

Chapitre 6

Le processus de Poisson

Introduction

Dans ce chapitre et les deux suivants, on va étudier des processus de Markov définis sur \mathbb{R}_+ , à valeurs dans un ensemble fini ou dénombrable E , qui sont constants entre leurs sauts qui se produisent à des instants aléatoires : on les appelle “processus markoviens de saut”.

Dans ce chapitre, nous allons introduire le “prototype” des processus markoviens de saut, à savoir le processus de Poisson.

Ce processus modélise des répartitions aléatoires de point sur \mathbb{R}_+ , qui peuvent correspondre à des instants de collision de particules, mais aussi à des instants d’arrivée de clients dans une file d’attente, d’appels à un central téléphonique.

6.1 Processus ponctuels et f.a. de comptage

Un **processus ponctuel** sur \mathbb{R}_+ se décrit par la donnée d’une suite croissante de points aléatoires

$$0 < T_1 < T_2 < \dots < T_n < \dots \text{ dans } \mathbb{R}_+$$

qui sont des v.a. définies sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Outre les inégalités ci-dessus, on suppose que $T_n \uparrow \infty, n \rightarrow \infty$.

On posera

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

les v.a. T_n sont les instants où se produisent un événement, les S_n sont les délais ou les temps d'attente entre deux événements successifs.

On définit la **f.a. de comptage** $\{N_t, t \geq 0\}$ du **processus ponctuel** $\{T_n, n \in \mathbb{N}\}$ de la façon suivante :

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

N_t est le nombre d'événement qui se sont produits avant l'instant t .

Notons que $N_0 = 0$, puisque $T_1 > 0$ et pour tout $t > 0$, $N_t < \infty$ puisque $T_n \uparrow \infty, n \rightarrow \infty$.

Pour $0 \leq s < t$, $N_t - N_s$ est le nombre d'événements qui se sont produits dans l'intervalle de temps $]s, t]$.

Une trajectoire type du processus $\{N_t, t \geq 0\}$ est dessinée à la figure 6.1
Précisons que les trajectoires de $\{N_t\}$ sont *continues à droite*.

Notons que la donnée de $\{N_t, t \geq 0\}$ est équivalente à celle de la suite $\{T_n, n \in \mathbb{N}\}$, et que l'on a les relations :

$$\begin{aligned} \{N_t \geq n\} &= \{T_n \leq t\} \\ \{N_t = n\} &= \{T_n \leq t < T_{n+1}\} \\ \{N_s < n \leq N_t\} &= \{s < T_n \leq t\} \end{aligned}$$

6.2 Le processus de Poisson

Définition 6.2.1. On dit que le processus ponctuel $\{T_n, n \in \mathbb{N}\}$ ou sa f.a. de comptage $\{N_t, t \geq 0\}$ est un **processus de Poisson** si $\{N_t, t \geq 0\}$ est une f.a. à accroissements indépendants et stationnaires i.e. si

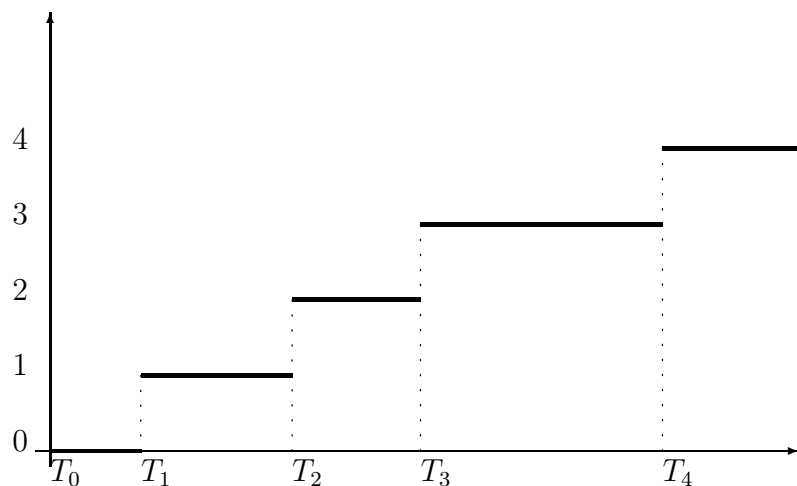


FIG. 6.1 – Trajectoire d'un processus de Poisson

- a) quels que soient $t_0 < t_1 < \dots < t_n$ dans \mathbb{R}_+ , les accroissements $\{N_{t_j} - N_{t_{j-1}}; 1 \leq j \leq n\}$ sont des v.a. indépendantes ;
- b) pour $0 \leq s < t$, la loi de $N_t - N_s$ ne dépend que de s et t que par la différence $t - s$.

La propriété b) est appelée la “stationarité des accroissements ” de $\{N_t\}$. Le nom “processus de Poisson” est justifié par la :

Proposition 6.2.2. Soit $\{N_t, t \geq 0\}$ la f.a. de comptage d'un processus de Poisson. Il existe $\lambda > 0$ tel que pour tous $0 \leq s < t$, la loi de $N_t - N_s$ est la loi de Poisson de paramètre $\lambda(t - s)$, i.e.

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

Remarque 6.2.3. Ce paramètre λ est appelé l'intensité du processus de Poisson $\{N_t, t \geq 0\}$. Il est égal au nombre moyen d'événements qui se produisent pendant un intervalle de temps de longueur unité, i.e.

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

PREUVE DE LA PROPOSITION 6.2.2 Pour tous $0 \leq s < t$, considérons la fonction génératrice de $N_t - N_s$, qui est l'application $u \rightarrow f_{t-s}(u)$ de $[0, 1]$ dans lui-même qui ne dépend que de $t - s$ et est définie par :

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

Par la propriété a) de la définition,

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

Il résulte de cette identité que

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

d'abord pour les t rationnels, puis pour tout t dans \mathbb{R}_+ car $t \rightarrow f_t(u)$ est décroissante.

Comme en outre

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

$f_1(u) \neq 0$, et donc il existe $\lambda(u)$ dans \mathbb{R}_+ tel que

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

Puisque $u \rightarrow \exp(-\theta(1-u))$ est la fonction génératrice de la loi de Poisson de paramètre θ , il reste à montrer que

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

Or clairement

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

Mais puisque $0 \leq u \leq 1$,

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

et le résultat découle de la formule

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

à condition que l'on ait

$$\frac{1}{t} \mathbb{P}(N_t \geq 2) \rightarrow 0, \text{ quand } t \downarrow 0.$$

Or

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

Comme $\mathbb{P}(N_t = 0) = f_t(0) = \exp(-\lambda(0)t)$, en utilisant en outre la propriété a) de la définition, on a :

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

Mais, quand $t \downarrow 0$,

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

et pour tout t suffisamment petit,

$$(\lambda(0)t)^{-1} < (1 - \exp(-\lambda(0)t))^{-1} < 2(\lambda(0)t)^{-1}.$$

Remarque 6.2.4. *Il résulte de la dernière partie de la preuve ci-dessus que*

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

Donc à des probabilités petites devant h près, $N(t+h) - N(t)$ est une v.a. de Bernoulli prenant la valeur 0 avec la probabilité $1 - \lambda h$ et la valeur 1 avec

probabilité λh . Cette propriété, jointe à l'indépendance des accroissements et à la formule

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

entraîne que $N_{t+s} - N_t$ suit approximativement une loi binomiale de paramètres $(n, \lambda s/n)$. Or, quand $n \rightarrow \infty$, cette loi converge vers la loi de Poisson de paramètre λs .

Notons que pour tout n , $0 < t_1 < t_2 < \dots < t_n$, la loi du vecteur aléatoire $(N_{t_1}, N_{t_2}, \dots, N_{t_n})$ est entièrement déterminée par la Proposition 6.2.2 et la condition a) de la définition 6.2.1.

Corollaire 6.2.5. *La loi de l'instant T_1 du premier événement est la loi exponentielle de paramètre λ (i.e. la loi sur \mathbb{R}_+ de densité $\lambda e^{-\lambda t}$). Il en est de même pour la loi du temps d'attente après s du premier événement après s , $T_{N_s+1} - s$, pour tout $s > 0$.*

PREUVE Il suffit de remarquer que pour tout $t > 0$,

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

et de même

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

6.3 Propriété de Markov

Soit $\{N_t, t \geq 0\}$ un processus de Poisson d'intensité λ . Pour tout $s, t > 0$, posons

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

Il résulte immédiatement de la Définition 6.2.1 que $\{N_t^s, t \geq 0\}$ est un processus de Poisson d'intensité λ , indépendant de $\{N_t, 0 \leq t \leq s\}$. Remarquons que la donnée de $\{N_t, 0 \leq t \leq s\}$ est équivalente à celle de $(N_s, T_1, T_2, \dots, T_{N_s})$. L'indépendance dont il est question ci-dessus

est équivalente à celle de des vecteurs aléatoires $(N_s, T_1, T_2, \dots, T_{N_s})$ et $(T_{N_s+1}, T_{N_s+2}, \dots, T_{N_s+p})$, pour tout entier p .

Puisque les accroissements futurs après s $\{N_{s+t} - N_s, t \geq 0\}$ sont indépendants du passé $\{N_t, 0 \leq t \leq s\}$, il est clair que le futur après s $\{N_{s+t}, t \geq 0\}$ ne dépend du passé $\{N_t, 0 \leq t \leq s\}$ que par l'intermédiaire du présent N_s ; ou encore, le futur et le passé sont conditionnellement indépendants, sachant le présent. C'est la propriété de Markov.

Nous reviendrons sur cette propriété au chapitre suivant.

Nous allons maintenant généraliser le résultat ci-dessus au cas où s est un certain type de temps aléatoire. Pour cela, il nous faut tout d'abord rappeler une notation, et poser une définition.

Rappelons tout d'abord qu'une tribu de parties d'un ensemble \mathcal{E} est une classe de parties de \mathcal{E} stable par passage au complémentaire, réunion et intersection dénombrable. On peut toujours parler de "la plus petite tribu contenant une certaine classe $\mathcal{C} \subset \mathcal{P}(\mathcal{E})$ ", car c'est l'intersection de toutes les tribus contenant \mathcal{C} (il existe au moins une telle tribu, à savoir $\mathcal{P}(\mathcal{E})$, et une intersection quelconque de tribus est une tribu, comme cela se vérifie aisément). Par exemple, la tribu borélienne de \mathbb{R}^d , notée \mathcal{B}_d , est la plus petite tribu de parties de \mathbb{R}^d contenant tous les ouverts.

Dans le cas d'une variable aléatoire à valeurs dans un espace dénombrable E , $\sigma(X) = \{X^{-1}(F); F \subset E\}$. Etant donné un vecteur aléatoire X de dimension d (i.e. une variable aléatoire à valeurs dans \mathbb{R}^d), on note $\sigma(X) = \{X^{-1}(B); B \in \mathcal{B}_d\}$ la plus petite sous tribu de parties de Ω qui rend X mesurable. C'est l'ensemble des événements dont on sait s'ils sont ou non réalisés au vu de la valeur prise par X . Etant donné $\{X_i, i \in I\}$ une collection quelconque de vecteurs aléatoires (de dimensions quelconques!), on note $\sigma\{X_i; i \in I\}$ la plus petite tribu contenant toutes les $\sigma(X_i), i \in I$.

Il sera commode d'utiliser dans la suite la notation suivante : pour $t \geq 0$,

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

Définition 6.3.1. *Etant donné un processus de Poisson $\{N_t, t \geq 0\}$, on appelle **temps d'arrêt** (de $\{N_t\}$) une v.a. S à valeurs dans $\mathbb{R}_+ \cup \{+\infty\}$ qui possède la propriété que pour tout t dans \mathbb{R}_+ ,*

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

Pour tout s dans \mathbb{R}_+ , $S \equiv s$ est un temps d'arrêt. Pour tout n , T_n est un temps d'arrêt. T_{N_s+1} est également un temps d'arrêt. Par contre, T_{N_s} n'est pas un temps d'arrêt, car si $t < s$,

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

A tout temps d'arrêt S de $\{N_t\}$, on associe la tribu des événements "déterminés par la trajectoire $\{N_{t \wedge S}, t \geq 0\}$ arrêtée à S " :

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

On a la :

Proposition 6.3.2. *Soit $\{N_t, t \geq 0\}$ un processus de Poisson d'intensité λ , et S un temps d'arrêt de $\{N_t\}$. Sur l'événement $\{S < \infty\}$ on pose pour $t \geq 0$*

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

Conditionnellement en $\{S < \infty\}$, $\{N_t^S, t \geq 0\}$ est un processus de Poisson d'intensité λ , indépendant de \mathcal{F}_S^N .

PREUVE On sait déjà que le résultat est vrai si S est constant. Supposons maintenant que S prend ses valeurs dans une suite croissante $(s_j, j \geq 1)$ de réels positifs. Notons que comme S est un temps d'arrêt,

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

Soit $A \in \mathcal{F}_S^N$, $0 < t_1 < t_2 < \dots < t_\ell$ et n_1, \dots, n_ℓ dans \mathbb{N} .

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

où on a utilisé la propriété $\{S = s_j\} \cap A \in \mathcal{F}_{s_j}^N$ pour la seconde égalité, et le fait que le second facteur de l'avant dernière expression ne dépend pas de s_j , par stationnarité des accroissements de $\{N_t\}$.

Le résultat est établi pour un temps d'arrêt prenant ses valeurs dans une suite croissante d'instant. Mais tout temps d'arrêt S peut être approché par une suite décroissante de temps d'arrêt de cette forme. En effet, pour tout n , posons

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

L'égalité ci-dessus est alors vraie avec S remplacé par S_n , puisque

$$S \leq S_n \Rightarrow \mathcal{F}_S^N \subset \mathcal{F}_{S_n}^N.$$

On passe aisément à la limite dans l'égalité avec S remplacé par S_n puisque par continuité à droite des trajectoires de $\{N_t, t \geq 0\}$,

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

Corollaire 6.3.3. *Soit $\{N_t; t \geq 0\}$ un processus de Poisson d'intensité λ , et $(T_n)_{n \geq 1}$ ses instants de saut. On pose $S_1 = T_1, S_2 = T_2 - T_1, \dots, S_n = T_n - T_{n-1}, \dots$. Les variables aléatoires $S_1, S_2, \dots, S_n, \dots$ sont indépendantes, toutes de loi exponentielle de paramètre λ .*

PREUVE On sait déjà que T_1 , le premier instant de saut d'un processus de Poisson d'intensité λ , suit une loi exponentielle de paramètre λ . Il résulte de la Proposition 6.3.2 avec $S = T_n$ que $S_{n+1} = T_{n+1} - T_n$ est le premier instant de saut d'un processus de Poisson d'intensité λ donc suit une loi exponentielle de paramètre λ , et est indépendant de T_1, T_2, \dots, T_n , donc aussi de S_1, S_2, \dots, S_n . Ceci étant vrai pour tout $n \geq 1$, le résultat est démontré. \square

Il est assez clair que l'on a la réciproque suivante :

Proposition 6.3.4. *Soit $\{S_n; n \geq 1\}$ une suite de v.a.r. indépendantes, toutes de loi exponentielle de paramètre $\lambda > 0$. On pose*

$$T_n = S_1 + \dots + S_n, \quad n \geq 1,$$

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

Alors $\{N_t, t \geq 0\}$ est un processus de Poisson d'intensité λ .

On a donc une façon de “construire” un processus de Poisson, ce qui en particulier démontre que la Définition 6.2.1 n’est pas vide! On a aussi une façon de *simuler* le processus de Poisson.

6.4 Comportement asymptotique

Soit à nouveau $\{N_t; t \geq 0\}$ un processus de Poisson d’intensité λ . On a :

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

Donc $\mathbb{E}[N_t/t] = \lambda$, $\text{Var}[N_t/t] = \frac{\lambda}{t}$, d’où $N(t)/t \rightarrow \lambda$ en moyenne quadratique, quand $t \rightarrow \infty$. En fait on a aussi la “loi forte des grands nombres” :

Proposition 6.4.1. *Soit $\{N_t; t \geq 0\}$ un processus de Poisson d’intensité λ .*

Alors $\frac{N_t}{t} \rightarrow \lambda$ p.s. quand $t \rightarrow \infty$.

PREUVE Remarquons tout d’abord que

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

est la somme de n variables aléatoires indépendantes, de même loi de Poisson de paramètre λ (donc intégrable). Il résulte donc de la loi forte des grands nombres que

$$\frac{N_n}{n} \rightarrow \lambda, \quad \text{p.s., quand } n \rightarrow \infty.$$

Or, avec la notation $[t] =$ partie entière de t ,

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

Il suffit donc de montrer que

$$\sup_{n < t < n+1} \frac{N_t - N_n}{n} \rightarrow 0, \quad \text{quand } n \rightarrow \infty.$$

Or si

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

les $\{\xi_n\}$ sont i.i.d. et intégrables. Donc $\frac{\xi_1 + \dots + \xi_n}{n} \rightarrow \lambda$ p.s., d'où

$$\frac{\xi_n}{n} \rightarrow 0 \text{ p.s.}$$

□

On a un "théorème de la limite centrale" :

Proposition 6.4.2. *Soit $\{N_t; t \geq 0\}$ un processus de Poisson d'intensité λ . Alors*

$$\frac{N_t - \lambda t}{\sqrt{\lambda t}} \rightarrow Z \text{ en loi, quand } t \rightarrow \infty,$$

où Z est une v.a.r de loi gaussienne centrée réduite (i.e. d'espérance 0 et de variance 1).

PREUVE On raisonne comme dans la preuve précédente.

$$\frac{N_n - \lambda n}{\sqrt{\lambda n}} \rightarrow Z \text{ en loi, quand } n \rightarrow \infty,$$

d'après le théorème de la limite centrale "classique". Et

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

qui tend en probabilité vers zéro quand $t \rightarrow \infty$ puisque

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

Donc a fortiori $\frac{N_t - N_{[t]}}{\sqrt{\lambda[t]}} \rightarrow 0$ en probabilité quand $t \rightarrow \infty$. Finalement :

$$\begin{aligned} \frac{N_t - \lambda t}{\sqrt{\lambda t}} &= \frac{N_{[t]} - \lambda[t]}{\sqrt{\lambda[t]}} \times \sqrt{\frac{[t]}{t}} \\ &\quad + \frac{N_t - N_{[t]}}{\sqrt{\lambda t}} \times \sqrt{\frac{[t]}{t}} + \sqrt{\lambda} \frac{[t] - t}{\sqrt{t}}, \end{aligned}$$

et on sait que si $X_n \rightarrow X$ en loi, $Y_n \rightarrow 0$ en probabilité, alors

$$X_n + Y_n \rightarrow X \text{ en loi.}$$

□

On peut en fait établir un “théorème de la limite centrale fonctionnel” que nous allons maintenant décrire. Une démonstration analogue à celle de la Proposition 6.4.2 montre que pour tout $t > 0$,

$$\frac{N_{tu} - \lambda tu}{\sqrt{\lambda u}} \rightarrow B_t \text{ en loi quand } u \rightarrow \infty,$$

où B_t est une v.a.r. gaussienne centrée de variance t . Remarquons que pour chaque u , $\{[N_{tu} - \lambda tu]/\sqrt{\lambda u}, t \geq 0\}$ est un processus à accroissements indépendants, dont les sauts sont d’amplitude $(\lambda u)^{-1/2}$. Il en résulte que l’on peut passer à la limite ci-dessus quand $u \rightarrow \infty$ “de façon coordonnée pour les différents t ”, de telle sorte que la limite $\{B_t, t \geq 0\}$ soit un processus gaussien centré, à accroissements indépendants, à trajectoires continues, et vérifiant $E[B_t] = t$. $\{B_t; t \geq 0\}$ est ce qu’on appelle le **mouvement brownien**.

6.5 Exercices

Exercice 6.5.1. Soit X une v.a.r. à valeurs dans \mathbb{R}_+ t.q. $\mathbb{P}(X > t) > 0, \forall t > 0$. On suppose en outre que $\forall s, t > 0$,

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

En déduire que la loi de X est une loi exponentielle.

Exercice 6.5.2. Trois personnes A, B et C arrivent à la poste en même temps pour téléphoner. Il y a deux cabines qu’occupent immédiatement A et B . C remplace le premier sorti. A, B et C quittent immédiatement la poste après avoir téléphoné.

On désigne par X, Y et Z les temps d’occupation de la cabine par A, B et C respectivement. Ces trois variables aléatoires sont supposées indépendantes et équidistribuées, de loi commune la loi sur \mathbb{R}_+ de densité $\lambda \exp(-\lambda t), t \geq 0$ (avec $\lambda > 0$).

a Calculer la probabilité que C sorte le dernier.

- b** Trouver la loi de probabilité du temps total T passé par C à la poste.
c L'instant 0 étant l'instant d'arrivée des trois personnes à la poste, donner la loi de probabilité de l'instant du dernier départ.

Indication : on cherchera à déterminer la loi du v.a. $(X \wedge Y, X \vee Y - X \wedge Y)$ ($\wedge = \inf$, $\vee = \sup$).

Exercice 6.5.3. Une machine possède une durée de vie τ_1 de loi exponentielle de paramètre θ . Lorsqu'elle tombe en panne, elle est instantanément remplacée par une machine semblable de durée de vie τ_2 et ainsi de suite. On suppose les durées de vie $(\tau_n; n \in \mathbb{N})$ indépendantes et équidistribuées. La première machine commence à travailler à l'instant 0; les instants T_n ($n \geq 1$) de défaillance des machines successives (soit $T_1 = \tau_1, T_2 = \tau_1 + \tau_2, \dots$) forment un processus de Poisson.

- a** Pour un instant $t > 0$ fixé, soit D_t la durée écoulée depuis la mise en fonctionnement de la machine en marche à l'instant t . Dans quel ensemble la v.a. D_t prend-elle ses valeurs? Quelle est la loi de D_t ? Montrer que lorsque $t \rightarrow \infty$, cette loi possède une limite.
b Soit S_t la v.a. positive telle que $t + S_t$ soit l'instant de défaillance de la machine en fonctionnement à l'instant t . Quelle est la loi de S_t ? Quelle est la loi du couple (D_t, S_t) et quelle est la limite de cette loi lorsque $t \rightarrow \infty$? Pourquoi les deux v.a. D_t et S_t ne sont-elles pas équidistribuées et pourquoi le deviennent-elles quand $t \rightarrow \infty$?
c Quelle est la loi de $D_t + S_t$, la durée de vie de la machine en fonctionnement à l'instant t ? Comparer la limite de cette loi quand $t \rightarrow \infty$ avec la loi commune des τ_n .

Exercice 6.5.4. **a)** Soit X_1, X_2, \dots, X_n des v.a.r. indépendantes, toutes de loi uniforme sur $[0, t]$, et Y_1, Y_2, \dots, Y_n la même suite ordonnée par ordre croissant, i.e. définie par

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

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

et ainsi de suite. Préciser la loi du vecteur aléatoire (Y_1, Y_2, \dots, Y_n) .

- b)** Soit $\{N_t, t \geq 0\}$ un processus de Poisson d'intensité λ . Montrer que la loi conditionnelle du vecteur aléatoire (T_1, T_2, \dots, T_n) , sachant que $N_t = n$, est la loi trouvée en a).

Exercice 6.5.5. Soit $\{N_t^1; t \geq 0\}$ et $\{N_t^2; t \geq 0\}$ deux processus de Poisson indépendants, d'intensité respective λ_1 et λ_2 . Montrer que $\{N_t^1 + N_t^2; t \geq 0\}$ est un processus de Poisson d'intensité $\lambda_1 + \lambda_2$.

Exercice 6.5.6. On admet que la population des individus infectés par le virus VIH croît selon un processus de Poisson d'intensité inconnue λ . On notera N_t le nombre d'individus infectés à l'instant t . On ne prendra pas en compte les décès.

Chaque individu infecté subit une période d'incubation entre le moment où il est infecté par le VIH et le moment où les symptômes du SIDA apparaissent. La durée de cette période d'incubation est aléatoire. Les périodes d'incubation pour les différents individus sont i. i. d., de loi commune la loi sur \mathbb{R}_+ de fonction de répartition connue G . On notera \bar{G} la fonction $\bar{G}(t) = 1 - G(t)$.

On note N_t^1 le nombre d'individus qui à l'instant t présentent les symptômes du SIDA, N_t^2 le nombre d'individus qui à l'instant t sont infectés par le VIH, mais ne présentent pas encore les symptômes du SIDA. On a bien sûr

$$N_t = N_t^1 + N_t^2,$$

et on demande de montrer que pour tout $t > 0$, N_t^1 et N_t^2 sont deux v. a. indépendantes, N_t^1 de loi de Poisson de paramètre $\lambda \int_0^t G(s) ds$, et N_t^2 de loi de Poisson de paramètre $\lambda \int_0^t \bar{G}(s) ds$.

Exercice 6.5.7. Programmation On appelle taux de panne d'une v. a. X à densité à valeurs dans \mathbb{R}_+ la fonction λ de \mathbb{R}_+ dans \mathbb{R}_+ définie par

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

où f désigne la densité de la loi de X , F sa fonction de répartition. L'exercice 6.5.1 dit que la loi exponentielle est la seule à avoir un taux de panne constant.

On appelle loi de Weibull de paramètres $\lambda, \alpha > 0$ la loi de fonction de survie

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

et de taux de panne

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

La loi de Weibull a un taux de panne croissant si $\alpha > 1$ et décroissant si $\alpha < 1$. On retrouve la loi exponentielle de paramètre λ dans le cas $\alpha = 1$.

On appelle loi $\Gamma(\alpha, \lambda)$ la loi de probabilités sur \mathbb{R}_+ de densité

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

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

A nouveau la loi Gamma a un taux de panne croissant si $\alpha > 1$ et décroissant si $\alpha < 1$. Notons que si l'on additionne n v. a. indépendantes, toutes exponentielles de même paramètre λ , on obtient une v. a. de loi $\Gamma(n, \lambda)$ à taux de panne croissant.

Supposons que deux machines travaillent en parallèle, et ont besoin pour leur fonctionnement d'une pièce M quelque peu fragile. On dispose d'une seule pièce de rechange, qui est placée instantanément sur la machine dont la pièce M tombe en panne la première. Les trois pièces M (les deux en place au début, et celle de rechange) ont des temps de vie i. i. d. La deuxième panne est fatale à la machine qui la subit. Si les temps de vie de ces pièces suivent une loi exponentielle, alors l'exercice 6.5.2 montre que les deux machines ont même probabilité de subir la deuxième panne.

On peut supposer que si l'on remplace la loi exponentielle par une loi à taux de panne croissant, la machine qui a vu sa pièce remplacée a plus de chance de vivre plus longtemps que l'autre, et que c'est l'inverse dans le cas d'une loi à taux de panne décroissant.

On demande d'illustrer à l'aide de tirages aléatoires le résultat de l'exercice 6.5.2, et les deux conjectures que nous avons formulées.

Successivement pour \mathbb{P} = la loi exponentielle de paramètre 1, la loi $\Gamma(3, 1)$, la loi de Weibull de paramètres $(1, 0.5)$ (que l'on simule aisément en inversant sa fonction de répartition), on simule une matrice $3 \times N$ de v. a. indépendantes de loi \mathbb{P} , notée X . On trace alors, en fonction de n variant de 1 à N , et sur le même graphique les trois quantités

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

On pourra choisir $N = 10^3$ ou $N = 10^4$.

Chapitre 7

Processus markoviens de saut

Introduction

Dans ce chapitre, nous allons présenter l'essentiel de la théorie des processus de Markov en temps continu, à valeurs dans un ensemble fini ou dénombrable E . Comme on le verra à la section 7.4, ces processus sont en quelque sorte une combinaison d'un processus de Poisson et d'une chaîne de Markov en temps discret (la "chaîne incluse"). Dans une seconde partie de ce chapitre, nous présenterons des applications à la phylogénie, aux équations aux dérivées partielles discrétisées, et à l'algorithme du recuit simulé. La preuve de la convergence de l'algorithme du recuit présentée ici est due à Francis Comets (communication privée). Les files d'attente seront traitées au chapitre suivant.

7.1 Généralités

Le but de ce chapitre est d'étudier les processus de Markov à valeurs dans un espace d'état E fini ou dénombrable. On supposera que les trajectoires de nos processus n'ont que des discontinuités de première espèce (des sauts), et pour fixer les idées que les trajectoires sont continues à droite et pourvues de limite à gauche en tout point. Les trajectoires d'un tel processus $\{X_t, t \geq 0\}$ sont nécessairement constantes entre ses sauts, lesquels se produisent à des instants aléatoires $T_1(\omega), T_2(\omega), \dots, T_n(\omega), \dots$. La différence avec le processus de Poisson du chapitre précédent est que, connaissant la position avant le saut, la position après le saut est aléatoire. Si l'on désigne par $Z_n(\omega)$ la

position de $\{X_t\}$ juste après la n -ième saut $T_n(\omega)$, $n \geq 1$, une trajectoire type de $\{X_t; t \geq 0\}$ est dessinée à la figure 7.1.

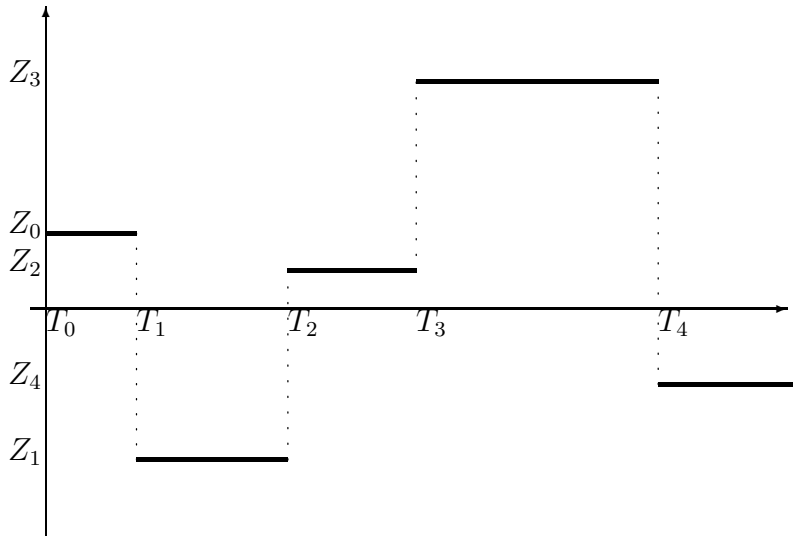


FIG. 7.1 – Trajectoire d'un processus markovien de sauts

La donnée de $\{X_t; t \geq 0\}$ est équivalente à celle de la double suite $\{T_n, Z_n; n \geq 0\}$.

Pour certaines applications, il convient de pouvoir rendre certains états absorbants (par exemple, dans le cas d'un modèle décrivant l'évolution de la taille d'une population sans possibilité d'immigration, l'état 0 est absorbant). $x \in E$ est dit absorbant si $X_{T_n}(\omega) = x \Rightarrow T_{n+1}(\omega) = +\infty$.

On supposera donc que les instants de saut forment une suite croissante

$$0 = T_0 \leq T_1 \leq T_2 \leq \dots \leq T_n \leq \dots$$

avec $T_n \in \mathbb{R}_+ \cup \{+\infty\}$, et

$$T_n(\omega) < T_{n+1}(\omega) \text{ si } T_n(\omega) < \infty,$$

et en outre qu'il n'y a pas d'explosion, i. e. que les instants de saut ne s'accumulent pas à distance finie, autrement dit que

$$T_n(\omega) \rightarrow +\infty \text{ p. s., quand } n \rightarrow \infty. \quad (7.1)$$

Une fonction aléatoire $\{X_t; t \geq 0\}$ à valeurs dans E est appelée fonction aléatoire de sauts si elle est de la forme :

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

où les v.a. Z_n prennent leurs valeurs dans E . Posons la :

Définition 7.1.1. Une fonction aléatoire de sauts $\{X_t, t \geq 0\}$ à valeurs dans E est appelée **processus markovien de sauts** (ou chaîne de Markov en temps continu) si pour tout $0 < s < t$, la loi conditionnelle de la v.a. X_t sachant $\{X_u; 0 \leq u \leq s\}$ ne dépend que de X_s , i.e. si pour tout $n \in \mathbb{N}$, $0 \leq t_0 < t_1 < \dots < t_n < s$, $x_0, x_1, \dots, x_n, x, y \in E$,

$$\mathbb{P}(X_t = y | X_{t_0} = x_0, X_{t_1} = x_1, \dots, X_{t_n} = x_n, X_s = x) = \mathbb{P}(X_t = y | X_s = x)^1.$$

On dira que le processus markovien $\{X_t, t \geq 0\}$ est **homogène** si la quantité $P(X_t = y | X_s = x)$ ne dépend de s et de t que par la différence $t - s$.

On n'étudiera dans la suite que des processus markoviens homogènes. On utilisera la notation :

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

où pour tout $t > 0$, $P(t)$ est une "matrice markovienne" sur $E \times E$, appelée matrice de transition dans le temps t . On notera ci-dessous $\mu(t)$ la loi de probabilité de X_t sur E , $t \geq 0$. $\mu(0)$ est appelée la "loi initiale" du processus $\{X_t; t \geq 0\}$.

Proposition 7.1.2. Soit $\{X_t, t \geq 0\}$ un processus markovien de sauts, de loi initiale μ et de matrices de transition $\{P(t), t > 0\}$. Pour tout n dans \mathbb{N} , $0 < t_1 < \dots < t_n$, la loi du vecteur aléatoire $(X_0, X_{t_1}, \dots, X_{t_n})$ est donnée par : pour tout x_0, x_1, \dots, x_n dans E ,

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

¹Cette condition n'a de sens que si

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

On ne tiendra pas compte de la condition pour les valeurs de $n, x_0, x_1, \dots, x_n, x$ pour lesquelles cette inégalité n'est pas satisfaite

Par conséquent, pour tout $t > 0$,

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

au sens où $\mu_y(t) = \sum_{x \in E} \mu_x(0)P_{xy}(t)$, et pour toute fonction positive ou bornée $g : E \rightarrow \mathbb{R}$,

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

En outre, les matrices de transition $\{P(t), t > 0\}$ vérifient la relation de semi-groupe (équation de Chapman–Kolmogorov) :

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

au sens où pour tous les x, y dans E

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

PREUVE Il résulte immédiatement de la définition des probabilités conditionnelles et de la propriété de Markov que

$$\begin{aligned} &\mathbb{P}(X_0 = x_0, X_{t_1} = x_1, \dots, X_{t_n} = x_n) \\ &= \mathbb{P}(X_0 = x_0)P(X_{t_1} = x_1|X_0 = x_0)\mathbb{P}(X_{t_2} = x_2|X_0 = x_0, X_{t_1} = x_1) \\ &\times \dots \times \mathbb{P}(X_{t_n} = x_n|X_0 = x_0, X_{t_1} = x_1, \dots, X_{t_{n-1}} = x_{n-1}) \\ &= \mu_{x_0}P_{x_0x_1}(t_1)P_{x_1x_2}(t_2 - t_1) \times \dots \times P_{x_{n-1}x_n}(t_n - t_{n-1}). \end{aligned}$$

Dans le cas $n = 1$, cette formule s'écrit :

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

et le second résultat s'en déduit en sommant sur $x \in E$. Par définition de $P(t)$,

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

le troisième résultat s'en déduit en multipliant par g_y et sommant en $y \in E$.

Enfin la formule ci-dessus dans le cas $n = 2$ donne, après division par μ_{x_0}

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

le dernier résultat s'en déduit en sommant en $z \in E$. \square

Nous allons maintenant présenter quelques exemples de processus markoviens de saut.

Exemple 7.1.3. *Un processus de Poisson $\{N_t; t \geq 0\}$ d'intensité λ est un processus de Markov à valeurs dans \mathbb{N} , de matrice de transition :*

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

Exemple 7.1.4. Processus du télégraphe *Etant donné un processus de Poisson $\{N_t\}$ d'intensité λ , et une v.a. X_0 à valeur dans $E = \{-1, +1\}$, indépendante de $\{N_t; t \geq 0\}$, on pose :*

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

$\{X_t, t \geq 0\}$ est un processus de Markov, de matrice de transition :

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

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

Exemple 7.1.5. *Soit $\{N_t; t \geq 0\}$ un processus de Poisson d'intensité λ , et d'instant de saut $0 < T_1 < T_2 < T_3 < \dots < T_n < \dots$. On se donne en outre une chaîne de Markov en temps discret $\{Z_n; n \in \mathbb{N}\}$ à valeurs dans E , de matrice de transition $\{P_{xy}; x, y \in E\}$, indépendante de $\{N_t, t \geq 0\}$. Alors on peut montrer (voir exercice 7.11.1 ci-dessous) que*

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

est un processus markovien de saut.

7.2 Générateur infinitésimal

La propriété de semi-groupe fait que $P(t)$ est connu pour tout t dès qu'il est connu pour t petit. En fait, on va voir qu'il est entièrement déterminé par sa dérivée à droite en $t = 0$ (on sait que $P(0) = I$).

Théorème 7.2.1. *Soit $\{P(t), t > 0\}$ le semi-groupe des matrices de transition d'un processus markovien de sauts $\{X_t, t \geq 0\}$.*

*Il existe une matrice $\{Q_{xy}; x, y \in E\}$ (appelée le **générateur infinitésimal** du semi-groupe $\{P(t); t \geq 0\}$ qui vérifie*

$$(i) \quad Q_{xy} \geq 0 \text{ si } x \neq y$$

$$(ii) \quad Q_{xx} = - \sum_{y \in E \setminus \{x\}} Q_{xy} \leq 0,$$

(cette dernière égalité étant stricte sauf si l'état x est absorbant) et telle que, lorsque $h \downarrow 0$,

$$P_{xy}(h) = hQ_{xy} + o(h) \text{ si } x \neq y$$

$$P_{xx}(h) = 1 + hQ_{xx} + o(h).$$

En outre, conditionnellement en $X_0 = x$, l'instant de premier saut T_1 et la position après le premier saut $Z_1 = X_{T_1}$ sont indépendants, T_1 de loi exponentielle de paramètre $q_x = -Q_{xx}$, et Z_1 de loi sur E donnée par $\{Q_{xy}/q_x; y \neq x\}$.

PREUVE Remarquons tout d'abord que

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

Comme $P(T_2 - T_1 \leq h) \rightarrow 0$ quand $h \rightarrow 0$, on a que si, $h \rightarrow 0$, $nh \rightarrow t$ (avec $nh \geq t$),

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

L'existence de cette dernière limite entraîne que

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

quand $h \rightarrow 0$, et donc

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

D'où nécessairement $q_x < \infty$ et $q_x = 0$ ssi x est absorbant. On pose $Q_{xx} = -q_x$.

La démonstration de l'existence des limites de $\frac{1}{h}P_{xy}(h)$ pour $x \neq y$ se fait de façon analogue :

$$\begin{aligned} & \{T_1 \leq t, Z_0 = x, Z_1 = y\} \\ &= \lim_{h \rightarrow 0, nh \rightarrow t} \cup_{1 \leq m \leq n} \{X_0 = X_h = \dots = X_{(m-1)h} = x, X_{mh} = y\} \end{aligned}$$

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

Donc $Q_{xy} = \lim \frac{1}{h} P_{xy}(h)$ existe pour $x \neq y$ et

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

d'où

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

et

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

Dans le cas $\text{card}E < \infty$, on déduit immédiatement du Théorème le :

Corollaire 7.2.2. (i) $\{P(t), t \geq 0\}$ est l'unique solution de l'équation de Kolmogorov "rétrograde"

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

En outre $u(t, x) := E[g(X_t) | X_0 = x]$ satisfait aussi une équation de Kolmogorov rétrograde

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

(ii) $\{P(t), t \geq 0\}$ est aussi l'unique solution de l'équation de Kolmogorov progressive :

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

En outre, la famille des lois de probabilité marginales $\{\mu(t), t \geq 0\}$ des v.a. $\{X_t; t \geq 0\}$ satisfait l'équation de Fokker–Planck :

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

PREUVE Pour établir l'équation de Kolmogorov rétrograde, il suffit de dériver $P_{xy}(t)$ en utilisant la propriété de semi-groupe sous la forme

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

L'équation pour u se déduit alors de l'équation obtenue en multipliant à droite par le vecteur colonne $\{g_x\}$.

L'équation progressive s'obtient en dérivant à partir de la formule :

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

L'équation de Fokker–Planck s'en déduit alors en multipliant à gauche par le vecteur ligne $\{\mu(0)\}$.

Remarque 7.2.3. Essayons d'expliquer la terminologie “équation progressive, équation rétrograde” (en Anglais “forward, backward equation”). L'équation rétrograde est une équation pour la fonction $(t, x) \rightarrow P_{xy}(t)$, à $y \in E$ fixé. C'est une équation en la variable t et en la variable “rétrograde” x . x désigne la position à l'instant initial, c'est la position dans le passé. Au contraire, l'équation progressive est une équation pour la fonction $(t, y) \rightarrow P_{xy}(t)$, à $x \in E$ fixé. La variable y désigne la position de la chaîne à l'instant t , c'est la position à l'instant présent.

Considérons maintenant l'équation rétrograde pour $u(t, x) = \mathbb{E}[g(X_t)|X_0 = x]$, choisissons $T > 0$, et posons pour $0 \leq t \leq T$, $v(t, x) = u(T - t, x) = \mathbb{E}[g(X_T)|X_t = x]$. v satisfait l'équation

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

L'équation pour v est une équation rétrograde au sens où elle se résout dans le sens négatif du temps, de $t = T$ vers $t = 0$. Notons que dans le cas non-homogène, où le générateur infinitésimal Q dépend de t , la quantité $v(t, x) = \mathbb{E}[g(X_T)|X_t = x]$ est bien solution de cette même équation, alors que l'on n'a plus l'équation pour u .

La démonstration ci-dessus n'est pas rigoureuse dans le cas $\text{card}E = \infty$, puisqu'elle implique la permutation d'une dérivation et d'une sommation infinie. Nous établirons l'équation de Kolmogorov "rétrograde" dans le cas général dans la section suivante.

7.3 Propriété de Markov forte

La notion de temps d'arrêt et la tribu \mathcal{F}_S^X sont définis comme à la section 6.3, en remplaçant $\{N_t; t \geq 0\}$ par $\{X_t; t \geq 0\}$.

Théorème 7.3.1. *Soit S un temps d'arrêt du processus markovien de sauts $\{X_t; t \geq 0\}$. Conditionnellement en $\{S < \infty\}$ et $\{X_S = x\}$, $\{X_{S+t}, t \geq 0\}$ est indépendant de \mathcal{F}_S^X , et sa loi est celle de $\{X_t; t \geq 0\}$ sachant que $X_0 = x$.*

PREUVE On va se contenter de faire la démonstration dans le cas d'un temps d'arrêt constant $S \equiv s$. Le cas général s'en déduit comme pour le processus de Poisson (cf. Preuve de la Proposition 6.3.2). Soit $0 \leq s_1 < s_2 < \dots < s_k < s$; $0 < t_1 < t_2 < \dots < t_\ell$; $x, x_1, \dots, x_z, y_1, \dots, y_\ell \in S$,

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

□

Nous allons maintenant établir l'équation de Kolmogorov rétrograde dans le cas général.

Théorème 7.3.2. *Pour tout $x, y \in E$, la fonction $t \rightarrow P_{xy}(t)$ est dérivable, et*

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

PREUVE Définissons pour tout $n \in \mathbb{N}$ la loi conditionnelle de (Z_n, T_n) sachant que $X_0 = Z_0 = x$:

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

Notons que

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

et qu'il résulte du Théorème 7.2.1 que

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

La propriété de Markov forte à l'instant T_m entraîne que

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

où nous avons utilisé la notation

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

Donc

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

Autrement dit,

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

soit

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

où la mesure $R_n(y; z, du - t)$ est définie par

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

Il est clair que les $\{R_n, n \geq 1\}$ sont entièrement déterminés par R_1 et cette équation fondamentale.

Remarquons que

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

où on a utilisé à la troisième égalité la propriété de Markov forte à l'instant T_m .

Donc d'après l'équation ci-dessus

$$\begin{aligned} P_{xy}(s) &= \delta_{xy} e^{-qxs} + \sum_{m \geq 1} \int_0^s e^{-qy(s-t)} R_m(x; y, dt) \\ P_{xy}(s) &= \delta_{xy} e^{-qxs} + \sum_{m \geq 0, z \in E} \int_0^s e^{-qy(s-t)} \int_0^t R_1(x; z, du) R_m(z; y, dt - u), \end{aligned}$$

soit

$$\begin{aligned} P_{xy}(t) &= \delta_{xy} e^{-qxt} + \sum_{z \in E} \int_0^t R_1(x; z, ds) P_{zy}(t - s) \\ e^{qxt} P_{xy}(t) &= \delta_{xy} + \int_0^t e^{qxs} \sum_{z \neq x} Q_{xz} P_{zy}(s) ds. \end{aligned}$$

Donc la fonction $t \rightarrow P_{xy}(t)$ est dérivable, et

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

□

Le raisonnement qui précède montre que

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

Cette formule se généralise à la loi de $((Z_1, T_1), \dots, (Z_n, T_n))$. Partant de cette loi jointe, on pourrait en déduire que le processus de saut correspondant $\{X_t, t \geq 0\}$ est markovien.

Remarque 7.3.3. *Si on se donne un générateur quelconque Q , on peut définir R_1 , et donc la loi des (Z_x, T_x) . Mais la suite $\{T_n\}$ correspondante ne vérifie pas forcément la condition de non-explosion (7.1), i.e. le processus $\{X_t\}$ correspondant n'est pas forcément défini pour tout $t \geq 0$. On donnera à la section suivante des conditions suivantes sur Q pour que ce soit le cas.*

7.4 Chaîne de Markov incluse

Soit $\{X_t; t \geq 0\}$ un processus de Markov dont les temps de saut $T_1, T_2, \dots, T_n, \dots$ vérifient la condition de non-explosion (7.1). La suite $\{Z_n; n \in \mathbb{N}\}$ définie par

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

est une chaîne de Markov en temps discret (c'est une conséquence de la propriété de Markov forte de $\{X_t\}$), appelée la "chaîne incluse", qui a la particularité que $Z_{n+1} \neq Z_n$ p.s., $\forall n \geq 0$. Sa matrice de transition P se calcule aisément en fonction du générateur Q de $\{X_t\}$:

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

Posons, pour $n \geq 1$,

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

et pour $t \geq 0$,

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

Alors $\{N_t; t \geq 0\}$ est un processus de Poisson d'intensité 1 (utiliser la propriété de Markov forte de $\{X_t\}$, et le fait que si $U \simeq$ exponentielle (λ) , $\lambda U \simeq$ exponentielle (1)).

Réciproquement, soit Q un générateur infinitésimal, i. e. une matrice indécrite par $E \times E$, telle que pour tout $x \in E$,

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

On pose $q_x = -Q_{xx}$, et on définit la matrice de transition P par

$$P_{xy} = \begin{cases} \frac{Q_{xy}}{q_x}, & \text{si } y \neq x; \\ 0, & \text{si } y = x. \end{cases} \quad (7.2)$$

A toute condition initiale $x \in E$, on associe la chaîne de Markov $\{Z_n, n \geq 0\}$ de matrice de transition P . Soit maintenant $\{N_t; t \geq 0\}$ un processus de Poisson d'intensité 1, indépendant de la chaîne $\{Z_n; n \in \mathbb{N}\}$. On note $0 = T_0 < T_1 < T_2 < \dots$ les instants de saut du processus de Poisson, et on définit, pour $n \geq 1$,

$$S_n = \frac{T_n - T_{n-1}}{q(Z_{n-1})},$$

$$T'_n = S_1 + \dots + S_n.$$

Si la condition de non explosion (7.1) est satisfaite, alors

$$X_t \stackrel{\text{def}}{=} \sum_{n \geq 0} Z_n \mathbf{1}_{[T'_n, T'_{n+1}[}(t), \quad t \geq 0 \quad (7.3)$$

est un processus de Markov de générateur infinitésimal Q .

Il reste à répondre à la question : étant donné un générateur infinitésimal Q , quand est-ce que la suite de temps d'arrêt $\{T'_n, n \geq 0\}$ associée vérifie la condition de non-explosion, i. e. quand est-ce que (7.3) définit X_t pour tout $t \geq 0$? On va montrer la

Proposition 7.4.1. *La condition de non explosion (7.1) est satisfaite ssi*

$$\sum_{n \geq 0} q_{Z_n}^{-1} = +\infty \text{ p.s.} \quad (7.4)$$

Commençons par énoncer le

Corollaire 7.4.2. *Pour qu'un générateur infinitésimal Q soit le générateur infinitésimal d'un processus de Markov qui satisfait la condition (7.1), il suffit que l'une des deux conditions suivantes soit satisfaite*

1. $\sup_{x \in E} q_x < \infty$.
2. La chaîne de Markov $\{Z_n\}$ de matrice de transition P définie (7.2) est récurrente.

Il est clair que que chacune des deux conditions du Corollaire entraîne (7.4). La Proposition résulte maintenant du Lemme suivant, en posant

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

Lemme 7.4.3. *Soit $\{A_n, n \geq 1\}$ et $\{B_n, n \geq 1\}$ deux suites indépendantes de v. a. à valeurs dans \mathbb{R}_+^* , la suite des $\{A_n\}$ étant i. i. d. de loi commune la loi exponentielle de paramètre 1. Alors il y a équivalence entre*

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

PREUVE Il résulte de l'indépendance des deux suites que le Lemme est une conséquence du fait que pour toute suite $\{b_n, n \geq 1\}$ de nombres strictement positifs,

$$\sum_{n=1}^{\infty} A_n b_n = +\infty \text{ p. s.} \iff \sum_{n=1}^{\infty} b_n = +\infty. \quad (7.5)$$

Si $\sum_n b_n < \infty$, alors $\mathbb{E} \sum_n A_n b_n = \sum_n b_n < \infty$, donc a fortiori $\sum_n A_n b_n < \infty$ p. s. Il reste à montrer que si $\sum_n b_n = +\infty$, alors

$$\Lambda_n := \sum_{k=1}^n A_k b_k \rightarrow +\infty \text{ p. s., quand } n \rightarrow \infty.$$

Dans le cas où il existe une sous-suite n_j telle que $b_{n_j} \rightarrow +\infty$, clairement $\sum_n A_n b_n \geq \sum_j A_{n_j} b_{n_j} = +\infty$, car les A_{n_j} étant i. i. d. de loi exponentielle,

une infinité d'entre eux est plus grande que 1. Il reste donc à considérer le cas où $0 \leq b_n \leq C$ et $\sum_n b_n = +\infty$. Alors pour tout $M > 0$, si n est assez grand pour que $\mathbb{E}\Lambda_n > 2M$,

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

et donc $\Lambda_n \rightarrow +\infty$ en probabilité, et donc p. s. par monotonie de la suite. Le Lemme est établi.

Remarque 7.4.4. *Au chapitre suivant, nous préciserons des processus markoviens de sauts en indiquant leur générateur infinitésimal Q . On pourra vérifier que tous les exemples traités satisfont l'une des conditions (en fait en général la première) du Corollaire 7.4.2.*

7.5 Classification des états

Dans ce qui suit, on désignera comme dans le cas du temps discret par \mathbb{P}_x la loi conditionnelle de $\{X_t, t \geq 0\}$, sachant que $X_0 = x$. Les classes d'équivalence du processus de Markov $\{X_t; t \geq 0\}$ sont celles de la chaîne incluse. Notons que dès que $\{X_t; t \geq 0\}$ est irréductible,

$$P_{xy}(t) > 0, \quad \forall x, y \in E, t > 0. \quad (7.6)$$

En effet, pour tout $x, y \in E$, il existe $n \geq 1$ et $x_0 = x, x_1, \dots, x_{n-1}, x_n = y$ tels que $Q_{x_{k-1}x_k} > 0$, $1 \leq k \leq n$, et il résulte des propriétés de la loi exponentielle que $P_{xy}(t) \geq P_{xx_1}(t/n) \times \dots \times P_{x_{n-1}y}(t/n) > 0$.

Un état $x \in E$ est dit récurrent (resp. transitoire) pour $\{X_t; t \geq 0\}$ s'il est récurrent (resp. transitoire) pour la chaîne incluse. Donc en particulier dans le cas irréductible, tous les états sont soit récurrents soit transitoires.

On a, comme dans le cas des chaînes en temps discret, le

Théorème 7.5.1. *Soit $\{X_t; t \geq 0\}$ un processus de Markov irréductible et récurrent. Alors il existe une mesure strictement positive π sur E solution*

de l'équation $\pi Q = 0$, unique à une constante multiplicative près. En outre, une telle mesure est invariante par le semi-groupe $\{P(t)\}$, i. e. $\pi P(t) = \pi$, $\forall t \geq 0$.

PREUVE On remarque que si Q désigne le générateur infinitésimal du processus $\{X_t\}$ et P la matrice de transition de sa chaîne incluse, alors

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

où \mathbf{q} désigne la matrice diagonale définie par

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

Notons que l'hypothèse que le processus est irréductible entraîne que $q_x > 0$, $\forall x \in E$. Donc on va pouvoir sans problème diviser par \mathbf{q} . L'hypothèse est que la chaîne incluse est irréductible et récurrente. Donc la mesure γ^x définie dans la preuve du Théorème 2.5.3 est strictement positive, et elle est l'unique (à une constante multiplicative près) solution de l'équation $\gamma^x P = \gamma^x$. Donc la mesure strictement positive $\mu^x = \mathbf{q}^{-1} \gamma^x$ vérifie $\mu^x Q = 0$, et toute autre solution μ' de cette équation est telle que $\mathbf{q} \mu'$ est invariante par P , donc il existe une constante c telle que $\mu' = c \mu$. On a la formule suivante pour la mesure μ^x :

$$\begin{aligned} \mu_y^x &= \frac{\gamma_y^x}{q_y} = \mathbb{E}_x \int_0^{R_x} \mathbf{1}_{\{X_s=y\}} ds \\ &= \mathbb{E}_x \int_0^\infty \mathbf{1}_{\{X_s=y, s < R_x\}} ds. \end{aligned}$$

Mais si $t > 0$, par la propriété de Markov forte,

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

Donc

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

□

7.6 Le cas irréductible récurrent

Pour distinguer entre récurrence positive ou nulle (cette question ne se pose que dans le cas $|E| = +\infty$), il ne suffit pas de regarder ce qu'il en est de la chaîne incluse, comme on va maintenant le voir. Définissons l'instant du premier retour à l'état x comme :

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

Définition 7.6.1. *L'état x est dit récurrent positif s'il est récurrent et si $\mathbb{E}_x(R_x) < \infty$, récurrent nul s'il est récurrent et si $\mathbb{E}_x(R_x) = +\infty$.*

A nouveau, dans le cas irréductible récurrent, tous les états sont soit récurrents nuls, soit récurrents positifs, et suivant le cas on dit que le processus $\{X_t\}$ est transitoire, récurrent nul ou récurrent positif. Le cas récurrent positif est équivalent à l'existence d'une unique probabilité invariante, comme on va maintenant le montrer.

Théorème 7.6.2. *Soit $\{X_t, t \geq 0\}$ un processus markovien de sauts irréductible. Alors un état $x \in E$ est récurrent positif si et seulement si tous les états sont récurrents positifs, si et seulement si il existe une probabilité invariante π , et dans ce cas*

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

PREUVE Si x est récurrent positif pour $\{X_t\}$, alors x est récurrent pour la chaîne incluse $\{Z_n, n \geq 0\}$. On désigne par γ_y^x le nombre moyen de visites à l'état y lors d'une excursion de $\{Z_n\}$ partant de x . Il résulte de ce que la durée de chaque séjour à l'état y est indépendante de la chaîne incluse, et d'espérance q_y^{-1} que

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

Mais on a vu dans la preuve du Théorème 7.5.1 que la mesure μ^x définie par

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

vérifie $\mu^x Q = 0$. La condition x récurrent positif entraîne donc l'existence d'une mesure invariante de masse finie, donc d'une probabilité invariante. Supposons réciproquement qu'il existe une probabilité π solution de $\pi Q = 0$. Alors la mesure $q\pi$ est invariante par P , et pour tout $x, y \in E$,

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

est le nombre moyen de visites à l'état y au cours d'une excursion de $\{Z_n\}$ partant de x . Donc

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

et n'importe quel état $x \in E$ est récurrent positif. \square

Nous allons maintenant nous limiter au cas récurrent positif, et établir le théorème ergodique, puis établir la convergence des probabilités de transition vers la probabilité invariante.

Théorème 7.6.3. *Soit $\{X_t, t \geq 0\}$ un processus markovien de sauts à valeurs dans E irréductible, récurrent positif. On note Q son générateur infinitésimal, et π l'unique probabilité invariante. Alors si $f : E \rightarrow \mathbb{R}$ est bornée,*

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

p. s. quand $t \rightarrow \infty$.

PREUVE Il suffit de considérer le cas $f(y) = \mathbf{1}_{\{y=x\}}$ et de travailler sous \mathbb{P}_x (cf. la preuve du Théorème 2.5.7). Comme dans le cas du temps discret, les excursions successives partant de x sont i. i. d.. On note $N^x(t)$ le nombre de visites à l'état x entre l'instant 0 et l'instant t , T_k^x le temps passé à l'état x par le processus $\{X_t\}$ au cours de sa k -ième visite. Puisque la $N^x(t)$ -ième visite à l'état x n'est pas forcément terminée à l'instant t , on a

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

Mais clairement $T_{N^x(t)}^x/t \rightarrow 0$ p. s. quand $t \rightarrow \infty$, et

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

En effet, puisque la suite $\{T_k^x, k \geq 1\}$ est i. i. d., et par la récurrence $N^x(t) \rightarrow \infty$ quand $t \rightarrow \infty$, p. s.

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

Enfin la preuve du fait que

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

suit le même argument que dans la preuve du Théorème 2.5.7. \square

Dans le cas du temps continu, la convergence de la loi de X_t vers la probabilité invariante quand $t \rightarrow \infty$ est vraie dans le cas irréductible et récurrent positif, sans restriction supplémentaire.

Théorème 7.6.4. *Soit $\{X_t, t \geq 0\}$ un processus markovien de sauts à valeurs dans E irréductible, récurrent positif, et π l'unique probabilité solution de l'équation de l'équation de Fokker-Planck stationnaire. Alors pour toute probabilité μ sur E et $x \in E$, $(\mu P)_x(t) \rightarrow \pi_x$ quand $t \rightarrow \infty$.*

PREUVE On pourrait suivre la démarche de la preuve du Théorème 2.6.4, mais on va plutôt utiliser ce résultat.

Si l'on échantillonne le processus $\{X_t\}$ en posant $Y_n = X_{nh}$, $n = 0, 1, \dots$, où $h > 0$ est arbitraire, il est clair que $\{Y_n, n \in \mathbb{N}\}$ est une chaîne de Markov irréductible et apériodique d'après (7.6), dont l'unique probabilité invariante, qui ne dépend pas de h , est π . Admettons un instant le

Lemme 7.6.5. *Pour tout $t, h > 0$, $x, y \in E$,*

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

Fixons $\varepsilon > 0$ et $x, y \in E$. On choisit tout d'abord $h > 0$ suffisamment petit pour que

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

puis on choisit N assez grand pour que

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

On conclut que si $t \geq Nh$, en notant n l'entier tel que $nh \leq t < (n+1)h$,

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

Le théorème se déduit aisément de ce résultat, en découpant l'ensemble des points de départ en un ensemble fini qui supporte toute la mesure μ à δ près, et son complémentaire. Il reste à procéder à la

PREUVE DU LEMME 7.6.5 : On note que

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

Remarque 7.6.6. *La convergence du Théorème 7.6.4 a lieu au sens de la convergence étroite des probabilités sur E . En effet d'une part quand $t \rightarrow \infty$, $(\mu P)_x(t) \rightarrow \pi_x$ pour tout $x \in E$, donc aussi pour tout sous-ensemble fini $F \subset E$, $\sum_{x \in F} (\mu P)_x(t) \rightarrow \sum_{x \in F} \pi_x$. Comme en outre $\mu P(t)$ et π sont des probabilités sur E , il n'est pas trop difficile de montrer que si $f : E \rightarrow \mathbb{R}$ est bornée,*

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

Notons que si la probabilité invariante π est la loi de X_0 , le processus $\{X_t; t \geq 0\}$ est **stationnaire** au sens où pour tout $n \in \mathbb{N}$, $0 \leq t_1 < t_2 < \dots < t_n$, la loi du vecteur aléatoire $(X_{t_1+s}, X_{t_2+s}, \dots, X_{t_n+s})$ ne dépend pas de $s \geq 0$.

Remarque 7.6.7. L'équation $\pi Q = 0$ s'écrit $\forall x \in E$,

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

Le membre de gauche de cette égalité s'interprète comme le flux entrant dans l'état x à l'équilibre en provenance des différents états, et le membre de droite comme le flux sortant de x à l'équilibre, vers les divers états. L'équation $\pi Q = 0$ dit donc qu'à l'équilibre, les nombres moyen par unité de temps de départs et d'arrivées à chaque état sont égaux.

On a aussi dans ce contexte une généralisation du théorème de la limite centrale.

Théorème 7.6.8. Supposons que le processus markovien de sauts $\{X_t; t \geq 0\}$ est irréductible, et qu'il possède une probabilité invariante π . Soit $f \in L^2(E, \pi)$ de la forme $f = Qg$, avec $g \in L^2(E, \pi)$ [ceci entraîne que $\langle \pi, f \rangle = \sum_x \pi_x f_x = \sum_{xy} \pi_x Q_{xy} g_y = 0$].

On pose

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

que l'on supposera non nul (alors $C(f) > 0$). Alors

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

en loi, quand $t \rightarrow \infty$, où Z est une v.a.r. gaussienne centrée réduite.

On a aussi la convergence de $\left\{ \frac{1}{\sqrt{uC(f)}} \int_0^{tu} f(X_s) ds, t \geq 0 \right\}$ vers un mouvement brownien $\{B_t, t \geq 0\}$, quand $u \rightarrow \infty$.

7.7 Réversibilité

Etant donné un processus markovien de sauts $\{X_t; t \geq 0\}$, et $T > 0$, $\{\hat{X}_t^T = X_{T-t}, 0 \leq t \leq T\}$ est aussi markovien. Si la loi de X_0 est une probabilité invariante π , alors \hat{X}^T est homogène. Notons \hat{Q} son générateur infinitésimal. On a le

Théorème 7.7.1. $\hat{Q} = Q$ ssi la relation d'équilibre ponctuel

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

est satisfaite. Dans ce cas, on dit que le processus $\{X_t\}$ est réversible (pour la probabilité π , qui est nécessairement invariante).

PREUVE : On a, pour les mêmes raisons que la formule analogue pour le temps discret,

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

d'où l'on déduit en prenant la dérivée en $t = 0$

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

Le résultat est maintenant évident.

Remarque 7.7.2. Comme dans le cas des chaînes en temps discret, un processus markovien de saut irréductible récurrent positif n'est pas nécessairement réversible. A nouveau, un contre-exemple est donné par le cas où pour un certain couple $x \neq y$, $Q_{xy} = 0 \neq Q_{yx}$, ce qui n'est pas contradictoire avec l'irréductibilité dès que $\text{card}E \geq 3$.

Remarque 7.7.3. Comme dans le cas du temps discret, déterminer un générateur Q connaissant une probabilité invariante π n'est pas difficile. Le plus simple est de chercher Q telle que le processus correspondant soit réversible par rapport à π , donc de chercher Q générateur infinitésimal irréductible, tel que la quantité $\pi_x Q_{xy}$ soit symétrique en x, y .

Déterminer la probabilité invariante, connaissant le générateur infinitésimal irréductible, est en général plus difficile. On peut chercher à résoudre l'équation

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

mais celle-ci n'a de solution que dans le cas réversible. Dans le cas non réversible, il faut résoudre l'équation $\pi Q = 0$. Si on sait deviner π à une constante multiplicative près, on peut utiliser le résultat suivant

Théorème 7.7.4. *Étant donnée une probabilité π sur E , on pose pour $x, y \in E$*

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

Si

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

alors π est une probabilité invariante, et \hat{Q} est le générateur du processus retourné.

PREUVE : En utilisant la première, puis la seconde identité de l'énoncé, on obtient

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

soit l'égalité $\pi Q = 0$. La seconde partie de l'énoncé est alors une conséquence de la formule contenue dans la preuve du théorème 7.7.1. \square

Notons que si l'on sait deviner le générateur du processus retourné, on en déduit la probabilité invariante à une constante multiplicative près, qui se calcule à l'aide d'une seule sommation.

7.8 Modèles markoviens d'évolution et Phylogénie

On va présenter les processus de Markov sur les arbres, qui sont couramment utilisées comme modèle en phylogénie. On considèrera des arbres binaires, avec ou sans racine. On représente à la figure 7.2 un arbre binaire avec racine (avec la racine en haut et les feuilles en bas!), et à la figure 7.3 un arbre binaire sans racine.

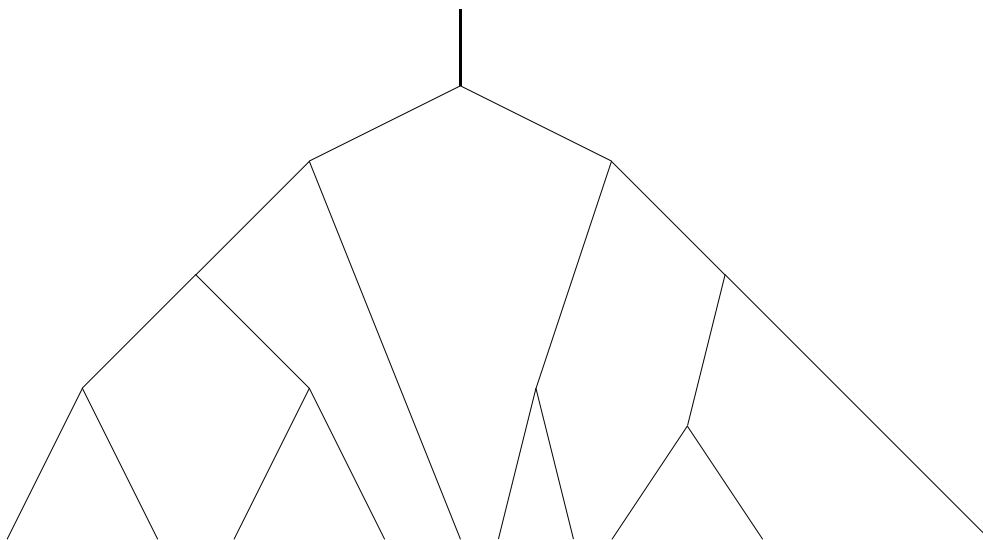


FIG. 7.2 – Arbre binaire avec racine

Processus de Markov sur un arbre avec racine Le processus part de la racine (qui joue le rôle de l'instant initial 0) dans un certain état, disons x . Il évolue jusqu'au premier noeud qui se trouve à la distance r de la racine, comme un processus markovien de sauts pendant l'intervalle de temps r . Notons y l'état du processus en ce noeud. Sur chaque branche qui part de ce noeud court un processus markovien de sauts, partant de y , de telle sorte que les deux processus sur les deux branches sont indépendants, jusqu'au prochain noeud, et ainsi de suite jusqu'aux feuilles de l'arbre. Notons que l'on considérera toujours un processus irréductible à valeurs dans un ensemble fini, donc ce sera un processus récurrent positif, et la loi du processus à la racine sera la probabilité invariante. On peut alors, sans modifier la loi du processus, supprimer la branche entre la racine et le premier noeud.

Processus de Markov sur un arbre sans racine Supprimons donc la branche entre la racine et le premier noeud. Cela signifie que le processus part de la racine (située au premier noeud) sous la probabilité invariante, et évolue indépendamment sur les deux branches, jusqu'à rencontrer les noeuds

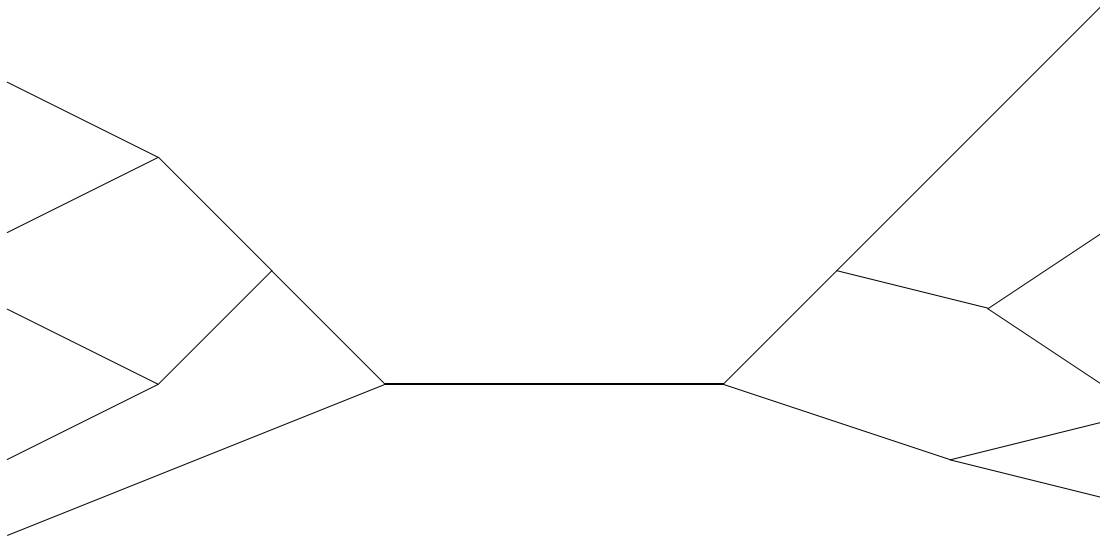


FIG. 7.3 – Le même arbre, sans racine

suivants, etc... Reportons-nous maintenant à la figure 7.3. La racine n'est plus marquée sur la branche centrale. On peut encore l'imaginer, et considérer que deux processus partent de ce point, un dans chaque sens, vers les deux noeuds aux extrémités de la branche centrale. Supposons maintenant que l'on déplace la racine sur la branche centrale, vers la droite ou vers la gauche. Il est assez facile de se convaincre de ce que la loi du processus résultant sur l'arbre n'est pas modifiée par un tel changement, à condition qu'il s'agisse d'un processus *réversible*. En effet, la différence entre les deux constructions avec la racine en deux points différents de la branche centrale est qu'une portion de cette branche est parcourue dans deux sens différents par le processus. Dans le cas réversible, la racine peut être placée équivalamment en une des deux extrémités de la branche centrale, ou même en n'importe quel noeud, ou n'importe quel point de l'arbre. On voit donc que l'on peut définir un processus markovien réversible sur un arbre sans racine, en plaçant le point de départ n'importe où sur l'arbre.

7.8.1 Modèles d'évolution

Pour préciser la vraisemblance d'un arbre au vu des données, il nous faut préciser un modèle d'évolution, qui indique comment ces données ont été "fabriquées" par l'évolution le long des branches de l'arbre, pour chaque site sur l'ADN. On va décrire plusieurs modèles markoviens d'évolution de l'ADN, en précisant les taux de transition d'un nucléotide à l'autre. C'est à dire que l'on précisera la matrice Q sous la forme

$$Q = \begin{array}{c|cccc} & \mathbf{a} & \mathbf{c} & \mathbf{g} & \mathbf{t} \\ \hline \mathbf{a} & x & x & x & x \\ \mathbf{c} & x & x & x & x \\ \mathbf{g} & x & x & x & x \\ \mathbf{t} & x & x & x & x \end{array}$$

Le modèle de Jukes–Cantor (1969) C'est le plus simple, qui suppose que toutes les mutations se font au même taux, i.e. pour un certain $\alpha > 0$,

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

La probabilité invariante associée est la probabilité uniforme sur les 4 nucléotides. Les probabilités de transition se calculent explicitement. $P(t) =$

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

Les modèles de Kimura (1980 et 1981) Parmi les quatre types de nucléotides, la cytosine et la thymine sont des pyrimidines, alors que l'adénine et la guanine sont des purines. Il est raisonnable de supposer que les transitions (remplacement d'une purine par une autre, ou d'une pyrimidine par une autre) sont plus fréquentes que les transversions (remplacement d'une purine par une pyrimidine ou vice versa). Donc on est amené à supposer que les taux de substitution entre \mathbf{a} et \mathbf{g} ou entre \mathbf{c} et \mathbf{t} sont plus élevés que tous

les autres, d'où le modèle (avec $\beta > \alpha$)

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

La probabilité invariante est encore la probabilité uniforme. Les probabilités de transition sont données par

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

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

si $x \neq y$ sont soit tous deux des purines, soit tous deux des pyrimidines,

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

dans le dernier cas.

Kimura a proposé un second modèle, de la forme

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

pour lequel la probabilité invariante est encore la probabilité uniforme.

Le modèle de Felsenstein Etant donné une loi de probabilité π sur l'espace $E = \{\mathbf{a}, \mathbf{c}, \mathbf{g}, \mathbf{t}\}$, et un nombre positif u , Felsenstein a proposé le modèle

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

Notons que clairement pour $x \neq y$,

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

donc π est la probabilité invariante, et la chaîne est réversible. La matrice Q possède deux valeurs propres : $-u$, dont l'espace propre associé est constitué

des vecteurs orthogonaux à π dans \mathbb{R}^4 , et 0, dont l'espace propre associé est constitué des vecteurs colinéaires au vecteur $(1, 1, 1, 1)$. Passant à l'exponentielle, on montre aisément que

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

Dans le cas particulier où $\pi = (1/4, 1/4, 1/4, 1/4)$, ce modèle se réduit à celui de Jukes–Cantor.

Le modèle d'Hasegawa, Kishino, Yano (1985) Il s'agit d'une généralisation à la fois du premier modèle de Kimura, et de celui de Felsenstein. Soit à nouveau π une probabilité sur E , et u, v deux nombres positifs.

$$Q = \begin{pmatrix} -u\pi_g - v\pi_2 & v\pi_c & u\pi_g & v\pi_t \\ v\pi_a & -u\pi_t - v\pi_1 & v\pi_g & u\pi_t \\ u\pi_a & v\pi_c & -u\pi_a - v\pi_2 & v\pi_t \\ v\pi_a & u\pi_c & v\pi_g & -u\pi_c - v\pi_1 \end{pmatrix},$$

où $\pi_1 = \pi_a + \pi_g$, $\pi_2 = \pi_c + \pi_t$. A nouveau π est la probabilité invariante. On peut encore donner une expression explicite pour $P(t)$.

Il y a de bonnes raisons de supposer que $\pi_c = \pi_g$ et $\pi_a = \pi_t$, puisque l'ADN est une molécule à deux brins, constitué de paires $c : g$ et $a : t$. L'égalité ci-dessus est une conséquence de l'hypothèse assez naturelle que l'évolution est la même sur les deux brins. Le modèle HKY, avec cette restriction, devient un modèle avec trois paramètres, à savoir u, v et $\theta = \pi_c + \pi_g$, qui a été proposé par Tamura en 1992. Il s'écrit

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

Le modèle réversible général Comme le cardinal de E est très petit, on peut chercher à utiliser le modèle le plus général. Tavaré a proposé une paramétrisation du modèle le plus général, qui prend la forme

$$Q = \begin{pmatrix} -uW & uA\pi_c & uB\pi_g & uC\pi_t \\ uD\pi_a & -uX & uE\pi_g & uF\pi_t \\ uG\pi_a & uH\pi_c & -uY & uI\pi_t \\ uJ\pi_a & uK\pi_c & uL\pi_g & -uZ \end{pmatrix},$$

où u est un paramètre positif, π la probabilité invariante,

$$\begin{aligned} W &= A\pi_c + B\pi_g + C\pi_t \\ X &= D\pi_a + E\pi_g + F\pi_t \\ Y &= G\pi_a + H\pi_g + I\pi_t \\ Z &= J\pi_a + K\pi_c + L\pi_g, \end{aligned}$$

et les paramètres A, B, \dots, L sont douze paramètres libres. Comme on le verra ci-dessous, il est important pour le calcul du maximum de vraisemblance que le modèle soit réversible. La contrainte de réversibilité impose six contraintes, à savoir

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

Il reste donc six paramètres à choisir, par exemple A, B, C, E, F et I . Il y a en outre les 3 paramètres de la probabilité invariante, cela fait donc en tout 9 paramètres à choisir.

Modèles de codons Un codon est un triplet de nucléotides qui code pour un acide aminé. Parmi les $4^3 = 64$ codons possible, 3 sont des codons STOP possibles, les 61 autres codent pour les 20 acides aminés. Notons au passage que le code génétique (la règle de traduction des codons en amino-acides) est *dégénéré*, au sens où plusieurs codons codent pour le même amino-acide. Donc parmi les mutations possibles de codons, il faut distinguer les mutations synonymes (qui transforment un codon en un autre qui code pour le même acide aminé) des mutations non synonymes. Les dernières sont soit freinées, soit favorisées par la sélection, alors que les changements synonymes s'accumulent au taux des mutations. En général, le rapport mutations synonymes / mutation non synonymes est plus grand que 1.

Goldman et Yang ont proposé en 1994 un modèle comportant 63 paramètres, à savoir 60 paramètres pour les fréquences π_{xyz} , le taux de transition α , le taux de transversion β , et le rapport

$$\omega = \text{taux des mutations non synonymes} / \text{taux des mutations synonymes}.$$

Le modèle GY s'écrit

$$Q_{(x_1y_1z_1)(x_2y_2z_2)} = \begin{cases} 0 & \text{si 1 et 2 diffèrent par plus d'une base,} \\ \alpha\pi_{x_2y_2z_2} & \text{pour une transition synonyme,} \\ \beta\pi_{x_2y_2z_2} & \text{pour une transversion synonyme,} \\ \omega\alpha\pi_{x_2y_2z_2} & \text{pour une transition non synonyme,} \\ \omega\beta\pi_{x_2y_2z_2} & \text{pour une transversion non synonyme.} \end{cases}$$

Notons que sur les 63 paramètres à estimer, les 60 paramètres de la probabilité invariante π ne sont en général pas estimés par le maximum de vraisemblance, mais par les fréquences empiriques des divers codons dans les données. Une autre possibilité est d'estimer π_{xyz} par le produit $\pi_x^1\pi_y^2\pi_z^3$ des fréquences des divers nucléotides aux positions 1, 2 et 3 du codon.

Modèles non homogènes Une hypothèse implicite dans les modèles markoviens considérés jusqu'ici est la stationnarité. Le générateur infinitésimal est le même sur les diverses branches de l'arbre phylogénétique. Donc la probabilité invariante est la même sur les diverses branches, autrement dit les diverses séquences doivent avoir approximativement la même composition en bases. Certaines données contredisent nettement cette situation. On peut alors relâcher l'hypothèse d'homogénéité du processus de Markov sur tout l'arbre. Galtier et Gouy adoptent le modèle de Tamura, avec des paramètres α et β homogènes sur tout l'arbre, et un paramètre θ (qui règle la proportion de $\mathbf{g} + \mathbf{c}$) qui peut varier d'une branche à l'autre de l'arbre.

Dépendance ou indépendance entre les sites La plupart des modèles markoviens supposent que le comportement des divers sites au cours de l'évolution est un comportement i.i.d. Evidemment cette hypothèse n'est pas raisonnable, mais elle simplifie grandement les calculs.

Il y a à ce jour très peu de travaux qui proposent des modèles markoviens où les évolutions des différents sites sont corrélées. Citons le travail de G. Didier, qui propose un modèle où l'évolution de chaque site dépend des sites voisins. Indiquons une autre approche, utilisée par Pollock, Taylor et Goldman pour modéliser l'évolution de séquences de protéines.

Considérons un modèle du type

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

qui est un modèle réversible, si $s_{xy} = s_{yx}$. On propose alors de modéliser l'évolution d'une paire de protéines en choisissant un générateur infinitésimal de la forme

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

où $\bar{\pi}$ est une probabilité invariante sur l'ensemble des paires de protéines.

Variation du taux d'évolution entre branches Etant donné un générateur infinitésimal Q , pour tout $u > 0$, uQ est encore un générateur infinitésimal. Supposons que Q est constant sur tout l'arbre. Si u est lui aussi constant sur l'arbre, puisque les feuilles (les espèces d'aujourd'hui) sont équidistantes (les distances sont mesurées en temps) de l'ancêtre commun situé à la racine, alors on est dans la situation de l'hypothèse d'une "horloge moléculaire". Certains jeux de données sont incompatibles avec une telle hypothèse. On doit alors, pour utiliser un modèle cohérent avec de telles données, permettre au paramètre u de prendre une valeur différente sur chaque branche de l'arbre. On a donc un nouveau paramètre par branche, ce qui fait au total beaucoup de paramètres.

Un autre point de vue est de supposer que u est la valeur prise par un processus stochastique, qui évolue sur l'arbre comme un processus de Markov, soit en temps continu, soit en temps discret (et alors la valeur du processus est constante sur chaque branche, les changements se produisant aux noeuds). Conditionnellement en les valeurs prises par ce processus, les divers nucléotides évoluent comme des processus de Markov non homogènes sur l'arbre. On est alors dans un cadre bayésien, qui se prête à des calculs grâce aux méthodes de simulation dites "Monte Carlo par Chaînes de Markov" (voir plus loin).

Variation du taux d'évolution entre sites Le modèle le plus fréquent de variation de taux entre sites est de supposer que les taux associés aux divers sites sont i.i.d., de loi commune une loi gamma (ou une discrétisation de cette loi).

Une autre approche, dûe à Felsenstein et Churchill consiste à supposer que les taux forment, le long de la séquence d'ADN considérée, une chaîne

de Markov (qui est en fait “cachée”), à valeurs dans un ensemble de cardinal petit (pour des raisons pratiques).

Modèles dit “covarion” Il s’agit de modèles où le taux d’évolution est non seulement différent d’un site à l’autre, mais aussi, pour un site donné, d’une branche à l’autre de l’arbre. Covarion est un acronyme pour “CONcomitantly VARIABLE codON”.

Posons $E = \{a, c, g, t\}$, $G =$ l’ensemble des valeurs possibles pour le taux u . Galtier considère en chaque site un processus de Markov indépendant, à valeurs dans $E \times G$.

7.8.2 Vraisemblance en phylogénie

La comparaison des génomes de diverses espèces est maintenant le principal outil pour tenter de reconstruire des arbres phylogénétiques. Il existe plusieurs algorithmes qui construisent de tels arbres. Nous allons donner des indications sur la méthode du maximum de vraisemblance.

Notons que l’on peut comparer des gènes (i.e. des collections d’acides aminés), ou bien des séquences d’ADN. Nous nous limiterons pour fixer les idées aux séquences d’ADN.

Calcul de la vraisemblance d’un arbre Pour fixer les idées, supposons que l’on utilise le modèle de Felsenstein. Le temps t correspond ici à une distance sur l’arbre. Notons que le seul paramètre d’intérêt est le produit $u \times t$. Quitte à modifier en conséquence les longueurs des branches de l’arbre, on peut toujours se ramener à $u = 1$, ce que nous supposons dorénavant.

Nous ne considérerons dans la suite que des *arbres binaires*.

On va supposer dans cette section que les différents sites évoluent indépendamment les uns des autres, et tous au même taux, ce taux étant également constant dans tout l’arbre. Cette hypothèse n’est pas très réaliste, et beaucoup de travaux récents se concentrent sur la détection des sites qui évoluent plus vite que les autres, éventuellement dans une partie seulement de l’arbre, mais pour démarrer l’étude et construire un premier arbre, il est naturel de faire l’hypothèse simplificatrice que nous venons d’énoncer. Une autre hypothèse assez utilisée est que les taux d’évolution des différents sites sont des v.a. i. i. d., de loi commune une loi Gamma.

L'information à notre disposition, les *données*, est constituée d'un jeu de k séquences alignées, de longueur m , i.e. pour chaque site s , $1 \leq s \leq m$, on a k lettres dans l'alphabet \mathbf{a} , \mathbf{c} , \mathbf{g} , \mathbf{t} , une pour chaque feuille de l'arbre. A chaque arbre binaire enraciné T possédant k feuilles, on va associer la vraisemblance $L(T)$, fonction des données. La vraisemblance $L(T)$ est un produit de $s = 1$ à m des vraisemblances associées à chaque site s :

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

Chaque $L_s(T)$ se calcule en utilisant la propriété de Markov, comme nous allons maintenant le voir.

Soit T un arbre enraciné. On peut par exemple coder les noeuds d'un tel arbre comme suit, en partant de la racine, vers les feuilles (cf. figure 7.4) :

- 0 désigne la racine ;
- 1, 2 sont les “fils” de la racine, i.e. les noeuds qui sont directement reliés à la racine par une branche ;
- 1.1, 1.2 désignent les fils de 1 ; 2.1, 2.2 les fils de 2 ;
- et ainsi de suite jusqu'aux feuilles.

Pour tout noeud $\alpha \in T \setminus \{0\}$, on note ℓ_α la longueur de la branche qui joint le “père” de α à α , et on associe à α l'ensemble Λ_α des feuilles du sous-arbre dont α est la racine. En particulier, Λ_0 désigne l'ensemble des feuilles de l'arbre. Si $\alpha \in \Lambda_0$, $\Lambda_\alpha = \{\alpha\}$. Si $\alpha \in T \setminus \Lambda_0$, on note $\Gamma_\alpha = \{\alpha.1, \alpha.2\}$ les “fils” de α .

On note $\{X_\alpha, \alpha \in T\}$ les nucléotides aux noeuds de l'arbre. On suppose qu'ils constituent les valeurs aux noeuds de l'arbre d'un processus de Markov sur l'arbre de générateur infinitésimal Q . Seules les valeurs des $\{X_\alpha, \alpha \in \Lambda_0\}$ sont observées. On note x_α la valeur observée de X_α , pour $\alpha \in \Lambda_0$. La vraisemblance de l'arbre, au vu des nucléotides au site s , est

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

On va expliciter le calcul de cette quantité, ce qui mettra en évidence sa dépendance par rapport à l'arbre T .

Pour tout $\alpha \in T$, $x \in E$, on définit $L_{s,x}^{(\alpha)}$, la vraisemblance conditionnelle du sous-arbre dont α est la racine, conditionnée par $X_\alpha = x$, par la récurrence montante suivante.

- Si $\alpha \in \Lambda_0$,

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

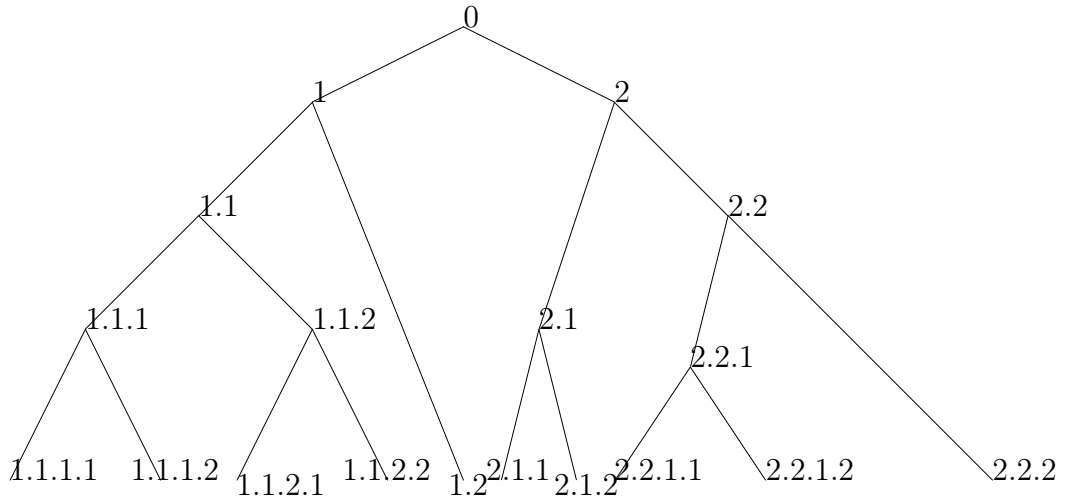


FIG. 7.4 – Arbre binaire avec racine, et codage des noeuds

– Sinon,

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

Ce calcul conduit finalement à préciser les quantités $L_{s,x}^{(0)}$, $x \in E$. Enfin

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

et

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

On aurait pu tout aussi bien écrire chaque $L_s(T)$ comme une somme de $4^{|T \setminus \Lambda_0|}$ termes. Mais les formules ci-dessus constituent l'algorithme qu'il faut utiliser en pratique.

Maximum de vraisemblance Le calcul du maximum de vraisemblance sur tous les arbres possibles est complexe. La partie la moins difficile consiste à maximiser par rapport aux longueurs des branches. Encore utilise-t-on un algorithme dont il n'est pas clair qu'il conduit à un maximum global. Celui-ci consiste à maximiser successivement par rapport à chaque longueur de branche, et à itérer la succession des maximisations tant que la vraisemblance augmente. On va voir maintenant que chaque maximisation par rapport à une longueur de branche se fait assez aisément.

Dans la mesure où les $\{X_\alpha, \alpha \in T\}$ sont issus d'un processus de Markov réversible sur l'arbre, la loi des $\{X_\alpha\}$ ne dépend pas du choix de la racine en n'importe quel noeud de l'arbre (ou plus généralement n'importe où sur une branche arbitraire).

Considérons deux noeuds voisins α et β de l'arbre. Désignons cette fois par $\ell_{\alpha\beta}$ la longueur de la branche qui les relie. Si l'on place la racine n'importe où sur cette branche, on définit comme ci-dessus des quantités $L_{s,x}^{(\alpha)}$ et $L_{s,y}^{(\beta)}$, $x, y \in E$. Attention cependant qu'il s'agit de la vraisemblance du sous-arbre dont α est la racine, dans un arbre dont on a déplacé la racine. Par exemple dans l'arbre de la figure 7.4, si $\alpha = 2$, $\beta = 2.2$, et on place la racine entre α et β , le sous-arbre dont α est la racine contient les noeuds 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, et le sous-arbre dont β est la racine contient les noeuds 2.2.1, 2.2.1.1, 2.2.1.2, 2.2.2.

Alors

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

Cette procédure permet d'expliciter la dépendance de $L_s(T)$ et de $L(T)$ par rapport à la longueur d'une branche donnée, et de calculer le maximum par rapport à cette longueur. La recherche de ce maximum est en tout cas assez simple dans le cas du modèle d'évolution que nous avons décrit ci-dessus (on maximise le logarithme de $L(T)$, ce qui remplace le produit des $L_s(T)$ par une somme, et simplifie la maximisation).

Remarque 7.8.1. *Les modèles d'évolution ne sont pas tous réversibles. Il est encore possible d'expliciter la dépendance de la vraisemblance par rapport à la longueur d'une branche donnée, mais il faut alors prendre soin d'utiliser*

la probabilité de transition du processus retourné lorsque le déplacement de la racine conduit à ce que le processus partant de la nouvelle racine parcourt une branche dans le sens inverse du sens initial.

7.8.3 L'approche bayésienne en phylogénie

Reprenons l'écriture de la vraisemblance. Notons D le vecteur des variables aléatoires qui sont observées, et d le vecteur des valeurs observées (d comme "données"), i. e. d est constitué des diverses séquences génomiques alignées.

Précisons maintenant les paramètres dont dépend la vraisemblance. Dans les paramètres inconnus (que l'on cherche à préciser), il y

- d'une part la forme de l'arbre, que nous noterons τ , qui est une inconnue dans un ensemble fini \mathcal{T} (de cardinal $(2n-3)!!$ dans le cas d'un arbre enraciné avec n feuilles, $(2n-5)!!$ dans le cas sans racine – si k est un entier impair, $k!! = 1 \times 3 \times 5 \times \dots \times k$),

- d'autre part les longueurs des diverses branches, et la matrice de transition Q du modèle d'évolution (ou du moins les paramètres de cette matrice autres que la probabilité invariante). Les longueurs de branche et les paramètres inconnus de la matrice Q varient dans une partie d'un espace euclidien $V \subset \mathbb{R}^d$. On notera cet ensemble de paramètres λ .

Le paramètre inconnu est donc le couple $\theta = (\tau, \lambda)$, dont la valeur est arbitraire dans $\Theta = \mathcal{T} \times V$, et la vraisemblance est la fonction

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

La vraisemblance de la valeur θ du paramètre inconnu est la probabilité d'observer les données que nous avons sous les yeux, si θ est la vraie valeur de ce paramètre.

Dans le point de vue bayésien, le paramètre inconnu θ est la réalisation d'une variable aléatoire, autrement dit (τ, λ) est la réalisation d'un vecteur aléatoire (T, Λ) . Prendre ce point de vue impose de se donner une *loi de probabilité a priori*, dont le bayésien nous dit qu'elle permet d'intégrer des informations a priori sur le paramètre inconnu, ce que refusent les anti-bayésiens, pour qui la seule information que l'on a le droit d'utiliser est celle contenue dans les données.

On va donc se donner une loi *a priori* pour le vecteur (T, Λ) , que l'on prendra de la forme suivante :

• on se donne la loi de T , qui est une loi sur un ensemble fini \mathcal{T} , donc on se donne des $\alpha_\tau = \mathbb{P}(T = \tau)$, $\tau \in \mathcal{T}$;

• on se donne la loi conditionnelle de Λ , sachant la valeur de T , et on suppose que pour tout $\tau \in \mathcal{T}$, la loi conditionnelle de Λ , sachant que $T = \tau$, admet une densité $q_\tau(\lambda)$, autrement dit pour toute fonction mesurable

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

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

à condition de noter $p_\tau(\lambda) = \alpha_\tau \times q_\tau(\lambda)$.

Dans ce contexte, on a un couple aléatoire, formé d'une part du "paramètre" (T, Λ) , et d'autre part des données D . La loi de ce couple est précisé par

- d'une part la loi a priori de (T, Λ) ;
- d'autre part la loi conditionnelle des données, sachant le paramètre.

Plus précisément, la vraisemblance s'interprète dans le contexte bayésien comme la probabilité conditionnelle :

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

On va alors chercher à calculer la loi *a posteriori*, qui est la loi conditionnelle du "paramètre" (T, Λ) , sachant les données, i. e. sachant que $D = d$. Cette loi conditionnelle est donnée par la célèbre "formule de Bayes", qui dans notre situation précise la loi de jointe de (T, Λ) sachant que $D = d$ sous la forme

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

Autrement dit, à nouveau si

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

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

Par exemple, on peut s'intéresser à la loi de probabilité *a posteriori* de la forme de l'arbre, i. e. de la v. a. T . Celle-ci est donnée par la formule suivante : pour tout $\tau \in \mathcal{T}$,

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

Méthode de calcul MCCM Supposons que l'on veuille calculer cette dernière quantité, pour un petit nombre de valeurs de τ . Un calcul explicite est sans espoir, compte tenu de la taille des données (nombre d'espèces considérées) et de la complexité croissante des modèles utilisés. On est donc amené à utiliser une méthode de type Monte Carlo, utilisant des tirages aléatoires. Cependant, il se présente une sérieuse difficulté pour réaliser des tirages aléatoires suivant la loi a posteriori de (T, Λ) , sachant les données, c'est que pour identifier cette loi, il est nécessaire de calculer le dénominateur dans les formules ci-dessus. Pour peu en particulier que le cardinal de \mathcal{T} soit gigantesque, ce calcul peut se révéler totalement impossible.

On se trouve exactement dans la même situation que celle décrite au début de la section 3.1.

Rappelons l'algorithme de Metropolis-Hastings. L'idée est la suivante. Soit Q une matrice de transition sur F (qui n'a a priori rien à voir avec la probabilité π), dont on soit capable de simuler aisément les transitions. On choisit comme matrice P la matrice de transition dont les termes hors diagonaux sont donnés par

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

et les termes diagonaux sont donnés par

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

sous réserve que P ainsi définie soit irréductible, ce qui est par exemple vrai si Q est irréductible et vérifie la propriété : pour tout x, y , $Q_{xy} > 0 \Leftrightarrow Q_{yx} > 0$. Il est clair que cette matrice P est bien une matrice de transition ($P_{xy} \leq Q_{xy}$, $x \neq y$ implique que $P_{xx} \geq 0$), et que π est invariante par P , puisque la relation d'équilibre ponctuel

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

est bien satisfaite. Pour expliquer comment on simule les transitions de la chaîne de matrice de transition P , posons pour $x, y \in F$,

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

Une façon de simuler une transition de la chaîne $\{X_k\}$ de matrice de transition P est la suivante. Supposons que $X_k = x$, et on veut simuler X_{k+1} . On simule

d'abord une transition de la chaîne $\{Y_k\}$ de matrice de transition Q , partant de $Y_k = x$. Supposons que le résultat de ce tirage soit $Y_{k+1} = y$. On *accepte cette transition* (et dans ce cas $X_{k+1} = y$) avec la probabilité $r(x, y)$; on *refuse cette transition* (et dans ce cas $X_{k+1} = x$) avec la probabilité $1 - r(x, y)$. Remarquons $r(x, x) = 1$, donc dans le cas $y = x$, la “non-transition” est toujours acceptée.

Autrement dit, pour passer de $X_k = x$ à X_{k+1} , on effectue

- le tirage de Y_{k+1} , suivant la probabilité Q_x ;
- le tirage de U_{k+1} , de loi uniforme sur $[0, 1]$;

et on pose

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

Mise en oeuvre de la méthode MCCM La mise en oeuvre de la méthode MCCM pose des problèmes délicats, pour lesquels on n'a essentiellement pas de réponse satisfaisante, en particulier dans le cas de la phylogénie. Nous avons déjà discuté cette question dans un cadre général à la section 3.3. Rappelons qu'il convient d'éliminer le début de la simulation (*burn-in*). En outre, pour obtenir un échantillon de la loi a posteriori, on ne retient qu'une itération toutes les n , où n est choisi en fonction de la vitesse de décorrélation de la chaîne, laquelle peut être estimée à partir des simulations. Certaines mises en oeuvre consistent à simuler plusieurs chaînes en parallèle, dont éventuellement certaines sont “chauffées”, voir la section 3.1.4.

7.9 Application aux EDP discrétisées

Soit D un domaine borné de \mathbb{R}^2 (le résultat qui suit est en fait vrai en dimension quelconque), de frontière ∂D lipschitzienne. On suppose pour fixer les notations que l'origine $0 \in D$. On considère le problème de Dirichlet :

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

avec $f \in C(D^e)$. Il est bien connu que cette équation admet une solution unique u élément de $C(D)$.

Etant donné $h > 0$, soit $h\mathbf{Z}^2$ l'ensemble des points du plan dont les coordonnées sont des multiples entiers relatifs de h . On pose $D_h = D \cap h\mathbf{Z}^2$.

∂D_h est constitué des points de $D^c \cap h\mathbf{Z}^2$ qui sont à distance h d'au moins un point de D_h , et $\bar{D}_h = D_h \cup \partial D_h$. e_1 et e_2 désignant les deux vecteurs de base, on définit l'opérateur approché Δ_h par :

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

D'après l'Exercice 7.11.4, la solution du problème de Dirichlet discrétisé

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

est donnée par la formule $u_h(x) = \mathbb{E}[f(X_{T_{D_h^c}^h}^h) | X_0^h = x]$, où $\{X_t^h; t \geq 0\}$ est un processus markovien de sauts de générateur infinitésimal $\frac{1}{2}\Delta_h$, à valeurs dans $h\mathbf{Z}^2$, et

$$T_{D_h^c}^h = \inf\{t \geq 0; X_t^h \in D_h^c\}.$$

Remarquons que $\{X_t^h, t \geq 0\}$ a même loi que $\{hX_{h^{-2}t}^1, t \geq 0\}$, et que d'après le début de la section 2.3 ce dernier processus converge vers un mouvement brownien standard unidimensionnel, quand $h \rightarrow 0$. Il n'est pas trop difficile (à condition tout de même notamment de bien comprendre cette notion de convergence de processus!) d'en déduire que

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

(où $T_{D^c} = \inf\{t \geq 0; B_t \in D^c\}$), cette dernière formule donnant la solution du problème de Dirichlet. On a ainsi une preuve (alternative aux méthodes classiques d'analyse numérique) de la convergence de u_h vers u .

Notons enfin que le problème de Dirichlet discrétisé pourrait aussi s'interpréter en terme d'une chaîne de Markov en temps discret, de matrice de transition $\Delta_h + I$.

Ces interprétations probabilistes permettent aussi d'utiliser des méthodes de Monte Carlo pour le calcul approché de solutions d'EDP. Ces méthodes sont très utilisées dans certains cas où les méthodes d'analyse numérique "classique" ne sont pas efficaces, ou sont inutilisables (en particulier dans les cas de dimension trop grande), cf. [29]. Elles sont par ailleurs très appréciées pour leur simplicité de programmation. Quelques lignes de code suffisent pour programmer le calcul de la solution approchée d'une équation aux dérivées

partielles! Même s'il faut laisser tourner l'ordinateur un peu plus longtemps pour avoir la même précision qu'avec une méthode de différences, éléments ou volumes finis, le fait que la programmation soit facile, rapide, et surtout permette facilement la reprise du programme par un autre programmeur, quand le premier aura quitté le service ou l'entreprise.

7.10 Le recuit simulé (suite)

Dans cette section, E est supposé fini. Rappelons que que l'on cherche à maximiser une fonction

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

telle que, pour fixer les idées,

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

Autrement dit, on cherche un des x tels que $U_x = 0$.

A chaque $\beta > 0$ on associe le générateur infinitésimal $Q = \{Q_{xy}, x, y \in E\}$ donné pour $x \neq y$ par

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

où G est un graphe non orienté dans E , i.e. une collection de paires de point de E , choisie de telle sorte que le processus de générateur Q_{xy} soit irréductible (ce qui revient à dire que $\forall x, j \in E, \exists n$ et x_1, x_2, \dots, x_n tels que $(x, x_1) \in G, (x_1, x_2) \in G, \dots, (x_n, y) \in G$).

Le processus markovien de générateur infinitésimal Q est clairement réversible par rapport à sa probabilité invariante π_β définie par

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

On définit la **forme de Dirichlet** associée à Q comme la forme bilinéaire sur \mathbb{R}^E :

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

où on a utilisé la réversibilité et deux fois l'identité $\sum_y Q_{xy} = 0$, donc $-Q : \ell^2(\pi) \rightarrow \ell^2(\pi)$ est autoadjoint semi défini positif.

Définition 7.10.1. On appelle **trou spectral** de Q la quantité :

$$\lambda \stackrel{\text{def}}{=} \inf_{\varphi \text{ non constante}} \frac{\mathcal{E}(\varphi, \varphi)}{\text{Var}_\pi(\varphi)},$$

$$\text{où } \text{Var}_\pi(\varphi) = \sum_{x \in E} \varphi_x^2 \pi_x - \left(\sum_{x \in E} \varphi_x \pi_x \right)^2.$$

Lemme 7.10.2. Q étant le générateur infinitésimal d'un processus de Markov irréductible à valeurs dans un espace fini E , son trou spectral vérifie

$$\lambda > 0.$$

PREUVE D'après la formule ci-dessus pour $\mathcal{E}(\varphi, \varphi)$, le quotient

$$\frac{\mathcal{E}(\varphi, \varphi)}{\text{Var}_\pi(\varphi)}$$

reste inchangé si l'on ajoute à φ une constante. On peut donc se contenter de minimiser ce rapport pour φ tel que $\mathbb{E}_\pi(\varphi) = 0$, d'où

$$\lambda = \inf_{\varphi \neq 0; \mathbb{E}_\pi(\varphi) = 0} \frac{\langle \varphi, -Q\varphi \rangle_\pi}{\langle \varphi, \varphi \rangle_\pi},$$

et λ est la plus petite valeur propre de $-Q$, considéré comme opérateur linéaire sur $\ell^2(\pi)$, restreint à l'orthogonal des constantes. Comme Q est autoadjoint semi défini positif, il suffit de montrer que l'espace propre associé à la valeur propre zéro est restreint aux constantes. Mais si φ appartient à cet espace propre,

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

donc a fortiori

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

ce qui, puisque Q est irréductible, entraîne bien que φ est constante. \square

Soit maintenant $\{X_t, t \geq 0\}$ un processus de Markov de générateur infinitésimal Q . Pour tout $t > 0$, on note $\mu(t) = (\mu_x(t); x \in E)$ la loi de X_t . On pose

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

et on remarque que $\varepsilon(t) = 0$ ssi $\mu(t) = \pi$.

Lemme 7.10.3. λ désignant le trou spectral de Q ,

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

PREUVE On remarque tout d'abord que

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

Donc

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

Soit

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

□

On a montré que $\mu_t \rightarrow \pi$ à vitesse exponentielle (comparer avec le théorème 2.6.7).

On en arrive maintenant au “recuit”. On va faire dépendre β de t , et le faire tendre vers l’infini (et donc la “température”, son inverse, vers zéro) quand $t \rightarrow \infty$. Plus précisément, Δ étant une constante à déterminer, on choisit

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

donc $\beta(0) = 0$ et $\beta(t) \rightarrow +\infty$ quand $t \rightarrow \infty$. Bien sûr, la chaîne n’est plus homogène, le générateur infinitésimal Q , le trou spectral λ , la mesure invariante π , et la constante de normalisation Z deviennent des fonctions de t : $Q(t)$, $\lambda(t)$, $\pi(t)$, $Z(t)$. Notons que $\pi(0)$ est la mesure uniforme sur E , $Z(0) = |E|^{-1}$, alors que $\pi(\infty) = \lim_{t \rightarrow \infty} \pi(t)$ est la mesure uniforme sur les zéros de U (i.e. sur les maxima de U).

Notre but est de montrer le $[M \stackrel{\text{def}}{=} \sup_{i \in E} (-U_i)]$:

Théorème 7.10.4. *Si $\Delta > M$, $\mu(t) \rightarrow \pi(\infty)$ quand $t \rightarrow \infty$.*

Nous allons tout d’abord montrer le :

Lemme 7.10.5.

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

PREUVE Choisissons φ tel que $\mathbb{E}_{\pi(0)}[\varphi] = 0$. Alors par définition de $\lambda(0)$,

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

Par ailleurs,

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

Donc

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

Les deux termes extrêmes de cette inégalité étant invariants par addition d'une constante à φ , l'inégalité résultante est vraie sans la restriction que $\mathbb{E}_{\pi(0)}[\varphi] = 0$. Le lemme est établi, puisque

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

PREUVE DU THÉORÈME : Il suffit de montrer que $\varepsilon(t) \rightarrow 0$, où

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

Notons que $\varepsilon(t)$ est un majorant du carré de la norme L^1 de la différence $\mu(t) - \pi(t)$. En effet, par Cauchy-Schwarz,

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

On a

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

Comme $\Delta > M$, $(1+t)^{-M/\Delta} \gg (1+t)^{-1}$ quand $t \rightarrow \infty$, il existe $c > 0$ tel que, pour t assez grand :

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

et le théorème résulte du

Lemme 7.10.6. Soit $x, b \in C^1(\mathbb{R}_+; \mathbb{R}_+)$, $a \in C(\mathbb{R}_+; \mathbb{R}_+)$ tels que

- (i) $\int_0^\infty a(t) dt = +\infty$;
- (ii) $b(t) \searrow 0$ quand $t \rightarrow \infty$;
- (iii) $\frac{dx}{dt}(t) \leq -a(t)(x(t) - b(t))$.

Alors $x(t) \rightarrow 0$ quand $t \rightarrow \infty$.

PREUVE

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

Donc, en intégrant, on obtient :

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

Le second membre de l'inégalité ci-dessus est la solution d'une EDO linéaire, qui majore $x(t)$. Donc il suffit d'établir le résultat pour $x(t)$ solution de l'EDO, c'est à dire que l'on peut supposer que :

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

On pose $y(t) = x(t) - b(t)$. Il suffit de montrer que $y(t) \rightarrow 0$ quand $t \rightarrow \infty$.

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

Notons que $b'(t) \leq 0$, et $-\int_t^\infty b'(s)ds = b(t) < \infty$. Donc pour $t \geq N$,

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

Soit $\delta > 0$ arbitraire. On choisit N assez grand pour que la somme des deux premiers termes du membre de droite soit plus petit que $\frac{\delta}{2}$. En choisissant alors t assez grand, le troisième terme à son tour est plus petit que $\frac{\delta}{2}$. Le lemme est établi.

Remarque 7.10.7. La fonction $\beta(t) = \frac{1}{\Delta} \log(1+t)$ tend trop lentement vers l'infini quand $t \rightarrow \infty$, pour être utilisée en pratique. On peut montrer des résultats plus faibles que $\mu(t) \rightarrow \pi(\infty)$ avec une fonction β qui croît plus vite qu'un logarithme (fonction puissance). D'un autre côté, si on se pose la question d'atteindre le meilleur résultat possible sur un horizon fini fixé, on peut montrer qu'il existe des programmes de croissance de β à vitesse exponentielle qui sont proches de l'optimum.

7.11 Exercices

Exercice 7.11.1. Soit $\{T_n, n \geq 1\}$ un processus ponctuel de Poisson d'intensité λ , et $\{Z_n, n \geq 0\}$ une chaîne de Markov indépendante de $\{T_n, n \geq 1\}$ à valeurs dans E , de matrice de transition $P_{xy}, x, y \in E$. On pose

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

Montrer que $\{X_t, t \geq 0\}$ est un processus markovien de sauts dont on précisera la matrice de transition, le générateur infinitésimal, et la loi de l'instant du premier saut.

Exercice 7.11.2. Soit $\{X_t; t \geq 0\}$ un processus markovien de sauts à valeurs dans l'ensemble fini ou dénombrable E , de générateur infinitésimal $\{Q_{xy}; x, y \in E\}$. On suppose que $\lambda := \sup_x -Q_{xx} < \infty$. Soit $\{N_t; t \geq 0\}$ le processus de comptage des sauts de $\{X_t\}$, et $\{N'_t, t \geq 0\}$ un processus de Poisson d'intensité λ .

Comparer $\mathbb{P}(N_t \geq n)$ et $\mathbb{P}(N'_t \geq n)$, ainsi que $\mathbb{E}[f(N_t)]$ et $\mathbb{E}[f(N'_t)]$, avec f une fonction croissante de \mathbb{N} dans \mathbb{R} . Montrer que l'exercice 7.11.1 fournit une autre démonstration de ce résultat.

Exercice 7.11.3. Soit $\{N_t, t \geq 0\}$ et $\{P_t, t \geq 0\}$ deux processus de Poisson indépendants, d'intensité respective λ et μ .

a Montrer que $\{X_t, t \geq 0\}$ défini par

$$X_t = N_t - P_t$$

est un processus markovien de sauts irréductible à valeurs dans \mathbf{Z} , dont on précisera le générateur infinitésimal.

b On suppose $\lambda \neq \mu$. Montrer que $\{X_t/t\}$ et $\{X_t\}$ convergent p.s. dans $\bar{\mathbb{R}}$ quand $t \rightarrow \infty$. Préciser la limite de $\{X_t\}$ suivant le signe de $\lambda - \mu$. Montrer que $\{X_t\}$ est transitoire.

c On suppose que $\lambda = \mu$. Préciser la loi de la chaîne incluse. Dédurre des exercices 2.9.7 et 2.9.9 que $\{X_t\}$ est récurrent nul.

Exercice 7.11.4. a Soit $\{X_t, t \geq 0\}$ un processus markovien de sauts à valeurs dans E , de générateur infinitésimal $\{Q_{xy}; x, y \in E\}$. Soit F

une partie de E . On définit

$$T_F : \begin{cases} \inf\{t; X_t \in F\}, & \text{si un tel } t \text{ existe;} \\ \infty, & \text{sinon,} \end{cases}$$

la fonction $u : E \rightarrow \mathbb{R}$ par

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

où h est une application bornée de F dans \mathbb{R} , et la fonction $v : E \rightarrow \mathbb{R} \cup \{+\infty\}$ par

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

Montrer que T_F est un temps d'arrêt.

Montrer que u et v vérifient respectivement les équations :

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

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

(on introduira le conditionnement par $(T_1, X(T_1))$).

- b** On considère maintenant le cas d'un processus de naissance et de mort sur $E = \mathbf{Z}$, i.e. le générateur infinitésimal Q satisfaisant $Q_{x,x+1} = \alpha(x)$, $Q_{x,x-1} = \beta(x)$, $Q_{x,x} = -\alpha(x) - \beta(x)$, dans le cas particulier $\alpha(x) = \alpha$, $\beta(x) = \beta$, $x \in \mathbf{Z}$ ($\alpha, \beta > 0$). On pose $F = \{1, 2, \dots, N-1\}^c$, où N est un entier positif. Calculer $u(x) = \mathbb{E}[X_{T_F} | X_0 = x]$, $x \in \mathbf{Z}$. Montrer que T_F est p.s. fini. Déterminer la loi conditionnelle de la v.a. X_{T_F} , sachant que $X_0 = x$.

Exercice 7.11.5. Etant donné un espace de probabilité $(\Omega, \mathcal{F}_t, P)$ muni d'une filtration $\{\mathcal{F}\}$ (i.e. une collection croissante indexée par $t \in \mathbb{R}_+$ de sous-tribus de \mathcal{A}), on appelle **martingale** (par rapport à la filtration $\{\mathcal{F}_t\}$) un processus stochastique $\{M_t, t \in \mathbb{R}_+\}$ qui vérifie :

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

- a** Soit $\{M_t, t \in \mathbb{R}_+\}$ une martingale continue à droite et S un temps d'arrêt borné par une constante t . Montrer que $\mathbb{E}[M_S] = \mathbb{E}[M_t] = \mathbb{E}[M_0]$.
- b** Soit $\{X_t, t \geq 0\}$ un processus markovien de sauts à valeurs dans E , de générateur infinitésimal $\{Q_{xy}; x, y \in E\}$, et f une application bornée de E dans \mathbb{R} . Montrer que $\{M_t, t \in \mathbb{R}_+\}$ définie par

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

est une martingale par rapport à la filtration $\{\mathcal{F}_t^X\}$;

- c** On reprend les notations de la deuxième partie de l'exercice précédent. Calculer $\mathbb{E}(T_F | X_0 = x)$ (pour le cas $\alpha \neq \beta$, on utilisera les résultats des deux questions précédentes avec la fonction $f(x) = x$ et le temps d'arrêt $S = \inf(T_F, t)$, puis on fera tendre t vers l'infini; pour le cas $\alpha = \beta$, on fera le même calcul, avec $f(x) = x^2$). On admettra que le résultat de la question **b** s'applique à ces deux fonctions, bien qu'elles ne soient pas bornées.

7.12 Problèmes corrigés

7.12.1 Enoncés

Problème 7.12.1. On considère à la fois la chaîne de Markov en temps discret $\{X_n; n \in \mathbb{N}\}$ à valeurs dans $E = \mathbb{N}$, de matrice de transition

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

et le processus markovien de sauts en temps continu $\{X_t; t \geq 0\}$ à valeurs dans $E = \mathbb{N}$, de générateur infinitésimal

$$Q = \begin{pmatrix} -p & p & 0 & 0 & \dots \\ q & -1 & p & 0 & \dots \\ 0 & q & -1 & p & \dots \\ 0 & 0 & q & -1 & p\dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$