Chapter 6

The Poisson process

Introduction

In this chapter and the two following, we will study Markov processes which are indexed by $\mathbb{R}_+$, with values in a finite or countable set E, and which are constant between their jumps, which happen at random times. These are called "jump Markov processes".

In this chapter, we shall introduce the "prototype" of jump Markov processes, namely the Poisson process.

This process models random distributions of points on $\mathbb{R}_+$, which could be times of collisions of particles, times of arrivals of customers in a queue, times of arrivals of telephone calls, etc...

6.1 Point Processes and counting processes

A point process on $\mathbb{R}_+$ can be described as an increasing sequence of random points

$$0 < T_1 < T_2 < \cdots < T_n < \cdots$$

which are r. v.'s defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In addition to the above inequalities, we assume that $T_n \uparrow \infty$, $n \to \infty$.

Let

$$\begin{aligned} S_1 &= T_1 \\ S_2 &= T_2 - T_1 \\ &\cdots\cdots\cdots \\ S_n &= T_n - T_{n-1} \\ &\cdots\cdots\cdots \end{aligned}$$

The T_n's are the times where some events happen, the S_n's are waiting times between successive events.

We define the random **counting function** $\{N_t, t \geq 0\}$ of the **point process** $\{T_n, n \in \mathbb{N}\}$ as follows :

$$\begin{aligned} N_t &= \sup\{n; T_n \leq t\} \\ &= \sum_{j \geq 1} \mathbf{1}_{\{T_j \leq t\}} \end{aligned}$$

N_t is the number of events which have happened before time t.

Note that $N_0 = 0$, since $T_1 > 0$ and for all $t > 0$, $N_t < \infty$ since $T_n \uparrow \infty$, $n \to \infty$.

For $0 \leq s < t$, $N_t - N_s$ is the number of events which have happened during the time interval $]s, t]$.

A typical trajectory of the process $\{N_t, t \geq 0\}$ is drawn on figure 6.1

Note that the trajectories of $\{N_t\}$ are *right continuous*.

The knowledge of $\{N_t,\ t \geq 0\}$ is equivalent to that of the sequence $\{T_n, n \in \mathbb{N}\}$, and we have the identities :

$$\{N_t \geq n\} = \{T_n \leq t\}$$

$$\{N_t = n\} = \{T_n \leq t < T_{n+1}\}$$

$$\{N_s < n \leq N_t\} = \{s < T_n \leq t\}$$

6.2 The Poisson process

Definition 6.2.1. *We will say that the point process* $\{T_n, n \in \mathbb{N}\}$ *or its counting function* $\{N_t, t \geq 0\}$ *is a* **Poisson process** *if* $\{N_t, t \geq 0\}$ *is a process with stationary independent increments, i.e. whenever*

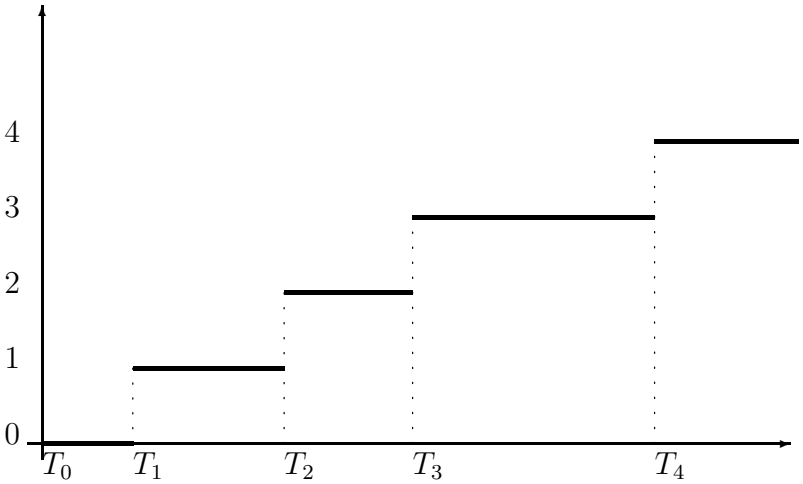

Figure 6.1: Trajectory of a Poisson process

a) for all $n \geq 2$, $0 \leq t_0 < t_1 < \cdots < t_n$, the increments $\{N_{t_j} - N_{t_{j-1}}; 1 \leq j \leq n\}$ are mutually independent;

b) for all $0 \leq s < t$, the law of $N_t - N_s$ depends upon the pair (s,t) only through the difference $t - s$.

Property *b*) is called the "stationarity of the increments " of $\{N_t\}$.
The terminology 'Poisson process" is justified by :

Proposition 6.2.2. *Let $\{N_t, t \geq 0\}$ be the counting function of a* **Poisson process**. *There exists $\lambda > 0$ such that for all $0 \leq s < t$, the law of $N_t - N_s$ is the Poisson distribution with parameter $\lambda(t-s)$, i.e.*

$$\mathbb{P}(N_t - N_s = k) = e^{-\lambda(t-s)}[\lambda(t-s)]^k/k\,!, \;\; k \in \mathbb{N}.$$

Remark 6.2.3. *The parameter λ is called the* **intensity** *of the Poisson process $\{N_t, t \geq 0\}$. It is equal to the mean number of events which happen during an time interval of unit length, i.e.*

$$\mathbb{E}[N_{t+1} - N_t] = \lambda.$$

PROOF OF PROPOSITION 6.2.2 For all $0 \leq s < t$, consider the generating function of the r. v. $N_t - N_s$, which is the mapping $u \to f_{t-s}(u)$ from $[0,1]$ into itself and is defined by :

$$\begin{aligned} f_{t-s}(u) &= \mathbb{E}[u^{N_t - N_s}] \\ &= \sum_{k \geq 0} \mathbb{P}(N_t - N_s = k) u^k. \end{aligned}$$

From property a) of the definition,

$$f_t(u) = f_s(u) f_{t-s}(u),\ 0 \leq s < t,\ \ u \in [0,1].$$

It follows from this identity that

$$f_t(u) = [f_1(u)]^t$$

first for t rational, then for all t in $\mathbb{R}_+$ since $t \to f_t(u)$ is decreasing.

Since moreover

$$\begin{aligned} f_t(u) &\geq P(N_t = 0) \\ &= P(T_1 > t) \\ &\nearrow 1, \text{ as } t \downarrow 0, \end{aligned}$$

$f_1(u) \neq 0$, hence there exists $\lambda(u) \in \mathbb{R}_+$ such that

$$f_t(u) = e^{-t\lambda(u)}.$$

Since $u \to \exp(-\theta(1-u))$ is the generating function of the Poisson distribution with parameter θ, it just remains to show that

$$\lambda(u) = \lambda(0)(1-u).$$

But clearly

$$\begin{aligned} \lambda(u) &= \lim_{t \downarrow 0} \frac{1}{t}(1 - f_t(u)) \\ &= \lim_{t \downarrow 0} \sum_{k \geq 1} \frac{1}{t} \mathbb{P}(N_t = k)(1 - u^k) \end{aligned}$$

Since $0 \leq u \leq 1$,

$$0 \leq \sum_{k\geq 2} \frac{1}{t}\mathbb{P}(N_t = k)(1-u^k) \leq \frac{1}{t}\mathbb{P}(N_t \geq 2)$$

and the result follows from the identity

$$\lambda(u) = \lim_{t\downarrow 0}\left[\frac{1}{t}\mathbb{P}(N_t = 1)\right](1-u),$$

provided we have

$$\frac{1}{t}\mathbb{P}(N_t \geq 2) \to 0, \text{ as } t \downarrow 0. \tag{6.1}$$

But

$$\bigcup_{n\in\mathbb{N}} \{N_{nt} = 0,\ N_{(n+1)t} \geq 2\} \subset \{T_2 < T_1 + t\}$$

Since $\mathbb{P}(N_t = 0) = f_t(0) = \exp(-\lambda(0)t)$, we deduce from this inclusion and property a) from the definition that :

$$\begin{aligned}\sum_{n\in\mathbb{N}} \exp(-\lambda(0)nt)\mathbb{P}(N_t \geq 2) &= [1-\exp(-\lambda(0)t)]^{-1}\mathbb{P}(N_t \geq 2)\\ &\leq \mathbb{P}(T_2 < T_1 + t)\end{aligned}$$

As $t \downarrow 0$,

$$P(T_2 < T_1 + t) \to P(T_2 \leq T_1) = 0$$

and for all t sufficiently small,

$$(\lambda(0)t)^{-1} < (1-\exp(-\lambda(0)t))^{-1},$$

hence (6.1) is established.

Remark 6.2.4. *We can give an intuitive interpretation of the preceding result. From the last part of the above proof,*

$$\begin{aligned}\mathbb{P}(N_{t+h} - N_t = 0) &= 1 - \lambda h + \circ(h)\\ \mathbb{P}(N_{t+h} - N_t = 1) &= \lambda h + \circ(h)\\ \mathbb{P}(N_{t+h} - N_t \geq 2) &= \circ(h)\end{aligned}$$

Then up to probabilities which are small compared with h, $N(t+h) - N(t)$ is a Bernoulli r. v. taking the value 0 *with probability* $1-\lambda h$ *and the value* 1

with probability λh. This, together with the independence of the increments and the formula

$$N_{t+s} - N_t = \sum_{j=1}^{n} [N_{t+jh} - N_{t+(j-1)h}], \text{ with } h = \frac{s}{n},$$

implies that $N_{t+s} - N_t$ is approximatively a binomial r. v. with parameters $(n, \lambda s/n)$. But as $n \to \infty$, that law converges towards Poisson with parameter λs.

Note that for all $n \geq 2$, $0 < t_1 < t_2 < \cdots < t_n$, the law of the random vector $(N_{t_1}, N_{t_2} \ldots, N_{t_n})$ is determined by Proposition 6.2.2 and condition a) from Definition 6.2.1.

Corollary 6.2.5. *The law of the time T_1 of the first event is exponential with parameter λ (i.e. the law on $\mathbb{R}_+$ with density $\lambda e^{-\lambda t}$). The same is true for the law of $T_{N_s+1} - s$, which is the waiting time after s of the next event, for all $s > 0$.*

Proof It suffices to note that for $t > 0$,

$$\begin{aligned} \mathbb{P}(T_1 > t) &= \mathbb{P}(N_t = 0) \\ &= e^{-\lambda t} \end{aligned}$$

and similarly

$$\begin{aligned} \mathbb{P}(T_{N_s+1} - s > t) &= \mathbb{P}(N_{s+t} - N_s = 0) \\ &= P(N_t = 0). \end{aligned}$$

□

6.3 The Markov property

Let $\{N_t, t \geq 0\}$ be a Poisson process with intensity λ. For all $s, t > 0$, let

$$N_t^s = N_{s+t} - N_s.$$

It follows from Definition 6.2.1 that $\{N_t^s, t \geq 0\}$ is a Poisson process with intensity λ, independent of $\{N_r, 0 \leq r \leq s\}$. Note that the knowledge

of $\{N_t, 0 \leq t \leq s\}$ is equivalent to that of $(N_s, T_1, T_2, \ldots, T_{N_s})$. The above independence is equivalent to that of the random vectors $(N_s, T_1, T_2, \ldots, T_{N_s})$ and $(T_{N_s+1}, T_{N_s+2}, \ldots, T_{N_s+p})$, for all p.

Since the increments $\{N_{s+t} - N_s, t \geq 0\}$ after s are independent from the past $\{N_t, 0 \leq t \leq s\}$, clearly the future $\{N_{s+t}, t \geq 0\}$ after s depends upon the past $\{N_t, 0 \leq t \leq s\}$ only through the present value N_s; in other words the past and the future are conditionally independent, given the present. This is the Markov property.

We shall discuss this property again in the next chapter.

We shall now generalize the above property to the case where s is a certain type of random time. Let us first recall some notation, and state a definition.

A σ–algebra of subsets of a set $\mathcal{E}$ is a class of subsets of $\mathcal{E}$ which is closed under complement, countable unions and intersections. One can always speak of the "smallest σ–algebra containing the class $\mathcal{C} \subset \mathcal{P}(\mathcal{E})$", since it is the intersection of all σ–algebras containing $\mathcal{C}$ (there exists at least one such σ algebra, namely $\mathcal{P}(\mathcal{E})$, and an arbitrary intersection of σ–algebras is a σ–algebra, as can be easily checked). For example, the Borel σ–algebra of subsets of $\mathbb{R}^d$, denoted $\mathcal{B}_d$, is the smallest σ–algebra of subsets of $\mathbb{R}^d$ containing all the open sets.

In the case of a random variable with values in a countable state space E, $\sigma(X) = \{X^{-1}(F);\ F \subset E\}$. Given a d–dimensional random vector X (i.e. an $\mathbb{R}^d$–valued random variable), we denote by $\sigma(X) = \{X^{-1}(B); B \in \mathcal{B}_d\}$ the smallest σ–algebra of subsets of Ω which makes X measurable. This is the set of events for which we know whether or not they are realized as soon as we know the value of X. Given an arbitrary collection $\{X_i, i \in I\}$ (of arbitrary dimensions), we denote by $\sigma\{X_i; i \in I\}$ the smallest σ–algebra containing $\sigma(X_i)$, for all $i \in I$.

It will be convenient to use below the following notation : for any $t \geq 0$,

$$\begin{aligned} \mathcal{F}_t^N &= \sigma\{N_s; 0 \leq s \leq t\} \\ &= \sigma\{N_t, T_1, T_2, \ldots, T_{N_t}\}. \end{aligned}$$

Definition 6.3.1. *Given a Poisson process $\{N_t, t \geq 0\}$, a* **stopping time** *(associated to $\{N_t\}$) is an $\mathbb{R}_+ \cup \{+\infty\}$–valued r. v. S such that for all t in $\mathbb{R}_+$,*

$$\{S \leq t\} \in \mathcal{F}_t^N.$$

For all s in $\mathbb{R}_+$, $S \equiv s$ is a stopping time. For all n, T_n is a stopping time. T_{N_s+1} is also a stopping time. But T_{N_s} is not a stopping time, since

whenever $t < s$,

$$\{T_{N_s} \leq t\} = \{N_s - N_t = 0\} \notin \mathcal{F}_t^N, \ 0 \leq t < s.$$

To any stopping time S associated to $\{N_t\}$, we associate the σ–algebra of those events which are "determined by the trajectory $\{N_{t\wedge S},\, t \geq 0\}$ stopped at time S" :

$$\mathcal{F}_S^N \stackrel{\text{def}}{=} \{A \in \mathcal{F}_\infty^N;\ A \cap \{S \leq t\} \in \mathcal{F}_t^N,\ \forall t \geq 0\}.$$

We have :

Proposition 6.3.2. *Let $\{N_t, t \geq 0\}$ be a Poisson process with intensity λ, and S a stopping time associated to $\{N_t\}$. On the event $\{S < \infty\}$ we define for $t \geq 0$*

$$N_t^S = N_{S+t} - N_S.$$

Conditionally upon $\{S < \infty\}$, $\{N_t^S, t \geq 0\}$ is a Poisson process with intensity λ, independent of $\mathcal{F}_S^N$.

Proof We already know that the result holds if S is constant. Suppose next that S takes its values in an increasing sequence $(s_j, j \geq 1)$ of positive real numbers. Note that since S is a stopping time,

$$\{S = s_j\} = \{S \leq s_j\} \backslash \{S \leq s_{j-1}\} \in \mathcal{F}_{s_j}^N$$

Let $A \in \mathcal{F}_S^N$, $0 < t_1 < t_2 < \ldots < t_\ell$ and $n_1, \ldots, n_\ell$ belong to $\mathbb{N}$.

$$\begin{aligned}
&\mathbb{P}\left(A \cap (\bigcap_{k=1}^{\ell} \{N_{t_k}^S = n_k\})\right) \\
&= \sum_j \mathbb{P}\left(\{S = s_j\} \cap A \cap \left(\bigcap_{k=1}^{\ell} \{N_{s_j+t_k} - N_{s_j} = n_k\}\right)\right) \\
&= \sum_j \mathbb{P}(\{S = s_j\} \cap A)\mathbb{P}\left(\bigcap_{k=1}^{\ell} \{N_{s_j+t_k} - N_{s_j} = n_k\}\right) \\
&= \mathbb{P}(A)\mathbb{P}\left(\bigcap_{k=1}^{\ell} \{N_{t_k} = n_k\}\right),
\end{aligned}$$

where we have used the property $\{S = s_j\} \cap A \in \mathcal{F}^N_{s_j}$ for the second equality, and the fact that the second factor of the for–to–last expression does not depend upon s_j, by the stationarity of the increments of $\{N_t\}$.

The result is thus established in the case of a stopping time which takes its values in an increasing sequence. But any stopping time S can be approximated by a decreasing sequence of stopping times of this form. Indeed for all n, define

$$S_n = \sum_{k \in \mathbb{N}} k 2^{-n} \mathbf{1}_{\{(k-1)2^{-n} < S \leq k2^{-n}\}}.$$

The above identity is true with S replaced by S_n, since

$$S \leq S_n \Rightarrow \mathcal{F}^N_S \subset \mathcal{F}^N_{S_n}.$$

We can now easily take the limit in the above identity, with S replaced by S_n since from the right continuity of the trajectories of $\{N_t, t \geq 0\}$,

$$\mathbb{P}\left(A \cap (\bigcap_{k=1}^{\ell} \{N^{S_n}_{t_k} = n_k\})\right) \to \mathbb{P}\left(A \cap (\bigcap_{k=1}^{\ell} \{N^{S}_{t_k} = n_k\})\right).$$

□

Corollary 6.3.3. *Let $\{N_t; t \geq 0\}$ be a Poisson process with intensity λ, and $(T_n)_{n \geq 1}$ its jump times. We let $S_1 = T_1, S_2 = T_2 - T_1, \ldots, S_n = T_n - T_{n-1}, \ldots$. The random variables $S_1, S_2, \ldots, S_n, \ldots$ are i. i. d., their common law being exponential with the parameter λ.*

Proof We already know that the law of T_1, the first jump time of a Poisson process with intensity λ, is exponential with parameter λ. It follows from Proposition 6.3.2 with $S = T_n$ that $S_{n+1} = T_{n+1} - T_n$ is the first jump time of a Poisson process with intensity λ, hence its law is exponential with parameter λ, and is independent of $T_1, T_2, \ldots, T_n$, hence also of $S_1, S_2, \ldots, S_n$. The result follows from the fact that this statement is true for all $n \geq 1$. □

In the other direction, we have the following result :

Proposition 6.3.4. *Let $\{S_n; n \geq 1\}$ be a sequence of i. i. d. r. v.'s, their common law being the exponential law with parameter $\lambda > 0$. We define*

$$T_n = S_1 + \cdots + S_n, \ n \geq 1,$$

$$N_t = \sup\{n; T_n \leq t\}, \ t \geq 0.$$

Then $\{N_t, t \geq 0\}$ *is a Poisson process with intensity* λ.

We then have a way to "construct" a Poisson process, which in particular shows that there does exist a process satisfying the requirements of Definition 6.2.1 ! We also have here a way to *simulate* a Poisson process.

6.4 Large time behaviour

Again, let $\{N_t; t \geq 0\}$ be a Poisson process with intensity λ. Then

$$\mathbb{E}[N_t] = \lambda t, \ Var[N_t] = \lambda t.$$

In particular $\mathbb{E}[N_t/t] = \lambda$, $\mathrm{Var}[N_t/t] = \frac{\lambda}{t}$, hence $N(t)/t \to \lambda$ in mean square, as $t \to \infty$. In fact we have the "strong law of large numbers" :

Proposition 6.4.1. *Let* $\{N_t; t \geq 0\}$ *be a Poisson process with intensity* $\lambda > 0$. *Then* $\dfrac{N_t}{t} \to \lambda$ *a. s. as* $t \to \infty$.

Proof First note that

$$N_n = \sum_{1 \leq i \leq n} [N_i - N_{i-1}]$$

is the sum of n i. i. d. r. v.'s, their common law being Poisson with parameter λ (hence they are integrable). It then follows from the strong law of large numbers that

$$\frac{N_n}{n} \to \lambda, \quad \text{a. s., as } n \to \infty.$$

But with the notation $[t]$ = integer part of t,

$$\frac{N_t}{t} = \frac{N_{[t]}}{[t]} \times \frac{[t]}{t} + \frac{N_t - N_{[t]}}{t}.$$

It then suffices to show that

$$\sup_{n<t<n+1} \frac{N_t - N_n}{n} \to 0, \text{ as } n \to \infty.$$

Let

$$\begin{aligned}\xi_n &\stackrel{\text{def}}{=} \sup_{n<t\leq n+1} N_t - N_n,\\ &= N_{n+1} - N_n.\end{aligned}$$

The $\{\xi_n\}$'s are i. i. d. and integrable. Then $\frac{\xi_1 + \cdots + \xi_n}{n} \to \lambda$ a. s., consequently

$$\frac{\xi_n}{n} \to 0 \text{ a. s.}$$

□

We have the following "central limit theorem" :

Proposition 6.4.2. *Let $\{N_t; t \geq 0\}$ be a Poisson process with intensity λ. Then*

$$\frac{N_t - \lambda t}{\sqrt{\lambda t}} \to Z \text{ in law, as } t \to \infty,$$

where Z is a centered Gaussian r. v. with unit variance.

Proof We essentially argue as in the preceding proof.

$$\frac{N_n - \lambda n}{\sqrt{\lambda n}} \to Z \text{ in law, as } n \to \infty,$$

from the "classical" central limit theorem. And

$$\frac{N_t - N_{[t]}}{\sqrt{\lambda [t]}} \leq \xi_{[t]} / \sqrt{\lambda [t]},$$

converges to 0 in probability as $t \to \infty$ since

$$\begin{aligned}P\left(\xi_n/\sqrt{\lambda n} > \varepsilon\right) &= P\left(\xi_n > \varepsilon\sqrt{\lambda n}\right)\\ &= P\left(\xi_1 > \varepsilon\sqrt{\lambda n}\right)\\ &\to 0, \text{ as } n \to \infty.\end{aligned}$$

Then also $\frac{N_t - N_{[t]}}{\sqrt{\lambda[t]}} \to 0$ in probability as $t \to \infty$. Finally :

$$\frac{N_t - \lambda t}{\sqrt{\lambda t}} = \frac{N_{[t]} - \lambda[t]}{\sqrt{\lambda[t]}} \times \sqrt{\frac{[t]}{t}} + \frac{N_t - N_{[t]}}{\sqrt{\lambda t}} \times \sqrt{\frac{[t]}{t}} + \sqrt{\lambda}\frac{[t] - t}{\sqrt{t}},$$

and we know that whenever $X_n \to X$ in law, $Y_n \to 0$ in probability, then

$$X_n + Y_n \to X \text{ in law.}$$

□

One could in fact establish a "functional central limit theorem" which we now briefly describe. A proof very similar to that of Proposition 6.4.2 shows that for all $t > 0$,

$$\frac{N_{tu} - \lambda tu}{\sqrt{\lambda u}} \to B_t \text{ in law as } u \to \infty,$$

where B_t is a centered Gaussian r. v. with variance t. Note that for each $u > 0$, $\{[N_{tu} - \lambda tu]/\sqrt{\lambda u},\ t \geq 0\}$ is a process with independent increments, whose jumps are of size $(\lambda u)^{-1/2}$. It follows that one can take the above limit as $u \to \infty$ jointly for the various t's, in such a way that the limit $\{B_t, t \geq 0\}$ is a centered Gaussian process with independent increments and continuous trajectories, satisfying $E[B_t^2] = t$. $\{B_t; t \geq 0\}$ is called a *Brownian motion*, which will be discussed in chapter 9 below.

6.5 Exercises

Exercise 6.5.1. *Let X be an $\mathbb{R}_+$–valued r. v. such that $\mathbb{P}(X > t) > 0,\ \forall t > 0$. We assume moreover that $\forall s, t > 0$,*

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

Conclude that the law of X is exponential with some parameter $\lambda > 0$.

Exercise 6.5.2. *Three persons A, B and C get to a post office at the same time. They want to make a telephone call. There are two telephone booths, which are immediately occupied by A and B. C replaces the one who finishes first. A, B and C leave the Post Office as soon as they have completed their calls.*

We denote by X, Y and Z the length of the telephone calls made by A, B and C respectively. These three r. v.'s are assumed to be i. i. d., their common law being exponential with parameter $\lambda > 0$.

1. *Compute the probability that C leaves last.*

2. *Give the probability distribution of the total time T spent by C in the Post Office.*

3. *With 0 being the time of arrival of the three persons at the Post Office, give the probability distribution of the time of the last departure.*

Hint : first give the probability distribution of the random vector $(X \wedge Y, X \vee Y - X \wedge Y)$, $(\wedge = \inf,\ \vee = \sup)$.

Exercise 6.5.3. *A machine has a lifespan τ_1 whose law is exponential with parameter θ. As soon as it breaks down, it is instantly replaced by an identical machine with lifespan τ_2, and so on. We assume that the r. v.'s $(\tau_n;\, n \in \mathbb{N})$ are i. i. d. The first machine starts running at time 0; the times T_n $(n \geq 1)$ of successive failure of the machines (in other words $T_1 = \tau_1$, $T_2 = \tau_1 + \tau_2, \ldots$) constitute a Poisson point process.*

1. *Given $t > 0$, let D_t denote the elapsed time since the machine running at time t started to run. In which set does the r. v. D_t take its values ? What is the law of D_t ? Show that as $t \to \infty$, this law has a limit.*

2. *Let S_t be such that $t + S_t$ is the time of failure of the machine which functions at time t. What is the law of S_t ? What is the law of the pair (D_t, S_t) and what is the limit of that law as $t \to \infty$? Why aren't D_t and S_t identically distributed, and why do they tend to be identically distributed as $t \to \infty$?*

3. *What is the law of $D_t + S_t$, the lifespan of the machine which is running at time t ? Compare the limit of that law as $t \to \infty$ with the joint law of the τ_n's. Explain the apparent contradiction.*

Exercise 6.5.4. *1. Let $X_1, X_2, \ldots, X_n$ be i. i. d. r. v.'s, whose common law is the uniform law on $[0,t]$, and $Y_1, Y_2, \ldots, Y_n$ be the same sequence, but in increasing order. In other words, the Y_k's are defined by*

$$Y_1 = \inf_{1\leq i\leq n} X_i = X_{i_1}$$
$$Y_2 = \inf_{1\leq i\leq n, i\neq i_1} X_i$$

and so on. Give the probability distribution of the random vector $(Y_1, Y_2, \ldots, Y_n)$.

2. Let $\{N_t, t \geq 0\}$ be a Poisson process with intensity λ. Show that the conditional law of the random vector $(T_1, T_2, \ldots, T_n)$, given that $N_t = n$, is the distribution which was identified in the previous question.

Exercise 6.5.5. *Let $\{N_t^1; t \geq 0\}$ and $\{N_t^2; t \geq 0\}$ be two independent Poisson processes, of intensity λ_1 and λ_2 respectively. Show that $\{N_t^1 + N_t^2;\ t \geq 0\}$ is a Poisson process with intensity $\lambda_1 + \lambda_2$.*

Exercise 6.5.6. *Suppose that the number of individuals infected by HIV follows a Poisson process of a given intensity λ. We denote by N_t the number of individuals infected at time t. We do not take deaths into account.*

Each infected individual has an incubation period between the time of infection and the time when the symptoms of AIDS appear. The length of this incubation period is random. The incubation periods for various individuals are i. i. d., their common law on $\mathbb{R}_+$ having a given distribution function G. We denote by $\bar{G}$ the function $\bar{G}(t) = 1 - G(t)$.

Let N_t^1 denote the number of individuals who have AIDS symptoms at time t, and let N_t^2 denote the number of individuals who at time t are infected by HIV, but do not yet have AIDS symptoms. Of course

$$N_t = N_t^1 + N_t^2.$$

Show that for each $t > 0$, N_t^1 and N_t^2 are independent, the probability distribution of N_t^1 is Poisson with parameter $\lambda \int_0^t G(s)ds$, and that of N_t^2 Poisson with parameter $\lambda \int_0^t \bar{G}(s)ds$. You can make use of the result from the exercise 6.5.4, which says that conditionally upon $N_t = n$, the times of infection between 0 and t which are counted by a Poisson process have the same law as an i. i. d. sequence of n uniform $[0,t]$ random variables.

Exercise 6.5.7 (Programming). *Define the failure rate of an $\mathbb{R}_+$–valued r. v. X with the density f and the distribution function F to be the function $\lambda : \mathbb{R}_+ \to \mathbb{R}_+$ defined by*

$$\lambda(t) = \frac{f(t)}{1 - F(t)}.$$

Exercise 6.5.1 proves that the only probability distribution with a constant failure rate is the exponential.

The Weibull distribution with parameters $\lambda, \alpha > 0$ is the distribution on $\mathbb{R}_+$ with survival function

$$\bar{F}(t) = 1 - F(t) = e^{-(\lambda t)^\alpha},$$

and failure rate

$$\lambda(t) = \alpha\lambda(\lambda t)^{\alpha-1}.$$

The Weibull distribution has an increasing failure rate if $\alpha > 1$ and a decreasing rate if $\alpha < 1$. It reduces to the exponential distribution with parameter λ when $\alpha = 1$.

The $\Gamma(\alpha, \lambda)$ distribution is the distribution on $\mathbb{R}_+$ with density

$$f(t) = \frac{\lambda}{\Gamma(\alpha)} e^{-\lambda t}(\lambda t)^{\alpha-1},$$

where $\Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha-1} dt$.

Again the Gamma distribution has an increasing failure rate if $\alpha > 1$ and a decreasing rate if $\alpha < 1$. Note that the sum of n i. i. d. exponential (λ) r. v.'s, follows the $\Gamma(n, \lambda)$ distribution, with an increasing failure rate.

Suppose that two machines function in parallel, and need a fragile part M. Suppose that we have only one spare part, which immediately replaces the one which breaks down first. The three parts M (the two in place at the beginning, plus the spare) have i. i. d. life times. The second failure is fatal for the machine which experiences it. If the life times are exponential, then exercise 6.5.2 shows that the two machines have the same probability of experiencing the fatal failure.

Suppose that we replace the exponential law by a law with an increasing failure rate, then the machine whose part has been already replaced has a better chance of functioning longuer than the other one, and the reverse is true in the case of a decreasing failure rate.

Illustrate by a Monte Carlo computation the result from exercise 6.5.2, and the two conjectures which we have just formulated.

More precisely, successively with $\mathbb{P}$ = *exponential with parameter 1, the* $\Gamma(3,1)$ *distribution, the Weibull (1,0.5) distribution (easy to simulate by inversion of its distribution function), simulate a matrix* $3 \times N$ *of i. i. d. r. v.'s with the law* $\mathbb{P}$*, denoted by* X*. Graph, for* n *from* 1 *to* N*, the three quantities*

$$n^{-1} \sum_{k=1}^{n} \{min[X(1,k), X(2,k)] + X(3,k) - max[X(1,k), X(2,k)]\}.$$

You can choose $N = 10^3$ *or* $N = 10^4$*.*

Chapter 7

Jump Markov processes

Introduction

In this chapter, we will present the theory of continuous time jump Markov processes, with values in a finite or countable state space E. As we will see in section 7.4, these processes are in a sense combinations of a Poisson process and a discrete time Markov chain (the "embedded chain"). In a second part of this chapter, we will present applications to phylogeny, to discretized partial differential equations, and to the annealing algorithm. The proof of convergence of the annealed algorithm which we present here is due to Francis Comets (private communication). Applications to queues will be developed in the next chapter.

7.1 General facts

We want to study continuous time Markov processes with values in a finite or countable state space E. We will assume that their trajectories are constant between their jumps, and that the latter are isolated. Moreover we will assume that the trajectories are right continuous. They will have left limits at each point. The jumps of such a process $\{X_t,\, t \geq 0\}$ happen at random times $T_1(\omega), T_2(\omega), \ldots, T_n(\omega), \ldots$. The main difference with the Poisson process of the preceding chapter is that, given the time of the jump and the position before the jump, the position after the jump is random. If we denote by $Z_n(\omega)$ the value of $\{X_t\}$ just after the n–th jump $T_n(\omega)$, $n \geq 1$, a typical trajectory of the process $\{X_t;\, t \geq 0\}$ is graphed on the figure 7.1.

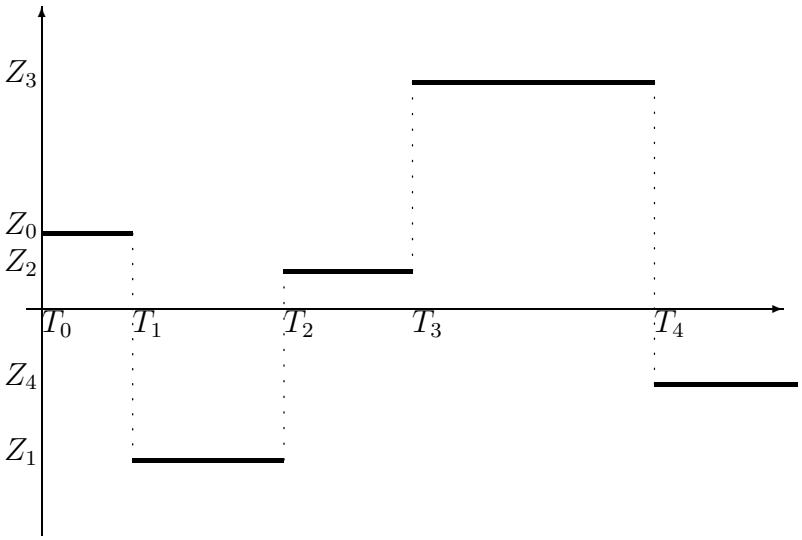

Figure 7.1: Trajectory of a continuous time jump Markov process

The knowledge of $\{X_t;\, t \geq 0\}$ is equivalent to that of the double sequence $\{T_n, Z_n;\, n \geq 0\}$.

For certain applications, it is convenient to make certain states absorbing (for instance, in a model describing the evolution of the size of a population without immigration, 0 is an absorbing state). $x \in E$ is absorbing if $X_{T_n}(\omega) = x \Rightarrow T_{n+1}(\omega) = +\infty$.

We will then assume that the jump times constitute an increasing sequence

$$0 = T_0 < T_1 \leq T_2 \leq \cdots \leq T_n \leq \cdots \tag{7.1}$$

with $T_n \in \mathbb{R}_+ \cup \{+\infty\}$, and

$$T_n(\omega) < T_{n+1}(\omega) \text{ whenever } T_n(\omega) < \infty. \tag{7.2}$$

We assume moreover that there is no explosion, i. e. jump times do not accumulate at finite distance, in other words

$$T_n(\omega) \to +\infty \text{ p. s.}, \quad \text{as } n \to \infty. \tag{7.3}$$

In the sequel, (7.1), (7.2) and (7.3) will be assumed to be in force, without any further notice.

An E–valued random function $\{X_t;\ t \geq 0\}$ is called a random jump function if it is of the form :

$$X_t(\omega) = \sum_{\{n\geq 0;\ T_n(\omega)<\infty\}} Z_n(\omega)\mathbf{1}_{[T_n(\omega),T_{n+1}(\omega)[}(t)$$

where the r. v. Z_n's take their values in E. We state

Definition 7.1.1. *An E–valued random jump function $\{X_t,\ t \geq 0\}$ is called a* **jump Markov proces** *(or a continuous time Markov chain) if for all $0 < s < t$, the conditional law of the r. v. X_t given $\{X_u;\ 0 \leq u \leq s\}$ depends upon X_s only , i. e. for all $n \in \mathbb{N}$, $0 \leq t_0 < t_1 < \cdots < t_n < s$, $x_0, x_1, \ldots, x_n,\ x, y \in E$,*
$\mathbb{P}(X_t = y | X_{t_0} = x_0, X_{t_1} = x_1, \ldots, X_{t_n} = x_n,\ X_s = x) = \mathbb{P}(X_t = y | X_s = x)$[1].

We shall say that the jump Markov process $\{X_t, t \geq 0\}$ is **homogeneous** *if the quantity $P(X_t = y | X_s = x)$ depends upon s and t only through the difference $t - s$.*

We shall limit ourselves to the study of homogeneous Markov processes. We will make use of the notation (in the case $s < t$) :

$$\mathbb{P}(X_t = y | X_s = x) = P_{xy}(t - s)$$

where for all $t > 0$, $P(t)$ is a "Markovian matrix" on $E \times E$, which is called the transition matrix during the time t. We shall denote below by $\mu(t)$ the probability distribution of X_t on E, $t \geq 0$. $\mu(0)$ is called the "initial law" of the process $\{X_t;\ t \geq 0\}$.

Proposition 7.1.2. *Let $\{X_t,\ t \geq 0\}$ be a jump Markov process, with the initial law μ and the transition matrices $\{P(t),\ t > 0\}$. For all n in $\mathbb{N}$, $0 < t_1 < \cdots < t_n$, the law of the random vector $(X_0, X_{t_1}, \ldots, X_{t_n})$ is given by : for all $x_0, x_1, \ldots, x_n$ in E,*

$$\begin{aligned}&\mathbb{P}(X_0 = x_0, X_{t_1} = x_1, X_{t_2} = x_2, \ldots, X_{t_n} = x_n)\\ &= \mu_{x_0} P_{x_0 x_1}(t_1) P_{x_1 x_2}(t_2 - t_1) \times \cdots \times P_{x_{n-1} x_n}(t_n - t_{n-1}).\end{aligned}$$

[1]This condition makes sense only when

$$\mathbb{P}(X(t_0) = x_0,\ X(t_1) = x_1, \ldots, X(t_n) = x_n,\ X(s) = x) > 0.$$

In this condition we shall disregard the values $n, x_0, x_1, \ldots, x_n, x$ for which that inequality does not hold

Consequently, for all $t > 0$,

$$\mu(t) = \mu(0)P(t)$$

in the sense that $\mu_y(t) = \sum_{x \in E} \mu_x(0) P_{xy}(t)$, *and for any positive or bounded function* $g : E \to \mathbb{R}$,

$$\begin{aligned} \mathbb{E}[g(X_t)|X_0 = x] &= (P(t)g)_x \\ &= \sum_{y \in E} P_{xy}(t) g_y. \end{aligned}$$

Moreover, the transition matrices $\{P(t), t > 0\}$ *satisfy the semigroup property (also called the Chapmann–Kolmogorov equation) :*

$$P(s+t) = P(s)P(t),$$

in the sense that for all x, y *in* E

$$P_{xy}(t+s) = \sum_{z \in E} P_{xz}(t) P_{zy}(s)$$

Proof It follows from the definition of conditional probability and the Markov property that

$$\begin{aligned} &\mathbb{P}(X_0 = x_0, X_{t_1} = x_1, \ldots, X_{t_n} = x_n) \\ &= \mathbb{P}(X_0 = x_0) P(X_{t_1} = x_1 | X_0 = x_0) \mathbb{P}(X_{t_2} = x_2 | X_0 = x_0, X_{t_1} = x_1) \\ &\times \cdots \times \mathbb{P}(X_{t_n} = x_n | X_0 = x_0, X_{t_1} = x_1, \ldots, X_{t_{n-1}} = x_{n-1}) \\ &= \mu_{x_0} P_{x_0 x_1}(t_1) P_{x_1 x_2}(t_2 - t_1) \times \cdots \times P_{x_{n-1} x_n}(t_n - t_{n-1}). \end{aligned}$$

In the case $n = 1$, this formula reduces to :

$$\mathbb{P}(X_0 = x,\ X_t = y) = \mu_x P_{xy}(t)$$

and the second result follows by summing over $x \in E$. From the definition of $P(t)$,

$$\mathbb{P}(X_t = y | X_0 = x) = P_{xy}(t),$$

the third result follows by multiplying by g_y and summing over $y \in E$.

Finally the above formula in the case $n = 2$ gives, after division by μ_{x_0}

$$\mathbb{P}(X_s = z,\ X_{s+t} = y | X_0 = x) = P_{xz}(s) P_{zy}(t).$$

The last result follows by summing over $z \in E$. □

We now present some examples of jump Markov processes.

Example 7.1.3. *A Poisson processes $\{N_t;\ t \geq 0\}$ of intensity λ is an $\mathbb{N}$–valued Markov process, with the transition matrix :*

$$P_{xy}(t) = \begin{cases} e^{-\lambda t}(\lambda t)^{y-x}/(y-x)!, & \text{if } y \geq x; \\ 0, & \text{otherwise.} \end{cases}$$

Example 7.1.4. The telegraph process *Given a Poisson process $\{N_t\}$ of intensity λ, and an $E = \{-1, +1\}$–valued r. v. X_0, independent of $\{N_t;\ t \geq 0\}$, we define :*

$$X_t = X_0(-1)^{N_t},\ t \geq 0.$$

$\{X_t,\ t \geq 0\}$ is a Markov process, with the transition matrix :

$$P_{+1+1}(t) = P_{-1-1}(t) = e^{-\lambda t} \sum_{n \geq 0} \frac{(\lambda t)^{2n}}{(2n)!}$$

$$P_{-1+1}(t) = P_{+1-1}(t) = e^{-\lambda t} \sum_{n \geq 0} \frac{(\lambda t)^{2n+1}}{(2n+1)!}$$

Example 7.1.5. *Let $\{N_t;\ t \geq 0\}$ be a Poisson process with the intensity λ, and jump times $0 < T_1 < T_2 < T_3 < \cdots < T_n < \cdots$. Let also $\{Z_n;\ n \in \mathbb{N}\}$ be an E–valued discrete time Markov chain, with transition matrix $\{P_{xy};\ x, y \in E\}$, independent of $\{N_t,\ t \geq 0\}$. One can show (see exercise 7.11.1 below) that*

$$X_t = \sum_{n=0}^{\infty} Z_n \mathbf{1}_{[T_n, T_n+1[}(t),\ t \geq 0$$

is a jump Markov process.

7.2 Infinitesimal generator

From the semigroup property it follows that $P(t)$ is known for all $t > 0$ as soon as it is known for all small enough t. In fact, we will see that it is completely determined by its right derivative at $t = 0$ (we know that $P(0) = I$).

Theorem 7.2.1. *Let $\{P(t),\ t > 0\}$ be the semigroup of transition matrices of a jump Markov process $\{X_t,\ t \geq 0\}$.*

There exists a matrix $\{Q_{xy}; x, y \in E\}$ *(called the* **infinitesimal generator** *of the semigroup* $\{P(t);\ t \geq 0\}$*, or of the Markov process* $\{X_t,\ t \geq 0\}$*) which satisfies*

$$\begin{aligned} &(i) \quad Q_{xy} \geq 0 \text{ if } x \neq y \\ &(ii) \quad Q_{xx} = -\sum_{y \in E \setminus \{x\}} Q_{xy} \leq 0, \end{aligned}$$

(this last inequality being strict, unless the state x *is absorbing) and such that, as* $h \downarrow 0$*,*

$$\begin{aligned} P_{xy}(h) &= hQ_{xy} + \circ(h) \text{ if } x \neq y \\ P_{xx}(h) &= 1 + hQ_{xx} + \circ(h). \end{aligned}$$

Moreover, conditioned upon $X_0 = x$*, the time* T_1 *of the first jump and the position* $Z_1 = X_{T_1}$ *after that jump are independent, the law of* T_1 *being exponential with parameter* $q_x = -Q_{xx}$*, and the law of* Z_1 *on* E *being given by* $\{Q_{xy}/q_x; y \neq x\}$*.*

Proof First note that

$$\{T_1 > nh\} \subset \{X_0 = X_h = \cdots = X_{nh}\} \subset \{T_1 > nh\} \cup \{T_2 - T_1 \leq h\}.$$

Since $P(T_2 - T_1 \leq h) \to 0$ as $h \to 0$, we have that as $h \to 0$, $nh \to t$ (with $nh \geq t$),

$$\begin{aligned} \mathbb{P}(T_1 > t | X_0 = x) &= \lim \mathbb{P}(X_0 = X_h = \cdots = X_{nh} | X_0 = x) \\ &= \lim [P_{xx}(h)]^n \end{aligned}$$

Existence of this limit implies that

$$\frac{1}{h}[1 - P_{xx}(h)] \to q_x \in [0, +\infty],$$

as $h \to 0$, and consequently

$$\mathbb{P}(T_1 > t | X_0 = x) = e^{-q_x t}.$$

Hence necessarily $q_x < \infty$ and $q_x = 0$ if and only if x is absorbing. Define $Q_{xx} = -q_x$.

The proof of existence of the limits of $\frac{1}{h}P_{xy}(h)$ for $x \neq y$ is done similarly :

$$\{T_1 \leq t,\ Z_0 = x,\ Z_1 = y\}$$
$$= \lim_{h\to 0, nh\to t} \cup_{1\leq m\leq n} \{X_0 = X_h = \cdots = X_{(m-1)h} = x, X_{mh} = y\}$$

$$\mathbb{P}(T_1 \leq t,\ Z_1 = y | X_0 = x) = \lim \frac{1 - P_{xx}(h)^n}{1 - P_{xx}(h)} P_{xy}(h)$$
$$= \frac{1 - e^{-q_x t}}{q_x} \lim \frac{1}{h} P_{xy}(h)$$

Hence $Q_{xy} = \lim \frac{1}{h} P_{xy}(h)$ exists for $x \neq y$ and

$$\mathbb{P}(T_1 \leq t,\ Z_1 = y | X_0 = x) = (1 - e^{-q_x t}) \frac{Q_{xy}}{q_x}$$

whence

$$\mathbb{P}(T_1 \leq t,\ Z_1 = y | X_0 = x) = \mathbb{P}(T_1 \leq t | X_0 = x) \mathbb{P}(Z_1 = y | X_0 = x)$$

and

$$\mathbb{P}(Z_1 = y | X_0 = x) = \frac{Q_{xy}}{q_x}.$$

□

In the case where E is a finite set, we immediately deduce from the Theorem :

Corollary 7.2.2. *(i) $\{P(t),\ t \geq 0\}$ is the unique solution of Kolmogorov's backward equation*

$$\frac{dP}{dt}(t) = QP(t)\ \ t > 0;\ P(0) = I.$$

Moreover $u(t,x) := E[g(X_t)|X_0 = x]$ also solves a Kolmogorov backward equation

$$\begin{cases} \dfrac{\partial u}{\partial t}(t,x) = \displaystyle\sum_{y\in E} Q_{xy} u(t,y),\ \ t > 0,\ x \in E; \\ \ \ u(0,x) = g(x),\ x \in E. \end{cases}$$

(ii) $\{P(t), t \geq 0\}$ *is also the unique solution of the forward Kolmogorov equation :*

$$\frac{dP(t)}{dt} = P(t)Q,\ t > 0;\ P(0) = I.$$

Moreover the family of the marginal probability distributions $\{\mu(t),\ t \geq 0\}$ *of the r. v.* $\{X_t; t \geq 0\}$*'s satisfies the Fokker–Planck equation :*

$$\frac{\partial \mu_x(t)}{\partial t} = \sum_{y \in E} \mu_y(t) Q_{yx},\ t > 0,\ x \in E.$$

PROOF The Kolmogorov backward equation follows by differentiating $P_{xy}(t)$, exploiting the semigroup property in the form

$$P(t + h) = P(h)P(t).$$

The equation for u then follows from the equation just obtained by multiplying it on the right by the column vector $\{g_x\}$.

The forward equation is obtained by differentiating, starting with the identity :

$$P(t + h) = P(t)P(h).$$

The Fokker–Planck equation then follows by multiplying on the left by the row vector $\{\mu(0)\}$. □

Remark 7.2.3. *Let us explain the terminology "forward equation, backward equation". The backward equation is an equation for the function* $(t, x) \to P_{xy}(t)$*, where* $y \in E$ *is fixed. The variables are* t *and the "backward" variable* x*.* x *is the position at the initial time, it is the position in the past. In contrast, the forward equation is an equation for the function* $(t, y) \to P_{xy}(t)$*, with* $x \in E$ *fixed. The variable* y *denotes the position of the process at time* t*, it is the position at the present time.*

Consider now the backward equation for the quantity $u(t, x) = \mathbb{E}[g(X_t)|X_0 = x]$*. Fix* $T > 0$*, and define for* $0 \leq t \leq T$*,* $v(t, x) = u(T - t, x) = \mathbb{E}[g(X_T)|X_t = x]$*.* v *satisfies the equation*

$$\begin{cases} \dfrac{\partial v}{\partial t}(t, x) + \displaystyle\sum_{y \in E} Q_{xy} u(t, y) = 0,\ t > 0,\ x \in E; \\ \qquad\qquad v(T, x) = g(x),\ x \in E. \end{cases}$$

The equation for v is a backward equation in the sense that it is solved in the backward direction of time, from $t = T$ to $t = 0$. Note that in the non–homogeneous case, where the infinitesimal generator Q depends upon t, the quantity $v(t, x) = \mathbb{E}[g(X_T)|X_t = x]$ solves that same equation, while we no longer have an equation for u.

The proof of the Corollary is not rigorous in the case where E is countable, since it implies interchanging a derivation and an infinite sum. The backward Kolmogorov equation will be established in the general case in the next section.

7.3 The strong Markov property

The notion of a stopping time S and the associated σ–field $\mathcal{F}_S^X$ are defined as in section 6.3, replacing $\{N_t; t \geq 0\}$ by $\{X_t; t \geq 0\}$.

Theorem 7.3.1. *Let S be a stopping time of the jump Markov process $\{X_t;\ t \geq 0\}$. Conditionally upon $\{S < \infty\}$ and $\{X_S = x\}$, $\{X_{S+t},\ t \geq 0\}$ is independent of $\mathcal{F}_S^X$, and its law is that of $\{X_t;\ t \geq 0\}$ given that $X_0 = x$.*

Proof It suffices to prove the theorem in the case of a constant stopping time $S \equiv s$. The general case then follows as for the Poisson process (see the proof of Proposition 6.3.2). Let $0 \leq s_1 < s_2 < \cdots < s_k < s$; $0 < t_1 < t_2 < \cdots < t_\ell$; $x, x_1, \ldots, x_z, y_1, \ldots, y_\ell \in S$,

$$\begin{aligned}
&\mathbb{P}(X_{s+t_1} = y_1, \ldots, X_{s+t_\ell} = y_\ell | X_{s_1} = x_1, \ldots, X_{s_k} = x_k, X_s = x) \\
&= \frac{\mathbb{P}(X_{s_1} = x_1, \ldots, X_{s_k} = x_k, X_s = x, X_{s+t_1} = y_1, \ldots, X_{s+t_\ell} = y_\ell)}{\mathbb{P}(X_{s_1} = x_1, \ldots, X_{s_k} = x_k, X_s = x)} \\
&= P_{xy_1}(t_1) P_{y_1 y_2}(t_2 - t_1) \times \cdots \times P_{y_{\ell-1} y_\ell}(t_\ell - t_{\ell-1}) \\
&= \mathbb{P}(X_{t_1} = y_1, \ldots, X_{t_\ell} = y_\ell | X_0 = x)
\end{aligned}$$

□

We now establish the backward Kolmogorov equation in the general case.

Theorem 7.3.2. *For all $x, y \in E$, the function $t \to P_{xy}(t)$ is differentiable, and*

$$\frac{d}{dt} P_{xy}(t) = (QP)_{xy}(t).$$

PROOF Define for all $n \in \mathbb{N}$ the conditional law of (Z_n, T_n), given that $X_0 = Z_0 = x$:

$$R_n(x; y, B) = \mathbb{P}(Z_n = y, T_n \in B | Z_0 = x),\ B \text{ Borel subset of } \mathbb{R}_+.$$

Note that

$$R_0(x; y, B) = \begin{cases} 1, & \text{if } x = y,\ 0 \in B; \\ 0, & \text{otherwise.} \end{cases}$$

and it follows from Theorem 7.2.1 that

$$R_1(x; y, B) = \begin{cases} Q_{xy} \displaystyle\int_B e^{-q_x t} dt, & \text{if } x \neq y; \\ 0, & \text{if } x = y. \end{cases}$$

The strong Markov property at time T_m implies that

$$\mathbb{P}(Z_{m+n} = z, T_{m+n} \in B | \mathcal{F}^X_{T_m}) = R_n(X_{T_m};\ z, B - T_m),$$

where we have used the notation

$$B - t = \{s \in \mathbb{R}_+;\ s + t \in B\}$$

Hence

$$\begin{aligned} \mathbb{P}(Z_{m+n} = z, T_{m+n} \in B | X_0 = x) &= \mathbb{E}[R_n(Z_m; z, B - T_m) | X_0 = x] \\ &= \sum_{y \in E} \int_{\mathbb{R}_+} R_m(x; y; dt) R_n(y; z, B - t), \end{aligned}$$

In other words,

$$R_{m+n}(x; z, B) = \sum_{y \in E} \int_B \int_{R_+} R_m(x; y, dt) R_n(y; z, du - t),$$

hence also

$$R_{m+n}(x; z, du) = \sum_y \int_0^u R_m(x; y, dt) R_n(y; z, du - t),$$

where the measure $R_n(y; z, du - t)$ is defined by

$$\int_{\mathbb{R}_+} R_n(y; z, du - t) f(u) = \int_{\mathbb{R}_+} R_n(y; z, du) f(t + u).$$

Clearly the $\{R_n, n \geq 1\}$'s are completely determined by R_1 and this equation.

Note that

$$\begin{aligned}
P_{xy}(s) &= \sum_{m\geq 0} \mathbb{P}(Z_m = y, T_m \leq s < T_{m+1} | Z_0 = x) \\
&= \sum_{m\geq 0} P(Z_m = y, T_m \leq s, T_{m+1} - T_m > s - T_m | Z_0 = x) \\
&= \sum_{m\geq 0} \mathbb{E}[\mathbb{P}(T_{m+1} - T_m > s - T_m | Z_m, T_m) \mathbf{1}_{\{Z_m = y, T_m \leq s\}} | Z_0 = x] \\
&= \sum_{m\geq 0} \mathbb{E}[e^{-q_{Z_m}(s-T_m)} \mathbf{1}_{\{Z_m = y, T_m \leq s\}} | Z_0 = x] \\
&= \sum_{m\geq 0} \int_0^s e^{-q_y(s-t)} R_m(x; y, dt),
\end{aligned}$$

where we have used the strong Markov property at time T_m for the third equality.

Hence from the above equation

$$P_{xy}(s) = \delta_{xy} e^{-q_x s} + \sum_{m\geq 1} \int_0^s e^{-q_y(s-t)} R_m(x; y, dt)$$

$$P_{xy}(s) = \delta_{xy} e^{-q_x s} + \sum_{m\geq 0, z\in E} \int_0^s e^{-q_y(s-t)} \int_0^t R_1(x; z, du) R_m(z; y, dt - u),$$

or equivalently

$$P_{xy}(t) = \delta_{xy} e^{-q_x t} + \sum_{z\in E} \int_0^t R_1(x; z, ds) P_{zy}(t - s)$$

$$e^{q_x t} P_{xy}(t) = \delta_{xy} + \int_0^t e^{q_x s} \sum_{z\neq x} Q_{xz} P_{zy}(s) ds.$$

Hence the function $t \to P_{xy}(t)$ is differentiable, and

$$\begin{aligned}
\frac{d}{dt} P_{xy}(t) &= \sum_{z\neq x} Q_{xz} P_{zy}(t) - q_x P_{xy}(t) \\
&= \sum_z Q_{xz} P_{zy}(t).
\end{aligned}$$

□

The above argument shows that

$$\begin{aligned}\mathbb{P}(Z_1 = y, T_1 \in B, Z_2 = z, T_2 - T_1 \in C | Z_0 = x) \\ = \int_B \int_C R_1(x, y, dt) R_1(y, z, du)\end{aligned}$$

This formula generalizes to the law of $((Z_1, T_1), \ldots, (Z_n, T_n))$. One could deduce the Markov property of the corresponding process $\{X_t,\ t \geq 0\}$ from that joint law.

Remark 7.3.3. *If we allow an arbitrary generator Q, one can always define R_1, and hence the law of the (Z_n, T_n)'s. But that sequence $\{T_n\}$ does not necessarily satisfy the non–explosion condition* (7.3)*, i. e. the corresponding process $\{X_t\}$ need not be defined for all $t \geq 0$. In the next section we shall give sufficient conditions on Q for the no explosion to hold.*

7.4 Embedded Markov chain

Let $\{X_t; t \geq 0\}$ be a jump Markov process, whose jump times $T_1, T_2, \ldots, T_n, \ldots$ satisfy the non–explosion condition (7.3). The sequence $\{Z_n; n \in \mathbb{N}\}$ defined by

$$Z_n = X_{T_n} \quad (\text{with } T_0 = 0)$$

is a discrete time Markov chain (this is a consequence of the strong Markov property of $\{X_t\}$), called the "embedded chain", which has the property that $Z_{n+1} \neq Z_n$ a. s., $\forall n \geq 0$. Its transition matrix P is easily computed in terms of the infinitesimal generator Q of $\{X_t\}$:

$$P_{xy} = \begin{cases} (-Q_{xx})^{-1} Q_{xy}, & \text{if } y \neq x; \\ 0, & \text{if } y = x. \end{cases}$$

Define, for $n \geq 1$,

$$S_n = q_{Z_{n-1}}(T_n - T_{n-1}) \quad (\text{where } q_x = -Q_{xx}),$$

and for $t \geq 0$,

$$N_t = \sup\{n; \sum_{k=1}^{n} S_k \leq t\}.$$

Then $\{N_t;\ t \geq 0\}$ is a Poisson process with intensity 1 (this follows from the strong Markov property of $\{X_t\}$, and the fact that if $U \simeq$ exponential (λ), then $\lambda U \simeq$ exponential (1)).

Now let Q be an infinitesimal generator, i. e. a matrix indexed by $E \times E$, such that for all $x \in E$,

$$Q_{xy} \geq 0, \quad y \neq x; \quad Q_{xx} = -\sum_{y \neq x} Q_{xy} < 0.$$

We let $q_x = -Q_{xx}$, and we define the transition matrix P by

$$P_{xy} = \begin{cases} \frac{Q_{xy}}{q_x}, & \text{if } y \neq x; \\ 0, & \text{if } y = x, \end{cases} \tag{7.4}$$

with the convention that $P_{xy} = 0\ \forall y \neq x$ and $P_{xx} = 1$, if $Q_{xx} = 0$. To any initial condition $x \in E$, we associate the Markov chain $\{Z_n,\ n \geq 0\}$ with the transition matrix P. Now let $\{N_t;\ t \geq 0\}$ be a Poisson process with intensity 1, independent of the chain $\{Z_n;\ n \in \mathbb{N}\}$. Denote by $0 = T_0 < T_1 < T_2 < \cdots$ the times of the jumps of the Poisson process, and define for $n \geq 1$,

$$S_n = \frac{T_n - T_{n-1}}{q(Z_{n-1})},$$
$$T'_n = S_1 + \ldots + S_n.$$

If the non–explosion condition (7.3) is satisfied by the sequence $\{T'_n\}$, then

$$X_t \stackrel{\text{def}}{=} \sum_{n \geq 0} Z_n \mathbf{1}_{[T'_n, T'_{n+1}[}(t),\ \ t \geq 0 \tag{7.5}$$

is a jump Markov process with the infinitesimal generator Q.

It remains to answer the question : given an infinitesimal generator Q, when does the associated sequence of stopping times $\{T'_n,\ n \geq 0\}$ satisfy the non–explosion condition, i. e. when does (7.5) define X_t for all $t \geq 0$? Let us establish

Proposition 7.4.1. *The non–explosion condition* (7.3) *is satisfied if and only if*

$$\sum_{n \geq 0} q_{Z_n}^{-1} = +\infty \ \ a.\ s. \tag{7.6}$$

Let us first state

Corollary 7.4.2. *A sufficient condition for the infinitesimal generator Q to be the infinitesimal generator of a Markov process which satisfies condition* (7.3) *is that one of the two following conditions holds :*

1. $\sup_{x\in E} q_x < \infty$.
2. *The Markov chain $\{Z_n\}$ with the transition matrix P defined by* (7.4) *is recurrent.*

It is clear that each of the two conditions in Corollary implies (7.6). The Proposition follows from the following Lemma, if we let

$$A_n = T_{n+1} - T_n, \quad B_n = \frac{1}{q_{Z_n}}; \quad n \geq 0.$$

Lemma 7.4.3. *Let $\{A_n,\ n \geq 1\}$ and $\{B_n,\ n \geq 1\}$ be two mutually independent sequences of $\mathbb{R}_+^*$–valued r. v.'s, the sequence $\{A_n\}$ being i. i. d., the common law being the exponential distribution with parameter 1. Then the following two statements are equivalent*

1. $\sum_{n=1}^{\infty} A_n B_n = +\infty$ *a. s.*
2. $\sum_{n=1}^{\infty} B_n = +\infty$ *a. s.*

Proof Since the two sequences are mutually independent, the Lemma will follow from the fact that for any sequence $\{b_n,\ n \geq 1\}$ of strictly positive real numbers,

$$\sum_{n=1}^{\infty} A_n b_n = +\infty \quad \text{a. s.} \quad \Longleftrightarrow \quad \sum_{n=1}^{\infty} b_n = +\infty. \tag{7.7}$$

If $\sum_n b_n < \infty$, then $\mathbb{E}\sum_n A_n b_n = \sum_n b_n < \infty$, and a fortiori $\sum_n A_n b_n < \infty$ a. s. It remains to prove that if $\sum_n b_n = +\infty$, then

$$\Lambda_n := \sum_{k=1}^{n} A_k b_k \to +\infty \ \text{ a. s., as } n \to \infty.$$

In the case where there is a subsequence n_j such that $b_{n_j} \to +\infty$, clearly $\sum_n A_n b_n \geq \sum_j A_{n_j} b_{n_j} = +\infty$, since the A_{n_j}'s being i. i. d. exponential r.

v.'s, infinitely many of them is greater than 1. It thus remains to consider the case where $0 \leq b_n \leq C$ and $\sum_n b_n = +\infty$. In this case for all $M > 0$, if n is big enough such that $\mathbb{E}\Lambda_n > 2M$, then

$$\begin{aligned}\mathbb{P}(\Lambda_n \leq M) &\leq \mathbb{P}\left(|\Lambda_n - \mathbb{E}\Lambda_n| \geq \frac{\mathbb{E}\Lambda_n}{2}\right) \\ &\leq 4\frac{Var(\Lambda_n)}{(\mathbb{E}\Lambda_n)^2} = 4\frac{\sum_1^n b_k^2}{(\sum_1^n b_k)^2} \\ &\leq \frac{4C}{\sum_1^n b_k} \to 0,\end{aligned}$$

hence $\Lambda_n \to +\infty$ in probability, and also a. s. since the sequence is monotone. □

Remark 7.4.4. *In the next chapter, we shall specify jump Markov processes by describing their infinitesimal generator Q. The reader can check that in each of the examples considered, one (usually the first) of the two sufficient conditions of the Corollary 7.4.2 is satisfied.*

7.5 Recurrent and transient states

In the sequel as in the discrete time case we shall denote by $\mathbb{P}_x$ the conditional law of $\{X_t,\ t \geq 0\}$, given that $X_0 = x$. The equivalence classes of the jump Markov process $\{X_t; t \geq 0\}$ are those of the embedded chain. Note that as soon as $\{X_t; t \geq 0\}$ is irreducible,

$$P_{xy}(t) > 0, \quad \forall x, y \in E,\ t > 0. \tag{7.8}$$

Indeed, for all $x, y \in E$, there exists $n \geq 1$ and $x_0 = x, x_1, \ldots, x_{n-1}, x_n = y$ such that $Q_{x_{k-1}x_k} > 0$, $1 \leq k \leq n$, and it follows from the property of the exponential law that $P_{xy}(t) \geq P_{xx_1}(t/n) \times \cdots \times P_{x_{n-1}y}(t/n) > 0$.

A state $x \in E$ is called recurrent (resp. transient) for $\{X_t; t \geq 0\}$ if it is recurrent (resp. transient) for the embedded chain. Then in particular in the irreducible case, either all states are recurrent, or all are transient.

As in the case of discrete time Markov chains, we have

Theorem 7.5.1. *Let $\{X_t; t \geq 0\}$ be an irreducible and recurrent jump Markov process. Then there exists a strictly positive invariant measure π*

on E which solves the equation $\pi Q = 0$, and is unique up to a multiplicative constant. Moreover such a measure is invariant for the semigroup $\{P(t)\}$, in the sense that $\pi P(t) = \pi$, $\forall t \geq 0$.

PROOF We note that if Q is the infinitesimal generator of the jump Markov process $\{X_t\}$ and P is the transition matrix of its embedded chain, then

$$Q = \mathbf{q}(P - I),$$

where $\mathbf{q}$ is the diagonal matrix defined by

$$\mathbf{q}_{xy} = \delta_{xy} q_x, \quad x, y \in E.$$

Note that the assumption that the process is irreducible implies that $q_x > 0$, $\forall x \in E$. Hence we can multiply by $\mathbf{q}^{-1}$. Our assumption is that the embedded chain is irreducible and recurrent. Hence the measure γ^x defined in the proof of Theorem 2.5.3 is strictly positive, and it is the unique (up to a multiplicative constant) solution of the equation $\gamma^x P = \gamma^x$. Hence the strictly positive measure $\mu^x = \mathbf{q}^{-1}\gamma^x$ satisfies $\mu^x Q = 0$, and any other solution μ' of the same equation is such that $\mathbf{q}\mu'$ is P–invariant, hence there exists a constant c such that $\mu' = c\mu^x$. It remains to check that μ^x is invariant by $P(t)$, $\forall t \geq 0$.

Since γ^x_y is the expectation of the number of visits to state y by the embedded chain during an excursion starting from x, q_y^{-1} is the expectation of the time spent at state y by the process $\{X_t\}$ at each of the visits to y of the embedded chain, and the embedded chain is independent from the times spent at each state by the jump Markov process,

$$\begin{aligned} \mu^x_y = \frac{\gamma^x_y}{q_y} &= \mathbb{E}_x \int_0^{R_x} \mathbf{1}_{\{X_s = y\}} ds \\ &= \mathbb{E}_x \int_0^{\infty} \mathbf{1}_{\{X_s = y, s < R_x\}} ds, \end{aligned}$$

where $R_x = \inf\{t > T_1;\ X_t = x\}$ denotes the time of the first return at x. But if $t > 0$, by the strong Markov property,

$$\mathbb{E}_x \int_0^t \mathbf{1}_{\{X_s = y\}} ds = \mathbb{E}_x \int_{R_x}^{R_x + t} \mathbf{1}_{\{X_s = y\}} ds.$$

Hence

$$\begin{aligned}\mu_y^x &= \mathbb{E}_x \int_t^{R_x+t} \mathbf{1}_{\{X_s=y\}} ds \\ &= \mathbb{E}_x \int_0^{R_x} \mathbf{1}_{\{X_{t+s}=y\}} ds \\ &= \int_0^\infty \mathbb{P}_x(X_{t+s} = y, s < R_x) ds \\ &= \int_0^\infty \sum_z \mathbb{P}_x(X_s = z, s < R_x) P_{zy}(t) ds \\ &= \sum_z \mu_z^x P_{zy}(t).\end{aligned}$$

□

7.6 The irreducible recurrent case

In order to distinguish between the positive and null recurrent cases (this is an open question only if $|E| = +\infty$), it is not sufficient to consider the property of the embedded chain, as we shall now see. Define again the time of the first return at state x as :

$$R_x = \inf\{t \geq T_1; X_t = x\}$$

Definition 7.6.1. *The state x is said to be positive recurrent if it is recurrent and $\mathbb{E}_x(R_x) < \infty$, null recurrent if it is recurrent and $\mathbb{E}_x(R_x) = +\infty$.*

Again in the irreducible recurrent case, either all states are null recurrent, or they are all positive recurrent, and one says accordingly that the process $\{X_t\}$ is null recurrent or positive recurrent. We now prove that the positive recurrent case is equivalent to the existence of a unique invariant probability distribution.

Theorem 7.6.2. *Let $\{X_t,\ t \geq 0\}$ be an irreducible jump Markov process. A state $x \in E$ is positive recurrent if and only if all states are positive recurrent, if and only if there exists a unique invariant probability distribution π, and in that case*

$$\mathbb{E}_x R_x = \frac{1}{\pi_x q_x}, \quad \forall x \in E.$$

PROOF If state x is positive recurrent for $\{X_t\}$, then x is recurrent for the embedded chain $\{Z_n,\ n \geq 0\}$. Denote by γ_y^x the mean number of visits to state y during an excursion of $\{Z_n\}$ starting from x. Since the time spent at y by $\{X_t\}$ at each visit of $\{Z_n\}$ is independent of the embedded chain, and has expectation q_y^{-1},

$$\mathbb{E}_x R_x = \sum_{y \in E} \frac{\gamma_y^x}{q_y}.$$

But we saw in the proof of Theorem 7.5.1 that the measure μ^x defined by

$$\mu_y^x = \frac{\gamma_y^x}{q_y}$$

satisfies $\mu^x Q = 0$. The condition that x is positive recurrent therefore implies the existence of an invariant measure with finite mass, hence of an invariant probability distribution, whose uniqueness follows from Theorem 7.5.1. Suppose now that there exists a probability distribution π solution of $\pi Q = 0$. Then the measure $\mathbf{q}\pi$ is P–invariant, and for all $x, y \in E$,

$$\frac{q_y \pi_y}{q_x \pi_x}$$

is the mean number of visits to state y during an excursion of $\{Z_n\}$ starting from x. Hence

$$\mathbb{E}_x R_x = \sum_{y \in E} \frac{\pi_y}{q_x \pi_x} < \infty, \quad \forall x \in E,$$

and any state $x \in E$ is positive recurrent. □

Remark 7.6.3. *An invariant measure π of the jump Marov process $\{X_t\}$ is a solution of equation $\pi Q = 0$. An invariant mesure μ of the embedded chain is a solution of equation $\mu(P - I) = 0$. Hence π is invariant for $\{X_t\}$ if and only if $\mu = \pi\mathbf{q}$ is invariant for the embedded chain. It is then easy to choose $\mathbf{q}$ (which specifies the expectations of the lengths of the visits of $\{X_t\}$ to the various states) in such a way that π has a finite mass, while μ has an infinite mass (hence $\{X_t\}$ is positive recurrent while the embedded chain is null recurrent), or vice–versa. See the last question of exercise 7.11.8 and the explanations at the end of the correction.*

We now restrict ourselves to the positive recurrent case, and establish an ergodic theorem, and the convergence transition probabilities towards the invariant probability.

Theorem 7.6.4. *Let $\{X_t, t \geq 0\}$ be an E–valued irreducible, positive recurrent jump Markov process. Let Q be its infinitesimal generator, and π the unique invariant probability distribution. Then if $f : E \to \mathbb{R}$ is bounded,*

$$\frac{1}{t}\int_0^t f(X_s)ds \to \sum_{x\in E} f(x)\pi_x$$

a. s. as $t \to \infty$.

PROOF It suffices to consider the case $f(y) = \mathbf{1}_{\{y=x\}}$ and to work under $\mathbb{P}_x$ (see the proof of Theorem 2.5.7). As in the discrete time case, the successive excursions starting from x are i. i. d.. Let $N^x(t)$ denote the number of visits to state x between time 0 and time t, and let T_k^x denote the time spent at state x by the process $\{X_t\}$ during its k–th visit. Since the $N^x(t)$–th visit to state x need not be terminated at time t, we have

$$\frac{1}{t}\sum_{k=1}^{N^x(t)-1} T_k^x < \frac{1}{t}\int_0^t \mathbf{1}_{\{X_s=x\}}ds \leq \frac{1}{t}\sum_{k=1}^{N^x(t)} T_k^x.$$

But clearly $T^x_{N^x(t)}/t \to 0$ a. s. as $t \to \infty$, and

$$\begin{aligned}\frac{1}{t}\sum_{k=1}^{N^x(t)} T_k^x &= \frac{N^x(t)}{t} \times \frac{1}{N^x(t)}\sum_{k=1}^{N^x(t)} T_k^x \\ &\to \frac{1}{\mathbb{E}_x R_x} \times \frac{1}{q_x} \\ &= \pi_x.\end{aligned}$$

Indeed, since the sequence $\{T_k^x,\ k \geq 1\}$ is i. i. d., and since from recurrence $N^x(t) \to \infty$ a. s. as $t \to \infty$,

$$\frac{1}{N^x(t)}\sum_{k=1}^{N^x(t)} T_k^x \to \mathbb{E}_x(T_1^x) = \frac{1}{q_x},$$

and the proof of the fact that

$$\frac{t}{N^x(t)} \to \mathbb{E}_x(R_x)$$

follows the same argument as in the proof of Theorem 2.5.7. □

In the continuous time case, the convergence of the probability distribution of X_t towards the invariant distribution as $t \to \infty$ holds in the irreducible and positive recurrent case, without any further restriction.

Theorem 7.6.5. *Let $\{X_t, t \geq 0\}$ be an E–valued irreducible, positive recurrent jump Markov process, and π its unique invariant probability distribution, solving the stationary Fokker–Planck equation $\pi Q = 0$. Then for any probability distribution μ on E and $x \in E$, $(\mu P)_x(t) \to \pi_x$ as $t \to \infty$.*

Proof One could imitate the proof of Theorem 2.6.4, but instead we shall use that result.

If we sample the process $\{X_t\}$ by letting $Y_n = X_{nh}$, $n = 0, 1, \ldots$, where $h > 0$ is arbitrary, then clearly $\{Y_n,\ n \in \mathbb{N}\}$ is an irreducible and aperiodic (see (7.8)) Markov chain, whose unique invariant probability distribution, which does not depend upon h, is π. Let us assume for a moment

Lemma 7.6.6. *For all $t, h > 0$, $x, y \in E$,*

$$|P_{xy}(t+h) - P_{xy}(t)| \leq 1 - e^{-q_x h}.$$

Fix $\varepsilon > 0$ and $x, y \in E$. We first choose $h > 0$ sufficiently small, in such a way that

$$1 - e^{-q_x s} \leq \varepsilon/2, \quad \text{if } 0 \leq s \leq h,$$

then we choose N large enough such that

$$|P_{xy}(nh) - \pi_y| \leq \varepsilon/2, \quad \text{if } n \geq N.$$

We conclude that if $t \geq Nh$, denoting by n the integer such that $nh \leq t < (n+1)h$,

$$|P_{xy}(t) - \pi_y| \leq |P_{xy}(t) - P_{xy}(nh)| + |P_{xy}(nh) - \pi_h| \leq \varepsilon.$$

The Theorem follows easily from this result, if we decompose the set of starting points into a finite set which supports the mass of μ except for δ, and its complementary. □

It remains to proceed to

PROOF OF LEMMA 7.6.6 It suffices to note that

$$\begin{aligned}
|P_{xy}(t+h) - P_{xy}(t)| &= |\sum_z P_{xz}(h)P_{zy}(t) - P_{xy}(t)| \\
&= \left|\sum_{z\neq x} P_{xz}(h)P_{zy}(t) - (1-P_{xx}(h))P_{xy}(t)\right| \\
&= \left|\sum_{z\neq x} P_{xz}(h)(P_{zy}(t) - P_{xy}(t))\right| \\
&\leq \sum_{z\neq x} P_{xz}(h) \\
&= 1 - P_{xx}(h).
\end{aligned}$$

Remark 7.6.7. *The convergence in Theorem 7.6.5 is in the sense of weak convergence of probability distributions on E. This follows from the fact that as $t \to \infty$, $(\mu P)_x(t) \to \pi_x$ for all $x \in E$, hence also for any finite subset $F \subset E$, $\sum_{x\in F}(\mu P)_x(t) \to \sum_{x\in F}\pi_x$. Since moreover $\mu P(t)$ and π are probability distributions on E, it is not hard to show that for any bounded $f : E \to \mathbb{R}$,*

$$\sum_x (\mu P)_x(t) f(x) \to \sum_x \pi_x f(x), \quad \text{as } t \to \infty.$$

Note that if we choose the invariant probability distribution π as the law of X_0, then the process $\{X_t; t \geq 0\}$ is **stationary** in the sense that for all $n \in \mathbb{N}$, $0 \leq t_1 < t_2 < \cdots < t_n$, the law of the random vector $(X_{t_1+s}, X_{t_2+s}, \ldots, X_{t_n+s})$ does not depend upon $s \geq 0$.

Remark 7.6.8. *The equation $\pi Q = 0$ reads $\forall x \in E$,*

$$\sum_{y\neq x} \pi_y Q_{yx} = \pi_x \sum_{y\neq x} Q_{xy}.$$

The left hand side of this equality is the mean flux entering state x at equilibrium, coming from the various other states, and the right hand side is the mean flux leaving x at equilibrium, towards the other states. The equation $\pi Q = 0$ says that at equilibrium, the mean numbers of departures and arrivals per unit of time are equal.

We also have a generalization of the central limit theorem. The next result is a special case of Theorem 2.1 in [4].

Theorem 7.6.9. *Suppose that the jump Markov process $\{X_t; t \geq 0\}$ is irreducible, and that it has an invariant probability distribution π. Let $f \in L^2(E, \pi)$ be of the type $f = Qg$, where $g \in L^2(E, \pi)$ [this implies that $< \pi, f >= \sum_{x \in E} \pi_x f_x = \sum_{x,y \in E} \pi_x Q_{xy} g_y = 0$].*

Let

$$C(f) := -2 \sum_{x \in E} f_x g_x \pi_x,$$

which we suppose not to be equal to zero (consequently $C(f) > 0$). Then

$$\frac{1}{\sqrt{tC(f)}} \int_0^t f(X_s) ds \to Z,$$

in law, as $t \to \infty$, where Z is a centered Gaussian r. v. with unit variance.

We also have the convergence of $\left\{ \frac{1}{\sqrt{uC(f)}} \int_0^{tu} f(X_s) ds,\ t \geq 0 \right\}$ towards a Brownian motion $\{B_t,\ t \geq 0\}$, as $u \to \infty$.

7.7 Reversibility

Given a jump Markov process $\{X_t;\ t \geq 0\}$, and $T > 0$, $\{\hat{X}_t^T = X_{T-t},\ 0 \leq t \leq T\}$ is also a Markov process. If the law of X_0 is an invariant probability distribution π, then $\hat{X}^T$ is time–homogeneous. Denote by $\hat{Q}$ its infinitesimal generator. We have

Theorem 7.7.1. *$\hat{Q} = Q$ if and only if the detailed balance equation*

$$\pi_x Q_{xy} = \pi_y Q_{yx},\ \ \forall x, y \in E,$$

is satisfied. In this case, we say that the process $\{X_t\}$ is reversible (with respect to the probability distribution π, which then is invariant).

Proof The same argument as in the discrete time case implies that for all $t > 0$, $x, y \in E$,

$$\hat{P}_{xy}(t) = \frac{\pi_y}{\pi_x} P_{yx}(t),$$

from which follows, by taking the t–derivative at $t = 0$

$$\hat{Q}_{xy} = \frac{\pi_y}{\pi_x} Q_{yx}.$$

The result is now obvious. □

Remark 7.7.2. *As in the case of discrete time Markov chains, a jump Markov process which is irreducible and positive recurrent need not be reversible. Again, a counter–example is provided by a Q–matrix such that for a given pair $x \neq y$, $Q_{xy} = 0 \neq Q_{yx}$, which does not contradict irreductibility as long as $|E| \geq 3$.*

Remark 7.7.3. *As in the case of discrete time Markov chains, to find a generator Q such that a given distribution π is Q–invariant is not difficult. The easiest approach is to look for Q such that the associated process is reversible with respect to π, hence to look for an infinitesimal generator Q such that the quantity $\pi_x Q_{xy}$ be symmetric in x, y.*

To find the invariant probability distribution, given an irreducible generator, is in general more difficult. One can try to solve the equation

$$\pi_x Q_{xy} = \pi_y Q_{yx}, \quad \forall x, y \in E,$$

but this equation has a solution only in the reversible case. In the non–reversible case, one should solve the stationary Fokker–Planck equation $\pi Q = 0$. If one can guess π up to a multiplicative constant, then one can take advantage of the following result

Theorem 7.7.4. *Given a probability distribution π on E, define for $x, y \in E$*

$$\hat{Q}_{xy} = \frac{\pi_y}{\pi_x} Q_{yx}.$$

If

$$\sum_{y \neq x} \hat{Q}_{xy} = \sum_{y \neq x} Q_{xy},$$

then π is an invariant probability distribution, and $\hat{Q}$ is the generator of the time–reversed process.

PROOF Using the first and then the second identity from the statement, we deduce

$$\begin{aligned}\sum_{y\neq x}\pi_y Q_{yx} &= \pi_x \sum_{y\neq x}\hat{Q}_{xy}\\ &= \pi_x \sum_{y\neq x} Q_{xy}\\ &= -\pi_x Q_{xx},\end{aligned}$$

which implies that $\pi Q = 0$. The second part of the statement is then a consequence of the formula which appears in the proof of Theorem 7.7.1. □

Note that if we can guess the generator of the time reversed process, we can deduce the invariant probability distribution up to the normalization constant.

7.8 Markov models of evolution and Phylogeny

We shall define Markov processes on trees, which is a model frequently used in phylogeny. We shall consider rooted and unrooted binary trees. We present in figure 7.2 a rooted binary tree (the root is at the top and the leaves are at the bottom !), and in figure 7.3 an unrooted binary tree.

Markov process on a rooted binary tree The process starts from the root (which plays the role of the initial time 0) in a certain state, say x. It evolves up to the first node which is located at distance r from the root, as a continuous time jump Markov process during a time interval of length r. Denote by y the state of the process at that node. On each branch which starts from that node a jump Markov process runs, starting from y, in such a way that the processes on the two branches are independent, up to the next node, and so on down to the leaves of the tree. Note that we shall only consider irreducible processes with values in a finite state space, hence the process will be positive recurrent, and we shall choose the invariant probability distribution as the law at the root. We can then, without changing the law of the process, suppress the branch between the root and the first node.

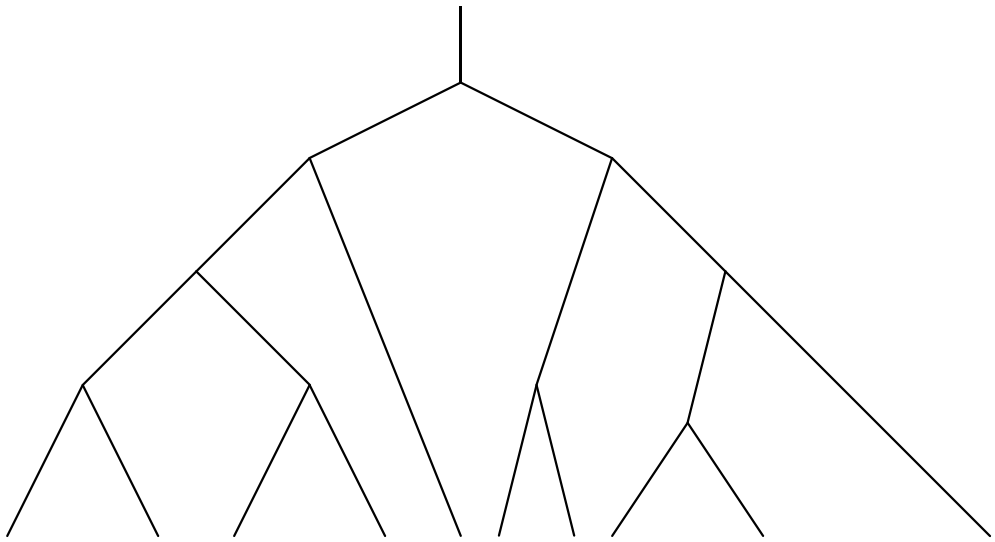

Figure 7.2: Rooted binary tree

Markov process on an unrooted tree Let us suppress the branch between the root and the first node. This means that the process starts from the root (located at the first node) under the invariant probability distribution, and evolves independently on the two branches, until it encouters the next node, etc... Consider figure 7.3. The root is no longer indicated on the central branch. One can still imagine that it is there, and consider that two processes start from that point, one in each direction, towards the two nodes located at the two ends of the central branch. Suppose now that we move the root on the central branch, either to the right or to the left. It is easy to convince oneself that the law of the resulting process on the tree is not modified, provided the process is *reversible*. Indeed the difference between the two constructions with the root at two different points on the central branch is that a portion of that branch is run in the two different directions by the two versions of the process. In the reversible case, the root can equivalently be located at either end of the central branch, or even at any node of the tree, or at any point of any branch of the tree. This means that one can define a jump Markov process on an unrooted tree, provided the dynamics

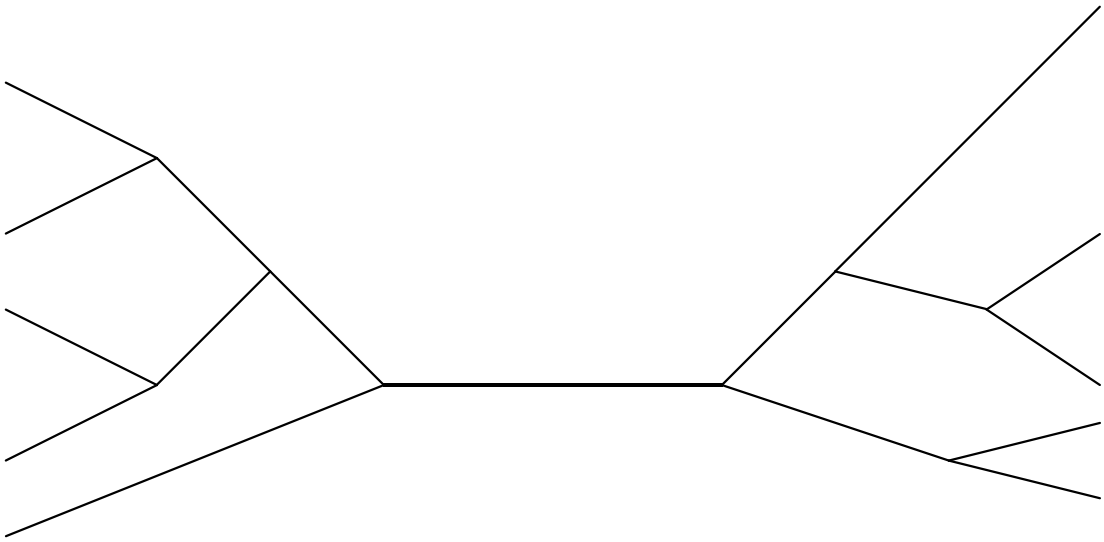

Figure 7.3: The unrooted version of the same tree

is time reversible, by locating the starting point anywhere on the tree.

7.8.1 Models of evolution

In order to compute the likelihood of a given tree as a function of the data, we need to choose a model of evolution, which tells us how the data have been "manufactured" by evolution along the branches of the tree, for each site of the DNA. We shall describe several Markovian models of evolution of the DNA, by describing the transition rate from each nucleotide to any other one. This means that we shall prescribe the matrix Q in the form

$$Q = \begin{array}{c} \\ \mathtt{a} \\ \mathtt{c} \\ \mathtt{g} \\ \mathtt{t} \end{array} \begin{array}{cccc} \mathtt{a} & \mathtt{c} & \mathtt{g} & \mathtt{t} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{array}$$

The Jukes–Cantor model (1969) This is the simplest one, which assumes that all mutations happen at the same rate, i. e. with a given $\alpha > 0$,

$$Q = \begin{pmatrix} -3\alpha & \alpha & \alpha & \alpha \\ \alpha & -3\alpha & \alpha & \alpha \\ \alpha & \alpha & -3\alpha & \alpha \\ \alpha & \alpha & \alpha & -3\alpha \end{pmatrix}.$$

The associated invariant probability distribution is uniform over the 4 nucleotides. The transition probabilities are easy to compute. $P(t) =$

$$\begin{pmatrix} 0,25+0,75e^{-4\alpha t} & 0,25-0,25e^{-4\alpha t} & 0,25-0,25e^{-4\alpha t} & 0,25-0,25e^{-4\alpha t} \\ 0,25-0,25e^{-4\alpha t} & 0,25+0,75e^{-4\alpha t} & 0,25-0,25e^{-4\alpha t} & 0,25-0,25e^{-4\alpha t} \\ 0,25-0,25e^{-4\alpha t} & 0,25-0,25e^{-4\alpha t} & 0,25+0,75e^{-4\alpha t} & 0,25-0,25e^{-4\alpha t} \\ 0,25-0,25e^{-4\alpha t} & 0,25-0,25e^{-4\alpha t} & 0,25-0,25e^{-4\alpha t} & 0,25+0,75e^{-4\alpha t} \end{pmatrix}.$$

The Kimura models (1980 et 1981) Among the four nucleotides, cytosine and thymine are pyrimidines, while adenine and guanine are purines. It is reasonable to assume that transitions (replacement of a purine by the other, or of a pyrimidine by the other) are more frequent than transversions (replacement of a purine by a pyrimidine or vice versa). One then is led to assume that the sustitution rates between `a` and `g` or between `c` and `t` are greater than all others, hence the model (with $\beta > \alpha$)

$$Q = \begin{pmatrix} -2\alpha-\beta & \alpha & \beta & \alpha \\ \alpha & -2\alpha-\beta & \alpha & \beta \\ \beta & \alpha & -2\alpha-\beta & \alpha \\ \alpha & \beta & \alpha & -2\alpha-\beta \end{pmatrix}.$$

The invariant probability distribution is again uniform. The transition probabilities are given by

$$P_{xx}(t) = 0,25+0,25e^{-4\beta t}+0,5e^{-2(\alpha+\beta)t},$$

$$P_{xy}(t) = 0,25+0,25e^{-4\beta t}-0,5e^{-2(\alpha+\beta)t},$$

if $x \neq y$ are either both purines or both pyrimidines,

$$P_{xy}(t) = 0,5-0,5e^{-4\beta t}$$

in the other case.

Kimura proposed a second model, of the form

$$Q = \begin{pmatrix} -\alpha-\beta-\gamma & \alpha & \beta & \gamma \\ \alpha & -\alpha-\beta-\gamma & \gamma & \beta \\ \beta & \gamma & -\alpha-\beta-\gamma & \alpha \\ \gamma & \beta & \alpha & -\alpha-\beta-\gamma \end{pmatrix},$$

for which the invariant probability distribution is still uniform.

The Felsenstein model Given a probability distribution π on $E = \{\mathtt{a}, \mathtt{c}, \mathtt{g}, \mathtt{t}\}$, and a positive number u, Felsenstein has proposed the model

$$Q = \begin{pmatrix} u(\pi_{\mathtt{a}}-1) & u\pi_{\mathtt{c}} & u\pi_{\mathtt{g}} & u\pi_{\mathtt{t}} \\ u\pi_{\mathtt{a}} & u(\pi_{\mathtt{c}}-1) & u\pi_{\mathtt{g}} & u\pi_{\mathtt{t}} \\ u\pi_{\mathtt{a}} & u\pi_{\mathtt{c}} & u(\pi_{\mathtt{g}}-1) & u\pi_{\mathtt{t}} \\ u\pi_{\mathtt{a}} & u\pi_{\mathtt{c}} & u\pi_{\mathtt{g}} & u(\pi_{\mathtt{t}}-1) \end{pmatrix}.$$

Note that clearly for $x \neq y$,

$$\pi_x Q_{xy} = \pi_y Q_{yx},$$

hence π is the invariant probability distribution, and the chain is reversible. The matrix Q has two eigenvalues : $-u$, whose associated eigenspace is composed of those vectors which are orthogonal to π in $\mathbb{R}^4$, and 0, whose associated eigenspace is composed of those vectors which are colinear to $(1,1,1,1)$. One can then show that

$$P_{xy}(t) = \left(e^{tQ}\right)_{xy} = e^{-ut}\delta_{xy} + (1-e^{-ut})\pi_y.$$

In the particular case where $\pi = (1/4, 1/4, 1/4, 1/4)$, this model reduces to the Jukes–Cantor model.

The Hasegawa, Kishino, Yano model (1985) This is a generalisation of both the first Kimura model and Felsenstein's. Given again π an arbitrary probability distribution on E, and u, v two positive numbers, let

$$Q = \begin{pmatrix} -u\pi_{\mathtt{g}} - v\pi_2 & v\pi_{\mathtt{c}} & u\pi_{\mathtt{g}} & v\pi_{\mathtt{t}} \\ v\pi_{\mathtt{a}} & -u\pi_{\mathtt{t}} - v\pi_1 & v\pi_{\mathtt{g}} & u\pi_{\mathtt{t}} \\ u\pi_{\mathtt{a}} & v\pi_{\mathtt{c}} & -u\pi_{\mathtt{a}} - v\pi_2 & v\pi_{\mathtt{t}} \\ v\pi_{\mathtt{a}} & u\pi_{\mathtt{c}} & v\pi_{\mathtt{g}} & -u\pi_{\mathtt{c}} - v\pi_1 \end{pmatrix},$$

where $\pi_1 = \pi_{\mathtt{a}} + \pi_{\mathtt{g}}$, $\pi_2 = \pi_{\mathtt{c}} + \pi_{\mathtt{t}}$. Again π is the invariant probability distribution. Again it is possible to deduce an explicit expression for $P(t)$.

There are good reasons to assume that $\pi_{\mathtt{c}} = \pi_{\mathtt{g}}$ and $\pi_{\mathtt{a}} = \pi_{\mathtt{t}}$, since DNA is a molecule with two strands, made of pairs $\mathtt{c} : \mathtt{g}$ et $\mathtt{a} : \mathtt{t}$. The above identities are a consequence of the effect of this constraint on evolution. With this restriction, the HKY model becomes a three–parameter model, namely u, v and $\theta = \pi_{\mathtt{c}} + \pi_{\mathtt{g}}$, which has been proposed by Tamura in 1992. It takes the form

$$Q = \frac{1}{2} \begin{pmatrix} -u\theta - v & v\theta & u\theta & v(1-\theta) \\ v(1-\theta) & -u(1-\theta) - v & v\theta & u(1-\theta) \\ u(1-\theta) & v\theta & -u(1-\theta) - v & v(1-\theta) \\ v(1-\theta) & u\theta & v\theta & -u\theta - v \end{pmatrix}.$$

The general reversible model Since $|E|$ is very small, one can try to use the most general model. Tavare has proposed a parametrization of the most general model, of the form

$$Q = \begin{pmatrix} -uW & uA\pi_{\mathtt{c}} & uB\pi_{\mathtt{g}} & uC\pi_{\mathtt{t}} \\ uD\pi_{\mathtt{a}} & -uX & uE\pi_{\mathtt{g}} & uF\pi_{\mathtt{t}} \\ uG\pi_{\mathtt{a}} & uH\pi_{\mathtt{c}} & -uY & uI\pi_{\mathtt{t}} \\ uJ\pi_{\mathtt{a}} & uK\pi_{\mathtt{c}} & uL\pi_{\mathtt{g}} & -uZ \end{pmatrix},$$

where u is a positive number, π the invariant probability distribution,

$$\begin{aligned} W &= A\pi_{\mathtt{c}} + B\pi_{\mathtt{g}} + C\pi_{\mathtt{t}} \\ X &= D\pi_{\mathtt{a}} + E\pi_{\mathtt{g}} + F\pi_{\mathtt{t}} \\ Y &= G\pi_{\mathtt{a}} + H\pi_{\mathtt{g}} + I\pi_{\mathtt{t}} \\ Z &= J\pi_{\mathtt{a}} + K\pi_{\mathtt{c}} + L\pi_{\mathtt{g}}, \end{aligned}$$

and the parameters $A, B, \ldots, L$ are to be chosen. As we shall see below, it is useful for the computation of the likelihood that the model be reversible. The constraint of reversibility imposes six relations, namely

$$A = D,\ B = G,\ C = J,\ E = H,\ F = K,\ I = L.$$

There remain six parameters to choose, for example A, B, C, E, F and I. There are moreover 3 free parameters describing the invariant probability distribution. Thus 9 parameters need to be chosen.

Codon models A codon is a triplet of nucleotides, which codes for an amino acid. Among the $4^3 = 64$ possible codons, 3 are possible STOP codons, the 61 others code for the 20 amino acids. Note that the genetic code (the translation rule from codons to amino–acids) is *degenerate*, in the sense that several distinct codons code for the same amino–acid. Hence among the possible codon mutations, one must distinguish the synonymous mutations (which transform a codon into another one which codes for the same amino acid) from the non–synonymous mutations. The latter are either slowed down or favoured by selection, while the former accumulate at the rate of the mutations. In general, the ratio synonymous mutations / non–synonymous mutations is greater than 1.

Goldman and Yang proposed in 1994 a model with 63 parameters, namely 60 parameters for the frequencies π_{xyz}, the 3 remaining for the transition rate α, the transversion rate β, and the ratio

$$\omega = \text{non–synonymous mutation rate / synonymous mutation rate.}$$

The GY model can be written

$$Q_{(x_1y_1z_1)(x_2y_2z_2)} = \begin{cases} 0 & \text{if 1 and 2 differ by more than one base,} \\ \alpha\pi_{x_2y_2z_2} & \text{for a synonymous transition,} \\ \beta\pi_{x_2y_2z_2} & \text{for a synonymous transversion,} \\ \omega\alpha\pi_{x_2y_2z_2} & \text{for a non–synonymous transition,} \\ \omega\beta\pi_{x_2y_2z_2} & \text{for a non–synonymous transversion.} \end{cases}$$

Note that among the 63 parameters to be estimated, the 60 parameters which determine the invariant probability distribution π are usually estimated not by a maximum likelihood procedure, but from the empirical frequencies of the various codons present in our data. Another possibility is to estimate π_{xyz} by the product $\pi_x^1\pi_y^2\pi_z^3$ of the frequencies of the various nucleotides at positions 1, 2 and 3 of the codons.

Nonhomogeneous models An implicit assumption in all the Markovian models which we have considered so far is their stationarity. The infinitesimal generator is the same on all the branches of the phylogenetic tree. Hence the invariant probability distribution is the same on the various branches, which implies that the various sequences must have roughly the same composition in bases. Some data contradict this assumption. One can then relax the

homogeneity assumption of the Markov process on the whole tree. For example, Galtier and Gouy adopt the Tamura model, with parameters α and β which are constant on the tree, and a parameter θ (which regulates the proportion of $\mathsf{g} + \mathsf{c}$) which is allowed to vary from one branch to another.

Dependence or independence between sites Almost all Markovian models assume that the mutations happen at each site independently from the others. This assumption is of course not reasonable, but it makes the computations (in particular of the likelihood, see below) feasible.

There is so far very few work on Markovian models where the evolutions at the various sites are correlated.

Consider a model of the type

$$Q_{xy} = s_{xy}\pi_y,$$

which is a reversible model, provided $s_{xy} = s_{yx}$. Pollock, Taylor and Goldman model the evolution of a pair of proteins by choosing an infinitesimal generator of the form

$$\begin{aligned} Q_{xx',yx'} &= s_{xy}\overline{\pi}_{yx'}, \\ Q_{xx',xy'} &= s_{x'y'}\overline{\pi}_{xy'}, \\ Q_{xx',yy'} &= 0, \text{ if } x \neq y \text{ and } x' \neq y', \end{aligned}$$

where $\overline{\pi}$ is an invariant probability distribution on the set of pairs of proteins.

Variation of the evolution rate between the branches Given an infinitesimal generator Q, for all $u > 0$, uQ is also an infinitesimal generator. Suppose that Q is constant on the tree. If u is constant as well, since the leaves (the species living today) all are at the same distance from the common ancestor, located at the root of the tree (distances are measured in elapsed time), this means that we are making the assumption of a "molecular clock". Certain sets of data are incompatible with such an assumption. One should then, in order to use a model which is coherent with such data, allow the parameter u to take a different value on each branch of the tree. We then have a new parameter on each branch of the tree, which all together means a lot of parameters.

Another point of view consists in assuming that the u's are the values taken by a stochastic process, which evolves on the tree as a Markov process,

either in continuous time, or else in discrete time (in which case the process is constant on each branch, the transitions taking place at the nodes). Conditionally upon the values taken by that process, the various nucleotides evolve as a nonhomogeneous Markov process along the tree. We then are in a Bayesian framework, for which the "MCMC" algorithm (see sections 3.1 and 7.8.3) makes the requested simulations feasible.

Variation of the evolution rate between sites The most popular model for the variation of rate between sites is to assume that the rates associated to the various sites are i.i.d., the common distribution being a gamma distribution (or a discretized version of the same).

Another approach, due to Felsenstein and Churchill, consists in assuming that they form a Markov chain along the DNA sequence (which is "hidden"), taking its values in a set which, for practical reasons, is taken to be of very small cardinality.

"Covarion" models This is a name for model where the rate of evolution not only differs from one site to another, but also, at a given site, from one branch to another branch of the tree. Covarion stands for "COncomitantly VARIable codON".

Denote $E = \{\mathtt{a}, \mathtt{c}, \mathtt{g}, \mathtt{t}\}$, $G=$ the set of all possible values of the rate u. Galtier considers an independent $E \times G$–valued Markov process at each site.

7.8.2 Likelihood methods in phylogeny

Comparison of the genomes of the various species is at present the main tool for the reconstruction of phylogenetic trees. Several algorithms exist for the construction of such trees. We shall now give some comments concerning the maximum likelihood method.

Note that we can compare either genes (i. e. amino acid sequences) or DNA sequences. In order to fix ideas, we shall consider DNA sequences, which we assume to be already aligned.

Computation of the likelihood of a tree Suppose we use the Felsenstein model. Time t corresponds here to a distance along the tree. Note that the only parameter of interest is the product $u \times t$. If we modify the lengths of

the branches of the tree accordingly, we can and shall from now on assume that $u = 1$.

We shall consider below only *binary trees*.

For the remainder of this section, we shall assume that each of the various sites evolves independently from the others, and that all evolve at the same rate, this rate being constant along the tree. This assumption is not very realistic, and several recent works concentrate on the detection of those sites which evolve faster than the others, possibly only on a portion of the tree. However, this simplifying assumption is natural for starting the analysis and constructing a first tree. Another popular assumption would be that the evolution rates of the various sites are i. i. d. r. v.'s, with a common law gamma.

The information at our disposal, the *data*, consists in a set of k aligned sequences of length m, i. e. for each site s, $1 \leq s \leq m$, we have k letters from the alphabet `a, c, g, t`, one for each leaf of the tree. To each rooted binary tree T with k leaves, we shall associate its likelihood $L(T)$, which is a function of the data. The likelihood $L(T)$ is a product from $s = 1$ to m of the likelihoods associated to each site s :

$$L(T) = \prod_{s=1}^{m} L_s(T).$$

The computation of each factor $L_s(T)$ takes advantage of the Markov property, as we shall now see. Let T denote a rooted tree. We can for example code as follows the nodes of such a tree, starting from the root, towards the leaves (see figure 7.4) :

- 0 denotes the root;
- 1, 2 are the "sons" of the root, i. e. the nodes which are directly connected with the root by one branch;
- 1.1, 1.2 denote the sons of 1; 2.1, 2.2 those of 2;
- and so on up to the leaves.

For each node $\alpha \in T \backslash \{0\}$, denote by ℓ_α the length of the branch joining the "father" of α and α. We associate to α the set Λ_α of the leaves of the subtree whose root is α. In particular, Λ_0 denotes the set of the leaves of the tree.

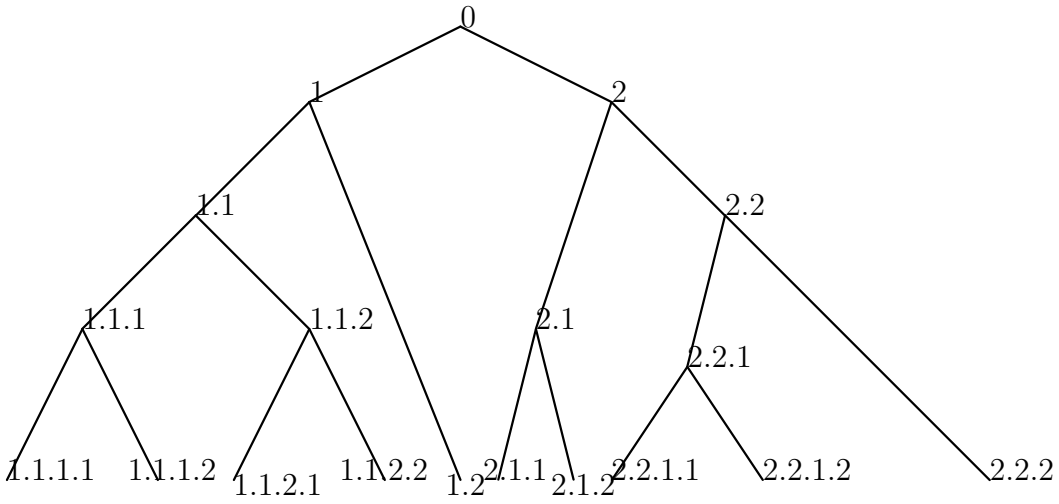

Figure 7.4: Rooted binary tree with coded nodes

If $\alpha \in \Lambda_0$, $\Lambda_\alpha = \{\alpha\}$. For $\alpha \in T \backslash \Lambda_0$, we denote by $\Gamma_\alpha = \{\alpha.1, \alpha.2\}$ the two "sons" of α.

Let $\{X_\alpha,\ \alpha \in T\}$ denote the nucleotides at the nodes of the tree. We assume that they are the values at those nodes of a Markov process on the tree whose infinitesimal generator is Q. Only the values of $\{X_\alpha,\ \alpha \in \Lambda_0\}$ are observed. We denote by x_α the observed value of X_α, for $\alpha \in \Lambda_0$. The likelihood of the tree, based upon the nucleotides at site s, is

$$L_s(T) = \mathbb{P}_T \left(\cap_{\alpha \in \Lambda_0} \{X_\alpha = x_\alpha\} \right).$$

We shall explain how to compute this quantity, and we shall exhibit how it depends upon the tree T.

For each $\alpha \in T$, $x \in E$, we define $L_{s,x}^{(\alpha)}$, the conditional likelihood of the subtree whose α is the root, conditioned upon $X_\alpha = x$, which we compute by the following upwards recurrence.

- For $\alpha \in \Lambda_0$,

$$L_{s,x}^{(\alpha)} = \begin{cases} 1, & \text{if } x = x_\alpha; \\ 0, & \text{otherwise.} \end{cases}$$

- In all other cases,

$$L_{s,x}^{(\alpha)} = \sum_{x_{\alpha.1}, x_{\alpha.2} \in E} P_{x x_{\alpha.1}}(\ell_{\alpha.1}) L_{s,x_{\alpha.1}}^{(\alpha.1)} \times P_{x x_{\alpha.2}}(\ell_{\alpha.2}) L_{s,x_{\alpha.2}}^{(\alpha.2)}.$$

This computation eventually specifies the quantities $L_{s,x}^{(0)}$, $x \in E$. Finally

$$L_s(T) = \sum_{x \in E} \pi_x L_{s,x}^{(0)},$$

and

$$L(T) = \prod_{s=1}^{m} L_s(T).$$

We could as well have described each $L_s(T)$ as a sum of $4^{|T \setminus \Lambda_0|}$ terms. But the above formulas describe the so–called pruning algorithm, due to Felsenstein, which should be used in practice.

Maximum likelihood The computation of a global maximum of the likelihood over all possible trees is complex. The easiest part is the maximization over branch lengths. However, it is not clear that the algorithm which is commonly used for that sub–problem leads to a global maximum. The idea is to maximize successively over each branch length, and to iterate as long as the likelihood increases. We shall now see explicitly how the likelihood depends upon a particular branch length. It is then easily seen how to maximize the likelihood with respect to that particular branch length.

We shall assume that the values $\{X_\alpha,\ \alpha \in T\}$ are the values at the nodes of the tree of a *reversible* Markov process. Hence the law of the $\{X_\alpha\}$'s does not depend upon the choice of of a root at any node of the tree (or more generally anywhere on the tree).

Consider two neighbouring nodes α and β in the tree. Denote by $\ell_{\alpha\beta}$ the length of the branch which connects them. If we put the root anywhere on that branch,we define as above the quantities $L_{s,x}^{(\alpha)}$ and $L_{s,y}^{(\beta)}$, $x, y \in E$. Note that the first term is the likelihood of a sub–tree whose root is α, taken as

a subtree whose root has been displaced. For example in the tree of figure 7.4, if $\alpha = 2$, $\beta = 2.2$, and we locate the root between α and β, the sub–tree whose root is α contains the nodes 2, 0, 1, 1.1, 1.1.1, 1.1.1.1, 1.1.1.2, 1.1.2, 1.1.2.1, 1.1.2.2, 1.2, 2.1, 2.1.1, 2.1.2, and the sub–tree whose root is β contains the nodes 2.2.1, 2.2.1.1, 2.2.1.2, 2.2.2.

Then

$$\begin{aligned} L_s(T) &= \sum_{x,y\in E} \pi_x P_{xy}(\ell_{\alpha\beta}) L_{s,x}^{(\alpha)} L_{s,y}^{(\beta)} \\ &= \sum_{x,y\in E} \pi_y P_{yx}(\ell_{\alpha\beta}) L_{s,x}^{(\alpha)} L_{s,y}^{(\beta)}. \end{aligned}$$

This formula makes explicit the dependence of $L_s(T)$ and $L(T)$ upon the length of a given branch, and allows us to compute the maximum of the likelihood with respect to that branch. The search for that maximum is rather easy in the case of the model of evolution which we have described above (on maximizes the logarithm of $L(T)$, which replaces the product of the $L_s(T)$'s by a sum, and thus simplifies the maximization).

Remark 7.8.1. *Not all models of evolution are reversible. It is still possible to give explicitly the dependence of the likelihood with respect to the length of a given branch, but one has to be careful using the transition probability of the time–reversed process whenever the displacement of the root makes the process starting from the new root run along a branch in the direction which the inverse of the inital one.*

7.8.3 The Bayesian approach to phylogeny

Let us go back to the expression for the likelihood. Denote by D the vector of the observed random variables, and by d the vector of the observed values (d like "data"), i. e. d contains the various aligned sequences.

We now describe the various parameters upon which the likelihood depends. Among the unknown parameters (which we want to estimate), we have

• on one hand the shape of the tree, which we shall denote by τ, which is an unknown in a finite set $\mathcal{T}$ (whose cardinal is $(2n-3)!!$ in the case of a rooted tree with n leaves, $(2n-5)!!$ in the case of an unrooted tree – where we have used the notation $k!! = 1 \times 3 \times 5 \times \cdots \times k$ for an odd number k),

• on the other hand the lengths of the various branches, and the infinitesimal generator Q of the model of evolution (at least the parameters of that matrix other than the invariant probability disribution). The branch lengths and the unknown parameters of the matrix Q vary in a subset of an Euclidian space $V \subset \mathbb{R}^d$. We shall denote by λ this set of parameters.

Thus the unknown parameter is the pair $\theta = (\tau, \lambda)$, whose value is arbitrary in the set $\Theta = \mathcal{T} \times V$, and the likelihood is the function

$$L(\theta) = \mathbb{P}_\theta(D = d).$$

The likelihood of the value θ of the unknown parameter is the probability of observing the data which we have in our computer, if θ is the true of that parameter.

In the Bayesian framework, the unknown parameter θ is the realization of a random variable, that is (τ, λ) is the realization of a random vector (T, Λ). This point of view forces us to choose an *a priori probability distribution*, which is an occasion to incorporate a priori information about the unknown parameter, which the anti–Bayesians refuse to do, since they claim that the only information which one should use is that contained in the data.

We shall thus give us an *a priori* probability distribution for the random vector (T, Λ), which we will take of the form :

• we specify the law of T, which is a probability distribution on the finite set $\mathcal{T}$, hence we specify the $\alpha_\tau = \mathbb{P}(T = \tau)$'s, $\tau \in \mathcal{T}$;

• we specify the conditional law of Λ, given T, and we assume that for all $\tau \in \mathcal{T}$, the conditional law of Λ, given that $T = \tau$, has a density $q_\tau(\lambda)$, in other words for any Borel measurable function

$$f : \mathcal{T} \times V \ \to \ \mathbb{R}_+,$$

$$\mathbb{E}[f(T, \Lambda)] = \sum_{\tau \in \mathcal{T}} \int_V f(\tau, \lambda) p_\tau(\lambda) d\lambda,$$

if we use the notation $p_\tau(\lambda) = \alpha_\tau \times q_\tau(\lambda)$.

In this context, we have a random pair, consisting of both a "parameter" (T, Λ), and the data D. The law of this pair is specified by

- on the one hand the a priori law of (T, Λ);

- on the other hand the conditional law of the data, given the parameter.

More precisely, in this Bayesian framework, the likelihood is interpreted as the conditional probability distribution of the data, given the parameters :

$$L(\tau, \lambda) = \mathbb{P}\left(D = d | (T, \Lambda) = (\tau, \lambda)\right).$$

The rule of the game is to compute the *a posteriori* law of the parameter, which is the conditional law of the "parameter" (T, Λ), given the data, i. e. given that $D = d$. This conditional probability distribution is given by the famous "Bayes formula", which in our case specifies the joint law of (T, Λ) given that $D = d$ in the form

$$p_\tau(\lambda | D = d) = \frac{\mathbb{P}\left(D = d | (T, \Lambda) = (\tau, \lambda)\right) p_\tau(\lambda)}{\sum_{\tau \in \mathcal{T}} \int_V \mathbb{P}\left(D = d | (T, \Lambda) = (\tau, \lambda)\right) p_\tau(\lambda) d\lambda}.$$

In other words, again if

$$f : \mathcal{T} \times V \ \rightarrow \ \mathbb{R}_+,$$

$$\mathbb{E}\left[f(T, \Lambda) | D = d\right] = \frac{\sum_{\tau \in \mathcal{T}} \int_V f(\tau, \lambda) \mathbb{P}\left(D = d | (T, \Lambda) = (\tau, \lambda)\right) p_\tau(\lambda) d\lambda}{\sum_{\tau \in \mathcal{T}} \int_V \mathbb{P}\left(D = d | (T, \Lambda) = (\tau, \lambda)\right) p_\tau(\lambda) d\lambda}.$$

For example, we might want to specify the *a posteriori* probability distribution of the shape of the tree, i. e. of the r. v. T. It is given by the next formula : for all $\tau \in \mathcal{T}$,

$$\mathbb{P}\left(T = \tau | D = d\right) = \frac{\int_V \mathbb{P}\left(D = d | (T, \Lambda) = (\tau, \lambda)\right) p_\tau(\lambda) d\lambda}{\sum_{\tau \in \mathcal{T}} \int_V \mathbb{P}\left(D = d | (T, \Lambda) = (\tau, \lambda)\right) p_\tau(\lambda) d\lambda}.$$

The MCMC algorithm Suppose that we want to compute the last quantity for a small number of values of τ. An explicit computation is hopeless, because of the size of the data (the number of species which we consider) and the complexity of the models which we might use. One is thus led to use a Monte Carlo type of method, using random draws. However, it is not really possible to simulate under the a posteriori probability distribution of (T, Λ), given the data. Indeed, in order to identify this probability distribution, it would be necessary to compute the denominator in the formula above. As soon as the cardinality of $\mathcal{T}$ is huge, this task becomes impossible.

We are exactly in the situation described at the beginning of section 3.1.

Let us recall the Metropolis–Hastings algorithm. Denote by π the a posteriori probability distribution. Let Q denote a transition matrix on F (which has nothing to do with the probability distribution π), whose transitions are easy to simulate. We choose as transition matrix P the matrix whose off–diagonal entries read

$$P_{xy} = \min\left(Q_{xy}, \frac{\pi_y}{\pi_x} Q_{yx}\right),$$

and whose diagonal entries are given by

$$P_{xx} = 1 - \sum_{y \neq x} P_{xy},$$

provided the matrix P thus defined is irreducible, which is the case if for instance Q is irreducible and satisfies the property : for all x, y, $Q_{xy} > 0 \Leftrightarrow Q_{yx} > 0$. This matrix P is clearly a transition matrix ($P_{xy} \leq Q_{xy}$, $x \neq y$ implies that $P_{xx} \geq 0$), and π is P–invariant, since the detailed balance equation

$$\pi_x P_{xy} = \pi_y P_{yx}, \quad \forall x \neq y$$

holds. For $x, y \in F$ let,

$$r(x, y) = \frac{P_{xy}}{Q_{xy}} = \min\left(1, \frac{\pi_y Q_{yx}}{\pi_x Q_{xy}}\right).$$

One way of simulating a transition of the chain $\{X_k\}$ with the transition matrix P is as follows. Suppose that $X_k = x$, and we wish to simulate X_{k+1}. We first simulate a transition of the chain $\{Y_k\}$ with the transition matrix Q, starting from $Y_k = x$. Suppose that the result of this simulation is $Y_{k+1} = y$. We *accept this transition* (and in this case $X_{k+1} = y$) with probability $r(x, y)$; we *reject this transition* (and in this case $X_{k+1} = x$) with probability $1 - r(x, y)$. Note that $r(x, x) = 1$, hence whenever $y = x$, $X_{k+1} = x$.

In other words the transition from $X_k = x$ to X_{k+1} is computed as follows

- we draw a realization Y_{k+1} of the probability distribution $Q_{x\cdot}$;
- we draw U_{k+1} with the uniform law on $[0, 1]$;

and we let

$$X_{k+1} = Y_{k+1} \mathbf{1}_{\{U_{k+1} \leq r(x, Y_{k+1})\}} + X_k \mathbf{1}_{\{U_{k+1} > r(x, Y_{k+1})\}}.$$

Implementation of the MCMC algorithm The implementation of the MCMC algorithm poses delicate questions, for which we essentially have no really satisfactory answer, in particular in the application to phylogeny. We have already discussed that issue in a general framework at section 3.3. Recall that one should eliminate the first simulations (*burn–in*). Moreover, in order to obtain a sample of the a posteriori law, one keeps only one iteration among n, where the choice of n depends upon the speed of decorrelation of the chain, which might be estimated from simulations. Some of the implementations involve the simulation of several chains in parallel, some of them being possibly "heated", see section 3.1.4.

7.9 Application to discretized partial differential equations

Let D be a bounded domain in $\mathbb{R}^2$ (we could as well treat a problem in higher dimensions), whose boundary ∂D is Lipschitz continuous. Suppose that $0 \in D$. Consider the Dirichlet problem :

$$\begin{cases} \Delta u(x) = 0, \ x \in D; \\ \quad u(x) = f(x), \ x \in \partial D; \end{cases}$$

where $f \in C(\partial D)$. It is well known that this equation has a unique solution u in $C(D)$.

Given $h > 0$, let $h\mathbf{Z}^2$ denote the set of points of the plane whose coordinates are multiples of h. Define $D_h = D \cap h\mathbf{Z}^2$. ∂D_h is made of those points in $D^c \cap h\mathbf{Z}^2$ which are at distance h from at least one point in D_h, and $\bar{D}_h = D_h \cup \partial D_h$. Let e_1 and e_2 the two vectors of an orthonormal basis. We define the approximate operator Δ_h as :

$$(\Delta_h v)(x) = \frac{1}{4} \sum_{i=1}^{2} (v(x + he_i) + v(x - he_i)) - v(x).$$

From the Exercise 7.11.4, the solution of the discretized Dirichlet problem

$$\begin{cases} \Delta_h u_h(x) = 0, \ x \in D_h; \\ \quad u_h(x) = f(x), \ x \in \partial D_h; \end{cases}$$

is given by the formula $u_h(x) = \mathbb{E}[f(X^h_{T^h_{D^c_h}})|X^h_0 = x]$, where $\{X^h_t; t \geq 0\}$ is an $h\mathbf{Z}^2$–valued jump Markov process with the infinitesimal generator $\frac{1}{2}\Delta_h$, and

$$T^h_{D^c_h} = \inf\{t \geq 0; X^h_t \in D^c_h\}.$$

Note that $\{X^h_t, t \geq 0\}$ has the same law as $\{hX^1_{h^{-2}t}, t \geq 0\}$, and we explained just after Theorem 7.6.9, that this process converges towards a standard one dimensional Brownian motion, as $h \to 0$. It is not too hard to deduce that

$$u_h(x) \to u(x) = \mathbb{E}[f(B_{T_{D^c}})|B_0 = x],$$

(where $T_{D^c} = \inf\{t \geq 0; B_t \in D^c\}$). This formula gives a probabilistic interpretation of the Dirichlet problem. We then have a proof (which is an alternative to classical arguments from numerical analysis) of the convergence of u_h towards u.

Note that the discretized Dirichlet problem could also be interpreted in terms of a discrete time Markov chain with the transition matrix $\Delta_h + I$.

Such a probabilistic interpretation justifies the use of Monte Carlo numerical methods for the approximate computation of solutions of PDE's. These methods are mainly used in those cases where "classical" numerical analysis algorithms cannot be used (in particular in cases of high dimensional problems), see [33]. They are also very much appreciated for the simplicity with which the associated programs can be written. A few lines of code are sufficient for programming the computation of the approximate solution of a partial differential equation ! Even if we need to let the computer run a bit longer than would be needed to obtain the same precision with a finite difference, finite element or finite volume method, the fact that the program is very easy to write is greatly appreciated by users, especially in situations where a program writen by one person might have to be modified later by another.

7.10 Simulated annealing (sequel to section 3.4)

In this section, E is assumed to be finite. Recall that we want to maximize a function

$$U : E \to \mathbb{R}_-,$$

such that

$$\max_{x \in E} U_x = 0.$$

In other words, we look for one of the x's which are such that $U_x = 0$.

To each $\beta > 0$ we associate the infinitesimal generator $Q = \{Q_{xy}, x, y \in E\}$, where for $x \neq y$

$$Q_{xy} = \mathbf{1}_{\{(x,y) \in G\}} \exp\left[\frac{\beta}{2}(U_y - U_x)\right],$$

G is a non–oriented graph in E, i. e. a collection of pairs of points from E, chosen in such a way that the jump Markov process with infinitesimal

generator Q is irreducible (which means that $\forall x, y \in E$, $\exists n$ and $x_1, x_2, \ldots, x_n$ such that $(x, x_1) \in G$, $(x_1, x_2) \in G, \ldots, (x_n, y) \in G$).

The jump Markov process with the infinitesimal generator Q is clearly reversible with respect to its invariant probability distribution π_β defined by

$$\pi_{\beta,x} = Z_\beta^{-1} e^{\beta U_x}, \quad x \in E.$$

We define the **Dirichlet form** associated to Q as the bilinear form on $\mathbb{R}^E$:

$$\begin{aligned}\mathcal{E}(\varphi, \varphi) &=< \varphi, -Q\varphi >_\pi \\ &= -\sum_{x,y} \varphi_x Q_{xy} \varphi_y \pi_x \\ &= \frac{1}{2} \sum_{x,y} |\varphi_x - \varphi_y|^2 Q_{xy} \pi_x,\end{aligned}$$

where we have exploited reversibility and twice the identity $\sum_y Q_{xy} = 0$, hence $-Q : \ell^2(\pi) \to \ell^2(\pi)$ is a selfadjoint positive semi definite operator.

Definition 7.10.1. *We call the quantity*

$$\lambda \stackrel{def}{=} \inf_{\varphi \ non\text{–}constant} \frac{\mathcal{E}(\varphi, \varphi)}{Var_\pi(\varphi)}$$

the **spectral gap** *of Q, where* $Var_\pi(\varphi) = \sum_{x \in E} \varphi_x^2 \pi_x - \left(\sum_{x \in E} \varphi_x \pi_x \right)^2$.

Lemma 7.10.2. *Since Q is the infinitesimal generator of an irreducible jump Markov process with values in a finite set E, its spectral gap is strictly positive.*

Proof From the above formula for $\mathcal{E}(\varphi, \varphi)$, the ratio

$$\frac{\mathcal{E}(\varphi, \varphi)}{Var_\pi(\varphi)}$$

is not modified if we add a constant to φ. We then can minimize this ratio over those φ which are such that $\mathbb{E}_\pi(\varphi) = 0$, whence

$$\lambda = \inf_{\varphi \not\equiv 0; \mathbb{E}_\pi(\varphi) = 0} \frac{\langle \varphi, -Q\varphi \rangle_\pi}{\langle \varphi, \varphi \rangle_\pi},$$

and λ is the smallest eigenvalue of $-Q$, considered as a linear operator on $\ell^2(\pi)$, restristed to the sub–vector space orthogonal to the constants. Since Q is a selfadjoint positive semi definite operator, it suffices to show that the eigenspace associated to the eigenvalue 0 is the set of constant functions. But if φ belongs to that eigenspace,

$$\sum_y Q_{xy}\varphi_y = 0, \ \forall x \in E,$$

then a fortiori

$$0 = -\sum_{x,y} \varphi_x Q_{xy}\varphi_y \pi_x = \frac{1}{2}\sum_{x,y} |\varphi_x - \varphi_y|^2 Q_{xy}\pi_x,$$

which, since Q is irreducible, does imply that φ is constant. □

Now let $\{X_t, t \geq 0\}$ be a jump Markov process with infinitesimal generator Q. For each $t > 0$, let $\mu(t) = (\mu_x(t); x \in E)$ denote the law of X_t. We let

$$\varepsilon(t) = \sum_{x\in E} \left(\frac{\mu_x(t)}{\pi_x} - 1\right)^2 \pi_x,$$

and remark that $\varepsilon(t) = 0$ if and only if $\mu(t) = \pi$.

Lemma 7.10.3. *λ denoting the spectral gap of Q,*

$$\varepsilon(t) \leq \varepsilon(0)e^{-2\lambda t}.$$

Proof We first remark that

$$\begin{aligned}\varepsilon(t) &= \sum_x \left(\frac{\mu_x(t)}{\pi_x} - 1\right)^2 \pi_x \\ &= \sum_x \left(\frac{\mu_x(t)}{\pi_x}\right)^2 \pi_x - 1.\end{aligned}$$

Hence

$$\begin{aligned}\frac{d\varepsilon}{dt}(t) &= 2\sum_x \frac{\mu_x(t)\mu_x'(t)}{\pi_x} \\ &= 2\sum_{x,y} \frac{\mu_x(t)\mu_y(t)Q_{yx}}{\pi_x} \\ &= 2\sum_{x,y} \frac{\mu_x(t)}{\pi_x} \times \frac{\mu_y(t)}{\pi_y} \times \pi_y Q_{yx} \\ &= 2 < Q\left(\frac{\mu(t)}{\pi}\right), \frac{\mu(t)}{\pi} >_\pi \\ &\leq -2\lambda Var_\pi\left(\frac{\mu(t)}{\pi}\right) \\ &= -2\lambda\varepsilon(t),\end{aligned}$$

and consequently

$$\frac{d}{dt}\log\varepsilon(t) \leq -2\lambda.$$

□

We have shown that $\mu_t \to \pi$ at exponential speed (compare with Theorem 2.6.8).

We now get to "annealing". We shall now let β depend upon t, and let it go to infinity (and hence the "temperature", its inverse, towards zero) as $t \to \infty$. More precisely, Δ being a constant which will be specified below, we choose

$$\beta(t) = \frac{1}{\Delta}\log(1+t),$$

hence $\beta(0) = 0$ and $\beta(t) \to +\infty$ as $t \to \infty$. Of course, the chain is no longer time homogeneous, since the infinitesimal generator Q, the spectral gap λ, the invariant measure π, and the normalization constant Z all become functions of t : $Q(t)$, $\lambda(t)$, $\pi(t)$, $Z(t)$. Note that $\pi(0)$ is the uniform measure on E, $Z(0) = |E|^{-1}$, while $\pi(\infty) = \lim_{t\to\infty} \pi(t)$ is the uniform measure on the zeros of U (i. e. on the maxima of U).

Let $M \stackrel{\text{def}}{=} \sup_{x\in E}(-U_x)$. Our goal is to show:

Theorem 7.10.4. *If $\Delta > M$, then $\mu(t) \to \pi(\infty)$ as $t \to \infty$.*

We first establish :

Lemma 7.10.5.

$$\lambda(t) \geq \lambda(0) \left(\frac{1}{1+t} \right)^{\frac{M}{\Delta}}.$$

PROOF Choose φ such that $\mathbb{E}_{\pi(0)}[\varphi] = 0$.Then from the definition of $\lambda(0)$,

$$\frac{1}{2} \sum_{x,y} |\varphi_x - \varphi_y|^2 Q_{xy}(0) \pi_x(0) \geq \lambda(0) \sum_x \varphi_x^2 \pi_x(0).$$

On the other hand,

$$\begin{aligned} Q_{xy}(t)\pi_x(t) &= Q_{xy}(0) \frac{e^{\beta(t)\frac{U_x+U_y}{2}}}{Z(t)} \\ &\geq Q_{xy}(0) \frac{e^{-\beta(t)M}|E|}{Z(t)} \pi_x(0) \end{aligned}$$

Hence

$$\begin{aligned} \mathcal{E}_t(\varphi,\varphi) &= \frac{1}{2} \sum_{x,y} |\varphi_x - \varphi_y|^2 Q_{xy}(t)\pi_x(t) \\ &\geq \frac{e^{-\beta(t)M}|E|}{2Z(t)} \sum_{x,y} |\varphi_x - \varphi_y|^2 Q_{xy}(0)\pi_x(0) \\ &\geq \lambda(0) e^{-\beta(t)M} \sum_x \varphi_x^2 \frac{1}{Z(t)} \\ &\geq \lambda(0) e^{-\beta(t)M} \sum_x \varphi_x^2 \pi_x(t) \\ &\geq \lambda(0) e^{-\beta(t)M} Var_{\pi(t)}(\varphi), \end{aligned}$$

where we have used for the third inequality the fact that $U_x \leq 0$, hence $Z^{-1}(t) \leq \pi_x(t)$. The two extreme expressions of this set of inequalities being invariant under the addition of a constant to φ, the resulting inequality still holds without the restriction that $\mathbb{E}_{\pi(0)}[\varphi] = 0$. The Lemma is etablished, since

$$e^{-\beta(t)M} = \left(\frac{1}{1+t} \right)^{\frac{M}{\Delta}}.$$

□

PROOF OF THE THEOREM It suffices to show that $\varepsilon(t) \to 0$, where

$$\begin{aligned}\varepsilon(t) &= \sum_x \left(\frac{\mu_x(t)}{\pi_x(t)} - 1\right)^2 \pi_x(t) \\ &= \sum_x \frac{\mu_x(t)^2}{\pi_x(t)} - 1.\end{aligned}$$

Note that $\varepsilon(t)$ is an upper bound on the square of the L^1– norm of the difference $\mu(t) - \pi(t)$. Indeed, from Cauchy–Schwarz,

$$\sum_x |\mu_x(t) - \pi_x(t)| = \sum_x \frac{|\mu_x(t) - \pi_x(t)|}{\sqrt{\pi_x(t)}} \sqrt{\pi_x(t)} \leq \sqrt{\varepsilon(t)}.$$

We have that

$$\begin{aligned}\frac{d\varepsilon}{dt}(t) &= \sum_x \frac{d}{dt}\left[\frac{\mu_x(t)^2}{\pi_x(t)}\right] \\ &= \sum_x \pi_x(t)^{-2}\left[2\mu_x(t)\frac{d\mu_x}{dt}(t)\pi_x(t) - \mu_x(t)^2\frac{d\pi_x}{dt}(t)\right] \\ &= -2\mathcal{E}_t\left(\frac{\mu(t)}{\pi(t)}, \frac{\mu(t)}{\pi(t)}\right) - \beta'(t)\sum_x U_x \frac{\mu_x^2(t)}{\pi_x^2(t)}\pi_x(t) \\ &+ \beta'(t)\sum_{x,y} \frac{U_y e^{\beta(t)U_y}}{Z^2(t)} \times \frac{\mu_x^2(t)}{\pi_x^2(t)} e^{\beta(t)U_x} \\ &\leq -2\lambda(t)\varepsilon(t) + \beta'(t)M(\varepsilon(t)+1) \\ &\leq -(2\lambda(t) - M(\beta'(t))\varepsilon(t) + M\beta'(t) \\ &\leq -\left(\frac{2\lambda(0)}{(1+t)^{M/\Delta}} - \frac{M}{\Delta(1+t)}\right)\varepsilon(t) + \frac{M}{\Delta(1+t)}\end{aligned}$$

Since $\Delta > M$, $(1+t)^{-M/\Delta} >> (1+t)^{-1}$ as $t \to \infty$, and there exists $c > 0$ such that, for t large enough,

$$\frac{d\varepsilon}{dt}(t) \leq -c(1+t)^{-M/\Delta}\varepsilon(t) + \frac{M}{\Delta}(1+t)^{-1},$$

and the Theorem follows from

Lemma 7.10.6. *Let $x, b \in C^1(\mathbb{R}_+;\mathbb{R}_+)$, $a \in C(\mathbb{R}_+;\mathbb{R}_+)$ be such that*

(i) $\displaystyle\int_0^\infty a(t)\,dt = +\infty;$

(ii) $b(t) \searrow 0$ *as* $t \to \infty;$

(iii) $\dfrac{dx}{dt}(t) \le -a(t)(x(t) - b(t)).$

Then $x(t) \to 0$ *as* $t \to \infty$.

Proof

$$\frac{d}{dt}\left(x(t)\exp\left(\int_0^t a(s)ds\right)\right) = e^{\int_0^t a(s)ds}\left(\frac{dx}{dt}(t) + a(t)x(t)\right)$$
$$\le e^{\int_0^t a(s)ds}a(t)b(t).$$

Integrating, we obtain

$$x(t) \le x(0)e^{-\int_0^t a(s)ds} + \int_0^t e^{-\int_s^t a(r)dr}a(s)b(s)ds.$$

The right hand side of this inequality is the solution of a linear ODE, which majorizes $x(t)$. Hence it suffices to establish the result with $x(t)$ solution of the ODE, in other words we can as well assume that :

$$\frac{dx}{dt}(t) = -a(t)(x(t) - b(t)).$$

Let $y(t) = x(t) - b(t)$. It suffices to show that $y(t) \to 0$ as $t \to \infty$.

$$\frac{dy}{dt}(t) = -a(t)y(t) - b'(t).$$

Note that $b'(t) \le 0$, and $-\displaystyle\int_t^\infty b'(s)ds = b(t) < \infty$. Hence for $t \ge N$,

$$y(t) = e^{-\int_0^t a(s)ds}y(0) - \int_0^t e^{-\int_s^t a(r)dr}b'(s)ds$$
$$\le e^{-\int_0^N a(s)ds}y(0) + \int_N^\infty |b'(s)|ds + e^{-\int_N^t a(r)dr}\int_0^N e^{-\int_s^N a(r)dr}|b'(s)|ds.$$

Let $\delta > 0$ be arbitrary. We choose N large enough such that the sum of the two first terms of the right hand side be smaller than $\frac{\delta}{2}$. Choosing now t large enough, the third term is smaller than $\frac{\delta}{2}$. The lemma is etablished. □

Remark 7.10.7. *The function $\beta(t) = \frac{1}{\Delta}\log(1+t)$ tends too slowly to infinity as $t \to \infty$ to be used in practice. One can prove some results which are weaker than $\mu(t) \to \pi(\infty)$ with a function β which grows faster than a logarithm (a power function). If on the other hand we ask how to achieve the best possible result on a fixed finite horizon, one can show that some β's growing at exponential speed are close to the optimum.*

7.11 Exercises

Exercise 7.11.1. *Let $\{T_n, n \geq 1\}$ be a Poisson point process with intensity λ, and $\{Z_n, n \geq 0\}$ be an E–valued Markov chain which is independent of $\{T_n, n \geq 1\}$, with transition matrix $P_{xy}, x, y \in E$. We let*

$$X_t = \sum_{n=0}^{\infty} Z_n \mathbf{1}_{[T_n, T_{n+1}[}(t), \;\; t \geq 0.$$

Show that $\{X_t, t \geq 0\}$ is a jump Markov process, and give its transition matrices, its infinitesimal generator, and the law of its first jump time.

Exercise 7.11.2. *Let $\{X_t; t \geq 0\}$ be a jump Markov process with values in a finite or countable state space E, with infinitesimal generator $\{Q_{xy}; x, y \in E\}$. Assume that $\lambda := \sup_x -Q_{xx} < \infty$. Let $\{N_t; t \geq 0\}$ denote the counting process of the jumps of $\{X_t\}$, and $\{N'_t, t \geq 0\}$ a Poisson process with intensity λ.*

Compare $\mathbb{P}(N_t \geq n)$ and $\mathbb{P}(N'_t \geq n)$, $\mathbb{E}[f(N_t)]$ and $\mathbb{E}[f(N'_t)]$, where the function f is increasing from $\mathbb{N}$ into $\mathbb{R}$. Show that exercise 7.11.1 gives another proof of that result.

Exercise 7.11.3. *Let $\{N_t, t \geq 0\}$ and $\{P_t, t \geq 0\}$ be two mutually independent Poisson processes, with intensities λ and μ respectively.*

1. *Show that* $\{X_t,\, t \geq 0\}$ *defined by*

$$X_t = N_t - P_t$$

is a $\mathbf{Z}$*– valued irreducible jump Markov process, and give its infinitesimal generator.*

2. *We assume that* $\lambda \neq \mu$*. Show that* $\{X_t/t\}$ *and* $\{X_t\}$ *converge a. s. in* $\bar{\mathbb{R}}$ *as* $t \to \infty$*. What is the limit of* $\{X_t\}$*, depending upon the sign of* $\lambda - \mu$ *? Show that* $\{X_t\}$ *is transient.*

3. *Assume that* $\lambda = \mu$*. Give the transition matrix of the embedded chain. Deduce from exercises 2.10.11 and 2.10.13 that* $\{X_t\}$ *is null recurrent.*

Exercise 7.11.4. *1. Let* $\{X_t, t \geq 0\}$ *be an* E*–valued jump Markov process, with infinitesimal generator* $\{Q_{xy}; x, y \in E\}$*. Let* $F \subset E$*. Define*

$$T_F = \begin{cases} \inf\{t; X_t \in F\}, & \text{if such a } t \text{ exists;} \\ \infty, & \text{otherwise,} \end{cases}$$

the function $u : E \to \mathbb{R}$ *by*

$$u(x) = \mathbb{E}[h(X_{T_F})\mathbf{1}_{\{T_F<\infty\}}|X_0 = x],$$

where h *is a bounded mapping from* F *into* $\mathbb{R}$*, and the function* $v : E \to \mathbb{R} \cup \{+\infty\}$ *by*

$$v(x) := \mathbb{E}[T_F|X_0 = x].$$

Show that T_F *is a stopping time.*

Show that u *and* v*solve respectively the equations :*

$$\begin{aligned} Qu(x) &= 0, x \in E\backslash F \\ u(x) &= h(x), x \in F; \end{aligned}$$

$$\begin{aligned} Qv(x) + 1 &= 0, x \in E\backslash F \\ v(x) &= 0, x \in F; \end{aligned}$$

(Hint : condition upon $(T_1, X(T_1))$*).*

2. *Consider next the case of an $E = \mathbf{Z}$–valued birth and death process,whose infinitesimal generator Q satisfies $Q_{x,x+1} = \alpha(x)$, $Q_{x,x-1} = \beta(x)$, $Q_{x,x} = -\alpha(x) - \beta(x)$, in the particular case $\alpha(x) = \alpha$, $\beta(x) = \beta$, $x \in \mathbf{Z}$ $(\alpha, \beta > 0)$. Let $F = \{1, 2, \ldots, N-1\}^c$, where N is a positive integer. Compute $u(x) = \mathbb{E}[X_{T_F}|X_0 = x], x \in \mathbf{Z}$. Show that T_F is a. s. finite. Give the conditional law of the r. v. X_{T_F}, given that $X_0 = x$.*

Exercise 7.11.5. *Given a probability space $(\Omega, \mathcal{F}_t, P)$ equipped with a filtration $\{\mathcal{F}\}$ (i. e. an increasing collection indexed by $t \in \mathbb{R}_+$ of sub–σ–fields of $\mathcal{A}$), a* **martingale** *(with respect to the filtration $\{\mathcal{F}_t\}$) is a stochastic process $\{M_t, t \in \mathbb{R}_+\}$ which satisfies :*

$$M_t \text{ is integrable }, \forall t \geq 0; \ \mathbb{E}[M_t|\mathcal{F}_s] = M_s, \ \forall 0 \leq s \leq t.$$

1. *Let $\{M_t, t \in \mathbb{R}_+\}$ be a martingale which is continuous on the right, and S a stopping time which is bounded by a constant t. Show that $\mathbb{E}[M_S] = \mathbb{E}[M_t] = \mathbb{E}[M_0]$.*

2. *Let $\{X_t, t \geq 0\}$ be an E–valued jump Markov process, with the infinitesimal generator $\{Q_{xy}; x, y \in E\}$ satisfying $\sup_x Q_{xx} < \infty$, and f be a bounded mapping from E into $\mathbb{R}$. Show that $\{M_t, t \in \mathbb{R}_+\}$ defined by*

$$M_t = f(X_t) - \int_0^t Qf(X_s)ds$$

is a martingale with respect to the filtration $\{\mathcal{F}_t^X\}$ (accept the fact that for any bounded function f from E into $\mathbb{R}$, and any $r > 0$, $P_r Qf = QP_r f$).

3. *Use again the notation from the second part of the previous exercise. Compute $\mathbb{E}(T_F|X_0 = x)$ in terms of the law of X_{T_F} (Hint : in the case $\alpha \neq \beta$, use the results of the two preceding questions with the function $f(x) = x$ and the stopping time $S = \inf(T_F, t)$, the let t tend to infinity; in the case $\alpha = \beta$, do the same computations with $f(x) = x^2$). You should assume that the result in question 2. applies to those two functions, even though they are not bounded.*

Exercise 7.11.6. *Let $\{X_t;\, t \geq 0\}$ be an $\mathbb{N}$–valued jump Markov process, with infinitesimal generator*

$$Q = \begin{pmatrix} -\mu & \mu & 0 & 0 & 0 & \dots \\ \lambda & -(\lambda+\mu) & \mu & 0 & 0 & \dots \\ 0 & \lambda & -(\lambda+\mu) & \mu & 0 & \dots \\ 0 & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

where $\lambda,\, \mu > 0$.

1. *Specify the embedded chain. Show that $\{X_t;\, t \geq 0\}$ is irreducible.*
2. *Show that $\{X_t;\, t \geq 0\}$ is recurrent in the case $\lambda \geq \mu$, transient in the case $\lambda < \mu$.*
3. *Show that whenever $\lambda > \mu$, $\{X_t;\, t \geq 0\}$ possesses a unique invariant probability distribution π which you should specify.*
4. *Show that $\{X_t;\, t \geq 0\}$ is positive recurrent in the case $\lambda > \mu$, null recurrent in the case $\lambda = \mu$.*

Exercise 7.11.7. *With P denoting the transition matrix defined at exercise 2.10.9, we let $Q = P - I$, and consider a continuous time jump Markov process $\{X_t,\, t \geq 0\}$ whose infinitesimal generator is Q.*

1. *Give the transition matrix P' of the embedded chain.*
2. *Describe the trajectories of the process $\{X_t\}$, and specify the parameters of the exponential laws of the time spent in the various states.*
3. *Show that the process $\{X_t\}$ is irreducible and positive recurrent. Determine its invariant probability distribution.*
4. *Determine the invariant probability distribution of the embedded chain.*

Exercise 7.11.8. *Consider both the $E = \mathbb{N}$–valued discrete time Markov chain $\{X_n;\, n \in \mathbb{N}\}$ whose transition matrix is*

$$P = \begin{pmatrix} q & p & 0 & 0 & \dots \\ q & 0 & p & 0 & \dots \\ 0 & q & 0 & p & \dots \\ 0 & 0 & q & 0 & p \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$