb{N}$ and $j \in \mathcal{E}$, defined in (1.5). In order to be able to compute this probability, it is necessary to introduce the **initial distribution** of the process, that is:

$$\pi_0(j) \;=\; \mathbb{P}(X(0) = j), \qquad j \in \mathcal{E}.$$

We then have, conditioning on the initial state:

$$\pi_n(j) \;=\; \sum_{i\in\mathcal{E}} \pi_0(i)\; p(i,j;n)\;. \tag{1.7}$$

This can be expressed in matrix form. To this end, we denote with $\boldsymbol{\pi}_n$ the (row) vector with components $\pi_n(j)$. Then, Equation (1.7) is equivalent to:

Theorem 1.3. *The transient state probabilities are given by:*

$$\boldsymbol{\pi}_n \;=\; \boldsymbol{\pi}_0\; \mathbf{P}^n\;. \tag{1.8}$$

1.1.2.3 The number of visits

A random variable of interest in Markov chains is the number of visits made to one specific state.

Given a Markov chain $\{X(n), n \in \mathbb{N}\}$, given an integer n and some state $i \in \mathcal{E}$, define:

$$V_n^{(i)} \;=\; \sum_{m=0}^{n-1} \mathbf{1}_{\{X(m)=i\}}\;. \tag{1.9}$$

Recall that the indicator function $\mathbf{1}_{\{A\}}$ is 1 if the event A is true, 0 if not. Accordingly, the value $V_n^{(i)}$ counts the number of times the chain X has visited state i between steps 0 (initial state) and step $n-1$.

Taking the mathematical expectation in (1.9), we have:

$$\mathbb{E}V_n^{(i)} \;=\; \sum_{m=0}^{n-1} \mathbb{P}(X_m = i) \;=\; \sum_{m=0}^{n-1} \boldsymbol{\pi}_0\; \mathbf{P}^m\; \mathbf{e}_i\;, \tag{1.10}$$

where $\mathbf{e}_i$ is the column vector $(0,\ldots,1,\ldots,0)^T$, where the "1" is in row i. According to this formula, the *average* number of visits can be expressed using the initial distribution and the transition matrix.

1.1.3 Asymptotic behavior and the classification of states

In this section, the set $\mathcal{E}$ is still finite.

The next important question is to know how the sequences of numbers $p(i,j;n)$ and $\pi_n(j)$ behave when n tends to infinity. Given Theorem 1.1 and Theorem 1.3, the question is equivalent to determining the behavior of $\mathbf{P}^n$ and $\boldsymbol{\pi}_0\mathbf{P}^n$ when $n \to \infty$. The principal general results about powers of matrices are recalled in Section A.3. We shall apply them here to the stochastic matrix $\mathbf{P}$.

Two types of information are useful in practice: *a*) results which tell under which condition the probabilities converge, and what is their limit, and *b*) what is the *speed* of convergence.

1.1.3.1 Classification of states

Let $\mathcal{G}$ be the transition graph of the transition matrix $\mathbf{P}$. It is possible to perform a decomposition of $\mathcal{G}$ into *strongly connected components* (see Appendix A.2). In the language of Markov chain theory, these components are called the *communication classes* of the chain.

Next, consider the graph $\mathcal{C}$ of communication classes: this graph is constructed with one vertex for each communication class, and with edges between classes C_1 and C_2 if there exist one pair (i,j), $i \in C_1$, $j \in C_2$ with $P_{i,j} > 0$. In other words: if it is possible to go from C_1 to C_2 with positive probability.

By construction, the graph $\mathcal{C}$ has no cycle: it is called a DAG: directed acyclic graph. In such graphs, there are always vertices with no outgoing edge. These vertices of $\mathcal{C}$ correspond to communication classes that are called the *terminal communication class*.

We can now introduce the terminology for states:

Definition 1.3 (Classes of states)**.** *The states in terminal communication classes are called* ***recurrent****. The other states are called* ***transient****.*

When a terminal communication class consists in a single state, this state is called ***absorbing****. It is necessarily recurrent.*

Recall from Definition A.2 that a matrix $\mathbf{P}$ is *irreducible* if all states are in a single communication class. In other words: if $\mathbf{P}$ is irreducible, all states are recurrent.

The decomposition of the state space in communication classes has the following algebraic interpretation, in the case of finite matrices. If $\mathbf{P}$ is reducible, then it is equivalent, up to a renumbering of the states, to a **block triangular** matrix:

$$\begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \dots & \\ \mathbf{0} & \mathbf{P}_{22} & \ddots & \\ \vdots & \mathbf{0} & \ddots & \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{P}_{pp} \end{pmatrix} . \tag{1.11}$$

Each of these blocks correspond to a communication class, and one vertex in the graph $\mathcal{C}$. If the communication class c is terminal, there is only one non-zero block in the matrix: the diagonal block P_{cc}. This sub-matrix is a stochastic matrix: it is the matrix of a Markov chain evolving in the communication class C_c. This Markov chain is irreducible. For recurrent communication classes, there is more than one block on the corresponding line.

1.1.3.2 Convergence of probabilities

We give below several results concerning the convergence of probabilities. These results use increasingly stronger assumptions on the matrix $\mathbf{P}$.

The first result is an application of Theorem A.5 (Perron-Frobenius). The other results can be proved using the spectral decomposition of matrices (Theorem A.2) and the results of Section A.3. The details are omitted here.

Theorem 1.4 (Existence of a stationary distribution)**.** *Assume $\mathbf{P}$ is a stochastic, irreducible matrix. Then there exists a unique probability vector $\boldsymbol{\pi}$ such that $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$ and $\boldsymbol{\pi}\mathbf{1} = 1$. The vector $\boldsymbol{\pi}$ is strictly positive.*

This vector is called the vector of stationary probabilities of the Markov chain.

The distribution on the state space $\mathcal{E}$ corresponding to this vector $\boldsymbol{\pi}$ is called the **stationary distribution**. The name "stationary" comes from the probability that if the initial distribution of a Markov chain is $\boldsymbol{\pi}$, then the state distribution does not change over time: the chain is stationary. This can be checked algebraically: if $\boldsymbol{\pi}_0 = \boldsymbol{\pi}$, then for all n.

$$\boldsymbol{\pi}_n \;=\; \boldsymbol{\pi}_0\,\mathbf{P}^n \;=\; \boldsymbol{\pi}\,\mathbf{P}^n \;=\; (\boldsymbol{\pi}\mathbf{P})\mathbf{P}^{n-1} \;=\; \boldsymbol{\pi}\,\mathbf{P}^{n-1} \;=\; \boldsymbol{\pi}\;.$$

Theorem 1.5. *Let $\mathbf{P}$ be a stochastic, irreducible matrix and $\boldsymbol{\pi}$ its vector of stationary probabilities. Then:*

$$\sum_{m=0}^{n-1} \mathbf{P}^m \;=\; n\,\mathbf{1}.\boldsymbol{\pi} + \mathcal{O}(1)\;,$$

and in particular:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \mathbf{P}^m \;=\; \mathbf{1}.\boldsymbol{\pi}\;. \tag{1.12}$$

As a consequence, for any probability vector π_0, we have:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \boldsymbol{\pi}_m \;=\; \boldsymbol{\pi}\;. \tag{1.13}$$

This result provides the interpretation of the stationary probability: we have seen in (1.10) that the expected number of visits to state i is:

$$\mathbb{E}V_n^{(i)} \;=\; \sum_{m=0}^{n-1} \pi_m(i)\;.$$

Then, using (1.13), we have:

$$\pi_i \;=\; \lim_{n\to\infty} \frac{1}{n}\mathbb{E}V_n^{(i)}\;.$$

In words: the stationary distribution of state i is the average number of passages of the DTMC in state i per step.

If the matrix $\mathbf{P}$ is assumed to be *aperiodic* in addition (see Definition A.2 on p. 75), the convergence result is stronger.

Theorem 1.6. *Let* $\mathbf{P}$ *be a stochastic, irreducible and aperiodic matrix. Let* $\boldsymbol{\pi}$ *be its unique stationary probability vector. Then,*

$$\lim_{n\to\infty} \mathbf{P}^n \;=\; \mathbf{1}.\boldsymbol{\pi} \;. \tag{1.14}$$

As a consequence for any probability vector $\boldsymbol{\pi}_0$*, we have:*

$$\lim_{n\to\infty} \boldsymbol{\pi}_n \;=\; \boldsymbol{\pi} \;. \tag{1.15}$$

The convergence is geometric: for any vector norm,

$$||\boldsymbol{\pi}_n - \boldsymbol{\pi}|| \;=\; \mathcal{O}(\rho^n) \tag{1.16}$$

for all ρ *such that* $|\lambda_2| < \rho < 1$*, where* $|\lambda_2|$ *is the second largest eigenvalue of* $\mathbf{P}$*.*

Given the importance of Theorem 1.6 and the related conditions on $\mathbf{P}$, the following terminology is used:

Definition 1.4 (Ergodic matrix)**.** *A stochastic matrix which is irreducible and aperiodic is called* ***ergodic****.*

When the matrix is not irreducible, the classification of states introduced in Definition 1.3 becomes relevant.

Introduce the series of matrices $\mathbf{V} = \sum_{n=0}^{\infty} \mathbf{P}^n$. It is called the potential matrix and it has values in $\mathbb{R} \cup \{+\infty\}$. It can be deduced from Equation (1.10) that this matrix is related with the number of visits. Indeed, the element V_{ij} is the average *total number* of visits in state j of the chain when it starts in state i:

$$V_{ij} \;=\; \lim_{n\to\infty} \mathbb{E}\left[V_n^{(j)} \mid X(0) = i\right] \;. \tag{1.17}$$

Theorem 1.7. *Let* $\{X(n), n \in \mathbb{N}\}$ *be a DTMC over the finite state space* $\mathcal{E}$*. Let* $i \in \mathcal{E}$*. Then:*

i/ If i *is transient, then* $(\pi_n)_i = \mathcal{O}(\rho^n)$ *for some* $\rho < 1$*, and* V_{ii} *is finite;*

ii/ If i *is recurrent, then* $V_{ii} = +\infty$*.*

This result explains the terminology for the classification of states. Transient states are called this way because the Markov chain goes through such states only a finite number of times with probability one. On the other hand, the chain goes through a recurrent state: either an infinite number of times (if the chain reaches the terminal class containing this state) or never (if the chain reaches another terminal class).

Finally, we state the following result, which is implicit in Theorem 1.7 but deserves a separate statement:

Corollary 1.8. *Let* $\{X(n), n \in \mathbb{N}\}$ *be a DTMC over a finite state space* $\mathcal{E}$*, with transition matrix* $\mathbf{P}$*. If* $\mathbf{P}$ *is irreducible, then the Markov chain goes an infinite number of times through each state* $i \in \mathcal{E}$ *with probability 1, whatever its initial distribution.*

1.1.4 Markov chains with infinite state spaces

When the state space $\mathcal{E}$ is countably infinite, definitions (1.1) and (1.2) are still valid. The matrix/vector notations are also valid *although* the product of infinite matrices is *not* well defined in general. This is because infinite sums may not converge. However, in the case of probability matrices and probability vectors, the series which appear in products do converge, and the result is well-defined.

The transition graph is defined as in the case of a finite state space. Although this is an infinite graphs, the notions of irreducibility and (a)periodicity can be defined the same way.

The classification of states introduced in Definition 1.3 must be refined in the infinite case: the set of transient states is divided into two subsets. Recall the definition of the values V_{ij}: Equation (1.17) in Section 1.1.3.2.

Definition 1.5 (Refined classification of states)**.** *A state of $\mathcal{E}$ is called:*

- *transient if $V_{ii} < \infty$,*
- *null recurrent if $V_{ii} = \infty$ and $(\mathbf{P}^n)_{ii} \to 0$*
- *positive recurrent (or simply: recurrent) if $V_{ii} = \infty$ and $(\mathbf{P}^n)_{ii} \not\to 0$.*

When the state space is finite, the case "null recurrent" is not possible. Theorem 1.7 and Definition 1.3 give the same result as Definition 1.5.

The complication for infinite Markov chains is also that now the analysis of the graphs in communication classes is not enough to classify the states (with Definition 1.3 and Theorem 1.7). When there exist infinite terminal classes, the states of this class may be of any of the three types.

On the other hand, there is a "solidarity theorem" for states.

Theorem 1.9 (Solidarity)**.** *If $\mathbf{P}$ is irreducible, then all states have the same type (transient or null recurrent or positive recurrent) and they all have the same period.*

Therefore, in the case of irreducible chains (or for every terminal class of a reducible chain), if one of the states is, say, positive recurrent, all states have this property. The qualifier "positive recurrent" is then given to the Markov chain. Likewise for null recurrent and transient.

Finally, we have the following classification theorem.

Theorem 1.10 (Classification of irreducible infinite chains)**.** *Let $\mathbf{P}$ be an infinite, irreducible and aperiodic stochastic matrix. Two possibilities only occur:*

- *either the system of equations $\mathbf{xP} = \mathbf{x}$ has at least one strictly positive solution $\mathbf{x}$ such that $\sum_{i\in\mathcal{E}} x_i < \infty$.*

  Then there exists a vector of stationary probabilities $\boldsymbol{\pi}$ and $\boldsymbol{\pi}_n \to \boldsymbol{\pi}$ for all initial distribution $\boldsymbol{\pi}_0$. The chain is positive recurrent.

- *or there exist no solution of $\mathbf{xP} = \mathbf{x}$ such that $\sum_{i\in\mathcal{E}} |x_i| < \infty$.*

  In that case, $\boldsymbol{\pi}_n \to \mathbf{0}$ for all initial distribution $\boldsymbol{\pi}_0$. The chain is either null recurrent, or transient.

The classical proof of these properties uses arguments of renewal theory: see for instance [5, 22]. Methods from linear algebra that are used in the finite case do not work in general.

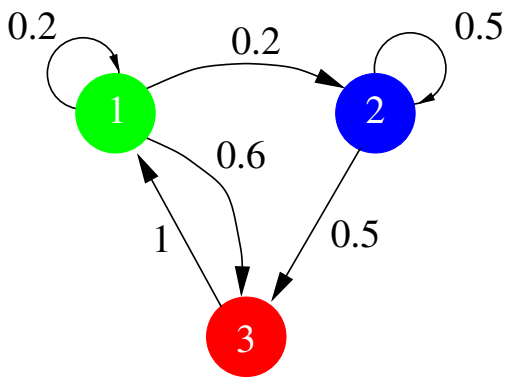

Figure 1.1: Transition diagram of a discrete-time Markov chain

1.1.5 An example

Consider the stochastic matrix:

$$\mathbf{P} = \begin{pmatrix} 0.2 & 0.2 & 0.6 \\ 0 & 0.5 & 0.5 \\ 1 & 0 & 0 \end{pmatrix} .$$

The corresponding transition diagram is represented in Figure 1.1.5.

Assuming an initial distribution $\boldsymbol{\pi}_0 = (1, 0, 0)$ and applying Theorem 1.3, we obtain the following sequence of transient probability vectors:

$$\begin{array}{lcl} \pi_0 & = & (1, 0, 0) \\ \pi_1 & = & (0.2, \quad 0.2, \quad 0.6) \\ \pi_2 & = & (0.64, \quad 0.14, \quad 0.22) \\ \pi_3 & = & (0.348, \quad 0.198, \quad 0.454) \\ \pi_4 & = & (0.5236, \quad 0.1686, \quad 0.3078) \\ \vdots & \vdots & \vdots \\ \pi_{100} & \simeq & (0.45454545, 0.18181818, 0.36363636) \\ \pi_{101} & \simeq & (0.45454545, 0.18181818, 0.36363636) \\ \vdots & \vdots & \vdots \\ \pi_\infty & = & (5/11, 2/11, 4/11) \ . \end{array}$$

The limit is obtained with Theorem 1.6. Indeed, the matrix is irreducible and aperiodic, so that the limit exists, and is the unique solution of the system of equations: $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$ and $\boldsymbol{\pi}\mathbf{1} = 1$.

1.2 Continuous-time Markov chains

The concepts introduced for discrete-time Markov chains have correspondences for continuous-time Markov chains.

1.2.1 Definitions

Continuous-time Markov chains are closely related to *Exponential distributions* and *Poisson processes*, which we introduce first.

The Exponential distribution

Definition 1.6 (Exponential distribution)**.** *X has an exponential distribution of parameter $\lambda > 0$ (denoted as: $X \sim Exp(\lambda)$) if for $x \geq 0$:*

$$F_X(x) \;=\; \mathbb{P}(X \leq x) \;=\; 1 \;-\; e^{-\lambda x} \;,$$

and $F_X(x) = 0$ for $x < 0$. Equivalently, X has the density $f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$ and $f_X(x) = 0$ for $x < 0$.

The distribution function $F_X(x)$ and the density $f_X(x)$ of the Exponential distribution are illustrated in Figure 1.2.1.

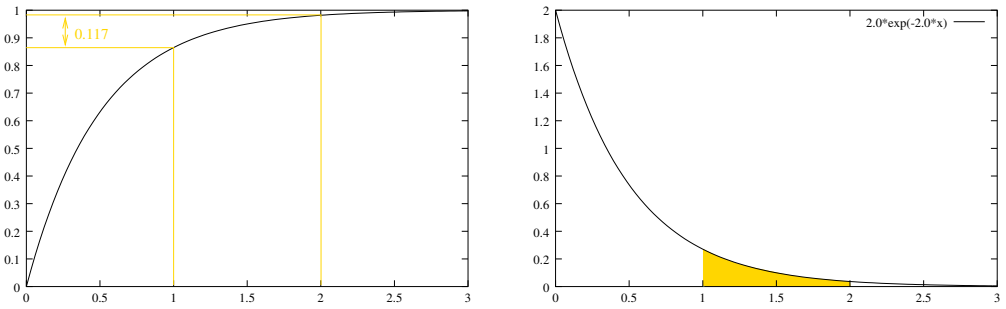

Figure 1.2: Cumulative distribution function and density of the exponential random variable

The exponential distribution enjoys several important properties:

- the moments of $X \sim \text{Exp}(\lambda)$, are given by:

  $$\mathbb{E}X \;=\; \lambda^{-1} \qquad \mathbb{E}X^k \;=\; k!\lambda^{-k}$$

  and in particular, $\text{var}(X) \;=\; \lambda^{-2}$.

- The exponential distribution is *memoryless*: $\forall s, t > 0$,

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

- the *failure rate* of the exponential distribution is constant: for all $t \geq 0$,

  $$\lim_{h \to 0} \frac{1}{h}\, \mathbb{P}(X \in [t, t+h[\mid X > t) \;=\; \lambda \;. \tag{1.18}$$

- The family of exponential distributions is stable under minimization: if $X_1 \sim \text{Exp}(\lambda_1)$, $X_2 \sim \text{Exp}(\lambda_2)$ and X_1 and X_2 are independent: then

$$\min\{X_1, X_2\} \ \sim \ \text{Exp}(\lambda_1 + \lambda_2) \ . \tag{1.19}$$

  Moreover:

$$\mathbb{P}(\min\{X_1, X_2\} = X_i) \ = \ \frac{\lambda_i}{\lambda_1 + \lambda_2} \ . \tag{1.20}$$

These properties can be generalized to any number of variables.

The Poisson process Given a sequence of time instants $T_0 \leq T_1 \leq \ldots \leq T_n \leq T_{n+1} \leq \ldots$, one defines the *counting process* of this sequence as:

$$N(a,b) \ = \ \#\{n \mid a \leq T_n < b\} \ = \ \sum_{n=0}^{\infty} \mathbf{1}_{\{a \leq T_n < b\}}$$

Definition 1.7 (Poisson process)**.** *A sequence of time instants $(T_n; n \in \mathbb{N})$ is a Poisson process of parameter λ if the sequence of time increments $(T_{n+1} - T_n; n \in \mathbb{N})$ is a i.i.d. sequence of variables Exp(λ).*

The Poisson process enjoys several important properties, among which:

- For all u:

$$\mathbb{P}(N(x, x+u) \ = \ k) \ = \ \frac{(\lambda u)^k}{k!} \, e^{-\lambda u} \ .$$

  In words: the distribution of the number of time instants in any interval of length u is a random variable with a Poisson distribution of parameter λ.[1]

  In particular, $\mathbb{E}N(x, x+u) = \lambda u$: λ is called the *arrival rate* of the process.

- For all $t_4 \leq t_3 < t_2 \leq t_1$,

$$\mathbb{P}(X(t_1) - X(t_2) = k \mid X(t_3) - X(t_4) = j) \ = \ \mathbb{P}(X(t_1) - X(t_2) = k) \ .$$

  In words: the number of events of the Poisson process in two disjoint intervals are independent random variables.

Continuous-time Markov chains

Definition 1.8 (Continuous-time Markov chain)**.** *Let $\mathcal{E}$ be a set at most countable. A process $\{X(t), t \in \mathbb{R}^+\}$ is a homogeneous continuous time Markov chain (or:* Markov process*) if and only if:*

i/ (Markov property) For all $n \in \mathbb{N}$, every $n+2$–uple of reals $t_0 < t_1, < \ldots < t_n < t_{n+1}$ and every $n+2$–uple $(j_0, j_1, \ldots, j_n, j_{n+1})$ of elements of $\mathcal{E}$:

$$\mathbb{P}(X(t_{n+1}) = j_{n+1} | X(t_n) = j_n, \ldots, X(t_0) = j_0) \ = \ \mathbb{P}(X(t_{n+1}) = j_{n+1} | X(t_n) = j_n) \ ;$$

[1] One should not confuse the *Poisson distribution* and the *Poisson process*.

ii/ (homogeneity) For all reals s, t and u, and every pair (i, j) of $\mathcal{E}$, independently of u we have:

$$\mathbb{P}(X(t+u) = j | X(s+u) = i) \;=\; \mathbb{P}(X(t) = j | X(s) = i) \;=\; P_{t-s}(i, j) \; .$$

We shall often use "CTMC" for Continuous-time Markov chain.

1.2.2 Transient probabilities and the Chapman-Kolmogorov equations

As in the discrete-time case, we are interested in the transition probabilities: for $t \in \mathbb{R}_+$, $i, j \in \mathcal{E}$,

$$p_t(i, j) \;=\; \mathbb{P}(X(t) = j \mid X(0) = i) \; , \tag{1.21}$$

and in the state probabilities:

$$p_t(i) \;=\; \mathbb{P}(X(t) = i) \; . \tag{1.22}$$

As in the discrete-time case, one forms the matrices $\mathbf{P}_t$ with the values $p_t(i, j)$, $(i, j) \in \mathcal{E}^2$, and the vectors $\boldsymbol{\pi}_t$ with the numbers $p_t(i)$.

From the Markov property and the homogeneity property, one deduces that for all pair of states i and j, the following result.

Theorem 1.11 (Chapman-Kolmogorov equations)**.** *For all $i, j \in \mathcal{E}$ and $t, s \in \mathbb{R}_+$, the transition probabilities satisfy the following equations:*

$$p_{t+s}(i, j) \;=\; \sum_{k \in \mathcal{E}} p_t(i, k) \; p_s(k, j) \; , \tag{1.23}$$

or, in algebraic form:

$$\mathbf{P}_{t+s} \;=\; \mathbf{P}_t \; \mathbf{P}_s \tag{1.24}$$

These are known as the Chapman-Kolmogorov equations.

These equations are not very useful to compute the numbers $p_t(i, j)$. We look for more explicit formulas, similar to that of Theorem 1.1 in the discrete-time case.

If the process $\{X(t)\}$ is "regular" enough, then there exists a matrix $\mathbf{Q}$ defined as:

$$\mathbf{Q} \;=\; \lim_{t \to 0} \frac{\mathbf{P}_t - \mathbf{I}}{t} \; . \tag{1.25}$$

Definition 1.9 (Infinitesimal generator)**.** *The matrix $\mathbf{Q}$ is called the **infinitesimal generator**, or more commonly, the generator, of the Markov chain.*

This essential matrix has several important properties:

Theorem 1.12. *Any infinitesimal generator $\mathbf{Q}$ has the following properties:*

i) for all $i \neq j$, $Q_{ij} \geq 0$, and for all i, $Q_{ii} \leq 0$;

ii) for all $i \in \mathcal{E}$:

$$\sum_{j\in\mathcal{E}} Q_{ij} \;=\; Q_{ii} + \sum_{i\neq j} Q_{ij} \;=\; 0\ . \tag{1.26}$$

In matrix form, this is written as: $\mathbf{Q1} = \mathbf{0}$.

iii) for all t,

$$\frac{d\mathbf{P}_t}{dt} \;=\; \mathbf{Q}\mathbf{P}_t \;=\; \mathbf{P}_t\mathbf{Q}\ . \tag{1.27}$$

The definition above of the infinitesimal generator does not give intuitive interpretations. Those will be the topic of Chapter 3 (Section 3.2 in particular).

The transient probabilities can be expressed using the infinitesimal generator. Recall the definition of the matrix exponential in Appendix A.4.

Theorem 1.13. *Let $\{X(t); t \in \mathbb{R}\}$ be a homogeneous, continuous-time Markov chain over a finite state space $\mathcal{E}$, and with generator $\mathbf{Q}$. Then:*

$$\mathbf{P}_t \;=\; e^{t\mathbf{Q}} \tag{1.28}$$

and

$$\boldsymbol{\pi}_t \;=\; \boldsymbol{\pi}_0\mathbf{P}_t \;=\; \boldsymbol{\pi}_0 e^{t\mathbf{Q}}\ , \tag{1.29}$$

where $\boldsymbol{\pi}_0$ is the initial distribution.

This result is the analogous of Theorem 1.3 in the discrete-time case, and it makes it possible, in principle, to compute numerically the transient probabilities. This is the topic of Chapter 2.

According to this theorem, the matrix knowledge of the matrix $\mathbf{Q}$ is enough to determine the complete family of matrices $\{\mathbf{P}_t; t \in \mathbb{R}\}$. This is the reason for the name "generator" attached to it. A continuous-time Markov chain is characterized by its generator.

As in the discrete-time case, we associate a graph $\mathcal{G}$ to every infinitesimal generator $\mathbf{Q}$: this graph is called the **transition diagram**. The convention is not to represent the "loop" arcs going from state i to state i. An example is shown in Section 1.1.5.

Note that when the state space $\mathcal{E}$ is countably infinite, it is not straight forward to generalize these matrix formulas. In the discrete-time case, this is possible because matrices and vectors have only positive and bounded entries. Here, the generator matrix $\mathbf{Q}$ has negative entries in the diagonal, and possibly unbounded ones off the diagonal. Nevertheless, the formula $e^{t\mathbf{Q}}$ is often used as the "formal" solution to the differential equation (1.27). This does not provide a practical way to compute probabilities.

1.2.3 Asymptotic behavior

Let us turn to the study of the behavior of the matrix function $\mathbf{P}_t$ or the vector function $\boldsymbol{\pi}_t$ when $t \to \infty$.

In the case where $\mathcal{E}$ is finite, there are results similar to Theorems 1.4 and 1.6.

Theorem 1.14 (Existence of a stationary distribution)**.** *Assume* $\mathbf{Q}$ *is an irreducible infinitesimal generator. Then there exists a unique probability vector* $\boldsymbol{\pi}$ *such that* $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$ *and* $\boldsymbol{\pi}\mathbf{1} = 1$*. The vector* $\boldsymbol{\pi}$ *is strictly positive.*

This vector is called the vector of stationary probabilities of the Markov chain.

This theorem states implicitly the property that $\mathbf{Q}$ has 0 as an eigenvalue, with $\boldsymbol{\pi}$ as a left-eigenvector. We knew already from Theorem 1.12 *ii*) that 0 is an eigenvalue, with 1 as a right-eigenvector.

Theorem 1.15. *Let* $\mathbf{Q}$ *be an irreducible infinitesimal generator. Let* $\boldsymbol{\pi}$ *be its unique stationary probability vector. Then:*

$$\lim_{t\to\infty} e^{\mathbf{Q}t} \;=\; \mathbf{1}.\boldsymbol{\pi} \;. \tag{1.30}$$

As a consequence, for any probability vector $\boldsymbol{\pi}_0$*, we have:*

$$\lim_{t\to\infty} \boldsymbol{\pi}_t \;=\; \boldsymbol{\pi} \;. \tag{1.31}$$

The convergence is geometric: for any vector norm,

$$||\boldsymbol{\pi}_t - \boldsymbol{\pi}|| \;=\; \mathcal{O}(e^{\phi t}) \tag{1.32}$$

for all real number ϕ *such that* $\Re(\lambda_2) < \phi < 0$*, where* λ_2 *is the eigenvalue of* $\mathbf{Q}$ *with the largest real part after 0.*

In the case where $\mathcal{E}$ is countably infinite, there is a result similar to Theorem 1.10:

Theorem 1.16. *Let* $\mathbf{Q}$ *be an infinite, irreducible infinitesimal generator. Two possibilities exactly occur:*

- *either the linear system of equations* $\mathbf{x}\mathbf{Q} = \mathbf{0}$ *admits a positive solution* $\mathbf{x}$ *such that* $\sum_{i\in\mathcal{E}} x_i < \infty$*.*

  Then it admits a unique stationary probability vector $\boldsymbol{\pi}$*, and* $\boldsymbol{\pi}_t \to \boldsymbol{\pi}$ *for all initial distribution* $\boldsymbol{\pi}_0$*.*

- *or there does not exist solutions of* $\mathbf{x}\mathbf{Q} = \mathbf{0}$ *which are such that* $\sum_{i\in\mathcal{E}} |x_i| < \infty$*.*

  In that case, $\boldsymbol{\pi}_t \to \mathbf{0}$ *for all initial distribution* $\boldsymbol{\pi}_0$*.*

In summary, there are many similarities between DTMC and CTMC, with the matrices $\mathbf{P}$ and $\mathbf{Q}$ playing a central role. There are differences too: the convergence results for CTMC do not involve any aperiodicity condition.

1.2.4 Uniformization and other embeddings

This section is devoted to the links that can be established between a CTMC and some DTMC.

The general idea is to "observe" the CTMC at particular time instants. This results in a discrete-time process, defined on the same state space. When these intants are chosen properly, the resulting process is a DTMC. The transition probabilities of this Markov chain depend on the way observation points have been chosen: to one CTMC correspond possibly many *embedded* DTMCs. There are

relationships between the transient and stationary distributions of the original CTMC and the embedded DTMCs, which is useful for proofs, computations and simulations.

The converse transformation of a DTMC into continuous-time processes is also possible: this gives rise to *semi-Markov processes*. Their study is outside the scope of the present document.

1.2.4.1 Uniformization

One useful connection between Markov chains in continuous time and in discrete time is through the process of uniformization. Roughly speaking, uniformizing a CTMC is constructing a *coupling* between this chain and the combination of a DTMC and a Poisson process.

As a preparation, consider the following Lemma. The state space $\mathcal{E}$ is at most countable.

Lemma 1.17. *Let $\{X(t), t \in \mathbb{R}\}$ be a CTMC of generator $\mathbf{Q}$, and let ν be a number such that*

$$\nu \geq q_m := \sup_i \{|Q_{ii}|\} \ . \tag{1.33}$$

Then, consider the matrix $\mathbf{P}$:

$$\mathbf{P} \ = \ \mathbf{I} \ + \ \frac{1}{\nu} \, \mathbf{Q} \ .$$

This matrix $\mathbf{P}$:

i) is stochastic;

ii) is irreducible, if and only if $\mathbf{Q}$ is irreducible;

iii) is irreducible and aperiodic (that is: ergodic), if $\mathbf{Q}$ is irreducible and in addition $\nu > q_m$.

Proof. The off-diagonal elements of $\mathbf{P}$ are: $P_{ij} = Q_{ij}/\nu$. They are therefore nonnegative. The diagonal elements are $P_{ii} = 1 + Q_{ii}/\nu$, and by definition of q_m, these numbers are nonnegative as well. Finally, we have: $\mathbf{P1} = \mathbf{I1} + \frac{1}{\nu}\mathbf{Q1} = \mathbf{1}$ because $\mathbf{Q1} = \mathbf{0}$. This proves i).

To see point i), simply observe that the transition diagrams associated to $\mathbf{Q}$ and $\mathbf{P}$ have the same arcs, except for the possible "loops" (i, i) which do not affect the irreducibility property.

Finally, if $\nu > q_m$, then there exists some state i such that $P_{ii} > 0$, and this implies aperiodicity. □

Since every stochastic matrix is the transition matrix of a DTMC, we can state the definition.

Definition 1.10 (Uniformized Markov Chain)**.** *Consider a CTMC with generator $\mathbf{Q}$ and $\nu \geq q_m$ as in Lemma 1.17. Any DTMC with transition matrix $\mathbf{P} = \mathbf{I} + \frac{1}{\nu}\mathbf{Q}$ is called a **uniformized** chain of the CTMC.*

We should actually indicate the dependence of the matrix $\mathbf{P}$ on the value of ν, but as we see below, the properties of $\mathbf{P}$ turn out to be independent from this value.

The stochastic relationship between the two processes is made explicit in the following *coupling* theorem.

Theorem 1.18. *Let $\{T_n, n \in \mathbb{N}\}$ be a Poisson process with parameter ν, and $\{Y(n), n \in \mathbb{N}\}$ be a DTMC with transition matrix $\mathbf{P}$. Consider the process $\{X(t), t \in \mathbb{R}^+\}$ defined as:*

$$X(t) \;=\; Y(n) \qquad T_n \leq t < T_{n+1} \;.$$

Then this process is a CTMC with generator $\mathbf{Q}$.

Proof. Let us compute $P(t)_{i,j}$ for the process $X(t)$. Conditioning on the number of events of the Poisson process $\{T_n\}$ in the interval $[0, t[$, we get:

$$\begin{aligned}
\mathbb{P}(X(t) = j | X(0) = i) &= \sum_{k=0}^{\infty} \frac{(\nu t)^k}{k!}\; e^{-\nu t}\; \mathbb{P}(X(t) = j \mid X(0) = i, k \text{ jumps of } Y \text{ in } [0,t)) \\
&= \sum_{k=0}^{\infty} \frac{(\nu t)^k}{k!}\; e^{-\nu t}\; \mathbb{P}(Y(k) = j \mid X(0) = i) \\
&= \sum_{k=0}^{\infty} \frac{(\nu t)^k}{k!}\; e^{-\nu t}\; (\mathbf{P}^k)_{i,j} \;.
\end{aligned}$$

Therefore:

$$P(t) \;=\; \sum_{k=0}^{\infty} \frac{(\nu t)^k}{k!}\; e^{-\nu t}\; \mathbf{P}^k \;=\; \sum_{k=0}^{\infty} \frac{(\nu t \mathbf{P})^k}{k!}\; e^{-\nu t} \;=\; e^{\nu \mathbf{P} t}\; e^{-\nu t \mathbf{I}} \;=\; e^{\nu(\mathbf{P}-\mathbf{I})t} \;.$$

So we indeed have $P(t) = e^{\mathbf{Q}t}$. □

In other words: the CTMC $\{X(t)\}$ can be seen as changing of state at each instant $T(n)$ of a Poisson process with intensity ν, the transition matrix being $\mathbf{P}$. Since in general $P_{ii} \neq 0$, it may be actually that the process does not change of state.

The construction of Definition 1.10 is always possible when $\mathcal{E}$ is finite, and actually, many constructions are possible since many values of ν are possible. If $\mathcal{E}$ is infinite, the construction is possible only if $q_m < +\infty$. Such CTMC are called **uniformizable**.

We conclude this paragraph with the important property. Recall Theorems 1.4 and 1.14 about stationary distributions.

Theorem 1.19. *Consider the infinitesimal generator $\mathbf{Q}$ of some CTMC over a finite state space $\mathcal{E}$. Let $\mathbf{P}$ be the If $\mathbf{Q}$ is irreducible, then its unique stationary probability vector $\boldsymbol{\pi}$ is also the unique stationary probability vector of $\mathbf{P}$.*

1.2.4.2 Embedding of a DTMC into a CTMC

Uniformization is a particular case of an *embedded Markov chain.*[2]

Consider a continuous-time process $\{X(t); t \in \mathbb{R}\}$, and assume that $\{T(n); n \in \mathbb{N}\}$ is an increasing sequence of time instants. It is always possible to define the **embedded** process:

$$Y(n) \;=\; X(T(n)) \;.$$

[2] This topic is not part of the "basic" theory, but is mentioned here for completeness.

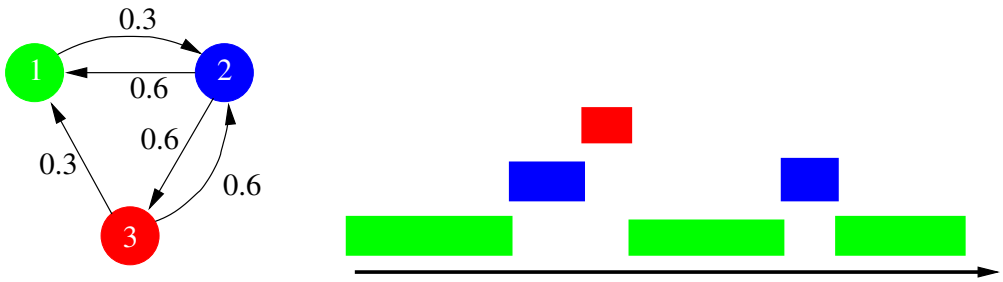

Figure 1.3: Transition diagram of a continuous-time Markov chain

This is a discrete-time process.

Under specific assumptions on the sequence $\{T(n); n \in \mathbb{N}\}$, it turns out that $\{Y(n); n \in \mathbb{N}\}$ is a CTMC. This chain is then called **embedded** into the chain $X(t)$.

Classical examples of this construction are:

- uniformized chains, as in the previous paragraph;
- the chain embedded at periodic instants $T(n) = n \times T_0$: its transition matrix is: $\mathbf{P} = e^{T_0 \mathbf{Q}}$;
- the chain embedded at the events of some independent Poisson process of rate λ: its transition matrix is $\mathbf{P} = (I - \frac{1}{\lambda}\mathbf{Q})^{-1}$;
- the chain embedded at jump times of $X(t)$;
- the chain embedded at times when $X(t)$ enters some sub-space $\mathcal{E}' \subset \mathcal{E}$.

In general, the embedded discrete-time Markov chains do not have the same properties (irreducibility, stationary distributions) as the CTMC in which they are embedded. The properties of uniformized chains (Theorem 1.19) is very particular in that respect.

1.2.5 An example

Consider the infinitesimal generator:

$$\mathbf{Q} = \begin{pmatrix} -0.3 & 0.3 & 0 \\ 0.6 & -1.2 & 0.6 \\ 0.3 & 0.6 & -0.9 \end{pmatrix} .$$

The corresponding transition diagram is represented in Figure 1.3, together with a possible evolution of the chain over time. It will be explained in Section 3.2.1 that the diagonal elements of $\mathbf{Q}$ are, up to the sign, the inverse of the average sojourn time in the respective states. Accordingly, the chain tends to stay longer periods in state 1 than it does in states 2 or 3.

If one chooses the value $\nu = 1.2$, the uniformized DTMC has the following transition matrix:

$$\mathbf{P} = \begin{pmatrix} 0.75 & 0.25 & 0 \\ 0.5 & 0 & 0.5 \\ 0.25 & 0.5 & 0.25 \end{pmatrix} .$$

If one chooses the value $\nu = 10$, the uniformized DTMC has the following transition matrix:

$$\mathbf{P} = \begin{pmatrix} 0.97 & 0.03 & 0 \\ 0.06 & 0.88 & 0.06 \\ 0.03 & 0.06 & 0.91 \end{pmatrix} .$$

In accordance with Theorem 1.19, both these matrices have the unique stationary probability vector

$$\boldsymbol{\pi} = \left(\frac{8}{13}, \frac{3}{13}, \frac{2}{13} \right) .$$

It can be checked that this is also the unique stationary probability vector of the matrix $\mathbf{Q}$.