Master de Mathématiques

Ingénierie Mathématique Informatique et Statistique (IMIS)

Mathématiques et Applications

Processus stochastiques

niveau 1

Michel Roussignol

Année 2007-2008

Introduction

Ce texte correspond à 12 heures de cours en première année de master de mathématiques appliquées. Il s'adresse à des étudiants ayant suivi un cours de probabilité en licence de mathématiques. Son objectif est d'étudier les chaînes de markov à valeurs dans un espace fini ou dénombrable et de donner une première vision du processus de Poisson. Ces deux processus sont utilisés dans de nombreuses applications et sont deux instruments de base de la boite à outils du mathématicien appliqué. Ils sont les deux processus les plus "simples" à étudier, c'est pourquoi ils apparaissent dans le cours de processus stochastiques niveau 1. Le cours de processus stochastiques niveau 2 permettra aux étudiants de rencontrer d'autres processus fondamentaux comme les processus markoviens de sauts, le mouvement Brownien, les diffusions.

Ce polycopié a été écrit à partir du polycopié de Christiane Cocozza-Thivent, qui enseignait ce cours il y a deux ans, en tenant compte d'apports de Marie-Claire Quenez, qui assurait ce cours l'an dernier.

Les séances de travaux dirigés associées à ce cours sont fondamentales pour assurer sa compréhension.

Table des matières

Chapitre 1

Chaîne de Markov : propriété de Markov

Nous allons étudier des suites $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires ayant une propriété particulière de dépendance entre ces variables aléatoires, la propriété de Markov. Nous désignerons ces suites sous le nom de chaîne de Markov. Le plus souvent ces suites représentent l'évolution dans le temps d'une quantité aléatoire : évolution dans le temps de l'état d'un composant d'un système, évolution dans le temps de la valeur d'un bien, évolution dans le temps de la taille d'une catégorie de population, évolution dans le temps d'indicateurs macro-économiques,... Ainsi on rencontre des chaînes de Markov dans de nombreux secteurs : économie, finance, biologie, physique, études de risque,...

Le nom Markov est celui d'un mathématicien russe, Andreï Andreïevitch Markov (1856-1922). On doit à cet élève de Tchebychev de très importants travaux en calcul des probabilités et en théorie du potentiel.

Dans ce cours nous supposerons que les variables aléatoires X_n prennent leurs valeurs dans un espace E fini ou dénombrable. L'espace E s'appelle l'espace d'états.

1.1 Définition d'une chaîne de Markov

La définition suivante exprime qu'une chaîne de Markov est une chaîne sans mémoire, c'est-à-dire que chaque variable aléatoire X_k dépend du passé $X_{k-1}, X_{k-2}, ..., X_0$ uniquement par l'intermédiaire de X_{k-1}.

Définition 1.1 *Une suite $(X_n)_{n\in N}$ de variables aléatoires sur un espace de probabilité à valeurs dans espace E fini ou dénombrable est une* **chaîne de Markov** *sur E si, pour tout entier strictement positif k et pour toute suite $x_0, x_1, \ldots, x_k$ d'éléments de E pour laquelle $\mathbb{P}(X_{k-1} = x_{k-1}, X_{k-2} = x_{k-2}, ..., X_1 = x_1, X_0 = x_0) \neq 0$, nous avons :*

$$\begin{aligned}&\mathbb{P}(X_k = x_k \,/\, X_{k-1} = x_{k-1}, X_{k-2} = x_{k-2}, ..., X_1 = x_1, X_0 = x_0)\\ &= \mathbb{P}(X_k = x_k \,/\, X_{k-1} = x_{k-1})\end{aligned}$$

La propriété dans cette définition est la propriété de Markov.

Proposition 1.2 *La suite $(X_n)_{n\in N}$ une chaîne de Markov sur E si et seulement si pour pour toute suite $x_0, x_1, \ldots, x_m$ d'éléments de E :*

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_m = x_m) = \\ \mathbb{P}(X_0 = x_0) \prod_{i=1}^{m} \mathbb{P}(X_i = x_i / X_{i-1} = x_{i-1})$$

Démonstration : Une généralisation facile de la formule de Bayes donne :

$$\begin{aligned}&\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_m = x_m)\\ &= \mathbb{P}(X_0 = x_0)\mathbb{P}(X_1 = x_1 \,/\, X_0 = x_0)\, \mathbb{P}(X_2 = x_2 \,/\, X_0 = x_0, X_1 = x_1)\\ &\quad \ldots \mathbb{P}(X_m = x_m \,/\, X_{m-1} = x_{m-1}, X_{m-2} = x_{m-2}, ..., X_1 = x_1, X_0 = x_0).\end{aligned}$$

En utilisant la définition d'une chaîne de Markov, les expressions de la forme

$$\mathbb{P}(X_i = x_i \,/\, X_{i-1} = x_{i-1}, X_{i-2} = x_{i-2}, \ldots, X_1 = x_1, X_0 = x_0)$$

se simplifient en $P(X_i = x_i \,/\, X_{i-1} = x_{i-1})$, ce qui donne la formule cherchée pour une chaîne de Markov.

Inversement, si la formule ci-dessus est vraie, il est très facile de vérifier la proriété de Markov.. ■

La loi d'une chaîne de Markov est donc entièrement caractérisée par la donnée de la loi de X_0 et des probabilités conditionnelles $\mathbb{P}(X_k = x/X_{k-1} = y)$ pour $k \geq 1$, x et y dans E. La loi de X_0 s'appelle la **loi initiale** de la chaîne.

Définition 1.3 *Une chaîne de Markov est dite* **homogène** *si pour tout $k \geq 1$ et tous x et y dans E, on a :*

$$\mathbb{P}(X_k = x/X_{k-1} = y) = \mathbb{P}(X_1 = x/X_0 = y)$$

Pour une chaîne de Markov homogène, on note $P(y,x) = \mathbb{P}(X_1 = x/X_0 = y)$. Cette fonction P définie sur $E \times E$ vérifie les propriétés suivantes :

- $\forall\, x \in E, \forall\, y \in E,\ 0 \leq P(x,y) \leq 1,$
- $\forall\, x \in E,\ \sum_{y \in E} P(x,y) = 1.$

Nous appellerons une telle fonction une fonction de transition.

Etant données deux fonctions de transition P et Q définies sur le même espace E, nous définissons la fonction de transition produit $P\,Q$ par :

$$P\,Q(x,y) = \sum_{z \in E} P(x,z)\,Q(z,y).$$

Les puissances successives (ou itérées) de la fonction de transition P s'écrivent :

$$P^2 = P\,P, \quad P^{n+1} = P\,P^n = P^n\,P \quad n \geq 1.$$

Plus généralement, on a :

$$P^{n+m} = P^n\,P^m = P^m\,P^n \quad n,m \geq 1.$$

Par convention, on pose :

$$P^0 = I,$$

c'est-à-dire :

$$P^0(x,y) = \begin{cases} 1 & \text{si } x = y, \\ 0 & \text{si } x \neq y. \end{cases}$$

Lorsque l'ensemble E est fini et comporte N éléments on peut toujours supposer que $E = \{1, 2, ..., N\}$, la fonction P s'identifie alors à une matrice $N \times N$ dont l'élément de la $i^{ème}$ ligne et de la $j^{ème}$ colonne est $P(i,j)$. Une telle matrice s'appelle une **matrice de transition** ou une **matrice stochastique**. Elle est caractérisée par le fait que tous ses éléments sont compris entre 0 et 1 et que la somme des éléments de chaque ligne est égale à 1. La notion de produit et de puissance de fonctions de transition correspond évidemment aux notions de produit et de puissance de matrices de transition dans le cas fini.

Pour une chaîne de Markov $(X_n)_{n \geq 0}$ de probabilité de transition $P(x,y)$, nous avons :

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_m = x_m) = \mathbb{P}(X_0 = x_0) \prod_{i=1}^{m} P(x_{i-1}, x_i)$$

La loi d'une chaîne de Markov homogène est donc entièrement caractérisée par la donnée de la loi initiale (loi de X_0) et de la fonction de transition P.

Lorsque l'on étudie une chaîne de Markov homogène, on est souvent amené à visualiser une partie du comportement de la chaîne en traçant ce que l'on appelle le graphe d'états ou graphe de Markov. Ce graphe d'états est un graphe orienté dont les sommets sont les différents états de la chaîne et dont les arcs orientés relient les sommets x et y tels $P(x,y) > 0$. Si on souhaite donner tous les renseignements sur le graphe, on construit un graphe valué en affectant à l'arc (x,y) le poids $P(x,y)$.

Pour simuler les N premières valeurs d'une chaîne de Markov homogène à valeurs dans $E = \{a_1, a_2, \ldots, a_p, \ldots\}$, de loi initiale μ, de fonction de transition P, il suffit d'appliquer la proposition 1.2, ce qui donne l'algorithme suivant :

1. tirer une v.a. U_0 de loi uniforme sur $[0,1]$,
2. si $\sum_{j=1}^{k-1} \mu(a_j) \leq U_0 < \sum_{j=1}^{k} \mu(a_j)$, poser $X_0 = a_k$,
3. pour $i = 1$ à N :
   (a) tirer une v.a. U_i de loi uniforme sur $[0,1]$, indépendante des précédentes,
   (b) si $\sum_{j=1}^{k-1} P(X_{i-1}, a_j) \leq U_i < \sum_{j=1}^{k} P(X_{i-1}, a_j)$, poser $X_i = a_k$.

Connaissant la loi initiale et la fonction de transition, il est facile de calculer la loi de X_n.

Proposition 1.4 *Si $(X_n)_{n\geq 0}$ est une chaîne de Markov homogène de fonction de transition P, nous avons pour $n \geq 1$:*

$$\mathbb{P}(X_n = x / X_0 = y) = P^n(y,x)$$

et

$$\mathbb{P}(X_n = x) = \sum_{y \in E} \mathbb{P}(X_0 = y)\ P^n(y,x)$$

Démonstration : Nous avons :

$$\begin{aligned}
&\mathbb{P}(X_0 = y, X_n = x) \\
&\quad = \sum_{x_1,\ldots,x_{n-1}} \mathbb{P}(X_0 = y, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x) \\
&\quad = \sum_{x_1,\ldots,x_{n-1}} \mathbb{P}(X_0 = y)\ P(y,x_1)\ P(x_1,x_2) \ldots P(x_{n-2},x_{n-1})\ P(x_{n-1},x) \\
&\quad = \mathbb{P}(X_0 = y)\ P^n(y,x)
\end{aligned}$$

Les deux résultats de la proposition suivent immédiatement. ■

1.2 Exemples

Dans ce chapitre, nous allons donner une série d'exemples de chaînes de Markov.

Exemple 1 : chaîne de Markov homogène sur $E = \{1, 2, 3, 4, 5, 6\}$ dont la matrice de transition est :

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.3 & 0 & 0.7 & 0 & 0 \\ 0.2 & 0.2 & 0.3 & 0.3 & 0 & 0 \\ 0 & 0.8 & 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.2 & 0.3 \\ 0 & 0 & 0.4 & 0 & 0.1 & 0.5 \end{pmatrix}$$

Pour visualiser le comportement de cette chaîne, on peut tracer le graphe de Markov associé :

On constate que la chaîne finira par atteindre et rester dans l'état 1 ou l'ensemble des deux états $\{2, 4\}$ quelque soit le point de départ.

Exemple 2 : chaîne à deux états

Considérons une machine qui peut être dans deux états : panne (noté 0) ou fonctionnement (noté 1). Supposons que, si la machine est en panne le matin du jour n, la probabilité pour qu'elle soit réparée ce jour-là et donc pour qu'elle soit en fonctionnement le lendemain matin (date $n + 1$) est a $(0 \leq a \leq 1)$, alors que si elle est en état de marche le matin du jour n la probabilité pour qu'elle soit en panne le lendemain matin est b $(0 \leq b \leq 1)$. Notons X_n l'état de la chaine le matin du jour n et μ "l'état" initial de la machine c'est-à-dire la probabilité pour qu'elle soit dans chacun des deux états à l'instant initial (jour 0). Puisque l'état de la chaîne à la date $n+1$ ne dépend que de l'état de la chaîne à la date n, la suite X_n forme une chaîne de Markov homogène d'espace d'état $\{0, 1\}$ et de matrice de transition :

$$P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$$

Cherchons $\mathbb{P}(X_n = i)$, $i = 0, 1$. Nous devons calculer P^n.

On va diagonaliser P. Cherchons ses valeurs propres :

$$\begin{aligned} \det(P - \lambda I) &= \begin{vmatrix} 1-a-\lambda & a \\ b & 1-b-\lambda \end{vmatrix} \\ &= (1-a-\lambda)(1-b-\lambda) - ab \\ &= \lambda^2 - \lambda(2-a-b) + 1 - a - b. \end{aligned}$$

La valeur $\lambda_1 = 1$ est racine, ce qui est normal puisque le fait que la somme des lignes de P valent 1 entraine que $(1, 1)$ est vecteur propre associé à la valeur propre 1.

On en déduit $\lambda_2 = 1 - a - b$. Les vecteurs propres (x, y) associés vérifient : $bx + ay = 0$. On choisit $x = a$ et $y = -b$. La matrice de changement de base est donc :

$$Q = \begin{pmatrix} 1 & a \\ 1 & -b \end{pmatrix}.$$

On en déduit :

$$Q^{-1} = \frac{1}{a+b} \begin{pmatrix} b & a \\ 1 & -1 \end{pmatrix}.$$

Par conséquent

$$D = \begin{pmatrix} 1 & 0 \\ 0 & 1-a-b \end{pmatrix} = Q^{-1}\, P\, Q,$$

et donc $P = Q\, D\, Q^{-1}$ et par suite

$$P^n = Q\, D^n\, Q^{-1} = \frac{1}{a+b} \begin{pmatrix} b + a(1-a-b)^n & a - a(1-a-b)^n \\ b - b(1-a-b)^n & a + b(1-a-b)^n \end{pmatrix}.$$

Par conséquent :

$$\begin{aligned} \begin{pmatrix} \mathbb{P}(X_n = 0) \\ \mathbb{P}(X_n = 1) \end{pmatrix} &= \begin{pmatrix} \mu(0) & \mu(1) \end{pmatrix} P^n \\ &= \frac{1}{a+b} \begin{pmatrix} \mu(0)(b + a(1-a-b)^n) + \mu(1)(b - b(1-a-b)^n)] \\ \mu(0)[a - a(1-a-b)^n] + \mu(1)[a + b(1-a-b)^n \end{pmatrix}, \end{aligned}$$

ce qui peut encore s'écrire :

$$\mathbb{P}(X_n = 0) = \frac{b}{a+b} + (1-a-b)^n(\mu(0) - \frac{b}{a+b})$$

et

$$\mathbb{P}(X_n = 1) = \frac{a}{a+b} + (1-a-b)^n(\mu(1) - \frac{a}{a+b})$$

Supposons que a et b ne soient ni tous deux égaux à 0 ni tous deux égaux à 1. Alors $0 < a + b < 2$ et donc $|1 - a - b| < 1$. Il s'ensuit que

$$lim_{n \to +\infty} \mathbb{P}(X_n = 0) = \frac{b}{a+b}, \quad \lim_{n \to +\infty} \mathbb{P}(X_n = 1) = \frac{a}{a+b}.$$

Les quantités $\frac{b}{a+b}$ et $\frac{a}{a+b}$ peuvent également être retrouvées d'une autre manière. Si l'on désire choisir $\mu(0)$ et $\mu(1)$ pour que $\mathbb{P}(X_n = 0)$ et $\mathbb{P}(X_n = 1)$ ne dépendent pas de n, on voit qu'il faut prendre $\mu(0) = \frac{b}{a+b}$ et $\mu(1) = \frac{a}{a+b}$

et alors nous aurons, pour tout n : $\mathbb{P}(X_n = 0) = \frac{b}{a+b}$ et $\mathbb{P}(X_n = 1) = \frac{a}{a+b}$. Nous retrouverons ce phénomène pour une classe très générale de chaînes de Markov.

Exemple 3 : considérons une suite de variables aléatoires $(X_n)_{n\geq 0}$ à valeurs dans un ensemble fini ou dénombrable E qui est définie par X_0 et par la formule de récurence pour tout $n \geq 0$:

$$X_{n+1} = f(X_n, U_n)$$

où f est une fonction de $E \times \mathbb{R}$ dans E et $(U_n)_{n\geq 0}$ est une suite de variables aléatoires réelles indépendantes entre elles et indépendantes de X_0.

La suite de variables aléatoires $(X_n)_{n\geq 0}$ est une chaîne de Markov homogène. En effet :

$$\begin{aligned}
&\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_m = x_m) \\
&\quad = \mathbb{P}(X_0 = x_0, f((x_0, U_0) = x_1, \ldots, f(x_{m-1}, U_{m-1}) = x_m) \\
&\quad = \mathbb{P}(X_0 = x_0) \prod_{i-1}^{m} \mathbb{P}(f(x_{i-1}, U_{i-1}) = x_i)
\end{aligned}$$

vu l'indépendance des variables aléatoires $X_0, U_0, U_1, \ldots, U_{m-1}$.

D'autre part

$$\begin{aligned}
\mathbb{P}(X_i = y / X_{i-1} = x) &= \frac{\mathbb{P}(X_i = y, X_{i-1} = x)}{\mathbb{P}(X_{i-1} = x)} \\
&= \frac{\mathbb{P}(f(x, U_{i-1}) = y, X_{i-1} = x)}{\mathbb{P}(X_{i-1} = x)} \\
&= \frac{\mathbb{P}(f(x, U_{i-1}) = y)\, \mathbb{P}(X_{i-1} = x)}{\mathbb{P}(X_{i-1} = x)} \\
&= \mathbb{P}(f(x, U_{i-1}) = y)
\end{aligned}$$

vu l'indépendance de U_{i-1} et X_{i-1}.

Ceci démontre que la suite X_n est une chaîne de Markov en appliquant la proposition 1.2. L'homogénéité est claire puisque les variables aléatoires U_n ont toutes même loi.

Exemple 4 : marche aléatoire

Soit $\xi_1, \xi_2, \ldots, \xi_m, \ldots$ une suite de v.a. à valeurs dans $\mathbb{Z}$, indépendantes et de même loi ν. Soit X_0 une v.a. à valeurs dans $\mathbb{Z}$ indépendante des précédentes et de loi μ. Posons :

$$X_n = X_0 + \xi_1 + \xi_2 + \cdots + \xi_n,$$

donc, pour $n \geq 1$:

$$X_n = X_{n-1} + \xi_n.$$

D'après l'exemple précédent, la suite $(X_n)_{n\in\mathbb{N}})$ est une chaîne de Markov homogène de fonction de transition :

$$P(i,j) = \nu(j-i) \quad (i,j) \in Z^2.$$

On appelle cette chaîne une marche aléatoire sur $\mathbb{Z}$.

En particulier si $\nu(1) = a$, $\nu(-1) = b$, $\nu(0) = c$ avec $a+b+c = 1$, on peut se représenter la marche aléatoire précédente comme la position d'une particule qui se "promène sur $\mathbb{Z}$" : lorsque la particule est au site i à l'instant n, elle avance (passe au site $i+1$ à l'instant $n+1$) avec probabilité a, recule (passe au site $i-1$ à l'instant $n+1$) avec probabilité b et ne bouge pas (reste au site i à l'instant $n+1$) avec probabilité c. On parle alors de marche aléatoire simple.

Exemple 5 : chaîne de naissance et mort

Cet exemple est une généralisation de la marche aléatoire simple. Il s'agit d'une chaîne à valeurs dans $E = \{0, 1, 2, ...d\}$ ou dans $E = \mathbb{N}$. Si la chaîne est dans l'état x à un certain instant, elle ne peut, à l'instant suivant qu'être dans les états $x-1$, x ou $x+1$ (à condition que celui-ci appartienne à E). Sa fonction de transition s'écrit

$$P(x,y) = \begin{cases} q_x & \text{si } y = x-1, \\ r_x & \text{si } y = x, \\ p_x & \text{si } y = x+1, \\ 0 & \text{sinon,} \end{cases}$$

avec p_x, q_x et r_x positifs et $p_x + q_x + r_x = 1$ (évidemment $q_0 = 0$ et $p_d = 0$ si $E = \{0, 1, 2, ...d\}$).

La terminologie "naissance et mort" provient du cas où cette chaîne représente l'évolution d'une population (une transition de x à $x+1$ correspond à une naissance, de x à $x-1$ à une mort) mais celle-ci peut servir à modéliser bien d'autres contextes. Lorsque p_x, q_x et r_x ne dépendent pas de x, on a une marche aléatoire simple.

1.3 Propriété de Markov faible

On peut généraliser la propriété de Markov de la définition. C'est l'objet des propositions suivantes.

Proposition 1.5 *Soit $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov à valeurs dans E. Pour tout n, pour tous x, y dans E et pour tout sous ensemble A de E^n, on a :*

$$\begin{aligned}&\mathbb{P}(X_{n+1} = y \,/\, (X_0, X_1, \ldots, X_{n-1}) \in A, X_n = x)\\ &\qquad = \mathbb{P}(X_{n+1} = y \,/\, X_n = x).\end{aligned}$$

Démonstration :

$$\begin{aligned}
&\mathbb{P}(X_{n+1} = y, (X_0, X_1, \ldots, X_{n-1}) \in A, X_n = x)\\
&= \sum_{(\alpha_0,\ldots,\alpha_{n-1})\in A} \mathbb{P}(X_{n+1} = y, X_0 = \alpha_0, X_1 = \alpha_1, \ldots X_{n-1} = \alpha_{n-1}, X_n = x)\\
&= \sum_{(\alpha_0,\ldots,\alpha_{n-1})\in A} \mathbb{P}(X_0 = \alpha_0, X_1 = \alpha_1, \ldots X_{n-1} = \alpha_{n-1}, X_n = x)\\
&\quad \times\mathbb{P}(X_{n+1} = y \,/\, X_0 = \alpha_0, X_1 = \alpha_1, \ldots X_{n-1} = \alpha_{n-1}, X_n = x)\\
&= \sum_{(\alpha_0,\ldots,\alpha_{n-1})\in A} \mathbb{P}(X_0 = \alpha_0, X_1 = \alpha_1, \ldots X_{n-1} = \alpha_{n-1}, X_n = x)\\
&\quad \times\mathbb{P}(X_{n+1} = y \,/\, X_n = x)\\
&= \mathbb{P}(X_0, X_1, \ldots, X_{n-1}) \in A, X_n = x)\\
&\quad \times\mathbb{P}(X_{n+1} = y \,/\, X_n = x)
\end{aligned}$$

■

Proposition 1.6 *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov à valeurs dans E. Pour tous n et k, pour tous $y_1, \ldots, y_k$ éléments de E, pour tout $A \subset E^n$, pour tout $B \subset E^k$ et pour toute fonction f de E^k dans $\mathbb{R}$ positive, on a :*

$$\begin{aligned}
&\mathbb{P}(X_{n+1} = y_1, \ldots, X_{n+k} = y_k \,/\, (X_0, \ldots, X_{n-1}) \in A, X_n = x)\\
&= \mathbb{P}(X_{n+1} = y_1, \ldots, X_{n+k} = y_k \,/\, X_n = x)\\
&= \mathbb{P}(X_{n+1} = y_1 \,/\, X_n = x)\, \mathbb{P}(X_{n+2} = y_2 \,/\, X_{n+1} = y_1) \ldots\\
&\quad \ldots \mathbb{P}(X_{n+k} = y_k \,/\, X_{n+k-1} = y_{k-1})
\end{aligned}$$

et

$$\begin{aligned}&\mathbb{P}((X_{n+1}, \ldots, X_{n+k}) \in B \,/\, (X_0, \ldots, X_{n-1}) \in A, X_n = x)\\ &\qquad = \mathbb{P}((X_{n+1}, \ldots, X_{n+k}) \in B \,/\, X_n = x)\end{aligned}$$

et

$$\begin{aligned}&\mathbb{E}(f(X_{n+1}, \ldots, X_{n+k}) \,/\, (X_0, \ldots, X_{n-1}) \in A, X_n = x)\\ &\qquad = \mathbb{E}(f(X_{n+1}, \ldots, X_{n+k}) \,/\, X_n = x)\end{aligned}$$

Démonstration : Notons C l'événement $\{(X_0, \ldots, X_{n-1}) \in A\}$. En utilisant la formule de Bayes et la proposition précédente on obtient :

$$\begin{aligned}
&\mathbb{P}(X_{n+1} = y_1, \ldots, X_{n+k} = y_k \,/\, C, X_n = x) \\
&\quad = \mathbb{P}(X_{n+k} = y_k \,/\, C, X_n = x, X_{n+1} = y_1, \ldots, X_{n+k-1} = y_{k-1}) \\
&\qquad \times \mathbb{P}(X_{n+k-1} = y_{k-1} \,/\, C, X_n = x, X_{n+1} = y_1, \ldots, X_{n+k-2} = y_{k-2}) \\
&\qquad \times \ldots \\
&\qquad \times \mathbb{P}(X_{n+1} = y_1 \,/\, C, X_n = x) \\
&\quad = \mathbb{P}(X_{n+k} = y_k \,/\, X_{n+k-1} = y_{k-1}) \\
&\qquad \times \mathbb{P}(X_{n+k-1} = y_{k-1} \,/\, X_{n+k-2} = y_{k-2}) \\
&\qquad \times \ldots \\
&\qquad \times \mathbb{P}(X_{n+1} = y_1 \,/\, X_n = x)
\end{aligned}$$

Le résultat est indépendant de l'événement C. Ceci démontre la première partie de la proposition.

Pour la deuxième partie, il suffit d'écrire :

$$\begin{aligned}
&\mathbb{P}((X_{n+1}, \ldots, X_{n+k}) \in B \,/\, (X_0, \ldots, X_{n-1}) \in A, X_n = x) \\
&\quad = \sum_{(y_1, \ldots, y_k) \in B} \mathbb{P}(X_{n+1} = y_1, \ldots, X_{n+k} = y_k \,/\, (X_0, \ldots, X_{n-1}) \in A, X_n = x)
\end{aligned}$$

Quant à la troisième partie, elle s'établit par combinaison linéaire de fonctions indicatrices, puis par passage à la limite croissante pour une suite de fonctions étagées approximant la fonction f. ■

Lorsque la chaîne de Markov est homogène, la proposition précédente prend la forme suivante.

Proposition 1.7 *Soient $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov homogène à valeurs dans E et $P(x, y)$ sa fonction de transition. Pour tous n et k, pour tous x, $y_1, \ldots, y_k$ éléments de E, pour tout $A \subset E^n$, pour tout $B \subset E^k$ et pour toute fonction f de E^k dans $\mathbb{R}$ positive,, on a :*

$$\begin{aligned}
\mathbb{P} \quad (\quad & X_{n+1} = y_1, \ldots, X_{n+k} = y_k \,/\, (X_0, \ldots, X_{n-1}) \in A, X_n = x) \\
= \quad & \mathbb{P}(X_{n+1} = y_1, \ldots, X_{n+k} = y_k \,/\, X_n = x) \\
= \quad & P(x, y_1)\, P(y_1, y_2) \ldots P(y_{k-1}, y_k) \\
= \quad & \mathbb{P}(X_1 = y_1, \ldots, X_k = y_k \,/\, X_0 = x)
\end{aligned}$$

et

$$\mathbb{P}((X_{n+1}, \ldots, X_{n+k}) \in B \,/\, (X_0, \ldots, X_{n-1}) \in A, X_n = x)$$

$$= \mathbb{P}((X_1, \ldots, X_k) \in B \,/\, X_0 = x)$$

et

$$\mathbb{E}(f(X_{n+1}, \ldots, X_{n+k}) \,/\, (X_0, \ldots, X_{n-1}) \in A, X_n = x)$$
$$= \mathbb{E}(f(X_1, \ldots, X_k) \,/\, X_0 = x)$$

On peut réécrire la formule précédente sous la forme :

$$\mathbb{P}((X_0, \ldots, X_{n-1}) \in A, X_n = x, (X_{n+1}, \ldots, X_{n+k}) \in B) =$$

$$\mathbb{P}((X_0, \ldots, X_{n-1}) \in A, X_n = x)\, \mathbb{P}((X_1, \ldots, X_k) \in B \,/\, X_0 = x)$$

ou sous la forme

$$\mathbb{E}\left(1_{\{(X_0,\ldots,X_{n-1}) \in A, X_n = x\}}\, f(X_{n+1}, \ldots, X_{n+k})\right) =$$

$$\mathbb{P}((X_0, \ldots, X_{n-1}) \in A, X_n = x)\, \mathbb{E}(f(X_1, \ldots, X_k) \,/\, X_0 = x)$$

Par combinaison linéaire et par passage à la limite croissante, on obtient alors la proposition suivante.

Proposition 1.8 *Soient $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène à valeurs dans E. Pour tous n et k, pour tout x dans E, pour toute fonction g de E^n dans $\mathbb{R}_+$ et pour toute fonction f de E^k dans $\mathbb{R}_+$, on a :*

$$\mathbb{E}\left(g(X_0, \ldots, X_{n-1})\, 1_{\{X_n = x\}}\, f(X_{n+1}, \ldots, X_{n+k})\right) =$$

$$\mathbb{E}\left(g(X_0, \ldots, X_{n-1})\, 1_{\{X_n = x)\}}\right)\, \mathbb{E}\left(f(X_1, \ldots, X_k) \,/\, X_0 = x\right)$$

Comme exemple d'application des résultats précédents, nous pouvons commencer à décrire le comportement d'une chaîne de Markov homogène.

Proposition 1.9 *Soient $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène à valeurs dans E et $P(x, y)$ sa fonction de transition. Pour tout x, nous avons :*

$$\mathbb{P}(X_{n+1} = x, \ldots, X_{n+k} = x, X_{n+k+1} \neq x \,/\, X_n = x)$$

$$= P(x, x)^k (1 - P(x, x)),$$

et pour tout $y \neq x$:

$$\mathbb{P}(X_{n+1} = y \,/\, X_n = x, X_{n+1} \neq x) = \frac{P(x, y)}{1 - P(x, x)}\,.$$

Démonstration : D'après la proposition 1.7

$$\begin{aligned}
&\mathbb{P}(X_{n+1} = x, ..., X_{n+k} = x, X_{n+k+1} \neq x \,/\, X_n = x) \\
&= \sum_{y \in E : y \neq x} \mathbb{P}(X_{n+1} = x, ..., X_{n+k} = x, X_{n+k+1} = y \,/\, X_n = x) \\
&= \sum_{y \in E : y \neq x} P(x,x)^k P(x,y) \\
&= P(x,x)^k \,(1 - P(x,x)).
\end{aligned}$$

D'autre part, pour $y \neq x$, on a :

$$\begin{aligned}
\mathbb{P}(X_{n+1} = y \,/\, X_n = x, X_{n+1} \neq x) &= \frac{\mathbb{P}(X_n = x, X_{n+1} = y)}{\mathbb{P}(X_n = x, X_{n+1} \neq x)} \\
&= \frac{\mathbb{P}(X_{n+1} = y \,/\, X_n = x)}{\mathbb{P}(X_{n+1} \neq x) \,/\, X_n = x)} \\
&= \frac{P(x,y)}{1 - P(x,x)}\,.
\end{aligned}$$

■

Cette proposition signifie que si la chaîne est à un instant dans un état donné, le temps nécessaire pour changer d'état suit une loi géométrique.

On va voir que l'on peut donner une formulation plus générale des propositions 1.6 et 1.7 grâce à des propriétés des mesures.

Les ensembles de la forme $\{(X_{n+1}, \ldots X_{n+k}) \in B\}$ avec $B \subset E^k$ décrivent des événements postérieurs à l'instant n. Mais tous les événements postérieurs à l'instant n ne sont pas de cette forme et ne sont pas réunion dénombrable d'événements de cette forme. Nous allons considérer l'ensemble $E^{\mathbb{N}}$ des suites à valeurs dans E. Cet ensemble n'est pas dénombrable. On le munit de la tribu $\mathcal{B}$ qui est engendrée par les parties de $E^{\mathbb{N}}$ de la forme $A_0 \times A_1 \times \ldots \times A_k \times E \times E \times \ldots$ avec k entier quelconque, et $A_0, \ldots, A_k$ parties de E. Un événement postérieur à n est alors un événement de la forme :

$$(X_{n+1}, X_{n+2}, \ldots) \in B$$

pour B élément de la tribu $\mathcal{B}$.

Théorème 1.10 (propriété de Markov) *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène à valeurs dans E. Pour tout n, pour tout x dans E, pour tout $A \subset E^n$, pour tout $B \in \mathcal{B}$ et pour tout fonction mesurable f de $(E^{\mathbb{N}}, \mathcal{B})$ dans $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$, on a :*

$$\mathbb{P}((X_{n+1}, X_{n+2}, \ldots) \in B \,/\, (X_0, \ldots, X_{n-1}) \in A, X_n = x)$$

$$= \mathbb{P}((X_1, X_2, \ldots) \in B \,/\, X_0 = x)$$

et

$$\mathbb{E}(f(X_{n+1}, X_{n+2}, \ldots) \,/\, (X_0, \ldots, X_{n-1}) \in A, X_n = x)$$
$$= \mathbb{E}(f(X_1, X_2, \ldots) \,/\, X_0 = x)$$

Démonstration : Pour n, $x \in E$ et $A \subset E^{\mathbb{N}}$ fixés, définissons les mesures m_1 et m_2 sur $(E^{\mathbb{N}}, \mathcal{B})$ par :

$$m_1(B) = \mathbb{P}((X_{n+1}, X_{n+2}, \ldots) \in B \,/\, (X_0, \ldots, X_{n-1}) \in A, X_n = x)$$

$$m_2(B) = \mathbb{P}((X_1, X_2, \ldots) \in B \,/\, X_0 = x).$$

D'après la proposition 1.7, les mesures finies m_1 et m_2 sont égales sur les ensembles B de la forme $B_1 \times \ldots \times B_k \times E \times E \times \ldots$ ($k \in \mathbb{N}, B_1, \ldots, B_k$ sous-ensembles de E). L'ensemble formé par de tels B contient $E^{\mathbb{N}}$ et est stable par intersection finie. Par conséquent, le théorème d'unicité des mesures entraine que $m_1 = m_2$, ce qui est le premier résultat cherché.

Le deuxième résultat s'obtient par combinaison linéaire et par passage à la limite croissante. ■

Toujours par combinaison linéaire et passage à la limite croissante, on obtient le corollaire suivant.

Corollaire 1.11 *Soient $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène à valeurs dans E. Pour tout n, pour tout x dans E, pour toute fonction g de E^n dans $\mathbb{R}_+$ et pour toute fonction mesurable f de $(E^{\mathbb{N}}, \mathcal{B})$ dans $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$, on a :*

$$\mathbb{E}\left(g(X_0, \ldots, X_{n-1})\, 1_{\{X_n=x\}}\, f(X_{n+1}, X_{n+2}, \ldots)\right) =$$

$$\mathbb{E}\left(g(X_0, \ldots, X_{n-1})\, 1_{\{X_n=x)\}}\right)\, \mathbb{E}\left(f(X_1, X_2, \ldots) \,/\, X_0 = x\right)$$

Remarque : Comme la probabilité conditionnelle sachant que $X_0 = x$ apparait souvent dans les formules, on utilise la notation $\mathbb{P}_x$ pour cette probabilité conditionnelle. De même on utilise la notation $\mathbb{E}_x$ pour l'espérance conditionnelle sachant que $X_0 = x$. Les formules du théorème 1.10 deviennent alors avec ces notations :

$$\mathbb{P}((X_{n+1}, X_{n+2}, \ldots) \in B \,/\, (X_0, \ldots, X_{n-1}) \in A, X_n = x)$$

$$= \mathbb{P}_x((X_1, X_2, \ldots) \in B)$$

et

$$\mathbb{E}(f(X_{n+1}, X_{n+2}, \ldots) \,/\, (X_0, \ldots, X_{n-1}) \in A, X_n = x)$$

$$= \mathbb{E}_x(f(X_1, X_2, \ldots))$$

Nous donnons maintenant un exemple typique d'application de la propriété de Markov. Il s'agit de calculer l'espérance du temps d'atteinte par la chaîne de Markov d'un sous-ensemble de E. Par exemple si la chaîne décrit l'évolution dans le temps de l'état d'un système et si une partie des états sont des états de panne, il s'agit de calculer le temps moyen de panne du système.

Proposition 1.12 *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène à valeurs dans E de fonction de transition $P(x, y)$. Soit A un sous-ensemble de E. On pose*

$$S_A = \min\{k \geq 0 : X_k \in A\},$$

avec la convention $\min \emptyset = +\infty$. *On suppose que pour tout $y \notin A$, $\mathbb{P}_y(S_A < +\infty) = 1$. Posons, pour $x \notin A$:*

$$h(x) = \mathbb{E}_x(S_A) \in \overline{\mathbb{R}}_+.$$

Alors la fonction h vérifie :

$$\forall\, x \notin A,\ \ h(x) = 1 + \sum_{y\notin A} P(x, y)\, h(y).$$

Démonstration : Nous supposons que $x \notin A$. Alors on a $S_A \geq 1$. On peut écrire :

$$h(x) = \mathbb{E}_x(S_A) = \sum_{m\geq 1} m\, \mathbb{P}_x(S_A = m) = \sum_{m\geq 1} m \sum_{y\in E} \mathbb{P}_x(S_A = m, X_1 = y)$$

Soit $m \geq 2$. Si $y \in A$ on a $\mathbb{P}(X_0 = x, X_1 = y, S_A = m) = 0$. Si $y \notin A$ en appliquant la propriété de Markov à l'instant 1, on obtient :

$$\begin{aligned}
&\mathbb{P}(X_0 = x, X_1 = y, S_A = m)\\
&\quad = \ \mathbb{P}(X_0 = x, X_1 = y, X_2 \notin A, \ldots, X_{m-1} \notin A, X_m \in A)\\
&\quad = \ \mathbb{P}(X_0 = x, X_1 = y)\, \mathbb{P}_y(X_1 \notin A, \ldots, X_{m-2} \notin A, X_{m-1} \in A)\\
&\quad = \ \mathbb{P}(X_0 = x, X_1 = y)\, \mathbb{P}_y(S_A = m - 1)
\end{aligned}$$

On obtient alors :

$$\begin{aligned}
h(x) &= \sum_{y\in E} \mathbb{P}_x(S_A = 1, X_1 = y) + \sum_{m\geq 2} m \sum_{y\in E} \mathbb{P}_x(S_A = m, X_1 = y)\\
&= \sum_{y\in A} \mathbb{P}_x(X_1 = y) + \sum_{m\geq 2} m \sum_{y\notin A} P(x, y) P_y(S_A = m - 1)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{y\in A} \mathbb{P}_x(X_1 = y) + \sum_{y\notin A} P(x,y) \sum_{m\geq 2} m\, \mathbb{P}_y(S_A = m-1) \\
&= \sum_{y\in A} P(x,y) + \sum_{y\notin A} P(x,y) \sum_{m\geq 2} \mathbb{P}_y(S_A = m-1) + \sum_{y\notin A} P(x,y)h(y) \\
&= 1 + \sum_{y\notin A} P(x,y)h(y)
\end{aligned}$$

■

Supposons que A^c (complémentaire de A) est fini. Pour ne pas alourdir les notations, nous écrivons A^c sous la forme $A^c = \{1, 2, \ldots, d\}$. La fonction h sur A^c s'identifie alors à un vecteur colonne u ($u_i = h(i)$) de dimension d. Notons P_1 la matrice P restreinte à A, c'est une matrice $d \times d$. La proposition 1.12 exprime que u est solution du système linéaire

$$(I - P_1)u = 1_d$$

où 1_d est le vecteur colonne de dimension d dont toutes les composantes sont égales à 1.

Ce système est un système de Cramer. En effet, pour montrer que $I - P_1$ est inversible, il suffit de montrer que $lim_{n\to+\infty} P_1^n = 0$ (voir [5] théorèmes 1.5.1 et 1.4.5). Or, pour $1 \leq i, j \leq d$:

$$\begin{aligned}
0 &\leq P_1^n(i,j) \\
&= \sum_{i_1=1}^{d} \sum_{i_2=1}^{d} \cdots \sum_{i_{n-1}=1}^{d} P(i,i_1)P(i_2,i_2)\ldots P(i_{n-1},j) \\
&= \sum_{i_1=1}^{d} \sum_{i_2=1}^{d} \cdots \sum_{i_{n-1}=1}^{d} \mathbb{P}_i(X_1 = i_1, X_2 = i_2, \ldots, X_{n-1} = i_{n-1}, X_n = j) \\
&= \mathbb{P}_i(X_1 \in A^c, X_2 \in A^c, \ldots, X_{n-1} \in A^c, X_n = j) \\
&\leq \mathbb{P}_i(X_1 \in A^c, X_2 \in A^c, \ldots, X_{n-1} \in A^c, X_n \in A^c) \\
&= \mathbb{P}_i(S_A > n).
\end{aligned}$$

Comme $\lim_{n\to+\infty} \mathbb{P}_i(S_A > n) = \mathbb{P}_i(S_A = +\infty) = 0$, on en déduit le résultat.

Dans le cas où A^c est fini, on peut donner une condition nécessaire et suffisante pour que pour tout $x \notin A$, $\mathbb{P}_x(S_A < +\infty) = 1$. C'est l'objet du lemme suivant.

Lemme 1.13 *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène à valeurs dans E de fonction de transition $P(x,y)$. Soit A un sous-ensemble de E. On suppose que A^c est fini.*

Alors $\mathbb{P}_x(S_A < +\infty) = 1$ si et seulement si pour tout $x \notin A$, il existe $n \geq 1$, $x_1, \ldots, x_{n-1} \notin A$ et $z \in A$ tels que

$$P(x, x_1)\, P(x_1, x_2) \ldots P(x_{n-2}, x_{n-1})\, P(x_{n-1}, z) > 0.$$

Démonstration : Supposons $\mathbb{P}_x(S_A < +\infty) = 1$ pour $x \notin A$. On a $\{S_A < +\infty\} = \cup_{n\geq 1}\{S_A = n\}$. Donc il existe $n \geq 1$ tel que

$$\begin{aligned}
0 &< \mathbb{P}_x(S_A = n) \\
&= \mathbb{P}_x(X_1 \notin A, \ldots X_{n-1} \notin A, X_n \in A) \\
&= \sum_{x_1 \notin A, \ldots x_{n-1} \notin A, z \in A} \mathbb{P}_x(X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = z) \\
&= \sum_{x_1 \notin A, \ldots x_{n-1} \notin A, z \in A} P(x, x_1) \ldots P(x_{n-2}, x_{n-1}) P(x_{n-1}, z),
\end{aligned}$$

donc au moins un des termes de la somme est > 0.

Réciproquement, supposons que pour tout $x \notin A$, il existe un chemin allant de x à A. Soit $x \notin A$. On a :

$$\begin{aligned}
\mathbb{P}(X_0 = x, S_A = +\infty) &= \sum_{y_1 \notin A} \mathbb{P}(X_0 = x, X_1 = y_1, X_2 \notin A, X_3 \notin A, \ldots) \\
&= \sum_{y_1 \notin A} \mathbb{P}(X_0 = x, X_1 = y_1)\, \mathbb{P}_{y_1}(X_1 \notin A, X_2 \notin A, \ldots) \\
&= \sum_{y_1 \notin A} \mathbb{P}(X_0 = x, X_1 = y_1)\, \mathbb{P}_{y_1}(S_A = +\infty)
\end{aligned}$$

d'où

$$\mathbb{P}_x(S_A = +\infty) = \sum_{y_1 \notin A} P(x, y_1)\, \mathbb{P}_{y_1}(S_A = +\infty) \tag{1.1}$$

Montrons par récurrence sur k, que pour tout $k \geq 1$

$$\mathbb{P}_x(S_0^A = +\infty) = \sum_{y_1 \notin A} \ldots \sum_{y_k \notin A} P(x, y_1) P(y_1, y_2) \ldots P(y_{k-1}, y_k) \mathbb{P}_{y_k}(S_0^A = +\infty). \tag{1.2}$$

On suppose que

$$\mathbb{P}_x(S_0^A = +\infty) = \sum_{y_1 \notin A} \ldots \sum_{y_{k-1} \notin A} P(x, y_1) P(y_1, y_2) \ldots P(y_{k-2}, y_{k-1}) \mathbb{P}_{y_{k-1}}(S_0^A = +\infty)$$

On applique la formule (1.1) avec $x = y_{k-1}$ et on obtient le résultat.

D'autre part, pour tout $x \notin A$, il existe $n = n_x \geq 1$ tel que

$$\sum_{y_1 \notin A} \sum_{y_2 \notin A} \dots \sum_{y_n \notin A} P(x, y_1) P(y_1, y_2) \dots P(y_{n-1}, y_n) < 1.$$

En effet par hypothèse il existe $n \geq 1$, $x_1, \dots, x_{n-1} \notin A$ et $z \in A$ tels que

$$P(x, x_1)\, P(x_1, x_2) \dots P(x_{n-2}, x_{n-1})\, P(x_{n-1}, z) > 0.$$

Or :

$$\begin{aligned}
1 &= \sum_{y_n \in E} P^n(x, y_n) \\
&= \sum_{y_1 \in E} \dots \sum_{y_{n-1} \in E} \sum_{y_n \in E} P(x, y_1) \dots P(y_{n-1}, y_n) \\
&\geq \sum_{y_1 \notin A} \dots \sum_{y_{n-1} \notin A} \sum_{y_n \in E} P(x, y_1) \dots P(y_{n-1}, y_n) \\
&\geq P(x, x_1) \dots P(x_{n-1}, z) + \sum_{y_1 \notin A} \dots \sum_{y_{n-1} \notin A} \sum_{y_n \notin A} P(x, y_1) \dots P(y_{n-1}, y_n).
\end{aligned}$$

Comme $P(x, x_1)\, P(x_1, x_2) \dots P(x_{n-2}, x_{n-1})\, P(x_{n-1}, z) > 0$, on a

$$\sum_{y_1 \notin A} \dots \sum_{y_{n-1} \notin A} \sum_{y_n \notin A} P(x, y_1) \dots P(y_{n-1}, y_n) < 1.$$

Puisque A^c est fini

$$\begin{aligned}
C &= \sup_{x \notin A} \sum_{y_1 \notin A} \dots \sum_{y_{n-1} \notin A} \sum_{y_n \notin A} P(x, y_1) \dots P(y_{n-1}, y_n) \\
&< 1.
\end{aligned}$$

En appliquant la formule (1.2) avec $k = n_x$, on obtient, pour $x \notin A$:

$$\begin{aligned}
&\mathbb{P}_x(S_A = +\infty) \\
&\quad \leq \sup_{y \notin A} \mathbb{P}_y(S_A = +\infty) \sum_{y_1 \notin A} \dots \sum_{y_{n_x} \notin A} P(x, y_1) P(y_1, y_2) \dots P(y_{n_x - 1}, y_{n_x}) \\
&\quad \leq C \sup_{y \notin A} \mathbb{P}_y(S_A = +\infty),
\end{aligned}$$

donc

$$\ell = \sup_{x \notin A} \mathbb{P}_x(S_A = +\infty) \leq C\, \ell,$$

avec $C < 1$, donc $\ell = 0$. $\blacksquare$

Ce résultat n'est plus vrai si A^c n'est pas fini. On peut donner un contre-exemple. Considérons $E = \mathbb{N} \cup \{a\}$, pour $i \in \mathbb{N}$, $P(i, a) = \epsilon_i$, $P(i, i+1) = 1 - \epsilon_i$, $P(a, a) = 1$. Alors :

$$\begin{aligned}
\mathbb{P}_x(S_{\{a\}} = +\infty) &= \lim_{n \to +\infty} \mathbb{P}_x(S_{\{a\}} > n) \\
&= \mathbb{P}_x(X_1 = x+1, X_2 = x+2, \ldots, X_n = x+n) \\
&= \prod_{i=x}^{x+n-1} (1 - \epsilon_i).
\end{aligned}$$

Si $\sum_i \epsilon_i < +\infty$, alors $\lim_{n \to +\infty} \prod_{i=1}^{n} (1 - \epsilon_i) > 0$.

1.4 Propriété de Markov forte

Une formulation encore plus générale de la propriété de Markov permet de remplacer l'instant n par un instant aléatoire à condition de l'instant aléatoire vérifie une propriété dite de temps d'arrêt. Cette formulation est ce qu'on appelle la propriété de Markov forte.

Pour définir un temp d'arrêt, nous avons besoin d'utiliser une suite $(\mathcal{B}_n)_{n\in\mathbb{N}}$ de tribus de l'espace de probabilité Ω sur lequel sont définies les variables aléatoires constituant la chaîne de Markov. On notera $\mathcal{B}_n$ la tribu engendrée par $X_0, X_1, \ldots, X_n$, c'est-à-dire la plus petite tribu rendant mesurable les variables aléatoires $X_0, X_1, \ldots, X_n$. Un élément de $\mathcal{B}_n$ s'écrit sour la forme $\{(X_0, \ldots, X_n) \in A\}$ pour une partie A de E^{n+1}. Cette suite de tribus est croissante, c'est-à-dire que $\mathcal{B}_n \subset \mathcal{B}_{n+1}$. Une suite croissante de tribus est appelée une filtration. On dit que la filtration $(\mathcal{B}_n)_{n\in\mathbb{N}}$ est la filtration naturelle associée à la chaîne $(X_n)_{n\in\mathbb{N}}$. Alors un temps d'arrêt pour la chaîne de Markov est défini de la manière suivante.

Définition 1.14 *Une variable aléatoire ν à valeurs dans $\mathbb{N} \cup \{+\infty\}$ est un temps d'arrêt pour la chaine de Markov $(X_n)_{n\in\mathbb{N}}$ (ou relativement à la filtration $(\mathcal{B}_n)_{n\in\mathbb{N}}$) si :*

$$\forall\, n \in \mathbb{N}, \quad \{\nu = n\} \in \mathcal{B}_n.$$

Exemple : soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov à valeurs dans E, A un sous-ensemble de E et S_A le premier temps d'atteinte de A : $S_A = \min\{k \geq 0 : X_k \in A\}$. On a :

$$\{S_A = n\} = \{X_0 \notin A, X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A\} \quad \in \quad \mathcal{B}_n$$

donc S_A est un temps d'arrêt.

Définition 1.15 *Si ν est un temps d'arrêt, on définit la tribu $\mathcal{B}_\nu$ des événements antérieurs à ν par :*

$$A \in \mathcal{B}_\nu \Longleftrightarrow A \cap \{\nu = n\} \in \mathcal{B}_n \ \forall n \in \mathbb{N}.$$

Si ν est un temps d'arrêt, alors X_ν est $\mathcal{B}_\nu$-mesurable. En effet, pour tout $A \subset E$, et $n \in \mathbb{N}$:

$$\{X_\nu \in A\} \cap \{\nu = n\} = \{X_n \in A\} \cap \{\nu = n\} \in \mathcal{B}_n.$$

L'événement $\{\nu < \infty\}$ appartient à $\mathcal{B}_\nu$.

Si ν_1 et ν_2 sont deux temps d'arrêt tels que $\nu_1 \leq \nu_2$, alors $\mathcal{B}_{\nu_1} \subset \mathcal{B}_{\nu_2}$.

Le théorème 1.10 se généralise de la manière suivante.

Théorème 1.16 (propriété de Markov forte) *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène à valeurs dans E et ν un temps d'arrêt relativement à celle-ci.*

Pour tout $x \in E$, si $A \in \mathcal{B}_\nu$ et si $B \in \mathcal{B}$, on a :

$$\mathbb{P}((X_{\nu+1}, X_{\nu+2}, \dots) \in B \ / \ A, X_\nu = x) = \mathbb{P}_x((X_1, X_2, \dots) \in B).$$

Pour tout $x \in E$, si $A \in \mathcal{B}_\nu$ et si f est une fonction mesurable de $(E^{\mathbb{N}}, \mathcal{B})$ dans $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$, on a :

$$\mathbb{E}(f(X_{\nu+1}, X_{\nu+2}, \dots) \ / \ A, X_\nu = x) = \mathbb{E}_x(f(X_1, X_2, \dots)).$$

Démonstration : Commençons par la première formule. Puisque les événements $\{\nu = n\}$ et $A \cap \{\nu = n\}$ appartiennent à $\mathcal{B}_n$, nous obtenons, en appliquant le théroème 1.10

$$\begin{aligned}
&\mathbb{P}(A, \{X_\nu = x\}, (X_{\nu+1}, X_{\nu+2}, \dots) \in B) \\
&= \sum_{n\geq 0} \mathbb{P}(A, \{X_\nu = x\}, (X_{\nu+1}, X_{\nu+2}, \dots) \in B, \nu = n) \\
&= \sum_{n\geq 0} \mathbb{P}(A, \nu = n\}, \{X_n = x\}, (X_{n+1}, X_{n+2}, \dots) \in B) \\
&= \sum_{n\geq 0} \mathbb{P}(A \cap \{\nu = n\} \cap \{X_n = x\})\, \mathbb{P}_x((X_1, X_2, \dots) \in B) \\
&= \sum_{n\geq 0} \mathbb{P}(A \cap \{\nu = n\} \cap \{X_\nu = x\})\, \mathbb{P}_x((X_1, X_2, \dots) \in B) \\
&= \mathbb{P}(A \cap \{X_\nu = x\})\, \mathbb{P}_x((X_1, X_2, \dots) \in B).
\end{aligned}$$

La deuxième formule s'obtient à partir de la première par combinaison linéaire et passage à la limite croissante. ■

Le corrollaire suivant qui est une simple réécriture du théorème 1.16.

Corollaire 1.17 *Soit* $(X_n)_{n\in\mathbb{N}}$ *une chaîne de Markov homogène à valeurs dans* E *et* ν *un temps d'arrêt relativement à celle-ci.*

Pour tout $x \in E$, $A \in \mathcal{B}_\nu$ *et* f *est fonction mesurable de* $(E^{\mathbb{N}}, \mathcal{B})$ *dans* $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$, *on a :*

$$\mathbb{E}\left(1_{\{A, X_\nu = x\}} f(X_{\nu+1}, X_{\nu+2}, \ldots)\right) = \mathbb{P}(A, X_\nu = x)\, \mathbb{E}_x(f(X_1, X_2, \ldots)).$$

Pour illustrer l'utilisation de la propriété de Markov forte, nous allons étudier les temps successifs d'atteinte par la chaîne d'un sous-ensemble A de E. Ces temps d'atteinte sont définis de la manière suivante. Le premier temps d'entrée dans A est noté $S_A^{(0)}$ et est défini par :

$$S_A^{(0)} = S_A = \min\{n \geq 0 : X_n \in A\}$$

Les instants successifs $S_A^{(p)}$, $p \geq 1$, de retour dans A sont définis par

$$S_A^{(p)} = \min\{n > S_A^{(p-1)}, X_{n-1} \notin A, X_n \in A\}$$

(avec la convention habituelle $\min \emptyset = +\infty$).

Lorsque $A = \{x\}$ pour un x dans E, on notera $S_x^{(p)}$ au lieu de $S_{\{x\}}^{(p)}$.

Commençons par établir que ces temps sont des temps d'arrêt.

Lemme 1.18 *Soient* A *un sous-ensemble de* E, $S_A^{(0)}$ *le premier temps d'entrée dans* A *et* $S_A^{(p)}$, $p \geq 1$, *les instants successifs de retour dans* A. *Ces temps aléatoires sont des temps d'arrêt, c'est-à-dire pour tous* p *et* m *dans* $\mathbb{N}$, *on a :*

$$\{S_A^{(p)} = m\} \in \mathcal{B}_m.$$

Démonstration : La démonstration se fait par récurrence sur p.

Nous avons déjà démontré dans l'exemple ci-dessus que $S_A^{(0)}$ est un temps d'arrêt.

Supposons maintenant que $\{S_A^{(p)} = k\} \in \mathcal{B}_k$ pour tout $k \in \mathbb{N}$ et regardons $\{S_A^{(p+1)} = m\}$. Nous avons :

$$\begin{aligned}\{S_A^{(p+1)} = m\} \quad = \quad & \bigcup_{\substack{k\geq 0, \ell \geq 1: \\ k+\ell \leq m-1}} \{S_A^{(p)} = k, X_{k+1} \in A, \ldots, X_{k+\ell-1} \in A, X_{k+\ell} \notin A, \\ & X_{k+\ell+1} \notin A, \ldots, X_{m-1} \notin A, X_m \in A\}. \end{aligned} \tag{1.3}$$

Par hypothèse de récurrence, $\{S_A^{(p)} = k\} \in \mathcal{B}_k$ et donc pour $k \leq m$, $\{S_A^{(p)} = k\} \in \mathcal{B}_m$. Il est clair que tous les autres événements intervenant dans le membre de droite de (1.3) appartiennent à $\mathcal{B}_m$, d'où le résultat. ∎

Soit x dans E. Nous allons nous intéresser à la durée du $p^{ième}$ temps de séjour en x que nous noterons D_p^x : c'est l'entier m qui vérifie $X_{S_x^{(p)}} = x, \ldots, X_{S_x^{(p)}+m-1} = x, X_{S_x^{(p)}+m} \neq x$. La variable aléatoire $X_{S_x^{(p)}+D_p^x}$ est le lieu où la chaîne saute en sortant de x après son $p^{ème}$ retour en x.

Proposition 1.19 *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène à valeurs dans E de fonction de transition $P(x,y)$. Pour tout $x \in E$, la durée du temps de séjour de la chaîne dans x est de loi géométrique sur $\mathbb{N}^*$ de paramètre $P(x,x)$ et il y a indépendance entre le temps de séjour en x et l'endroit où la chaîne saute lorsqu'elle quitte x.*

Plus précisément, avec les notations ci-dessus, pour tout $p \geq 0$, on a :

$$\mathbb{P}(D_p^x = k+1 \,/\, S_x^{(p)} < +\infty) = P(x,x)^k(1-P(x,x)),$$

et pour $y \neq x$ et $P(x,x) \neq 1$:

$$\mathbb{P}(X_{S_x^{(p)}+D_p^x} = y \,/\, S_x^{(p)} < +\infty) = \frac{P(x,y)}{1-P(x,x)},$$

$$\begin{aligned}
&\mathbb{P}(D_p^x = k+1, X_{S_x^{(p)}+D_p^x} = y \,/\, S_x^{(p)} < +\infty) \\
&\quad = \mathbb{P}(D_p^x = k+1 \,/\, S_x^{(p)} < +\infty)\,\mathbb{P}(X_{S_x^{(p)}+D_p^x} = y \,/\, S_x^{(p)} < +\infty).
\end{aligned}$$

Démonstration : Pour $y \neq x$, le théorème 1.16 et le lemme 1.18 entrainent :

$$\begin{aligned}
&\mathbb{P}(S_x^{(p)} < \infty, D_p^x = k+1, X_{S_x^{(p)}+D_p^x} = y) \\
&= \mathbb{P}(S_x^{(p)} < \infty, X_{S_x^{(p)}} = x, X_{S_x^{(p)}+1} = x, \cdots, X_{S_x^{(p)}+k} = x, X_{S_x^{(p)}+k+1} = y) \\
&= \mathbb{P}(S_x^{(p)} < \infty, X_{S_x^{(p)}} = x)\,\mathbb{P}_x(X_1 = x, \ldots, X_{k-1} = x, X_k = y) \\
&= \mathbb{P}(S_x^{(p)} < +\infty)\,P(x,x)^k\,P(x,y).
\end{aligned}$$

Par conséquent :

$$\mathbb{P}(D_p^x = k+1, X_{S_x^{(p)}+D_p^x} = y \,/\, S_x^{(p)} < \infty) = P(x,x)^k\,P(x,y).$$

On en déduit :

$$\begin{aligned}
&\mathbb{P}(D_p^x = k+1 \,/\, S_x^{(p)} < \infty) \\
&= \sum_{\substack{y\in E,\\ y\neq x}} \mathbb{P}(D_p^x = k+1, X_{S_x^{(p)}+D_p^x} = y \,/\, S_x^{(p)} < \infty) \\
&= P(x,x)^k \sum_{\substack{y\in E,\\ y\neq x}} P(x,y) \\
&= P(x,x)^k(1-P(x,x)),
\end{aligned}$$

$$
\begin{aligned}
&\mathbb{P}(X_{S_x^{(p)}+D_p^x} = y \,/\, S_x^{(p)} < \infty) \\
&= \sum_{k\geq 0} \mathbb{P}(D_p^x = k+1, X_{S_x^{(p)}+D_p^x} = y \,/\, S_x^{(p)} < \infty) \\
&= \sum_{k\geq 0} P(x,x)^k \, P(x,y) \\
&= \frac{P(x,y)}{1-P(x,x)}.
\end{aligned}
$$

Enfin :

$$
\begin{aligned}
&\mathbb{P}(D_p^x = k+1, X_{S_x^{(p)}+D_p^x} = y \,/\, S_x^{(p)} < \infty) \\
&= P(x,x)^k(1-P(x,x))\,\frac{P(x,y)}{1-P(x,x)} \\
&= \mathbb{P}(D_p^x = k+1 \,/\, S_x^{(p)} < \infty)\,\mathbb{P}(X_{S_x^{(p)}+D_p^x} = y \,/\, S_x^{(p)} < \infty).
\end{aligned}
$$

■

Plus généralement, on peut montrer que le comportement d'une chaîne de Markov homogène peut être caractérisé de la façon suivante : à l'instant initial, la "chaîne" choisit un état x selon la loi initiale et décide d'y séjourner pendant une durée de loi géométrique (sur $\mathbb{N}^*$) de paramètre $P(x,x)$. Lorsque cette durée est écoulée, la chaîne choisit un nouvel état y suivant la probabilité $P(x,y)/(1-P(x,x))$ et en "arrivant" en y, elle choisit la durée de son séjour suivant la loi géométrique de paramètre $P(y,y)$,... On pourrait penser que ceci fournit une bonne façon de simuler une chaîne de Markov. En fait la simulation d'une loi géométrique se faisant à partir de la simulation de variables aléatoires de Bernouilli, on ne gagne rien par rapport à la méthode de simulation naturelle décrite après la proposition 1.2.

Si A est un sous ensemble de E, nous allons nous intéresser à la "sous-chaîne" constituée des valeurs successives de la chaîne lors des instants de retour successifs dans A.

Proposition 1.20 *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène à valeurs dans E et A un sous-ensemble de E. On suppose que pour tout $x \in E$ et pour tout $n \geq 0$, les variables aléatoires $S_A^{(n)}$ sont finies presque-sûrement. La suite $(X_{S_A^{(n)}})_{n\in\mathbb{N}}$ est une chaîne de Markov homogène de loi initiale $\mu(x) = \mathbb{P}(X_{S_A^{(0)}} = x)$ et de fonction de transition Q donnée par :*

$$\forall x, y \in A, \quad Q(x,y) = \mathbb{P}_x(X_{S_A^{(1)}} = y).$$

Démonstration : Montrons que les $Y_n = X_{S_A^{(n)}}$ forment une chaîne de Markov. Soit $k \in \mathbb{N}$ et $y_0, \ldots y_k \in A$. Nous avons :

$$\begin{aligned}
&\mathbb{P}(Y_0 = y_0, \ldots, Y_k = y_k) \\
&= \mathbb{P}(X_{S_A^{(0)}} = y_0, \ldots X_{S_A^{(k-1)}} = y_{k-1}, X_{S_A^{(k)}} = y_k)
\end{aligned}$$

On veut appliquer la propriété de Markov au temps d'arrêt $S_A^{(k-1)}$. Pour cela il faut vérifier que

$$\{X_{S_A^{(0)}} = y_0, \ldots X_{S_A^{(k-1)}} = y_{k-1}\} \in \mathcal{B}_{S_A^{(k-1)}},$$

soit que

$$\{X_{S_A^{(0)}} = y_0, \ldots X_{S_A^{(k-1)}} = y_{k-1}, S_A^{(k-1)} = n\} \in \mathcal{B}_n,$$

ce qui est bien intuitif mais nécessite une démonstration.

On a :

$$\begin{aligned}
&\{X_{S_A^{(0)}} = y_0, \ldots X_{S_A^{(k-)1}} = y_{k-1}, S_A^{(k-1)} = n\} \\
&= \bigcup_{\substack{m_0,\ldots,m_{k-2}, \\ m_0<\ldots<m_{k-2}<n}} \{X_{S_A^{(0)}} = y_0, S_A^{(0)} = m_0, \ldots, X_{S_A^{(k-2)}} = y_{k-2}, S_A^{(k-2)} = m_{k-2}, \\
&\qquad X_{S_A^{(k-1)}} = y_{k-1}, S_A^{(k-1)} = n\} \\
&= \bigcup_{\substack{m_0,\ldots,m_{k-2}, \\ m_0<\ldots<m_{k-2}<n}} \{X_{m_0} = y_0, S_A^{(0)} = m_0, \ldots, X_{m_{k-2}} = y_{k-2}, S_A^{(k-2)} = m_{k-2}, \\
&\qquad X_n = y_{k-1}, S_A^{(k-1)} = n\}
\end{aligned}$$

D'après le lemme 1.18, pour tout $p \leq k-2$, $\{S_A^{(p)} = m_p\} \in \mathcal{B}_{m_p} \subset \mathcal{B}_{m_{k-2}} \subset \mathcal{B}_n$, et $\{S_A^{(k-1)} = n\} \in \mathcal{B}_n$. Par conséquent

$$\{X_{S_A^{(0)}} = y_0, \ldots X_{S_A^{(k-1)}} = y_{k-1}, S_A^{(k-1)} = n\} \in \mathcal{B}_n.$$

On applique la propriété de Markov forte au temps d'arrêt $S_A^{(k-1)}$. On obtient :

$$\begin{aligned}
&\mathbb{P}(Y_0 = y_0, \ldots, Y_k = y_k) \\
&= \sum_{\ell\geq 1} \sum_{p\geq 1} \mathbb{P}(X_{S_A^{(0)}} = y_0, \ldots X_{S_A^{(k-1)}} = y_{k-1}, X_{S_A^{(k-1)}+1} \in A, \ldots, \\
&\qquad X_{S_A^{(k-1)}+\ell-1} \in A, X_{S_A^{(k-1)}+\ell} \notin A, \ldots, X_{S_A^{(k-1)}+\ell+p-1} \notin A, X_{S_A^{(k-1)}+\ell+p} \in A) \\
&= \sum_{\ell\geq 1} \sum_{p\geq 1} \mathbb{P}(X_{S_A^{(0)}} = y_0, \ldots, X_{S_A^{(k-1)}} = y_{k-1}) \, \mathbb{P}_{y_{k-1}}(X_1 \in A, \ldots, \\
&\qquad X_{\ell-1} \in A, X_\ell \notin A, \ldots, X_{\ell+p-1} \notin A, X_{\ell+p} \in A) \\
&= \mathbb{P}(X_{S_A^{(0)}} = y_0, \ldots X_{S_A^{(k-1)}} = y_{k-1}) \, \mathbb{P}_{y_{k-1}}(X_{S_A^{(1)}} = y_k)
\end{aligned}$$

En faisant une récurrence sur k, on obtient :

$$\begin{aligned}&\mathbb{P}(Y_0 = y_0, \ldots, Y_k = y_k)\\ &\quad = \quad \mathbb{P}(X_{S_A^{(0)}} = y_0)\,\mathbb{P}_{y_0}(X_{S_A^{(1)}} = y_1)\ldots\mathbb{P}_{y_{k-1}}(X_{S_A^{(1)}} = y_k),\end{aligned}$$

ce qui montre que $(Y_n)_{n\in\mathbb{N}}$ est une chaîne de Markov de loi initiale $\mu(y) = P(X_{S_A^{(0)}} = y)$ et de matrice de transition $Q(x,y) = \mathbb{P}_x(X_{S_A^{(1)}} = y)$ $\blacksquare$

Chapitre 2

Chaîne de Markov : classification des états

Dans tout ce chapitre, nous étudierons le comportement d'une chaîne de Markov homogène, à valeurs dans un espace E (fini ou dénombrable), de fonction de transition $P(x, y)$.

2.1 Communication entre les états

Noue allons définir un certain nombre de notions qui décrivent la manière dont les états "communiquent" entre eux, c'est-à-dire si, à partir d'un état initial donné, la chaîne peut atteindre un autre état donné. Un première notion est celle d'état absorbant.

Définition 2.1 *Un état x est* **absorbant** *si $P(x, x) = 1$.*

Si x est absorbant, alors pour tout $n \geq 1$, $\mathbb{P}_x(X_1 = x, X_2 = x, \ldots, X_n = x) = P(x, x)^n = 1$, donc $\mathbb{P}_x(\forall\, n, X_n = x) = 1$, d'où la terminologie. Cet état ne communique avec aucun autre état.

Pour décrire cette notion de communication, on définit le temps d'atteinte T_y d'un état y :

$$T_y = \min\{n \geq 1 : X_n = y\}$$

Il faut noter que, lorsque l'état initial vaut y, la variable aléatoire T_y est différente de la variable alátoire $S_y = \min\{n \geq 0 : X_n = y\}$ rencontrée dans le chapitre précédent.

On montre que T_y est un temps d'arrêt pour la chaîne de Markov de la même manière que S_y est un temps d'arrêt.

Définition 2.2 *Nous dirons que x conduit à y si $\mathbb{P}_x(T_y < +\infty) > 0$.*

L'état x conduit à l'état y si et seulement si il existe $n \geq 1$ vérifiant $\mathbb{P}_x(X_n = y) > 0$. En effet, pour tout $k \geq 1$:

$$\mathbb{P}_x(X_k = y) \leq \mathbb{P}_x(T_y < +\infty) = \mathbb{P}_x(\cup_{n\geq 1}\{X_n = y\}) \leq \sum_{n\geq 1} \mathbb{P}_x(X_n = y).$$

et donc

$$\mathbb{P}_x(T_y < +\infty) > 0 \Longleftrightarrow \exists\, n \geq 1,\ \mathbb{P}_x(X_n = y) > 0.$$

Si l'on construit le graphe d'états, cela correspond à l'existence d'un chemin conduisant de x à y car

$$\mathbb{P}_x(X_n = y) = P^n(x, y)$$

$$= \sum_{y_1\in E} \dots \sum_{y_{n-1}\in E} P(x, y_1)\, P(y_1, y_2) \dots P(y_{n-2}, y_{n-1})\, P(y_{n-1}, y),$$

donc $\mathbb{P}_x(X_n = y) > 0$ équivant à l'existence de $x_1 \in E, \dots, x_{n-1} \in E$ tel que $P(x, x_1)\, P(x_1, x_2), \dots P(x_{n-2}, x_{n-1})\, P(x_{n-1}, y) > 0$ donc tel que $P(x, x_1) > 0$, $P(x_1, x_2) > 0, \dots P(x_{n-2}, x_{n-1}) > 0$, $P(x_{n-1}, y) > 0$.

Cela entraine que la relation "conduit" est transitive, c'est-à-dire que si x conduit à y et si y conduit à z, alors x conduit à z.

Une première notion est celle de classe fermée.

Définition 2.3 *Un sous-ensemble C de E est appelé une classe fermée si aucun état appartenant à C ne conduit à un état hors de C, c'est-à-dire si :*

$$\forall\, x \in C,\ \forall\, y \notin C,\ \mathbb{P}_x(T_y < +\infty) = 0$$

Proposition 2.4 *Soit $C \subset E$. Les conditions suivantes sont équivalentes :*

1. *C est une classse fermée,*
2. *Pour tout $x \in C$, la condition $\mathbb{P}_x(T_y < +\infty) > 0$ entraine $y \in C$,*
3. $$\forall\, x \in C,\ \forall\, y \notin C,\ P(x, y) = 0,$$
4. $$\forall\, x \in C,\ \mathbb{P}_x(\forall n \geq 1, X_n \in C) = 1.$$

Démonstration : Les conditions 1. et 2. sont équivalents, car C fermée équivant à : pour tout $x \in C$, on a $y \notin C \Rightarrow \mathbb{P}_x(T_y < +\infty) = 0$, ou encore : pour tout $x \in C$, on a $\mathbb{P}_x(T_y < +\infty) > 0 \Rightarrow y \in C$.

Voyons maintenant les autres équivalences.

Supposons la condition 1. vérifiée. Alors, pour $x \in C$ et $y \notin C$:

$$P(x,y) = \mathbb{P}_x(X_1 = y) \leq \mathbb{P}_x(T_y < +\infty) = 0,$$

donc $P(x,y) = 0$. Par conséquent, la condition 1. entraine la condition 3.

Supposons maintenant la condition 3. vérifiée. Montrons par récurrence sur n que pour tout $n \geq 1$,

$$\forall\, x \in C, \forall\, y \notin C,\ P^n(x,y) = 0.$$

Pour $n = 1$, c'est la condition 3. Supposons la propriété vraie à l'orde n, alors pour $x \in C$ et $y \notin C$:

$$P^{n+1}(x,y) = \sum_{z \in E} P^n(x,z)P(z,y) = \sum_{z \in C} P^n(x,z)P(z,y) = 0.$$

La récurrence est établie, donc sous la condition 3. pour tout n, si $x \in C$ et $y \notin C$, alors $P^n(x,y) = 0$; par conséquent $\mathbb{P}_x(T_y < +\infty) = 0$ et donc la condition 1. est vérifiée.

Regardons la condition 4. Elle équivant à :

$$\forall\, x \in C,\ \mathbb{P}_x(\exists n \geq 1, X_n \notin C) = 0.$$

Or :

$$\begin{aligned}
\mathbb{P}_x(\exists n \geq 1, X_n \notin C) &= \mathbb{P}_x(\exists n \geq 1,\ \exists y \notin C,\ X_n = y) \\
&= \mathbb{P}_x(\cup_{y \notin C}\{\exists n \geq 1,\ X_n = y\}) \\
&= \mathbb{P}_x(\cup_{y \notin C}\{T_y < +\infty\}).
\end{aligned}$$

Par conséquent $\mathbb{P}_x(\exists n \geq 1, X_n \notin C) = 0$ équivant à : pour tout $y \notin C$, $\mathbb{P}_x(T_y < +\infty) = 0$. ■

Une deuxième notion est celle de classe irréductible.

Définition 2.5 *Un sous-ensemble C de E est irréductible si $\forall\, x \in C$, $\forall\, y \in C$, x conduit à y (en langage des graphes, un sous-ensemble irréductible est une classe fortement connexe).*

La proposition suivante laisse entrevoir une décomposition en classes fermées irréductibles.

Proposition 2.6 *Soit C_1 et C_2 deux classes fermées irréductibles. Alors ou bien $C_1 \cap C_2 = \emptyset$, ou bien $C_1 = C_2$.*

Démonstration : Soit C_1 et C_2 deux classes fermées irréductibles. Supposons que $C_1 \cap C_2 \neq \emptyset$. Soit $x \in C_1 \cap C_2$. Soit $y \in C_1$. Puisque x et y sont dans C_1 et que C_1 est irréductible, x conduit à y. Puisque C_2 est fermée et que x conduit à y, on a donc $y \in C_2$ (condition 2. de la proposition 2.4). On a donc prouvé que $C_1 \subset C_2$. L'inclusion inverse se démontre de la même manière. ■

2.2 Etats récurrents et transients

Nous allons étudier les notions fondamentales pour les chaînes de Markov de récurrence et de transience.

Définition 2.7 *Un état y est dit* **récurrent** *si $\mathbb{P}_y(T_y < +\infty) = 1$, il est dit* **transient** *ou* **transitoire** *si $\mathbb{P}_y(T_y < +\infty) < 1$.*

Un premier exemple d'état récurrent est un état absorbant, mais nous allons évidemment voir qu'il existe d'autres états récurrents.

Notons

$$N_y = \sum_{n \geq 1} 1_{\{X_n = y\}}$$

le nombre de visites à l'état y (à partir de l'instant 1).

Remarquons que $\{T_y < +\infty\} = \{N_y \geq 1\}$.

Posons :

$$\rho_{xy} = \mathbb{P}_x(T_y < +\infty).$$

Proposition 2.8 *Pour tous x et y appartenant à E et $m \geq 1$:*

- $\mathbb{P}_x(N_y \geq m) = \rho_{xy}\, \rho_{yy}^{m-1}$,
- $\mathbb{P}_x(N_y = m) = \rho_{xy}\, \rho_{yy}^{m-1}(1 - \rho_{yy})$.

Démonstration : Pour $m \geq 1$, l'événement $\{N_y \geq m\}$ entraine $\{N_y \geq 1\}$ et donc $\{T_y < +\infty\}$.

En appliquant la propriété de Markov forte à l'instant T_y nous avons :

$$\begin{aligned}
\mathbb{P}_x(N_y \geq m) &= \mathbb{P}_x(T_y < +\infty, X_{T_y} = y, 1 + \sum_{n \geq 1} 1_{\{X_{T_y+n} = y\}} \geq m) \\
&= \mathbb{P}_x(T_y < +\infty, X_{T_y} = y)\, \mathbb{P}_y(N_y \geq m - 1) \\
&= \mathbb{P}_x(T_y < +\infty)\, \mathbb{P}_y(N_y \geq m - 1) \\
&= \rho_{xy}\, \mathbb{P}_y(N_y \geq m - 1).
\end{aligned}$$

En particulier, en prenant $x = y$ on obtient :

$$\mathbb{P}_y(N_y \geq m) = \rho_{yy}\,\mathbb{P}_y(T_y \geq m-1).$$

Par suite :

$$\mathbb{P}_y(N_y \geq m) = \rho_{yy}^{m-1}\,\mathbb{P}_y(N_y \geq 1) = \rho_{yy}^{m}.$$

Donc :

$$\mathbb{P}_x(N_y \geq m) = \rho_{xy}\,\mathbb{P}_y(N_y \geq m-1) = \rho_{xy}\,\rho_{yy}^{m-1}.$$

La deuxième formule de la proposition s'obtient en écrivant :

$$\mathbb{P}_x(N_y = m) = \mathbb{P}_x(N_y \geq m) - \mathbb{P}_x(N_y \geq m-1).$$

■

La dernière formule de la proposition est intuitive : une chaîne partant de x visite y exactement m fois si elle parvient à y puis y retourne $m-1$ fois et ensuite n'y revient plus.

Théorème 2.9

1. *Si y est transient, alors :*

$$\mathbb{P}_x(N_y < +\infty) = 1, \quad \mathbb{E}_x(N_y) = \frac{\rho_{xy}}{1-\rho_{yy}}.$$

2. *Si y est récurrent, alors :*

$$\mathbb{P}_y(N_y = +\infty) = 1, \quad \mathbb{P}_x(N_y = +\infty) = \rho_{xy} = \mathbb{P}_x(T_y < +\infty).$$

Démonstration : Reprenons la formule $\mathbb{P}_x(N_y \geq m) = \rho_{xy}\,\rho_{yy}^{m-1}$. Nous avons :

$$\begin{aligned}
\mathbb{P}_x(N_y = \infty) &= \lim_{m\to+\infty} \mathbb{P}_x(N_y \geq m) \\
&= \lim_{m\to+\infty} \rho_{xy}\,\rho_{yy}^{m-1} \\
&= \begin{cases} 0 & \text{si } \rho_{yy} < 1, \\ \rho_{xy} & \text{si } \rho_{yy} = 1, \end{cases}
\end{aligned}$$

d'où le résultat annoncé lorsque y est récurrent.

Dans le cas où y est transient, c'est-à-dire $\rho_{yy} < 1$, nous obtenons :

$$\begin{aligned}
\mathbb{E}_x(N_y) &= \sum_{m\geq 1} m\,\mathbb{P}_x(N_y = m) \\
&= \sum_{m\geq 1} m\,\rho_{xy}\,\rho_{yy}^{m-1}(1-\rho_{yy}) \\
&= \frac{\rho_{xy}(1-\rho_{yy})}{(1-\rho_{yy})^2} \\
&= \frac{\rho_{xy}}{1-\rho_{yy}}.
\end{aligned}$$

■

Si l'état y est transient, alors pour tout $x \in E$, $\mathbb{E}_x(N_y) < +\infty$ et donc $N_y < +\infty$ $\mathbb{P}_x$ preque sûrement. Par conséquent, quelle que soit la loi initiale $N_y < +\infty$, donc la chaîne ne passera qu'un nombre fini de fois par y. Si, au contraire, y est récurrent et si la chaîne part de y, elle y revient une infinité de fois ; tandis que si elle part d'un point $x \neq y$, elle peut ou non revenir en y mais si elle y revient au moins une fois, elle y revient alors une infinité de fois.

Corollaire 2.10

$$y \text{ récurrent } \iff N_y = +\infty \, \mathbb{P}_y \ p.s. \iff \mathbb{E}_y(N_y) = +\infty,$$

$$y \text{ transient } \iff N_y < +\infty \, \mathbb{P}_y \ p.s. \iff \mathbb{E}_y(N_y) < +\infty.$$

Définition 2.11 *Une chaîne dont tous les états sont transients est appelée "chaîne transiente", une chaîne dont tous les états sont récurrents est appelée "chaîne récurrente".*

La notions de récurrence et de transience se transmettent lorsque les états communiquent.

Proposition 2.12

- *Si x conduit à y et si x est un état récurrent alors y est également récurrent et :*

$$\mathbb{P}_x(T_y < +\infty) = \mathbb{P}_y(T_x < +\infty) = 1.$$

- *Si x conduit à y et si y est un état transient, alors x est aussi transient.*
- *Si x conduit à y et si y conduit à x, alors x et y sont de même nature, c'est-à-dire ou bien tous deux transients ou bien tous deux récurrents.*

Démonstration : Supposons que x soit un état récurrent et que x conduise à y.

Commençons par montrer que $\rho_{yx} = 1$. Intuitivement, comme x conduit à y la chaîne à une probabilité strictement positive d'aller de x à y sans repasser par x et si nous avions $\rho_{yx} < 1$, la chaîne partant de y aurait une probabilité $1 - \rho_{yx} > 0$ de ne jamais passer par x, par conséquent (d'après la propriété de Markov !) la chaîne partant de x aurait une probabilité strictement positive de ne jamais revenir en x, ce qui contredit le fait que x est récurrent.
Plus précisément, puisque $\rho_{xy} > 0$, nous avons vu qu'il existe $n \geq 1$ tel que $\mathbb{P}_x(X_n = y) > 0$. Soit n_0 le plus petit des $n \geq 1$ vérifiant $\mathbb{P}_x(X_n = y) > 0$. On

peut alors trouver des états $x_1, ..., x_{n_0-1}$ différents de x et de y pour lesquels $P(x, x_1)P(x_1, x_2).....P(x_{n_0-1}, y) > 0$. Par conséquent :

$$\begin{aligned}
\mathbb{P}_x(T_x = +\infty) &\geq \mathbb{P}_x(X_1 = x_1, X_2 = x_2, \ldots, X_{n_0-1} = x_{n_0-1}, X_{n_0} = y, \\
&\quad \forall\ k \geq 1, X_{n+k} \neq x) \\
&= \mathbb{P}_x(X_1 = x_1, X_2 = x_2, \ldots, X_{n_0-1} = x_{n_0-1}, X_{n_0} = y) \\
&\quad \mathbb{P}_y(\forall k \geq 1, X_k \neq x) \\
&= P(x, x_1)P(x_1, x_2).....P(x_{n_0-1}, y)(1 - \rho_{yx}).
\end{aligned}$$

L'état x étant récurrent, on a $\mathbb{P}_x(T_x = +\infty) = 0$. Comme $P(x, x_1)P(x_1, x_2) \ldots P(x_{n_0-1}, y) > 0$ et que $P(x, x_1)P(x_1, x_2).....P(x_{n_0-1}, y)(1 - \rho_{yx}) = 0$, on a nécessairement $1 - \rho_{yx} = 0$, c'est-à-dire $\rho_{yx} = 1$.

Montrons maintenant que y est récurrent. Puisque $\rho_{yx} = 1 > 0$, il existe $n_1 \geq 1$ pour lequel $\mathbb{P}_y(X_{n_1} = x) > 0$. D'autre part :

$$\begin{aligned}
\mathbb{P}_y(X_{n_1+n+n_0} = y) &\geq \mathbb{P}_y(X_{n_1} = x, X_{n_1+n} = x, X_{n_1+n+n_0} = y) \\
&= \mathbb{P}_y(X_{n_1} = x)\,\mathbb{P}_x(X_n = x)\,\mathbb{P}_x(X_{n_0} = y).
\end{aligned}$$

Par conséquent :

$$\begin{aligned}
\mathbb{E}_y(N_y) &= \sum_{p\geq 1} \mathbb{P}_y(X_p = y) \\
&\geq \sum_{p:p\geq n_1+1+n_0} \mathbb{P}_y(X_p = y) \\
&= \sum_{n\geq 1} \mathbb{P}_y(X_{n_1+n+n_0} = y) \\
&\geq \mathbb{P}_y(X_{n_1} = x)\,\mathbb{P}_x(X_{n_0} = y) \sum_{n\geq 1} \mathbb{P}_x(X_n = x) \\
&= \mathbb{P}_y(X_{n_1} = x)\,\mathbb{P}_x(X_{n_0} = y)\,\mathbb{E}_x(N_x).
\end{aligned}$$

Puisque x est récurrent, on a $\mathbb{E}_x(N_x) = +\infty$ et par suite $\mathbb{E}_y(N_y) = +\infty$ (car $\mathbb{P}_y(X_{n_1} = x) > 0$ et $\mathbb{P}_x(X_{n_0} = y) > 0$), ce qui prouve que y est récurrent.

Enfin puisque y est récurrent et que y conduit à x, le début de la démonstration montre que $\rho_{xy} = 1$.

Si y conduit à x et si x est transient alors y est également transient car si y était récurrent, d'après le résultat ci-dessus, x serait récurrent. ■

Si un sous-ensemble est irréductible alors, d'après la proposition 2.12, tous ses états sont de même nature : ou tous récurrents ou tous transients.

Du théorème 2.9 et de la proposition 2.12, on déduit immédiatement le résultat suivant

Proposition 2.13 *Soit C un sous-ensemble irréductible dont les états sont récurrents, alors pour tous x et y dans C, $\rho_{xy} = \mathbb{P}_x(T_y < +\infty) = 1$, $\mathbb{P}_x(N_y = +\infty) = 1$.*

Théorème 2.14 *Soit C une classe fermée irréductible. Si C est fini (c'est-à-dire comporte un nombre fini d'états) alors tous les états de C sont récurrents.*

Démonstration : Soit C est une classe fermée, pour tout $x \in C$, nous avons $\mathbb{P}_x(X_n \in C, \forall n) = 1$ (proposition 2.4). Par conséquent pour toute loi initiale μ portée par C, on a $\mathbb{P}_\mu$ presque- sûrement :

$$+\infty = \sum_{n\geq 1} 1_{\{X_n \in C\}} = \sum_{n\geq 1}\sum_{y\in C} 1_{\{X_n = y\}} = \sum_{y\in C}\sum_{n\geq 1} 1_{\{X_n = y\}} = \sum_{y\in C} N_y.$$

Si de plus C est supposée irréductible, tous ses états sont de même nature : tous récurrents ou tous transients. S'ils étaient tous transients, alors pour tout $y \in C$, on aurait $N_y < +\infty$ $\mathbb{P}_\mu$ presque-sûrement (pour toute loi initiale μ portée par C) et donc, C étant fini, $\sum_{y\in C} N_y < +\infty$ $\mathbb{P}_\mu$ presque-sûrement (pour toute loi initiale μ portée par C). D'où une contradiction. ■

Une classe fermée irréductible non finie n'est pas nécessairement formée d'états récurrents mais on a vu que tous ses états sont de même nature : ils sont soit tous récurrents soit tous transients. Par exemple une marche aléatoire simple sur $\mathbb{Z}$ pour laquelle a et b ne sont pas nuls est irréductible, elle est récurrente dans le cas $a = b$ et transiente si $a \neq b$.

Voyons dans le cas général, comment on peut regrouper les états récurrents.

Théorème 2.15 *Soit E_R l'ensemble des états récurrents. Si cet ensemble n'est pas vide, il est la réunion finie ou dénombrable de classes fermées irréductibles deux à deux disjointes appelées classes récurrentes irréductibles.*

Démonstration : Soit $x \in E_R$ et $C(x)$ l'ensemble des $y \in E_R$ tels que x conduise à y, c'est-à-dire :

$$C(x) = \{y \in E : \rho_{xy} > 0\} \ = \{y \in E_R : \rho_{xy} > 0\} \ = \{y \in E_R : \rho_{xy} = 1\}).$$

Puisque x est récurrent, nous avons par définition $\rho_{xx} = 1$ et par suite $x \in C(x)$.

Montrons que $C(x)$ est une classe fermée. Soit $y \in C(x)$ et supposons que y conduise à $z \in E$. Comme $y \in E_R$, la proposition 2.12 entraine que $z \in E_R$. Et comme x conduit à y et que y conduit à z, alors x conduit à z. Nous avons donc montré que si $y \in C(x)$ conduit à z, alors $z \in C(x)$, ce qui prouve bien que la classe $C(x)$ est fermée (condition 2. de la proposition 2.4).

Montrons que $C(x)$ est irréductible. Soit y et z appartenant à $C(x)$. Puisque x est récurrent et conduit à y, d'après la proposition 2.12, on a $\rho_{xy} = \rho_{yx} = 1$. Donc y conduit à x. Puisque y conduit à x et que x conduit à z, alors y conduit à z. Donc $C(x)$ est bien irréductible.

Enfin le lemme 2.6 entraine que deux classes ainsi construites sont soit disjointes soit égales. ■

Lorsque E est fini et pas trop grand, il est possible de lire sur le graphe d'états de la chaîne une partie des propriétés évoquées ci-dessus. Décomposons le graphe des états en composantes fortement connexes maximales. Rappelons qu'un (sous-)graphe est fortement connexe si, pour tout couple de sommets x et y, il existe un chemin conduisant de x à y. Il existe deux types de composantes fortement connexes maximales :

– celles desquelles ne sort aucun arc : ce sont des classes fermées (lorsqu'on est entré dans une telle classe, on ne peut en sortir), on les appelle "classes finales",
– celles desquelles sort au moins un arc.

Soit C une composante fortement connexe maximale d'où ne sort aucun arc. C'est donc une classe irréductible (car fortement connexe) et fermée (puisque qu'il ne sort aucun arc). Comme elle est finie, tous ses états sont récurrents. Donc, compte-tenu de la démonstration du théorème 2.15, C est une classe récurrente irréductible.

Si maintenant C est une composante fortement connexe maximale d'où sort au moins un arc. Il existe $y \in C$ et $z \notin C$ tels qu'il y ait un chemin de y à z, donc une probabilité strictement positive pour la chaîne partant de y d'aller en z sans repasser par y. Lorsque la chaîne est en $z \notin C$, elle a une probabilité égale à 1 de ne jamais revenir dans C (sinon il y aurait un chemin de z à C et donc, comme C est fortement connexe, de z à y et C ne serait pas une classe fortement connexe maximale). Donc la chaîne partant de y a une probabilité strictement positive de ne jamais revenir en y, donc y est un état transitoire. Donc une composante fortement connexe maximale d'où sort au moins un arc est formée d'états transitoires.

Pour "lire le comportement" de la chaîne sur le graphe d'états, on construit le graphe réduit qui fait apparaitre la décomposition en composantes fortement connexes maximales. On repère alors facilement les classes finales et les classes transitoires.

Exemples :

Reprenons l'exemple 1 du premier chapitre. Le graphe est le suivant :

Il y a deux classes finales donc des classes récurrentes irréductibles : la classe (1) et la classe (2,4). Les classes (5,6) et (3) sont des classes transitoires (elles sont formées d'états transitoires).

Dans l'exemple 2 d'une chaîne à deux états avec $a < 1$ et $b < 1$, le graphe possède une seule composante connexe et donc une seule classe récurrente irréductible.

2.3 Probabilités d'absorption et temps d'atteinte

Si A est un sous-ensemble de E, on note T_A le premier instant d'entrée dans A après l'instant 0 :

$$T_A = \min\{n \geq 1 : X_n \in A).$$

Remarquons que si $X_0 \notin A$, $T_A = S_A = \min\{n \geq 0 : X_n \in A\}$, tandis que si $X_0 \in A$, T_A est différent de S_A.

On démontre que T_A est un temps d'arrêt pour la chaîne de Markov de la même manière que S_A est un temps d'arrêt.

Soit C un sous-ensemble de E. Posons $\rho_C(x) = \mathbb{P}_x(T_C < +\infty)$: $\rho_C(x)$ représente la probabilité que la chaîne issue de x passe au moins une fois par C après l'instant 0.

Lemme 2.16 *La fonction $f(x) = \mathbb{P}_x(T_C < +\infty) = \rho_C(x)$ est solution du système d'équations :*

$$f(x) = \sum_{y \in C} P(x, y) + \sum_{y \notin C} P(x, y) f(y).$$

Démonstration : Nous avons :

$$\begin{aligned}
\mathbb{P}_x(T_C < +\infty) &= \mathbb{P}_x(\exists\, n \geq 1, X_n \in C) \\
&= \mathbb{P}_x(X_1 \in C) + \mathbb{P}_x(X_1 \notin C, \exists\, n \geq 2, X_n \in C) \\
&= \sum_{y \in C} P(x, y) + \sum_{y \notin C} \mathbb{P}_x(X_1 = y, \exists\, n \geq 1, X_{n+1} \in C) \\
&= \sum_{y \in C} P(x, y) + \sum_{y \notin C} \mathbb{P}_x(X_1 = y)\, \mathbb{P}_y(\exists\, n \geq 1, X_n \in C) \\
&= \sum_{y \in C} P(x, y) + \sum_{y \notin C} P(x, y)\, \rho_C(y).
\end{aligned}$$

■

Supposons que C soit une classe fermée irréductible formée d'états récurrents. Si $x \in C$ alors $\rho_C(x) = 1$. Si x est récurrent et n'appartient pas à C alors $\rho_C(x) = 0$ car x appartient à une classe fermée différente de C. Il ne reste donc à calculer $\rho_C(x)$ que pour x transient. Dans ce cas, le système linéaire précédent se simplifie et admet une unique solution.

Proposition 2.17 *Supposons que C soit une classe fermée irréductible formée d'état récurrents et notons E_T l'ensemble des états transients. Nous supposons que E_T est un ensemble fini. Alors la fonction f définie sur E_T par $f(x) = \rho_C(x)$ est l'unique solution du système d'équations :*

$$\forall\, x \in E_T, \quad f(x) = \sum_{y \in C} P(x,y) + \sum_{y \in E_T} P(x,y) f(y).$$

Démonstration : Comme C est une classe fermée irréductible formée d'états récurrents, nous avons vu que si y est un état récurrent qui n'appartient pas à C alors $\rho_C(y) = 0$. Donc les équations du lemme 2.16 se réduisent à :

$$\forall\, x \in E_T, \quad f(x) = \sum_{y \in C} P(x,y) + \sum_{y \in E_T} P(x,y) f(y).$$

Montrons que la seule solution de ces équations est $f = \rho_C$. Supposons que f vérifie ces équations et itérons la formule. Nous obtenons pour $x \in E_T$:

$$\begin{aligned}
f(x) &= \sum_{y \in C} P(x,y) + \sum_{y \in E_T} P(x,y) f(y) \\
&= \sum_{y \in C} P(x,y) + \sum_{y \in E_T} P(x,y) \left(\sum_{z \in C} P(y,z) + \sum_{z \in E_T} P(y,z)\, f(z) \right) \\
&= \sum_{y \in C} P(x,y) + \sum_{y \in E_T} P(x,y) \sum_{z \in C} P(y,z) \\
&\quad + \sum_{y \in E_T} \sum_{z \in E_T} P(x,y)\, P(y,z)\, f(z) \\
&= \mathbb{P}_x(T_C = 1) + \mathbb{P}_x(T_C = 2) + \sum_{y_1 \in E_T} \sum_{y_2 \in E_T} P(x,y_1)\, P(y_1,y_2)\, f(y_2) \\
&= \mathbb{P}_x(T_C \leq 2) + \sum_{y_1 \in E_T} \sum_{y_2 \in E_T} P(x,y_1)\, P(y_1,y_2)\, f(y_2).
\end{aligned}$$

En raisonnant par récurrence sur n, on obtient :

$$f(x) = \mathbb{P}_x(T_C \leq n) + \sum_{y_1 \in E_T} \ldots \sum_{y_n \in E_T} P(x,y_1)\, P(y_1,y_2) \ldots P(y_{n-1},y_n)\, f(y_n). \tag{2.1}$$

Posons $M = \max_{z \in E_T} |f(z)|$. Nous avons :

$$\begin{aligned}
&| \sum_{y_1 \in E_T} \dots \sum_{y_n \in E_T} P(x, y_1)\, P(y_1, y_2) \dots P(y_{n-1}, y_n)\, f(y_n)| \\
\leq\ & M \sum_{y_1 \in E_T} \dots \sum_{y_n \in E_T} P(x, y_1)\, P(y_1, y_2) \dots P(y_{n-1}, y_n) \\
=\ & M\, \mathbb{P}_x(\forall k, 1 \leq k \leq n,\ X_k \in E_T).
\end{aligned}$$

Notons N_{E_T} le nombre de visites à E_T :

$$N_{E_T} = \sum_{y \in E_T} \sum_{k \geq 1} 1_{\{X_k = y\}} = \sum_{y \in E_T} N_y.$$

Nous avons :

$$\mathbb{P}_x(\forall k \leq n,\ X_k \in E_T) \leq \mathbb{P}_x(N_{E_T} \geq n).$$

Pour $y \in E_T$, $N_y < +\infty$ $\mathbb{P}_x$ presque-sûrement. Comme E_T est supposé fini, $N_{E_T} < +\infty$ $\mathbb{P}_x$ presque-sûrement et par conséquent :

$$\lim_{n \to +\infty} \mathbb{P}_x(N_{E_T} \geq n) = \mathbb{P}_x(N_{E_T} = +\infty) = 0,$$

et donc :

$$\lim_{n \to +\infty} \mathbb{P}_x(\forall k \leq n,\ X_k \in E_T) = 0.$$

En faisant tendre n vers l'infini dans (2.1), nous obtenons :

$$f(x) = \mathbb{P}_x(T_C < +\infty) = \rho_C(x).$$

■

Exemple : Considérons la chaîne de Markov à valeurs dans

$$E = \{0, 1, 2, 3, 4, 5\}$$

dont la matrice de transition est :

$$\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\
0 & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & 0 & \frac{1}{5} \\
0 & 0 & 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \\
0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\
0 & 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4}
\end{pmatrix}$$

L'état 0 est absorbant, la classe $\{3,4,5\}$ est récurrente, les états 1 et 2 sont transients. Prenons $C = \{0\}$, et calculons $\rho_0(1) = \alpha_1, \rho_0(2) = \alpha_2$. Cela conduit au système :

$$\begin{aligned}
\alpha_1 &= \frac{1}{4} + \frac{1}{2}\alpha_1 + \frac{1}{4}, \alpha_2 \\
\alpha_2 &= \frac{1}{5}\alpha_1 + \frac{2}{5}\alpha_2,
\end{aligned}$$

dont la solution est $\alpha_1 = 3/5$, $\alpha_2 = 1/5$.

On montre également que $\rho_{\{3,4,5\}}(1) = 2/5$, $\rho_{\{3,4,5\}}(2) = 4/5$.

Si E_T est infini ce système possède une infinité d'équations et d'inconnues. Il possède au moins une solution (puisque ρ_C en est une !) mais celle-ci n'est pas nécessairement unique. On peut montrer qu'alors ρ_C est la plus petite solution positive du système.

2.4 Proportion de temps passé dans un état

La quantité $\sum_{k=1}^n 1_{\{X_k=y\}}$ est le nombre de visites à l'état y entre les instants 1 et n, ou encore le temps passé en y jusqu'à l'instant n.

La proportion de temps passé dans l'état y entre les instants 1 et n est donc :

$$\frac{1}{n}\sum_{k=1}^n 1_{\{X_k=y\}}.$$

Nous avons :

$$\mathbb{E}_x(\sum_{k=1}^n 1_{\{X_k=y\}}) = \sum_{k=1}^n P^n(x,y).$$

Si y est un état transient, nous avons vu que pour tout $x \in E$,

$$\lim_{n\to+\infty}\sum_{k=1}^n 1_{\{X_k=y\}} = N_y < +\infty$$

$\mathbb{P}_x$-presque-sûrement et que $\lim_{n\to+\infty}\mathbb{E}_x(\sum_{k=1}^n 1_{\{X_k=y\}}) = \mathbb{E}_x(N_y) < +\infty$. Par suite :

$$\frac{1}{n}\sum_{k=1}^n 1_{\{X_k=y\}} \xrightarrow[n\to+\infty]{\mathbb{P}_x\ p.s.} 0, \quad \frac{1}{n}\mathbb{E}_x\left(\sum_{k=1}^n 1_{\{X_k=y\}}\right) \xrightarrow[n\to+\infty]{} 0. \tag{2.2}$$

Regardons maintenant le cas où y est récurrent. Commençons par établir un lemme très intuitif.

Soit $T_y^{(1)} = T_y = \min\{n : n \geq 1, X_n = y\}$ et pour $k \geq 2$:

$$T_y^{(k)} = \min\{n > T_y^{(k-1)}, X_n = y\}.$$

Les $T_y^{(k)}$ sont les instants successifs de passage par l'état y à partir de l'instant 1. Il ne faut pas les confondre avec les $S_y^{(k)}$ qui sont les instants successifs de *retour* en y. Ainsi $S_y^{(0)} = S_y = \min\{n : n \geq 0, X_n = y\}$.

Les variables aléatoires $T_y^{(k)}$ sont des temps d'arrêt pour la chaîne de Markov (à démontrer en exercice).

Lemme 2.18 *Soit y un état récurrent. Sachant $\{S_y < +\infty\}$, les v.a. $T_y^{(p+1)} - T_y^{(p)}$ $(p \geq 1)$ sont des variables aléatoires réelles indépendantes et de même loi. Leur loi commune est la loi de $T_y^{(1)}$ sous $\mathbb{P}_y$.*

Démonstration : Comme y est récurrent, partant de y on y revient presque-sûrement une infinité de fois. Donc sur $\{S_y < +\infty\}$, les v.a. $T_y^{(p)}$ sont finies presque-sûrement, ce qui montre que les variables aléatoires $T_y^{(p+1)} - T_y^{(p)}$ sont définies et à valeurs réelles. Soit $m_0, m_1, \ldots, m_k$ des entiers positifs (ou nuls). En appliquant la propriété de Markov à l'instant $T_y^{(k)}$ on obtient :

$$\begin{aligned}
&\mathbb{P}(S_y < +\infty, T_y^{(2)} - T_y^{(1)} = m_1, \ldots, T_y^{(k)} - T_y^{(k-1)} = m_{k-1}, T_y^{(k+1)} - T_y^{(k)} = m_k) \\
&= \mathbb{P}(S_y < +\infty, T_y^{(2)} - T_y^{(1)} = m_1, \ldots, T_y^{(k)} - T_y^{(k-1)} = m_{k-1}, X_{T_y^{(k)}} = y, \\
&\quad X_{T_y^{(k)}+1} \neq y, \ldots, X_{T_y^{(k)}+m_k-1} \neq y, X_{T_y^{(k)}+m_k} = y) \\
&= \mathbb{P}(S_y < +\infty, T_y^{(2)} - T_y^{(1)} = m_1, \ldots, T_y^{(k)} - T_y^{(k-1)} = m_{k-1}, X_{T_y^{(k)}} = y) \\
&\quad \mathbb{P}_y(T_y^{(1)} = m_k).
\end{aligned}$$

Par récurrence sur k, on en déduit :

$$\mathbb{P}(S_y < +\infty, T_y^{(2)} - T_y^{(1)} = m_1, \ldots, T_y^{(k)} - T_y^{(k-1)} = m_{k-1}, T_y^{(k+1)} - T_y^{(k)} = m_k)$$

$$= \mathbb{P}(S_y < +\infty)\mathbb{P}_y(T_y^{(1)} = m_1) \ldots \mathbb{P}_y(T_y^{(1)} = m_k),$$

c'est-à-dire :

$$\mathbb{P}(T_y^{(2)} - T_y^{(1)} = m_1, \ldots, T_y^{(k+1)} - T_y^{(k)} = m_k \,/\, S_y < +\infty)$$

$$= \mathbb{P}_y(T_y^{(1)} = m_1) \ldots \mathbb{P}_y(T_y^{(1)} = m_k),$$

ce qui prouve le lemme. ■

Théorème 2.19 *Soit y est un état récurrent. Alors :*

$$\frac{1}{n}\sum_{k=1}^{n} 1_{\{X_k=y\}} \xrightarrow[n\to+\infty]{} \frac{1_{\{S_y<+\infty\}}}{\mathbb{E}_y(T_y)} \tag{2.3}$$

presque-sûrement quelle que soit la loi initiale. De plus :

$$\frac{1}{n}\mathbb{E}_x\left(\sum_{k=1}^{n} 1_{\{X_k=y\}}\right) \xrightarrow[n\to+\infty]{} \frac{\rho_{xy}}{\mathbb{E}_y(T_y)}.$$

La formule (2.3) est assez intuitive, car entre deux visites à y, s'écoule en moyenne une durée de $\mathbb{E}_y(T_y)$. Donc la proportion de temps passée en y doit être à peu près $\mathbb{E}_y(T_y)$.

La relation (2.3) est un cas de lois des grands nombres pour des v.a. non indépendantes.

Démonstration : Si $S_y = +\infty$, alors $\sum_{k=1}^{n} 1_{\{X_k=y\}} = 0$, et le résultat (2.3) est évident.

Regardons maintenant ce qui se passe sur $\{S_y < +\infty\}$. Le lemme 2.18 et la loi forte des grands nombres entrainent que sur $\{S_y < +\infty\}$,

$$\frac{T_y^{(k)}}{k} \xrightarrow[k\to+\infty]{} \mathbb{E}_y(T_y^{(1)}) = \mathbb{E}_y(T_y)$$

presque-sûrement. Posons $N_n(y) = \sum_{k=1}^{n} 1_{\{X_k=y\}}$. Nous avons :

$$N_n(y) = k \Longleftrightarrow T_y^{(k)} \le n < T_y^{(k+1)},$$

d'où :

$$T_y^{(N_n(y))} \le n < T_y^{(N_n(y)+1)}.$$

Par conséquent :

$$\frac{T_y^{(N_n(y))}}{N_n(y)} \le \frac{n}{N_n(y)} < \frac{T_y^{(N_n(y)+1)}}{N_n(y)+1}\,\frac{N_n(y)+1}{N_n(y)}. \tag{2.4}$$

Comme y est récurrent, $N_n(y)$ converge presque-sûrement vers l'infini lorsque n tend vers l'infini. En utilisant la convergence presque-sûre de $T_y^{(k)}/k$ vers $\mathbb{E}_y(T_y)$ lorsque k tend vers l'infini, on en déduit que les membres de droite et de gauche de (2.4) tendent presque-sûrement vers $\mathbb{E}_y(T_y)$ lorsque n tend vers l'infini. D'où le résultat (2.3) sur $\{S_y < +\infty\}$.

Le résultat (2.3) et le théorème de convergence dominée entrainent que :

$$\frac{1}{n}\mathbb{E}_x\left(\sum_{k=1}^{n} 1_{\{X_k=y\}}\right) \xrightarrow[n\to+\infty]{} \frac{\mathbb{P}_x(S_y<+\infty)}{\mathbb{E}_y(T_y)}.$$

Si $x \neq y$, $\mathbb{P}_x(S_y < +\infty) = \mathbb{P}_x(T_y < +\infty) = \rho_{xy}$. Si $x = y$, alors $\mathbb{P}_x(S_y < +\infty) = \mathbb{P}_y(S_y < +\infty) = 1 = \rho_{yy}$ puisque, y étant récurrent, on a par définition $\rho_{yy} = 1$. Donc dans tous les cas $\mathbb{P}_x(S_y < +\infty) = \rho_{xy}$. ■

Les résultats ci-dessus conduisent à regarder "d'un oeil différent" les états récurrents selon que $\mathbb{E}_y(T_y)$ est fini ou non.

Définition 2.20 *Un état récurrent y est dit* **récurrent positif** *si $\mathbb{E}_y(T_y) < +\infty$ et* **récurrent nul** *si $\mathbb{E}_y(T_y) = +\infty$. Une chaîne de Markov récurrente est dite récurrente positive (resp. récurrente nulle) si tous ses états sont récurrents positifs (resp. récurrents nuls).*

Proposition 2.21 *Si x conduit à y et si x est récurrent positif, alors y est également récurrent positif. Une classe récurrente irréductible est donc formée d'états tous récurrents positifs ou tous récurrents nuls.*

Démonstration : Si x est récurrent et conduit à y, la proposition 2.12 entraine que $\rho_{xy} = \rho_{yx} = 1 > 0$ donc x conduit à y et y conduit à x. Par suite il existe n_0 et n_1 tels que :

$$P^{n_0}(x,y) > 0, \quad P^{n_1}(y,x) > 0.$$

Pour tout $m \geq 0$, nous avons :

$$\begin{aligned} P^{n_1+m+n_0}(y,y) &= \mathbb{P}_y(X_{n_1+m+n_0} = y) \\ &\geq \mathbb{P}_y(X_{n_1} = x, X_{n_1+m} = x, X_{n_1+m+n_0} = y) \\ &= P^{n_1}(y,x)\, P^m(x,x) P^{n_0}(x,y). \end{aligned}$$

On en déduit :

$$\begin{aligned} &\frac{1}{n}\sum_{m=1}^{n} \mathbb{E}_y(1_{\{X_{n_1+m+n_0}=y\}}) \\ &\quad\geq P^{n_1}(y,x)\, P^{n_0}(x,y)\, \frac{1}{n}\sum_{m=1}^{n} P^m(x,x) \\ &\quad\geq P^{n_1}(y,x)\, P^{n_0}(x,y)\, \frac{1}{n}\, \mathbb{E}_x(\sum_{m=1}^{n} 1_{\{X_m=x\}}). \end{aligned}$$

Or :

$$\begin{aligned}
\sum_{m=1}^{n} \mathbb{E}_y(1_{\{X_{n_1+m+n_0}=y\}} &= \sum_{k=n_0+n_1+1}^{n_0+n_1+n} \mathbb{E}_y(1_{\{X_k=y\}}) \\
&= \mathbb{E}_y(\sum_{k=1}^{n_0+n_1+n} 1_{\{X_k=y\}}) - \mathbb{E}_y(\sum_{k=1}^{n_0+n_1} 1_{\{X_k=y\}}).
\end{aligned}$$

Le théorème 2.19 entraine que :

$$\frac{1}{n}\sum_{m=1}^{n} \mathbb{E}_y(1_{\{X_{n_1+m+n_0}=y\}} \xrightarrow[n\to+\infty]{} \frac{\rho_{yy}}{\mathbb{E}_y(T_y)} = \frac{1}{\mathbb{E}_y(T_y)},$$

et que

$$\frac{1}{n}\,\mathbb{E}_x(\sum_{m=1}^{n} 1_{\{X_m=x\}}) \xrightarrow[n\to+\infty]{} \frac{\rho_{xx}}{\mathbb{E}_x(T_x)} = \frac{1}{\mathbb{E}_x(T_x)}.$$

Par conséquent :

$$\frac{1}{\mathbb{E}_y(T_y)} \geq \frac{p^{n_1}(y,x)\,p^{n_0}(x,y)}{\mathbb{E}_x(T_x)} > 0,$$

ce qui entraine que $\mathbb{E}_y(T_y) < +\infty$. ■

Le cas "sympathique" sera celui pour lequel $\mathbb{E}_y(T_y)$ est fini. Le théorème suivant affirme que c'est le seul cas possible pour une chaîne ayant un nombre fini d'états.

Théorème 2.22 *Soit C une classe fermée irréductible. Si C est finie alors tous ses états sont récurrents positifs.*

Démonstration : D'après le théorème 2.14, tous les états de C sont récurrents. Soit $x_0 \in C$. Pour tout $y \in C$,

$$\lim_{n\to+\infty} \frac{1}{n}\,\mathbb{E}_{x_0}(\sum_{k=1}^{n} 1_{\{X_k=y\}}) = \frac{\rho_{x_0 y}}{\mathbb{E}_y(T_y)} = \frac{1}{\mathbb{E}_y(T_y)}.$$

Or C étant une classe fermée, si on part de C on reste dans C, donc

$$\sum_{y\in C} \frac{1}{n}\,\mathbb{E}_{x_0}(\sum_{k=1}^{n} 1_{\{X_k=y\}})) = \frac{1}{n}\,\mathbb{E}_{x_0}(\sum_{k=1}^{n} 1_{\{X_k\in C\}})) = 1.$$

Comme C est fini, on peut intervertir la somme en y et la limite en n et on obtient :

$$\sum_{y\in C} \frac{1}{\mathbb{E}_y(T_y)} = 1.$$

Par conséquent il existe un $y_0 \in C$ tel que $\mathbb{E}_{y_0}(T_{y_0}) < +\infty$. Donc il existe au moins un état $y_0 \in C$ qui est récurrent positif. Puisque C est irréductible et formée détats récurrents, d'après la proposition 2.21, tous les états de C sont donc récurrents positifs. ■

Bien que le cas récurrent positif soit le cas "sympathique", il ne faut pas croire que le cas récurrent nul soit "pathologique", il peut se produire pour des chaines "naturelles". La marche aléatoire simple sur $\mathbb{Z}$ est récurrente si $a = b$ $(ab > 0)$ et alors tous ses états sont récurrents nuls : elle revient une infinité de fois dans chaque état mais les intervalles de temps entre les passages sont "très longs" car la chaîne a pu aller très loin avant de revenir.

Chapitre 3

Chaîne de Markov : mesures stationnaires et théorèmes de convergence

Dans tout ce chapitre, nous étudierons le comportement d'une chaîne de Markov homogène, à valeurs dans un espace E (fini ou dénombrable), de fonction de transition $P(x, y)$.

3.1 Mesures stationnaires

Définition 3.1 *Une probabilité π sur E est* **invariante** *ou* **stationnaire** *pour la chaine de Markov $(X_n)_{n\geq 0}$ si, pour tout $n \geq 0$:*

$$(\forall\, x \in E,\, \mathbb{P}(X_n = x) = \pi(x)) \Longrightarrow (\forall\, x \in E,\, \mathbb{P}(X_{n+1} = x) = \pi(x)).$$

Proposition 3.2 *La probabilité π est stationnaire si et seulement si :*

$$\forall\, y \in E, \quad \sum_{x\in E} \pi(x) P(x, y) = \pi(y).$$

Démonstration : Nous avons :

$$\begin{aligned}
\mathbb{P}(X_{n+1} = y) &= \sum_{x\in E} \mathbb{P}(X_n = x, X_{n+1} = y) \\
&= \sum_{x\in E} \mathbb{P}(X_n = x)\, \mathbb{P}(X_{n+1} = y \,/\, X_n = x) \\
&= \sum_{x\in E} \mathbb{P}(X_n = x)\, P(x, y).
\end{aligned}$$

Supposons que pour tout $x \in E$, $\mathbb{P}(X_n = x) = \pi(x)$, alors π est stationnaire si et seulement si pour tout x dans E, $\mathbb{P}(X_{n+1} = x) = \pi(x)$, ce qui s'écrit, en vertu de la relation précédente, $\pi(y) = \sum_{x \in E} \pi(x) P(x, y)$. ■

En fait π est stationnaire si et seulement si, lorsque la loi initiale de la chaine est π (c'est-à-dire si $\mathbb{P}(X_0 = x) = \pi(x)$ pour tout x) alors, pour tout instant n, la loi de X_n est également π (c'est-à-dire $\mathbb{P}(X_n = x) = \pi(x)$ pour tout x).

Exemple : chaîne à deux états

Nous avons montré dans l'exemple 2 du chapitre 1 que la probabilité π définie par $\pi(0) = \frac{b}{a+b}, \pi(1) = \frac{a}{a+b}$ est stationnaire. De plus $\lim_{n \to +\infty} \mathbb{P}(X_n = 0) = \frac{b}{a+b} = \pi(0)$ et $\lim_{n \to +\infty} \mathbb{P}(X_n = 1) = \frac{a}{a+b} = \pi(1)$.

Si π est stationnaire et si la loi de X_0 est π, alors pour tout m et n, le vecteur $(X_n, X_{n+1}, \ldots, X_{n+m})$ a même loi que le vecteur $(X_0, X_1, \ldots, X_m)$. En effet :

$$\begin{aligned}
&\mathbb{P}(X_n = x_0, X_{n+1} = x_1, \ldots, X_{n+m} = x_m) \\
&\quad = \mathbb{P}(X_n = x_0)\, P(x_0, x_1)\, P(x_1, x_2) \ldots P(x_{m-1}, x_m) \\
&\quad = \pi(x_0)\, P(x_0, x_1)\, P(x_1, x_2) \ldots P(x_{m-1}, x_m) \\
&\quad = \mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_m = x_m).
\end{aligned}$$

Cela signifie que la chaîne de Markov regardée depuis l'instant initial a même loi que la chaîne de Markov regardée à partir de tout instant n.

Il est commode d'étendre la notion d'invariance aux mesures sur E et de ne pas la réserver aux probabilités sur E.

Définition 3.3 *Une mesure m sur E (c'est à dire une famille $(m(x))_{x \in E}$ de réels positifs ou nuls) est dite invariante (ou stationnaire) si la mesure m n'est pas la mesure identiquement nulle et si :*

$$\forall\, y \in E \quad \sum_{x \in E} m(x)\, P(x, y) = m(y). \tag{3.1}$$

On peut faire les remarques suivantes :

1. Une probabilité invariante est évidemment une mesure invariante.
2. Si m est une mesure invariante et si $\lambda > 0$, alors λm est une mesure invariante.
3. Si m est une mesure invariante et si $m(E) < +\infty$, alors $m/m(E)$ est une probabilité invariante.

En pratique, on cherche des mesures stationnaires et on les renormalise éventuellement pour avoir des probabilités stationnaires.

Si E est fini et comprend d éléments, m peut être représentée par un vecteur colonne de d composantes et les équations de stationnarité 3.1 s'écrivent :

$$m^t P = m^t.$$

La recherche d'une mesure stationnaire se ramène donc à la résolution d'un système linéaire homogène de d équations à d inconnues. Cela revient à chercher les vecteurs propres à gauche de P (c'est-à-dire les vecteurs propres de P^t) associés à la valeur propre 1. Comme la somme des lignes de P est égale à 1, le vecteur dont toutes les composantes sont égales à 1 est vecteur propre de P associé à la valeur propre 1. Donc 1 est valeur propre de P^t, mais il n'est pas évident qu'il existe un vecteur propre dont toutes les composantes soient positives ou nulles. Nous discuterons ce point un peu plus loin.

Exemple :
Prenons une chaîne de naissance et mort avec $d = 3$ de matrice de transition P égale à :

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

Les équations de stationnarité s'écrivent :

$$\frac{1}{3}\,\pi(1) = \pi(0), \quad \pi(0) + \frac{2}{3}\,\pi(2) = \pi(1),$$

$$\frac{2}{3}\,\pi(1) + \pi(3) = \pi(2), \quad \frac{1}{3}\,\pi(2) = \pi(3).$$

On obtient :

$$\pi(1) = 3\pi(0),\ \pi(2) = 3\pi(0),\ \pi(3) = \pi(0).$$

Puisqu'on cherche une probabilité, on doit avoir $\sum_{i=1}^{3} \pi(i) = 1$. On en déduit :

$$\pi(0) = \frac{1}{8},\ \pi(1) = \frac{3}{8},\ \pi(2) = \frac{3}{8},\ \pi(3) = \frac{1}{8}.$$

Pour cette chaîne, nous n'avons pas $\lim_{n\to+\infty} \mathbb{P}(X_n = 0) = \pi(0)$ et $\lim_{n\to+\infty} \mathbb{P}(X_n = 1) = \pi(1)$ car si n est impair $\mathbb{P}_x(X_n = x) = 0$. Cette chaîne a un comportement "périodique", nous y reviendrons à la fin du chapitre.

Il peut être lourd de chercher les mesures stationnaires à l'aide de la définition (3.1), la notion de réversibilité que nous allons définir est plus restrictive mais plus facile à manipuler.

Définition 3.4 *La mesure m sur E est réversible pour la chaîne de Markov de fonction de transition P si :*

$$m(x)\,P(x,y) = m(y)\,P(y,x), \forall\, x; y \in E.$$

Proposition 3.5 *Toute mesure réversible est stationnaire.*

Démonstration : La stationnarité de m s'écrit :

$$\sum_{x\in E} m(x)\,P(x,y) = m(y) = m(y)\sum_{x\in E} P(y,x), \quad \forall\, y \in E,$$

ou encore :

$$\sum_{x\in E}(m(x)\,P(x,y) - m(y)\,P(y,x)) = 0, \quad \forall\, y \in E.$$

On voit donc qu'une condition suffisante (mais non nécessaire) pour que m soit stationnaire est que :

$$m(x)P(x,y) - m(y)P(y,x) = 0, \quad \forall\, x, y \in E.$$

■

Si la mesure réversible m est une probabilité, la réversibilité s'interprète ainsi : si la loi de la chaîne à l'instant n est m, alors $\mathbb{P}(X_n = x, X_{n+1} = y) = \mathbb{P}(X_n = y, X_{n+1} = x)$. Autrement dit, le comportement de la suite des X_n lorsque le temps croit ou décroit est le même, d'où le terme "réversibilité" : on peut "retourner le temps".

Exemple : chaîne de naissance et mort

On suppose que $q_x > 0$ pour tout $x \geq 1$. Les équations de réversibilité s'écrivent :

$$\forall\, x \geq 0, x \in E : m(x-1)\,p_{x-1} = m(x)\,q_x.$$

On en déduit que m est une mesure réversible si et seulement si pour $x \geq 1, x \in E$:

$$m(x) = \frac{p_0 \cdots p_{x-1}}{q_1 \cdots\cdots q_x}\, m(0).$$

Si $E = \{0, 1, ...d\}$, la probabilité π donnée par :

$$\pi(x) = \frac{m(x)}{\sum_{y=0}^{d} m(y)} = \frac{\frac{p_0 \cdots p_{x-1}}{q_1 \cdots\cdots q_x}}{\sum_{y=0}^{d} \frac{p_0 \cdots p_{y-1}}{q_1 \cdots\cdots q_y}} \quad \text{pour } 0 \leq x \leq d,$$

est réversible donc stationnaire. Si $E = \mathbb{N}$, et si $\sum_{y\geq 0} m(y) =< +\infty$, c'est-à-dire si $\sum_{y\geq 0} \frac{p_0 \cdots p_{y-1}}{q_1 \cdots q_y} < +\infty$, il existe une et une seule probabilité réversible π donnée par :

$$\pi(x) = \frac{m(x)}{\sum_{y=0}^{d} m(y)} = \frac{\frac{p_0 \cdots p_{x-1}}{q_1 \cdots q_x}}{\sum_{y\geq 0} \frac{p_0 \cdots p_{y-1}}{q_1 \cdots q_y}} \quad \text{pour } x \in \mathbb{N}.$$

Si $E = \mathbb{N}$, et si $\sum_{y\geq 0} \frac{p_0 \cdots p_{y-1}}{q_1 \cdots q_y} = +\infty$, il n'existe pas de probabilité réversible.

Exemple : algorithme de Metropolis

Il s'agit d'un algorithme permettant de construire une chaîne de Markov de loi stationnaire π donnée (son expression peut n'être connue qu'à une constante multiplicative près). On suppose que $\pi(x) > 0$ pour tout x (sinon on réduit l'espace d'états !). On se donne une fonction de transition q irréductible et symétrique, c'est-à-dire telle que $q(x, y) = q(y, x)$. On simule la chaîne de la manière suivante : si $X_n = x$, on tire un point de E avec la probabilité $q(x, .)$. Supposons avoir tiré y, on prend $X_{n+1} = y$ avec la probabilité $\min(\frac{\pi(y)}{\pi(x)}, 1)$ et on prend $X_{n+1} = x$ avec la probabilité $1 - \min(\frac{\pi(y)}{\pi(x)}, 1)$. La fonction de transition de la chaîne ainsi construite vérifie, pour $x \neq y$:

$$P(x, y) = q(x, y) \min(\frac{\pi(y)}{\pi(x)}, 1).$$

Montrons que π est réversible pour cette chaîne. Il faut vérifier que pour $x \neq y$, $\pi(x) P(x, y) = \pi(y) P(y, x)$. Compte-tenue de la symétrie de la formule, on peut supposer que $\pi(x) \leq \pi(y)$. On a alors :

$$\pi(x) P(x, y) = \pi(x) q(x, y), \quad \pi(y) P(y, x) = \pi(y) q(y, x) \frac{\pi(x)}{\pi(y)} = \pi(x) q(y, x),$$

d'où le résultat.

Proposition 3.6 *Soit π une probabilité stationnaire. Si x est un état transient ou récurrent nul, alors $\pi(x) = 0$.*

Démonstration : Soit π une probabilité stationnaire. Pour tout $x \in E$, on a :

$$\pi(x) = \sum_{y\in E} \pi(y) P(y, x),$$

et en itérant :

$$\pi(x) = \sum_{y\in E} \pi(y) P^k(y, x) = \sum_{y\in E} \pi(y) \mathbb{E}_y(1_{\{X_k = x\}}).$$

Par conséquent :

$$\pi(x) = \sum_{y \in E} \pi(y) \, \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}_y(1_{\{X_k = x\}}).$$

Si x est transient ou si x est récurrent nul, $\frac{1}{n} \sum_{k=1}^{n} \mathbb{E}_y(1_{\{X_k = x\}})$ tend vers 0 lorsque n tend vers l'infini (formule (2.2) et théorème 2.19). Par le théorème de convergence dominée (π est une probabilité), on en déduit donc que $\pi(x) = 0$. ■

3.2 Cas d'une chaîne récurrente irréductible

Commençons par un lemme.

Lemme 3.7 *Soit m une mesure invariante (donc non identiquement nulle) d'une chaîne de Markov irréductible. Alors, pour tout $y \in E$, on a $m(y) > 0$.*

Démonstration : On a fait remarquer dans la proposition précédente que si m est invariante, on montre par récurrence sur k que :

$$\forall\, k \geq 1, \; \forall\, y \in E, \quad m(y) = \sum_{x \in E} m(x) \, P^k(x, y). \tag{3.2}$$

Puisque la mesure m n'est pas la mesure identiquement nulle, il existe $x_0 \in E$ tel que $m(x_0) > 0$. Soit $y \in E, y \neq x_0$. Puisque la chaîne est irréductible, x_0 conduit à y, donc il existe $n \geq 1$ tel que $P^n(x_0, y) > 0$. L'équation (3.2) donne pour $k = n$:

$$m(y) = \sum_{x \in E} m(x) \, P^n(x, y) \geq m(x_0) \, P^n(x_0, y) > 0.$$

■

Théorème 3.8 *Une chaîne de Markov récurrente irréductible possède une mesure invariante m. Cette mesure stationnaire est strictement positive en tout point et unique à une constante multiplicative près. En outre pour tout $x \in E$ on a :*

$$\forall\, y \in E, \; m(y) = c(x) \, \mathbb{E}_x \left(\sum_{k=1}^{T_x} 1_{\{X_k = y\}} \right) \quad (c(x) > 0).$$

Par suite la chaîne est récurrente positive si et seulement si ses mesures stationnaires sont de masse totale finie.

Démonstration : Pour x et y dans E, posons :

$$\lambda_x(y) = \mathbb{E}_x\left(\sum_{k=1}^{T_x} 1_{\{X_k=y\}}\right) = \sum_{k\geq 1} \mathbb{P}_x(T_x \geq k, X_k = y) \in \overline{\mathbb{R}}_+.$$

1) Nous allons montrer que pour tout x, λ_x est une mesure invariante strictement positive.

Montrons d'abord que pour tout $z \in E$, $\sum_{y\in E} \lambda_x(y)\, P(y,z) = \lambda_x(z)$ (les deux membres étant à valeurs dans $\bar{\mathbb{R}}_+$).

En utilisant le fait que $\lambda_x(x) = 1$, nous obtenons :

$$\begin{aligned}
\sum_{y\in E} \lambda_x(y)\, P(y,z) &= \lambda_x(x)\, P(x,z) + \sum_{y:y\neq x} \lambda_x(y)\, P(y,z) \\
&= P(x,z) + \sum_{y:y\neq x}\sum_{k\geq 1} \mathbb{P}_x(T_x \geq k, X_k = y)\, P(y,z).
\end{aligned}$$

On peut appliquer la propriété de Markov à l'instant k car

$$\{T_x \geq k\} = \{X_1 \neq x, \ldots, X_{k-1} \neq x\}.$$

On obtient, après interversion des sommes :

$$\begin{aligned}
\sum_{y\in E} \lambda_x(y)\, P(y,z) &= P(x,z) \\
&\quad + \sum_{k\geq 1}\sum_{y:y\neq x} \mathbb{P}_x(T_x \geq k, X_k = y, X_{k+1} = z) \\
&= P(x,z) \\
&\quad + \sum_{k\geq 1} \mathbb{P}_x(T_x \geq k, X_k \neq x, X_{k+1} = z) \\
&= P(x,z) + \sum_{k\geq 1} \mathbb{P}_x(T_x \geq k+1, X_{k+1} = z) \\
&= P(x,z) + \lambda_x(z) - \mathbb{P}_x(T_x \geq 1, X_1 = z) \\
&= P(x,z) + \lambda_x(z) - \mathbb{P}_x(X_1 = z) \\
&= \lambda_x(z).
\end{aligned}$$

Montrons maintenant que $\lambda_x(y_0) < +\infty$ pour tout $y_0 \in E$. Soit $y_0 \in E$. Puisque y_0 conduit à x, il existe n tel que $P^n(y_0,x) > 0$. D'autre part, en itérant l'égalité d'invariance de λ_x, on obtient pour tout $z \in E$:

$$\sum_{y\in E} \lambda_x(y)\, P^n(y,z) = \lambda_x(z).$$

En prenant $z = x$, on obtient :

$$\lambda_x(y_0)\, P^n(y_0, x) \le \sum_{y\in E} \lambda_x(y)\, P^n(y, x) = \lambda_x(x) = 1 < +\infty.$$

Par conséquent $\lambda_x(y_0) < +\infty$ car $\lambda_x(y_0)\, P^n(y_0, x) < +\infty$ et $P^n(y_0, x) > 0$.

Nous avons donc montré que λ_x est une mesure invariante.

2) Il reste à montrer que toutes les mesures invariantes sont proportionnelles à λ_x (x donné quelconque). Soit m une mesure invariante. Montrons d'abord par récurrence sur n que :

$$\forall\, z \in E, \quad m(z) \ge m(x) \sum_{k=1}^{n} \mathbb{P}_x(T_x \ge k, X_k = z). \tag{3.3}$$

Cette formule est vraie pour $n = 1$ car :

$$m(z) = \sum_{y\in E} m(y)\, P(y, z) \ge m(x)\, P(x, z) = m(x)\, \mathbb{P}_x(T_x \ge 1, X_1 = z).$$

Supposons donc (3.3) vraie et montrons-là pour n remplacé par $n + 1$. En utilisant l'hypothèse de récurrence et en appliquant la même méthode que ci-dessus, on obtient pour tout $z \in E$:

$$\begin{aligned}
m(z) &= \sum_{y\in E} m(y)\, P(y, z) \\
&= m(x)\, P(x, z) + \sum_{y: y\ne z} m(y)\, P(y, z) \\
&\ge m(x)\, P(x, z) + \sum_{y: y\ne z} m(x) \sum_{k=1}^{n} \mathbb{P}_x(T_x \ge k, X_k = y)\, P(y, z) \\
&= m(x)\, P(x, z) + \sum_{y: y\ne z} m(x) \sum_{k=1}^{n} \mathbb{P}_x(T_x \ge k, X_k = y, X_{k+1} = z) \\
&= m(x)\, P(x, z) + m(x) \sum_{k=1}^{n} \mathbb{P}_x(T_x \ge k, X_k \ne x, X_{k+1} = z) \\
&= m(x)\, P(x, z) + m(x) \sum_{k=1}^{n} \mathbb{P}_x(T_x \ge k + 1, X_{k+1} = z) \\
&= m(x)\, \mathbb{P}_x(T_x \ge 1, X_1 = z) + m(x) \sum_{m=2}^{n+1} \mathbb{P}_x(T_x \ge m, X_m = z) \\
&= m(x) \sum_{m=1}^{n+1} \mathbb{P}_x(T_x \ge m, X_m = z).
\end{aligned}$$

La récurrence est donc établie. En faisant tendre n vers l'infini dans (3.3) on obtient :

$$\forall\, z \in E, \quad m(z) \geq m(x) \sum_{k \geq 1} \mathbb{P}_x(T_x \geq k, X_k = z),$$

c'est-à-dire pour tout $x \in E$:

$$\forall\, z \in E, \quad m(z) \geq m(x) \lambda_x(z).$$

Par conséquent $m_1 = m - m(x)\, \lambda_x$ est une mesure positive. On voit immédiatement qu'elle vérifie, tout comme m et λ_x,

$$\forall\, y \in E, m_1(y) = \sum_{x \in E} m_1(x)\, P(x, y).$$

D'après le lemme 3.7, elle est donc soit strictement positive, soit identiquement nulle. Or

$$m_1(x) = m(x) - m(x)\, \lambda_x(x) = m(x) - m(x) = 0.$$

Donc la mesure m_1 est identiquement nulle. Donc

$$m = m(x)\, \lambda_x$$

ce qui montre que m est proportionnelle à λ_x. On a donc prouvé que toute mesure invariante est proportionnelle à λ_x.

On en déduit que $\sum_{y \in E} m(y) = m(x) \mathbb{E}_x(T_x)$. Donc il existe une probabilité invariante π si et seulement si $\mathbb{E}_x(T_x) < +\infty$ et alors $\pi(x) = 1/\mathbb{E}_x(T_x)$. ■

Théorème 3.9 *Une chaîne de Markov irréductible et récurrente positive possède une et une seule probabilité stationnaire π. Cette probabilité stationnaire π vérifie :*

$$\forall\, x_0 \in E, \forall\, y \in E, \quad \pi(y) = \frac{\mathbb{E}_{x_0}(\sum_{k=1}^{T_{x_0}} 1_{\{X_k = y\}})}{\mathbb{E}_{x_0}(T_{x_0})} = \frac{1}{\mathbb{E}_y(T_y)}.$$

En outre, pour tout $y \in E$:

$$\frac{1}{n} \sum_{k=1}^{n} 1_{\{X_k = y\}} \xrightarrow[n \to +\infty]{\text{p.s.}} \pi(y).$$

Plus généralement, pour toute fonction f positive ou π-intégrable :

$$\frac{1}{n} \sum_{k=1}^{n} f(X_k) \xrightarrow[n \to +\infty]{\text{p.s.}} \sum_{y \in E} f(y)\, \pi(y) = \int f d\pi.$$

Démonstration : L'existence d'une unique probabilité stationaire π qui vérifie

$$\forall\, x_0 \in E, \forall\, y \in E\ \pi(y) = \frac{\mathbb{E}_{x_0}(\sum_{k=1}^{T_{x_0}} 1_{\{X_k=y\}})}{\mathbb{E}_{x_0}(T_{x_0})}$$

est une conséquence immédiate du théorème 3.8. En prenant $x_0 = y$, on obtient :

$$\pi(y) = \frac{\mathbb{E}_y(\sum_{k=1}^{T_y} 1_{\{X_k=y\}})}{\mathbb{E}_y(T_y)} = \frac{1}{\mathbb{E}_y(T_y)}.$$

En appliquant le théorème 2.19, on obtient :

$$\frac{1}{n}\sum_{k=1}^{n} 1_{\{X_k=y\}} \xrightarrow[n\to+\infty]{} \frac{1_{\{S_y<+\infty\}}}{\mathbb{E}_y(T_y)}.$$

Or la chaîne étant récurrente irréductible, S_y est fini presque-sûrement, donc $1_{\{S_y<+\infty\}} = 1$ presque-sûrement, quelle que soit la loi initiale. Donc :

$$\frac{1}{n}\sum_{k=1}^{n} 1_{\{X_k=y\}} \xrightarrow[n\to+\infty]{} \frac{1}{\mathbb{E}_y(T_y)} = \pi(y).$$

Soit f une fonction positive sur E. Supposons $X_0 = y$. Définissons les variables aléatoires $Y_1 = \sum_{k=1}^{T_y^{(1)}} f(X_k), \ldots, Y_n = \sum_{k=T_y^{(n-1)}+1}^{T_y^{(n)}} f(X_k), \ldots$. Ces variables aléatoires sont indépendantes et de même loi. Leur espérance vaut :

$$\begin{aligned}
\mathbb{E}_y(Y_1) &= \mathbb{E}_y\left(\sum_{k=1}^{T_y^{(1)}}\sum_{x\in E} f(x)\, 1_{\{X_k=x\}}\right) \\
&= \sum_{x\in E} f(x)\, \mathbb{E}_y\left(\sum_{k=1}^{T_y^{(1)}} 1_{\{X_h=x\}}\right) \\
&= \sum_{x\in E} f(x)\, \frac{\pi(x)}{\pi(y)}
\end{aligned}$$

On a :

$$T_y^{(N_n(y))} \leq n \leq T_y^{(N_n(y)+1)}$$

d'où comme f est positive :

$$\frac{1}{n}\sum_{k=1}^{T_y^{(N_n(y))}} f(X_k) \leq \frac{1}{n}\sum_{k=1}^{n} f(X_k) \leq \frac{1}{n}\sum_{k=1}^{T_y^{(N_n(y)+1)}} f(X_k)$$

soit

$$\frac{N_n(y)}{n}\frac{1}{N_n(y)}\sum_{m=1}^{N_n(y)} Y_m \leq \frac{1}{n}\sum_{k=1}^{n} f(X_k) \leq \frac{N_n(y)+1}{n}\frac{1}{N_n(y)+1}\sum_{m=1}^{N_n(y)+1} Y_m$$

Si n tend vers l'infini, $\frac{N_n(y)}{n}$ et $\frac{N_n(y)+1}{n}$ convergent p.s. vers $\pi(y)$ tandis que $\frac{1}{N_n(y)}\sum_{m=1}^{N_n(y)} Y_m$ et $\frac{1}{N_n(y)+1}\sum_{m=1}^{N_n(y)+1} Y_m$ convergent p.s. vers $\mathbb{E}_y(Y_1)$. On en déduit que $\frac{1}{n}\sum_{k=1}^{n} f(X_k)$ converge p.s. vers $\pi(y)\mathbb{E}_y(Y_1) = \sum_{x\in E} f(x)\,\pi(x)$.

Si f est de signe quelconque et π-intégrable, le résultat s'obtient en appliquant le résultat précédent aux fonctions positives f_+ et f_-. ■

Le théorème 3.9 est un théorème ergodique : il relie une moyenne temporelle à une moyenne "spatiale" (moyenne sur l'espace $(\Omega, \mathcal{A}, \mathbb{P})$). C'est une généralisation de la loi des grands nombres à des variables aléatoires non indépendantes.

On peut appliquer le théorème 3.9 pour calculer $\int h\,d\pi$ par simulation, où π est une probabilité. On construit une chaîne de Markov $(X_n)_{n\geq 0}$ récurrente positive de loi stationnaire π et on approche $\int h\,d\pi$ par $\frac{1}{n}\sum_{k=1}^{n} h(X_k)$.

Corollaire 3.10 *Une chaîne irréductible est récurrente positive si et seulement si elle possède une probabilité stationnaire (et celle-ci est alors unique).*

Démonstration : La condition nécessaire est contenue dans l'énoncé du théorème 3.9.

Pour la condition suffisante, la chaine étant irréductible, soit tous ses états sont récurrents positifs, soit tous ses états sont récurrents nuls, soit tous ses états sont transients. On suppose qu'il existe une probabilité π stationnaire. D'après la proposition 3.6, $\pi(x) = 0$ si x est récurrent nul ou transient. Donc si ou bien tous les états sont récurrents nuls ou bien tous les états sont transients, alors $\pi(x) = 0$ pour tout x, ce qui contredit le fait que π est une probabilité. Par conséquent les états sont tous récurrents positifs. ■

Corollaire 3.11 *Une chaîne irréductible sur un espace d'états fini possède une et une seule probabilité stationnaire π et tous ses états sont récurrents positifs.*

En pratique : Pour trouver la probabilité invariante éventuelle d'une chaîne irréductible, on peut commencer par chercher des mesures réversibles, c'est-à-dire de suites $(\gamma(x))_{x\in E}$ vérifiant :

$$\forall\, x \in E, \quad \gamma(x) \geq 0,$$

$$\forall\, x, y \in E, \quad \gamma(x)p(x,y) = \gamma(y)p(y,x).$$

Si on en trouve une, alors la mesure $m = \gamma$ est invariante.

Si on n'en trouve pas, on ne peut conclure et il faut revenir à la recherche directe de mesures invariantes.

Si on ne trouve pas de mesure invariante non identiquement nulle, alors la chaîne est transiente.

Si on a trouvé une mesure invariante m non identiquement nulle :

– ou bien $\sum_x m(x) < +\infty$, , la chaîne est alors récurrente positive et la probabilité invariante est $\pi(x) = m(x)/\sum_y m(y)$,

– ou bien $\sum_x m(x) = +\infty$, la chaîne est alors récurrente nulle ou transiente.

Exemple : Considérons une chaîne de naissance et mort sur $\mathbb{N}$ telle que pour tout $n \in \mathbb{N}$, $p_n > 0$ et pour tout $n \geq 1$, $q_n > 0$.

On a mis en évidence une mesure réversible, donc invariante de la forme pour $n \geq 1$:

$$m(n) = \frac{p_0 \cdot \ldots \cdot p_{n-1}}{q_1 \cdot \ldots \cdot q_n}\, m(0).$$

D'après le corollaire 3.10, une condition nécessaire et suffisante pour que la chaîne soit récurrente positive est que :

$$\sum_{n \geq 1} \frac{p_0 \cdot \ldots \cdot p_{n-1}}{q_1 \cdot \ldots \cdot q_n} < +\infty.$$

On verra en TD que la chaîne est transiente si et seulement si :

$$\sum_{n \geq 1} \frac{q_1 \cdots q_n}{p_1 \cdots p_n} < +\infty.$$

Par conséquent, elle est récurrente nulle si et seulement si :

$$\sum_{n \geq 1} \frac{p_0 \cdot \ldots \cdot p_{n-1}}{q_1 \cdot \ldots \cdot q_n} = +\infty \quad \text{et} \quad \sum_{n \geq 1} \frac{q_1 \cdots q_n}{p_1 \cdots p_n} = +\infty.$$

En particulier si $p_0 = 1$ et pour tout $n \geq 1$, $p_n = p$ et $q_n = 1 - p$ avec $0 < p < 1$ (marche aléatoire simple sur $\mathbb{N}$), la chaîne est récurrente positive pour $p < 1/2$, elle est récurrente nulle pour $p = 1/2$ et transiente pour $p > 1/2$.

Par contre dans le cas d'une marche aléatoire simple sur $\mathbb{Z}$, (c'est-à-dire pour tout $i \in \mathbb{Z}$, $p(i, i+1) = p > 0$, $p(i, i-1) = q = 1 - p > 0$), la chaîne est transiente pour $p \neq 1/2$ et récurrente nulle pour $p = 1/2$ (voir TD).

3.3 Cas d'une chaîne non irréductible

Lemme 3.12 *Soit C une classe fermée et m une mesure portée par C. Alors m est invariante pour la chaîne de Markov initiale si et seulement si elle est invariante pour la chaîne de Markov restreinte à C.*

Démonstration : La mesure m est invariante pour la chaîne initiale si et seulement si

$$\forall\, y \in E, \quad m(y) = \sum_x m(x)\, P(x, y).$$

Comme m est portée par C, cela équivaut à

$$\forall\, y \in E, \quad m(y) = \sum_{x \in C} m(x)\, P(x, y).$$

Comme C est fermée et m portée par C, c'est équivalent

$$\forall\, y \in C, \quad m(y) = \sum_{x \in C} m(x)\, P(x, y)$$

car si $y \notin C$, $m(y) = 0$ et pour tout $x \in C$, $P(x, y) = 0$, donc $\sum_{x \in C} m(x)P(x, y) = 0$. ■

Théorème 3.13 *Soit C une classe fermée irréductible formée d'états récurrents positifs, alors la chaîne possède une et une seule probabilité stationnaire π concentrée sur C (c'est-à-dire telle que $\pi(C^c) = 0$). Elle est donnée par :*

$$\pi(x) = \begin{cases} \frac{1}{\mathbb{E}_x(T_x)} & \text{si } x \in C, \\ 0 & \text{sinon.} \end{cases}$$

Démonstration : Soit π_C la probabilité invariante pour la chaine restreinte à C. Alors pour tout x dans C, $\pi_C(x) = 1/\mathbb{E}_x(T_x)$ et T_x est le même pour la chaîne restreinte ou la chaîne initiale si on part d'un point de C.

D'après le lemme 3.12, la probabilité π donnée dans le théorème 3.13 est portée par C et invariante pour la chaîne initiale.

Réciproquement, si π_1 est une probabilité invariante pour la chaîne initiale et portée par C, d'après le lemme 3.12, sa restriction à C est invariante pour la chaîne restreinte à C donc elle est égale à π_C et par suite $\pi_1 = \pi$. ■

Théorème 3.14 *Soit $(C_k)_{k \in K}$ (K fini ou dénombrable) l'ensemble des classes récurrentes irréductibles qui sont récurrentes positives (c'est-à-dire formées* *d'états récurrents positifs). Notons π_k ($k \in K$) l'unique probabilité stationnaire concentrée sur C_k. Alors les probabilités stationnaires de la chaine sont les mesures de la forme :*

$$\pi = \sum_{k \in K} c_k \, \pi_k, \quad \text{avec } c_k \geq 0, \ \sum_{k \in K} c_k = 1.$$

Démonstration : Supposons que

$$\pi = \sum_{k \in K} c_k \, \pi_k, \quad \text{avec } c_k \geq 0, \ \sum_{k \in K} c_k = 1,$$

alors π est une probabilité et pour tout $y \in E$:

$$\begin{aligned}
\sum_{x \in E} \pi(x) \, P(x,y) &= \sum_{x \in E} \sum_{k \in K} c_k \, \pi_k(x) \, P(x,y) \\
&= \sum_{k \in K} c_k \sum_{x \in E} \pi_k(x) \, P(x,y) \\
&= \sum_{k \in K} c_k \, \pi(y) \\
&= \pi(y).
\end{aligned}$$

Réciproquement soit π une probabilité stationnaire. Posons :

$$m_k(x) = \begin{cases} \pi(x) & \text{si} \quad x \in C_k \\ 0 & \text{si} \quad x \notin C_k. \end{cases}$$

Montrons que la mesure m_k est invariante. Nous avons :

$$\sum_{x \in E} m_k(x) \, P(x,y) = \sum_{x \in C_k} \pi(x) \, P(x,y).$$

Si $y \notin C_k$, alors $m_k(y) = 0$, et pour $x \in C_k$, $P(x,y) = 0$ car C_k est fermée. On a donc dans ce cas

$$m_k(y) = 0 = \sum_{x \in C_k} \pi(x) \, P(x,y) = \sum_{x \in E} m_k(x) \, P(x,y).$$

Supposons maintenant que $y \in C_k$. Si x est transient, alors $\pi(x) = 0$ (proposition 3.6). Si x est récurrent et n'appartient pas à C_k, x et y ne sont pas dans la même classe récurrente irréductible, donc $P(x,y) = 0$. Donc si $x \notin C_k$, on a $\pi(x) \, P(x,y) = 0$. Par suite :

$$\sum_{x \in E} m_k(x) \, P(x,y) = \sum_{x \in C_k} \pi(x) \, P(x,y) = \sum_{x \in E} \pi(x) \, P(x,y).$$

En utilisant que π est invariante, on obtient finalement :

$$\sum_{x\in E} m_k(x)\, P(x,y) = \sum_{x\in E} \pi(x)\, P(x,y) = \pi(y) = m_k(y),$$

ce qui achève de prouver que m_k est invariante.

On a $m_k(C_k) = \pi(C_k) \leq 1 < +\infty$, donc ou bien $c_k = \pi(C_k) = 0$, et alors m_k est la mesure identiquement nulle, donc $m_k = c_k\,\pi_k$, ou bien $c_k = \pi(C_k) > 0$ et alors m_k/c_k est une probabilité invariante portée par C_k, et donc $m_k/c_k = \pi_k$ (théorème 3.13), ce qui s'écrit encore $m_k = c_k \pi_k$.

Puisque $\pi(x) = 0$ si $x \notin \cup_{k\in K} C_k$ (proposition 3.6), on a pour tout x, $\pi(x) = \sum_{k\in K} \pi(x) 1_{C_k}(x)$. Or $m_k = \pi 1_{C_k}$. Donc

$$\pi = \sum_{k\in C_k} m_k = \sum_{k\in K} c_k\, \pi_k.$$

En outre $c_k \geq 0$ et (toujours en utilisant la proposition 3.6) :

$$\sum_{k\in K} c_k = \sum_{k\in K} \pi(C_k) = \pi(E) = 1.$$

■

3.4 Convergence en loi vers la loi stationnaire

Etant donnée une mesure initiale μ, supposons que

$$\forall\, x \in E, \;\lim_{n\to+\infty} \mathbb{P}_\mu(X_n = x) = \nu(x).$$

Alors pour tout $y \in E$:

$$\begin{aligned}\nu(y) &= \lim_{n\to+\infty} \mathbb{P}_\mu(X_{n+1} = y) \\ &= \lim_{n\to+\infty} \sum_{x\in E} \mathbb{P}_\mu(X_n = x, X_{n+1} = y) \\ &= \lim_{n\to+\infty} \sum_{x\in E} \mathbb{P}_\mu(X_n = x)\, P(x,y).\end{aligned}$$

Si tout se passe bien,

$$\lim_{n\to+\infty} \sum_{x\in E} \mathbb{P}_\mu(X_n = x)\, P(x,y) = \sum_{x\in E} \lim_{n\to+\infty} \mathbb{P}_\mu(X_n = x)\, P(x,y)$$

et par suite
$$\nu(y) = \sum_{x\in E} \nu(x)\, P(x,y),$$
donc ν est stationnaire.

Voyons ce qui se passe sur deux exemples.

Exemple : Considérons la chaîne de Markov à deux états à valeurs dans $E = \{0,1\}$ de matrice de transition
$$P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$$
$0 < a, b < 1$. Elle est récurrente irréductible. Sa probabilité stationnaire est
$$\pi(0) = \frac{b}{a+b}, \quad \pi(1) = \frac{a}{a+b}.$$
Nous avons vu que si la loi initiale est μ, alors
$$\mathbb{P}_\mu(X_n = 0) = \frac{b}{a+b} + (1-a-b)^n(\mu(0) - \frac{b}{a+b}),$$
$$\mathbb{P}_\mu(X_n = 1) = \frac{a}{a+b} + (1-a-b)^n(\mu(1) - \frac{a}{a+b}),$$
donc :
$$\mathbb{P}_\mu(X_n = 0) \xrightarrow[n\to+\infty]{} \pi(0), \quad \mathbb{P}_\mu(X_n = 1) \xrightarrow[n\to+\infty]{} \pi(1).$$

Exemple : Considérons un processus de naissance et mort sur $E = \{0, 1, \ldots, d\}$ tel que $p(x,x) = 0$, $p(x, x+1) = p_x > 0$ pour $0 \leq x < d$ et $p(x, x-1) = q_x = 1-p_x > 0$ pour $1 \leq x \leq d$, $p(0,1) = 1$, $p(d, d-1) = 1$. C'est une chaîne irréductible sur un espace d'états fini, elle est donc récurrente positive et possède une unique probabilité stationnaire π qui vérifie $\pi(y) > 0$ pour tout $y \in E$. Soit $x \in E$, la quantité $\mathbb{P}_x(X_n = y)$ ne peut converger vers $\pi(y)$ quand n tend vers l'infini, car cette quantité est nulle si ou bien n est impair et x et y sont de même parité, ou bien n est pair et x et y ne sont pas de même parité.

Ces comportements différents vont être expliqués par les résultats ci-dessous.

Définition 3.15 *Etant donné $x \in E$ pour lequel $\mathbb{P}_x(T_x < +\infty) > 0$ (c'est-à-dire il existe $n \geq 1$ tel que $P^n(x,x) > 0$), sa période d_x est le plus grand commun diviseur de $\{n \geq 1 : P^n(x,x) > 0\}$.*

Il est clair que :

$$1 \leq d_x \leq \min\{n \geq 1 : P^n(x,x) > 0\}.$$

En particulier s'il existe x tels que $P(x,x) > 0$, alors $d_x = 1$.

Proposition 3.16 *Si x conduit à y et si y conduit à x, alors $d_x = d_y$ Par conséquent, tous les éléments d'une chaîne de Markov irréductible ont même période. On dit que la chaîne est* **périodique** *de période d si $d > 1$ et* **apériodique** *si $d = 1$.*

Démonstration : Comme x conduit à y et y conduit à x, il existent $n_0 \geq 1$ et $n_1 \geq 1$ tels que $P^{n_0}(x,y) > 0$ et $P^{n_1}(y,x) > 0$. Alors on a $P^{n_0+n_1}(x,x) \geq P^{n_0}(x,y)\,P^{n_1}(y,x) > 0$, donc d_x est un diviseur de $n_0 + n_1$.

Pour tout $n \geq 0$ on a $P^{n_0+n+n_1}(x,x) \geq P^{n_0}(x,y)\,P^n(y,y)\,P^{n_1}(y,x)$. Donc si $P^n(y,y) > 0$ on a $P^{n_0+n+n_1}(x,x) > 0$ et d_x est un diviseur de $n_0 + n + n_1$.

On en déduit que si $P^n(y,y) > 0$, d_x étant un diviseur de $n_0 + n_1$ et de $n_0 + n + n_1$, d_x est également un diviseur de n. Donc $d_x \leq d_y$.

On a évidemment de la même manière $d_y \leq d_x$. ∎

Exemple : Une chaîne de naissance et mort pour laquelle $r_x = 0$ pour tout x est périodique de période 2. Une chaîne de naissance et mort pour laquelle il existe x tel que $r_x > 0$ est apériodique.

Nous admettrons sans démonstration le théorème suivant.

Théorème 3.17 (théorème d'Orey) *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov récurrente irréductible.*

1. *Si elle est récurrente nulle, alors :*

$$\lim_{n\to+\infty} P^n(x,y) = \lim_{n\to+\infty} \mathbb{P}_x(X_n = y) = 0.$$

2. *Si elle est récurrente positive de loi stationnaire π :*

   (a) *soit elle est apériodique, et alors :*

$$\lim_{n\to+\infty} P^n(x,y) = \lim_{n\to+\infty} \mathbb{P}_x(X_n = y) = \pi(y).$$

   (b) *soit elle est périodique de période d, et pour tout couple x, y d'états de E, il existe un entier r $(0 \leq r < d)$ tel que :*

$$\lim_{n\to+\infty} P^n(x,y) = \lim_{n\to+\infty} \mathbb{P}_x(X_n = y) = 0$$

   si n n'est pas de la forme $n = md + r$ et

$$\lim_{m\to+\infty} P^{md+r}(x,y) = \lim_{m\to+\infty} \mathbb{P}_x(X_{md+r} = y) = d\,\pi(y).$$

Considérons une chaîne récurrente irréductible de période $d > 1$. Le théorème 3.17 affirme qu'étant donnés x et y il existe $r \in \mathbb{N}$, $0 \leq r < d$ tel que $P^n(x,y)$ converge vers 0 lorsque n tend vers l'infini, sauf si n est de la forme $n = md + r$, $m \in \mathbb{N}$ et $P^{md+r}(x,y)$ converge lorsque m tend vers l'infini vers une quantité non nulle. Montrons que cette quantité ne peut être que $d\pi(y)$. Posons $\ell(x,y) = \lim_{m\to+\infty} P^{md+r}(x,y)$. D'après le théorème 3.9, on a

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} P^k(x,y) = \pi(y).$$

Or tout entier k s'écrit d'une et d'une seule façon sous la forme $k = md + q$ avec $m \in \mathbb{N}$ et $0 \leq q < d$ (division euclidienne par d). Par conséquent :

$$\begin{aligned}
\frac{1}{n}\sum_{k=1}^{n} P^k(x,y) &= \sum_{q=0}^{d-1} \frac{1}{n} \sum_{\substack{m: \\ 1\leq md+q\leq n}} P^{md+q}(x;y) \\
&= \sum_{q=0}^{d-1} \frac{[(n-q)/d]}{n} \frac{1}{[(n-q)/d]} \sum_{m=1}^{[(n-q)/d]} P^{md+q}(x,y).
\end{aligned}$$

Pour $q \neq r$, $P^{md+q}(x,y)$ tend vers 0 quand m tend vers l'infini, donc sa moyenne de Césaro aussi et pour $q = r$ il y a convergence de $P^{md+q}(x,y)$, et donc de sa moyenne de Césaro vers $\ell(x,y)$.

D'autre part,

$$\frac{n-q}{d} - 1 < \left[\frac{n-q}{d}\right] \leq \frac{n-q}{d},$$

donc

$$\frac{n-q}{nd} - \frac{1}{n} < \frac{\left[\frac{n-q}{d}\right]}{n} \leq \frac{n-q}{nd},$$

et par suite $\frac{[(n-q)/d]}{n}$ tend vers $1/d$ lorsque n tend vers l'infini.

Finalement on obtient $\pi(y) = \frac{\ell(x,y)}{d}$, donc $\ell(x,y) = d\,\pi(y)$.

Exemple : Pour une chaîne de naissance et mort irréductible, récurrente positive et périodique de période 2 (c'est-à-dire $r_x = 0, p_x q_x > 0$), on a :

– si $y - x$ est pair,

$$\lim_{m\to+\infty} P^{2m}(x,y) = 2\,\pi(y),$$

$$\lim_{m\to+\infty} P^{2m+1}(x,y) = 0,$$

– si $y - x$ est impair,

$$\lim_{m\to+\infty} P^{2m+1}(x,y) = 2\,\pi(y),$$

$$\lim_{m \to +\infty} P^{2m}(x, y) = 0.$$

Nous allons démontrer une partie du théorème d'Orey dans le cas particulier d'une chaîne de Markov irréductible et apériodique sur un espace fini qui admet une probabilité réversible.

Proposition 3.18 *Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov irréductible sur un espace fini de dimension d qui admet une probabilité réversible π. On suppose qu'il existe x dans E tel que $P(x, x) > 0$. Alors si $\lambda_1 = 1, \lambda_2, \ldots, \lambda_d$ sont les valeurs propres de la matrice P, on a :*

$$\alpha = \sup_{2 \le k \le d} |\lambda_k| < 1$$

et il existe une constant K telle que pour tous x et y

$$|P^n(x, y) - \pi(y)| \le K\ \alpha^n$$

Démonstration : Soit la matrice $D = diag((\sqrt{\pi(x)}), x \in E)$. Comme π est réversible la matrice DPD^{-1} est symétrique, donc toutes ses valeurs propres sont réelles. Mais les valeurs propres de la matrice DPD^{-1} sont les mêmes que celles de la matrice P car ces deux matrice ont le même polynôme caractéristique. Donc toutes les valeurs propres de P sont réelles. Les formes quadratiques de matrices $I - DPD^{-1}$ et $I + DPD^{-1}$ sont positives. En effet pour tout vecteur u :

$$u^t(I - DPD^{-1})u = \frac{1}{2} \sum_{x, x'} \pi(x) P(x, x') \left(\frac{u(x)}{\sqrt{\pi(x)}} - \frac{u(x')}{\sqrt{\pi(x')}} \right)^2$$

$$u^t(I + DPD^{-1})u = \frac{1}{2} \sum_{x, x'} \pi(x) P(x, x') \left(\frac{u(x)}{\sqrt{\pi(x)}} + \frac{u(x')}{\sqrt{\pi(x')}} \right)^2$$

Donc les valeurs propres des matrices $I - DPD^{-1}$ et $I + DPD^{-1}$ sont positives ou nulles. On en déduit que les valeurs propres de la matrice DPD^{-1} sont comprises entre -1 et $+1$.

Si P admet la valeur propre -1, alors $I + DPD^{-1}$ admet la valeur propre 0, c'est à dire qu'il existe un vecteur u non nul tel que $u^t(I + DPD^{-1})u$ soit nul. D'après l'expression ci-dessus, pour un x tel que $P(x, x) > 0$, on a $u(x) = 0$. Mais alors si $P(x, x') > 0$ on a aussi $u(x') = 0$. Et du fait de l'irréductibilité, le vecteur u est nul. Donc la matrice P ne peut pas admettre -1 comme valeur propre.

La matrice markovienne P admet 1 comme valeur propre. Cela correspond à la valeur propre 0 pour la matrice $I - DPD^{-1}$. Un vecteur propre associé vérifie nécessairement

$$\frac{u(x)}{\sqrt{\pi(x)}} = \frac{u(x')}{\sqrt{\pi(x')}}$$

si $P(x, x') > 0$. Le sous espace propre est donc de dimension 1 puisque la chaîne est irréductible. Donc la valeur propre 1 est simple.

Donc la valeur propre 1 de P est simple et toutes les autres valeurs propres sont réelles et strictement comprises entre -1 et $+1$. Si on note $(\lambda_l), l = 1, \ldots, d$ les valeurs propres de P rangées par ordre décroissant, on a :

$$\lambda_1 = 1 > \lambda_2 \geq \ldots \geq \lambda_d > -1$$

La matrice symétrique DPD^{-1} est diagonalisable dans une base orthonormée $((\Phi_l), l = 1 \ldots, d)$ où Φ_l est un vecteur propre associé à la valeur propre λ_l. Pour Φ_1 on choisit $\Phi_1 = (\sqrt{\pi(x)})_{x \in E}$. On a alors pour tout couple de vecteurs u et v :

$$u^t D P^n D^{-1} v = \sum_{l=1}^{d} u^t \Phi_l \; v^t \Phi_l \; \lambda_l^n$$

Si on applique la formule précédente aux deux vecteurs $u = \frac{1}{\sqrt{\pi(x)}} 1_{\{x\}}$ et $v = \sqrt{\pi(y)} 1_{\{y\}}$, on obtient :

$$P^n(x, y) = \pi(y) + \frac{\sqrt{\pi(y)}}{\sqrt{\pi(x)}} \sum_{l=2}^{d} \Phi_l(x) \; \Phi_l(y) \lambda_l^n$$

On constate que $P^n(x, y)$ converge vers $\pi(y)$ exponentiellement vite. Si $\alpha = \sup_{2 \leq l \leq d} |\lambda_l|$, cette convergence se fait à la vitesse α^n. ■

Chapitre 4

Processus de Poisson

La suite des temps de passage de véhicules en un point donné d'un autoroute lorsque le traffic est fluide, les temps d'arrivée de clients à un guichet, les temps d'émission de photons d'un élément radioactif sont des suites de temps aléatoires qui ont les mêmes propriétés probabilistes : les temps d'interarrivée sont des variables aléatoires indépendantes de même loi exponentielle. On donne à ce modèle, rencontré fréquemment en pratique, le nom de processus de Poisson, car la loi de Poisson joue un rôle important dans les caractéristiques de ce processus.

Le nom de Poisson vient de Siméon Denis Poisson (21 juin 1781 à Pithiviers - 25 avril 1840 à Sceaux). Son œuvre est immense et touche à beaucoup de branches des mathématiques et de la physique, en particulier au calcul des probabilités.

4.1 Définition et propriétés

Définition 4.1 *La suite $(T_n)_{n\geq 1}$ est appelée un processus de Poisson homogène de paramètre λ ($\lambda > 0$) si les variables aléatoires $T_1, T_2 - T_1, \ldots, T_{n+1} - T_n, \ldots$ sont indépendantes et de même loi exponentielle de paramètre λ.*

On définit la fonction de comptage N associée au processus de Poisson par :

$$N(t) = \sum_{n\geq 1} 1_{\{T_n \leq t\}}.$$

pour $t \geq 0$.

La fonction $t \to N(t)$ est une fonction aléatoire croissante de $\mathbb{R}_+$ dans $\mathbb{N}$, constante sur chaque intervalle $[T_n, T_{n+1}[$. Il est clair qu'il est équivalent de se donner la suite $(T_n)_{n\in\mathbb{N}}$ ou la fonction de comptage N.

Proposition 4.2 *Soit $(T_n)_{n\geq 1}$, un processus de Poisson homogène de paramètre λ, et de fonction de comptage N. Alors :*

1. *pour tout $n \geq 1$, la loi du vecteur $(T_1, \dots T_n)$ a pour densité*
$$\lambda^n e^{-\lambda t_n} 1_{\{0<t_1<\dots<t_n\}}$$
par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$,
2. *pour tout $t > 0$, la variable aléatoire $N(t)$ est de loi de Poisson de paramètre λt,*
3. *pour tout $n > 0$, la loi conditionnelle de $(T_1, \dots, T_n)$ sachant l'événement $\{N(t) = n\}$ a pour densité*
$$\frac{n!}{t^n} 1_{\{0<t_1<\dots<t_n\leq t\}}$$
par rapport à la mesure de Lebesgue.

Démonstration : L'assertion 1 se montre facilement par changement de variables à partir de la loi de $(T_1, T_2 - T_1, \dots, T_n - T_{n-1})$.

Par suite, étant donnée une fonction φ borélienne positive définie sur $\mathbb{R}^n$, nous avons :

$$\begin{aligned}
&\mathbb{E}(\,\varphi(T_1, \dots, T_n) 1_{\{N(t)=n\}}\,) \\
&\quad = \mathbb{E}(\,\varphi(T_1, \dots, T_n) 1_{\{T_n\leq t<T_{n+1}\}}\,) \\
&\quad = \int_{\mathbb{R}^{n+1}} \varphi(t_1, \dots, t_n) \lambda^{n+1} e^{-\lambda t_{n+1}} 1_{\{0<t_1<\dots<t_n\leq t<t_{n+1}\}} dt_1 \cdots dt_{n+1} \\
&\quad = \int_{\mathbb{R}^n} \varphi(t_1, \dots, t_n) \lambda^n \, 1_{\{0<t_1<\dots<t_n\leq t\}} \left(\int_t^{+\infty} \lambda e^{-\lambda t_{n+1}} dt_{n+1} \right) dt_1 \cdots dt_n \\
&\quad = \int_{\mathbb{R}^n} \varphi(t_1, \dots, t_n) \lambda^n e^{-\lambda t} 1_{\{0<t_1<\dots<t_n\leq t\}} \, dt_1 \cdots dt_n.
\end{aligned}$$

En prenant φ identiquement égale à 1 et en utilisant le fait que :

$$\int_{\mathbb{R}^n} 1_{\{0<t_1<\dots<t_n\leq t\}} \, dt_1 \cdots dt_n = \frac{t^n}{n!},$$

nous trouvons que :

$$\mathbb{P}(\,N(t) = n\,) = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \ .$$

Nous avons donc :

$$\begin{aligned}
&\mathbb{E}(\,\varphi(T_1,\ldots,T_n)/\,N(t)=n) \\
&= \frac{1}{\mathbb{P}(\,N(t)=n\,)}\mathbb{E}(\,\varphi(T_1,\ldots,T_n)1_{\{T_n\le t<T_{n+1}\}}\,) \\
&= \int_{\mathbb{R}^n}\varphi(t_1,\ldots,t_n)\frac{n!}{t^n}1_{\{0<t_1<\ldots<t_n\le t\}}\,dt_1\cdots dt_n,
\end{aligned}$$

ce qui est bien la formule annoncée. ∎

Le dernier alinéa de la proposition dit que la loi de $(T_1,\ldots,T_n)$ sachant $\{N_t=n\}$ est la loi de la statistique d'ordre de n variables aléatoires indépendantes de loi uniforme sur $[0,t]$. La proposition suivante donne ce résultat.

Proposition 4.3 *Soient $X_1,\ldots,X_n$ n variables aléatoires réelles indépendantes de même loi de densité f par rapport à la mesure de Lebesgue. Notons $(X_{(1)},\ldots,X_{(n)})$ la statistique d'ordre associée, c'est-à-dire les variables aléatoires $X_1,\ldots,X_n$ rangées par ordre croissant.*

Alors la loi de $(X_{(1)},\ldots,X_{(n)})$ est de densité

$$n!\left(\prod_{i=1}^n f(x_i)\right)1_{\{x_1<x_2<\cdots<x_n\}}$$

par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$.

Démonstration : Soit φ une fonction borélienne positive définie sur $\mathbb{R}^n$ et Σ l'ensemble des permutations de $\{1,\ldots,n\}$. Les variables aléatoires X_i étant indépendantes et de loi diffuse, elles sont presque-sûrement deux à deux distinctes. Par suite, pour presque tout ω, il existe une permutation σ (qui dépend de ω) appartenant à Σ et telle que :

$$X_{(1)}(\omega)=X_{\sigma(1)}(\omega),\ldots,X_{(n)}(\omega)=X_{\sigma(n)}(\omega).$$

Il s'ensuit que :

$$\begin{aligned}
&\mathbb{E}(\,\varphi(X_{(1)},\ldots,X_{(n)}\,)) \\
&= \sum_{\sigma\in\Sigma}\mathbb{E}(\,\varphi(X_{(1)},\ldots,X_{(n)})1_{\{X_{\sigma(1)}<\ldots<X_{\sigma(n)}\}}\,) \\
&= \sum_{\sigma\in\Sigma}\mathbb{E}(\,\varphi(X_{\sigma(1)},\ldots,X_{\sigma(n)})1_{\{X_{\sigma(1)}<\ldots<X_{\sigma(n)}\}}\,) \\
&= \sum_{\sigma\in\Sigma}\int_{\mathbb{R}^n}\varphi(y_1,\ldots,y_n)1_{\{y_1<\ldots<y_n\}}\,f(y_1)\cdots f(y_n)\,dy_1\ldots dh_n \\
&= n!\int_{\mathbb{R}^n}\varphi(y_1,\ldots,y_n)1_{\{y_1<\ldots<y_n\}}\,f(y_1)\cdots f(y_n)\,dy_1\ldots dy_n,
\end{aligned}$$

ce qui donne bien la formule annoncée. ∎

Théorème 4.4 *Soit $(T_n)_{n\geq 1}$ une suite croissante de variables aléatoires à valeurs dans $\mathbb{R}_+$ et N la fonction de comptage associée :*

$$N(t) = \sum_{n\geq 1} 1_{\{T_n \leq t\}}.$$

C'est un processus de Poisson de paramètre λ si et seulement si les conditions suivantes sont vérifiées :

1. *N(0) = 0,*
2. *pour $0 \leq s < t$, la variable aléatoire $N(t) - N(s)$ est de loi de Poisson de paramètre $\lambda(t-s)$,*
3. *le processus N est à accroissements indépendants, c'est-à-dire que, pour tout n et $0 \leq t_1 < t_2 < \ldots < t_n$, les variables aléatoires*

$$N(t_1), N(t_2) - N(t_1), \ldots, N(t_n) - N(t_{n-1})$$

sont indépendantes.

Démonstration : Supposons que $(T_n)_{n\geq 1}$ soit un processus de Poisson homogène de paramètre λ. Donnons-nous une famille strictement croissante de p réels positifs $0 \leq t_1 < t_2 < \ldots < t_p$ et une famille de p entiers positifs $k_1, \ldots, k_p$. Posons $k = k_1 + \cdots + k_p$. Enfin soit $(X_{(1)}, \ldots, X_{(k)})$ la statistique d'ordre associée à k variables aléatoires indépendantes $X_1, \ldots X_k$ de même loi uniforme sur $[0, t_p]$. La proposition 4.2 entraine que :

$$\mathbb{P}(\, N(t_1) = k_1, N(t_2) - N(t_1) = k_2, \ldots, N(t_p) - N(t_{p-1}) = k_p \,)$$

$$\begin{aligned}
&= \mathbb{P}(\sum_{n=1}^{k} 1_{\{T_n \leq t_1\}} = k_1, \ldots, \sum_{n=1}^{k} 1_{\{t_{p-1} < T_n \leq t_p\}} = k_p \,/ N(t_p) = k\,) \\
&\quad \times \mathbb{P}(\, N(t_p) = k\,) \\
&= \mathbb{P}(\sum_{n=1}^{k} 1_{\{X_{(n)} \leq t_1\}} = k_1, \ldots, \sum_{n=1}^{k} 1_{\{t_{p-1} < X_{(n)} \leq t_p\}} = k_p\,)\, \mathbb{P}(\, N(t_p) = k\,) \\
&= \mathbb{P}(\sum_{n=1}^{k} 1_{\{X_n \leq t_1\}} = k_1, \ldots, \sum_{n=1}^{k} 1_{\{t_{p-1} < X_n \leq t_p\}} = k_p\,)\, \mathbb{P}(\, N(t_p) = k\,) \\
&= C_k^{k_1} C_{k-k_1}^{k_2} \ldots C_{k_p}^{k_p} \left(\frac{t_1}{t_p}\right)^{k_1} \left(\frac{t_2 - t_1}{t_p}\right)^{k_2} \ldots \left(\frac{t_p - t_{p-1}}{t_p}\right)^{k_p} e^{-\lambda t_p} \frac{(\lambda t_p)^k}{k!} \\
&= \frac{k!}{k_1! k_2! \ldots k_p!} \frac{t_1^{k_1} (t_2 - t_1)^{k_2} \ldots (t_p - t_{p-1})^{k_p}}{t_p^{k_1 + \cdots + k_p}} e^{-\lambda t_p} \frac{(\lambda t_p)^k}{k!}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{k_1!k_2!\dots k_p!} t_1^{k_1}(t_2-t_1)^{k_2}\dots(t_p-t_{p-1})^{k_p}\lambda^{k_1+\dots+k_p}e^{-\lambda(t_1+(t_2-t_1)+\dots(t_p-t_{p-1}))} \\
&= e^{-\lambda t_1}\frac{(\lambda t_1)^{k_1}}{k_1!}e^{-\lambda(t_2-t_1)}\frac{(\lambda(t_2-t_1))^{k_2}}{k_2!}\dots e^{-\lambda(t_n-t_{n-1})}\frac{(\lambda(t_n-t_{n-1}))^{k_n}}{k_n!}.
\end{aligned}
$$

Le calcul ci-dessus est écrit dans le cas où aucun des k_i n'est nul. Dans le cas contraire, il faut effectuer quelques modifications évidentes mais le résultat final reste valable. Nous avons donc démontré que les variables aléatoires $N(t_1)$, $N(t_2)-N(t_1),\dots,N(t_n)-N(t_{n-1})$ sont indépendantes et de loi de Poisson de paramètres respectifs $\lambda t_1, \lambda(t_2-t_1),\dots,\lambda(t_n-t_{n-1})$.

Réciproquement, supposons que $(T_n)_{n\geq 1}$ soit une suite croissante de variables aléatoires sur $\mathbb{R}_+$ dont la fonction de comptage N vérifie les conditions *1*, *2* et *3*, Pour tout $n\geq 1$, nous allons montrer que la loi de $(T_1,\dots,T_N)$ admet $\lambda^n e^{-\lambda t_n}1_{\{0<t_1<\dots<t_n\}}$ comme densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$, ce qui démontre le résultat (un simple changement de variable amène à la loi de $(T_1,T_2-T_1,\dots,T_n-T_{n-1})$).

Soient donc $n\geq 1$ et f fonction de $\mathbb{R}_+^n$ dans $\mathbb{R}_+$ continue à support compact $[0,K]^n$. Soit $r_0=0<r_1<\dots<r_{N-1}<r_N=K$ une subdivision de $[0,K]$ telle que pour tout $1\leq i\leq N$ on ait $r_i-r_{i-1}=\delta=K/N$.

On peut écrire :

$$
\begin{aligned}
&\mathbb{E}\left(f(T_1,\dots,T_n)\right) \\
&\quad = \sum_{\substack{(i_1,\dots,i_n): \\ 1\leq i_1<\dots i_n\leq N}} \mathbb{E}\left(f(T_1,\dots,T_n)\prod_{j=1}^n 1_{\{T_j\in[r_{i_j-1},r_{i_j}[\}}\right)+R_1
\end{aligned}
$$

et

$$
\begin{aligned}
R_1 &\leq \|f\|\sum_{i=1}^N \mathbb{P}\left(N(r_i)-N(r_{i-1})\geq 2\right) \\
&\leq \|f\|\,N\left(1-e^{-\lambda K/N}-e^{-\lambda K/N}\lambda\frac{K}{N}\right) \\
&\to 0 \quad si \quad N\to\infty
\end{aligned}
$$

Comme la fonction f est uniformément continue, pour tout $\epsilon>0$, si N est assez grand, si $\sup_{1\leq i\leq n}|x_i-y_i|\leq K/N$ alors $|f(x_1,\dots,x_n)-f(y_1,\dots,y_n)|\leq\epsilon$. On a alors :

$$
\begin{aligned}
&| \sum_{\substack{(i_1,\dots,i_n): \\ 1\leq r_1<\dots<r_n\leq N}} \mathbb{E}\left(f(T_1,\dots,T_n)\prod_{j=1}^n 1_{\{T_j\in[r_{i_j-1},r_{i_j}[\}}\right) \\
&- \sum_{\substack{(i_1,\dots,i_n): \\ 1\leq r_1<\dots<r_n\leq N}} \mathbb{E}\left(f(r_{i_1},\dots,r_{i_n})\prod_{j=1}^n 1_{\{T_j\in[r_{i_j-1},r_{i_j}[\}}\right)|\leq\epsilon
\end{aligned}
$$

Avec les hypothèses faites, nous avons :

$$\sum_{\substack{(i_1,\ldots,i_n):\\ 1\le r_1<\ldots<r_n\le N}} \mathbb{E}\left(f(r_{i_1},\ldots,r_{i_n}) \prod_{j=1}^{n} 1_{\{T_j\in[r_{i_j-1},r_{i_j}[\}} \right)$$

$$= \sum_{\substack{(i_1,\ldots,i_n):\\ 1\le r_1<\ldots<r_n\le N}} f(r_{i_1},\ldots,r_{i_n})\mathbb{P}(N(r_{i_1-1})=0, N(r_{i_1})-N(r_{i_1-1})=1,$$
$$N(r_{i_2-1})-N(r_{i_1})=0,\ldots,N(r_{i_n})-N(r_{i_n-1})=1)$$
$$= \sum_{\substack{(i_1,\ldots,i_n):\\ 1\le r_1<\ldots<r_n\le N}} f(r_{i_1},\ldots,r_{i_n})e^{-\lambda r_{i_n}}\ \lambda^n\ \delta^n$$
$$= \sum_{\substack{(i_1,\ldots,i_n):\\ 1\le r_1<\ldots<r_n\le N}} \int_{\mathbb{R}_+^n} f(r_{i_1},\ldots,r_{i_n})\ e^{-\lambda r_{i_n}}\ \lambda^n \prod_{j=1}^{n} 1_{\{t_j\in[r_{i_j-1},r_{i_j}[\}}\ dt_1\ldots dt_n$$

Grâce à la continuité de la fonction f, si N est assez grand, on a :

$$| \sum_{\substack{(i_1,\ldots,i_n):\\ 1\le r_1<\ldots<r_n\le N}} \int_{\mathbb{R}_+^n} f(r_{i_1},\ldots,r_{i_n})\ e^{-\lambda r_{i_n}}\ \lambda^n \prod_{j=1}^{n} 1_{\{t_j\in[r_{i_j-1},r_{i_j}[\}}\ dt_1\ldots dt_n$$
$$- \sum_{\substack{(i_1,\ldots,i_n):\\ 1\le r_1<\ldots<r_n\le N}} \int_{\mathbb{R}_+^n} f(t_1,\ldots,t_n)\ e^{-\lambda t_n}\ \lambda^n \prod_{j=1}^{n} 1_{\{t_j\in[r_{i_j-1},r_{i_j}[\}}\ dt_1\ldots dt_n| \le \epsilon$$

Or

$$\sum_{\substack{(i_1,\ldots,i_n):\\ 1\le r_1<\ldots<r_n\le N}} \int_{\mathbb{R}_+^n} f(t_1,\ldots,t_n)\ e^{-\lambda t_n}\ \lambda^n \prod_{j=1}^{n} 1_{\{t_j\in[r_{i_j-1},r_{i_j}[\}}\ dt_1\ldots dt_n$$

$$= \int_{\mathbb{R}_+^n} f(t_1,\ldots,t_n)\ e^{-\lambda t_n}\ \lambda^n 1_{A_N} dt_1\ldots dt_n$$

avec $A_N = \cup_{\{(i_1,\ldots,i_n):1\le r_1<\ldots<r_n\le N\}} \cap_{\{1\le j\le n\}} \{t_j \in [r_{i_j-1}, r_{i_j}[\}$.

Comme le sous-ensemble A_N de $\mathbb{R}_+^n$ tend en croissant vers le sous-ensemble $\{0 < t_1 < \ldots < t_n\}$ lorsque N tend vers l'infini, nous obtenons :

$$\lim_{N\to\infty} \int_{\mathbb{R}_+^n} f(t_1,\ldots,t_n)\ e^{-\lambda t_n}\ \lambda^n 1_{A_N} dt_1\ldots dt_n$$
$$= \int_{\mathbb{R}_+^n} f(t_1,\ldots,t_n)\ e^{-\lambda t_n}\ \lambda^n 1_{\{0<t_1<\ldots<t_n\}} dt_1\ldots dt_n$$

Tout ceci prouve que

$$\mathbb{E}(f(T_1,\ldots,T_n) = \int_{\mathbb{R}_+^n} f(t_1,\ldots,t_n)\; e^{-\lambda t_n}\; \lambda^n 1_{\{0<t_1<\ldots<t_n\}} dt_1\ldots dt_n$$

ce qui achève la démonstration. ■

Nous allons démontrer une propriété fondamentale du processus de Poisson : pour tout t, le processus des points arrivés après t est également un processus de Poisson de même paramètre. Cette propriété est une propriété de stationnarité.

Soit t un temps fixé. Le premier point du processus de Poisson qui arrive après t est $T_{N(t)+1}$, donc le temps qui s'écoule entre t et la premère arrivée vaut $W(t) = T_{N(t)+1} - t$. Il est facile d'identifier la loi de $W(t)$ car pour tout $s \geq 0$ on a

$$\mathbb{P}(W(t) > s) = \mathbb{P}\left(N(t+s) - N(t) = 0\right) = e^{-\lambda\, s},$$

ce qui prouve que $W(t)$ suit une loi exponentielle de paramètre λ.

Considérons maintenant le processus $(S_n)_{n\geq 1}$ des points arrivés après t :

$$(S_1, S_2, \ldots, S_n, \ldots) = (T_{N(t)+1} - t, T_{N(t)+2} - t, \ldots, T_{N(t)+n} - t, \ldots).$$

Pour tout $n \geq 1$ on a $S_n = T_{N(t)+n} - t$.

Proposition 4.5 *Soit $(T_n)_{n\geq 1}$ un processus de Poisson de paramètre λ. Alors pour tout t, le processus $(S_n)_{n\geq 1} = (T_{N(t)+n} - t)_{n\geq 1}$ est un processus de Poisson de paramètre λ.*

Démonstration : Nous allons utiliser la proposition 4.4. Appelons $\tilde{N}(s)$ la fonction de comptage associée à la suite $(S_n)_{n\geq 1}$. On a $\tilde{N}(0) = 0$. Soient $0 \leq s_1 < s_2$. On a $\tilde{N}(s_2) - \tilde{N}(s_1) = N(t+s_2) - N(t+s_1)$, donc cette variable aléatoire suit une loi de Poisson de parmètre $s_2 - s_1$. De même si $0 \leq s_1 < s_2 < \ldots < s_n$, on a

$$(\tilde{N}(s_1), \tilde{N}(s_2) - \tilde{N}(s_1), \ldots, \tilde{N}(s_n) - \tilde{N}(s_{n-1}) =$$

$$(N(t+s_1) - N(t), (N(t+s_2) - N(t+s_1), \ldots, (N(t+s_n) - N(t+s_{n-1}))$$

et l'indépendance de ces variables aléatoires est claire. Ceci prouve la proposition. ■

4.2 Résultats asymptotiques

Le comportement asymptotique de $N(t)$ lorsque t tend vers l'infini est décrit par la convergence presque sûre de $N(t)/t$ vers λ et la convergence en loi de $\sqrt{t}\left(\frac{N(t)}{t}-\lambda\right)$ vers une loi de Gauss. Nous allons prouver ces propriétés dans le théorème suivant.

Proposition 4.6 *Soit N la fonction de comptage d'un processus de Poisson homogène de paramètre λ, alors :*

1. *les variables aléatoires $\frac{N(t)}{t}$ convergent presque-sûrement vers λ lorsque t tend vers $+\infty$,*
2. *les variables aléatoires $\sqrt{t}\,(\frac{N(t)}{t}-\lambda)$ convergent en loi vers une variable aléatoire gaussienne centrée de variance λ.*

Démonstration :

1) Pour t donné, nous avons :

$$T_{N(t)} \leq t < T_{N(t)+1},$$

ce qui entraine :

$$\frac{N(t)}{T_{N(t)+1}} \leq \frac{N(t)}{t} < \frac{N(t)}{T_{N(t)}} .$$

Or $N(t)$ tend presque-sûrement vers $+\infty$ quand t tend vers l'infini, et la loi des grands nombres entraine la convergence presque-sûre de T_k/k vers $\mathbb{E}(T_1)=1/\lambda$ lorsque k tend vers l'infini, d'où le résultat.

2) L'assertion 2 n'est autre que la convergence de la loi de Poisson vers la loi gaussienne. Nous en rappelons brièvement la démonstration. Notons ψ_t la fonction caractéristique de la variable aléatoire $\sqrt{t}(\frac{N(t)}{t}-\lambda)$, alors :

$$\psi_t(u)=\mathbb{E}\left(\exp(iu\sqrt{t}(\frac{N(t)}{t}-\lambda))\right)=\exp\left(-iu\lambda\sqrt{t}-\lambda t\,(1-\exp(\frac{iu}{\sqrt{t}}))\right).$$

Or :

$$iu\lambda\sqrt{t}+\lambda t\,(1-\exp(\frac{iu}{\sqrt{t}}))=\frac{\lambda u^2}{2}+\varepsilon(\frac{1}{t}) \quad \text{avec} \quad \lim_{s\to 0}\varepsilon(s)=0,$$

donc :

$$\lim_{t\to+\infty}\psi_t(u)=\psi(u)=\exp(-\frac{\lambda u^2}{2}),$$

et la fonction ψ est la fonction caractéristique d'une loi gaussienne centrée de variance λ. ■

4.3 Transformée de Laplace

Il est possible de caractériser un processus de Poisson à l'aide d'une fonctionnelle, appelée transformé de Laplace.

Proposition 4.7 *La suite $(T_n)_{n\geq 1}$ est un processus de Poisson de paramètre λ si et seulement si, pour toute fonction f mesurable positive*

$$\mathbb{E}\left(e^{-\sum_{n\geq 1} f(T_n)}\right) = e^{-\lambda \int_0^\infty (1-e^{-f(x)})\, dx}.$$

Démonstration : Soit $(T_n)_{n\geq 1}$ un processus de Poisson de paramètre λ. Supposons d'abord que f est à support compact et soit A tel que f soit nulle sur $]A, +\infty[$. Nous reprenons la même technique que dans la démonstration du théorème 4.4. Soit $X_{(1)}^{(k)}, \ldots X_{(k)}^{(k)}$ la statistique d'ordre associée à k variables aléatoires $X_1, \ldots, X_k$ indépendantes de même loi uniforme sur $[0, A]$. Nous avons :

$$\begin{aligned}
\mathbb{E}\left(e^{-\sum_n f(T_n)}\right) &= \sum_{k\geq 0} \mathbb{E}\left(e^{-\sum_n f(T_n)} / \, N(A) = k\right) \mathbb{P}(N(A) = k) \\
&= \sum_{k\geq 0} \mathbb{E}\left(e^{-\sum_{n=1}^k f(T_n)} / \, N(A) = k\right) \mathbb{P}(N(A) = k) \\
&= \sum_{k\geq 0} \mathbb{E}\left(e^{-\sum_{n=1}^k f(X_{(n)}^{(k)})}\right) \mathbb{P}(N(A) = k) \\
&= \sum_{k\geq 0} \mathbb{E}\left(e^{-\sum_{n=1}^k f(X_n)}\right) \mathbb{P}(N(A) = k) \\
&= \sum_{k\geq 0} \mathbb{E}\left(\prod_{n=1}^k e^{-f(X_n)}\right) \mathbb{P}(N(A) = k) \\
&= \sum_{k\geq 0} \prod_{n=1}^k \mathbb{E}\left(e^{-f(X_n)}\right) \mathbb{P}(N(A) = k) \\
&= \sum_{k\geq 0} \left(\int_0^A e^{-f(x)} \frac{1}{A}\, dx\right)^k e^{-\lambda A} \frac{(\lambda A)^k}{k!} \\
&= e^{-\lambda A} \sum_{k\geq 0} \frac{1}{k!} \left(\lambda \int_0^A e^{-f(x)}\, dx\right)^k \\
&= e^{-\lambda A}\, e^{\lambda \int_0^A e^{-f(x)}\, dx} \\
&= e^{-\lambda \int_0^A (1-e^{-f(x)})\, dx}
\end{aligned}$$

Puisque f est nulle sur $]A, +\infty[$, on a $\int_A^{+\infty}(1 - e^{-f(x)})\, dx = 0$, et par suite

$$\int_0^A (1 - e^{-f(x)})\, dx = \int_0^{+\infty} (1 - e^{-f(x)})\, dx.$$

La formule est donc établie lorsque f est à support dans $[0, A]$. Dans le cas général, on applique la formule à $f1_{[0,n]}$, on fait tendre n vers l'infini et on applique la propriété de Beppo-Lévy.

Réciproquement, supposons que

$$\mathbb{E}\left(e^{-\sum_{n\geq 1} f(T_n)}\right) = e^{-\lambda \int_0^\infty (1-e^{-f(x)})\, dx}$$

pour toute fonction f borléienne positive. Soit $0 \leq t_1 < t_2 < \ldots < t_p$ et $s_1, \ldots s_p$ des éléments de $\mathbb{R}_+$. Prenons

$$f = s_1 1_{[0,t_1]} + s_2 1_{]t_1,t_2]} + \cdots + s_p 1_{]t_{p-1},t_p]}.$$

Nous avons

$$\begin{aligned}\sum_n f(T_n) &= s_1 \sum_n 1_{\{T_n \leq t_1\}} + s_2 \sum_n 1_{\{t_1 < T_n \leq t_2\}} + \cdot + s_p \sum_n 1_{\{t_{p-1} < T_n \leq t_p\}} \\ &= s_1 N(t_1) + s_2(N(t_2) - N(t_1)) + \cdots + s_p(N(t_p) - N(t_{p-1}))\end{aligned}$$

et

$$\begin{aligned}&\int_0^\infty (1 - e^{-f(x)})\, dx \\ &= \int_0^{t_1} (1 - e^{-s_1})\, dx + \int_{t_1}^{t_2} (1 - e^{-s_2})\, dx + \cdots + \int_{t_{p-1}}^{t_p} (1 - e^{-s_p})\, dx \\ &= t_1(1 - e^{-s_1}) + (t_2 - t_1)(1 - e^{-s_2}) + (t_p - t_{p-1})(1 - e^{-s_p}).\end{aligned}$$

On obtient donc

$$\begin{aligned}&\mathbb{E}\left(e^{-s_1 N(t_1) - s_2(N(t_2) - N(t_1)) - \cdots - s_p(N(t_p) - N(t_{p-1}))}\right) \\ &= e^{-\lambda t_1(1-e^{-s_1})}\, e^{-\lambda(t_2 - t_1)(1-e^{-s_2})} \ldots e^{-\lambda(t_p - t_{p-1})(1-e^{-s_p})}.\end{aligned}$$

En prenant $s_i = 0$ pour $i \neq k$ et $s_k = s$, on obtient

$$\mathbb{E}(e^{-s\,(N(t_1))}) = e^{-\lambda t_1(1-e^{-s})}$$

$$\mathbb{E}(e^{-s\,(N(t_k) - N(t_{k-1}))}) = e^{-\lambda(t_k - t_{k-1})(1-e^{-s})},$$

et on reconnait la transformée de Laplace de la loi de Poisson de paramètre $\lambda(t_k - t_{k-1})$. Ou bien on pose $u = e^{-s}$ et on reconnait la fonction génératrice de cette loi. D'où :

$$\begin{aligned}
&\mathbb{E}\left(e^{-s_1 N(t_1) - s_2(N(t_2)-N(t_1)) - \cdots - s_p(N(t_p)-N(t_{p-1}))}\right) \\
&\quad = \mathbb{E}(e^{-s_1 N(t_1)})\,\mathbb{E}(e^{-s_2(N(t_2)-N(t_1))}) \ldots \mathbb{E}(e^{-s_p(N(t_p)-N(t_{p-1}))}).
\end{aligned}$$

Cette forme produit donne l'indépendance des variables aléatoires. Et le théorème 4.4 permet de conclure qu'on a bien un processus de Poisson de paramètre λ. ■

La fonctionnelle qui à toute fonction mesurable positive f associe la quantité $\mathbb{E}\left(e^{-\sum_{n\geq 1} f(T_n)}\right)$ s'appelle la transformée de Laplace du processus de Poisson. On vient de voir que cette fonctionnelle caractérise ce processus. Cet outil est très utile pour démontrer que des suites de variables aléatoires sont des processus de Poisson, comme par exemple dans la proposition suivante.

Proposition 4.8 *La superposition de deux processus de Poisson indépendants de paramètres respectifs* λ_1 *et* λ_2 *est un processus de Poisson de paramètre* $\lambda_1 + \lambda_2$.

Démonstration : Soient $(T_n^{(1)})_{n\in\mathbb{N}}$ et $(T_n^{(2)})_{n\in\mathbb{N}}$ deux processus de Poisson indépendants de paramètres λ_1 et λ_2. En superposant ces deux suites, on obtient une nouvelle suite $(S_n)_{n\in\mathbb{N}}$. On peut écrire pour toute fonction f mesurable positive :

$$\begin{aligned}
\mathbb{E}\left(e^{-\sum_{n\geq 1} f(S_n)}\right) &= \mathbb{E}\left(e^{-\sum_{n\geq 1}\left(f(T_n^{(1)}) + f(T_n^{(2)})\right)}\right) \\
&= \mathbb{E}\left(e^{-\sum_{n\geq 1} f(T_n^{(1)})}\, e^{-\sum_{n\geq 1} f(T_n^{(2)})}\right) \\
&= \mathbb{E}\left(e^{-\sum_{n\geq 1} f(T_n^{(1)})}\right)\, \mathbb{E}\left(e^{-\sum_{n\geq 1} f(T_n^{(2)})}\right) \\
&= e^{-\lambda_1 \int (1-e^{-f(x)})\,dx}\, e^{-\lambda_2 \int (1-e^{-f(x)})\,dx} \\
&= e^{-(\lambda_1+\lambda_2) \int (1-e^{-f(x)})\,dx}
\end{aligned}$$

Le processus $(S_n)_{n\in\mathbb{N}}$ est donc un processus de Poisson de paramètre $\lambda_1 + \lambda_2$. ■

Bibliographie

[1] G. AULIAC, C. COCOZZA-THIVENT, S. MERCIER et M. ROUSSIGNOL : Intégration et probabilités. Ediscience 2005.

[2] P. BALDI : Martingales et chaînes de Markov : théorie élémentaire et exercices corrigés. Hermann 2001.

[3] N. BOULEAU : Processus stochastiques et applications. Hermann 1988.

[4] P. BREMAUD : Markov chains, Gibbs fields, Monte Carlo simulations and queues. Springer 1999.

[5] P. CIARLET : Introduction à l'analyse numérique matricielle et à l'optimisation. Masson 1997.

[6] J. F. DELMAS : Modèles aléatoires : applications aux sciences de l'ingénieur et du vivant. Springer 2006.

[7] M. DUFLO : Algorithmes stochastiques. Springer 1996.

[8] R. DURRETT : Essentials of stochastic processes. Springer 1999.

[9] D. FOATA, A. FUCHS : Processus stochastiques : Processus de Poisson, chaînes de Markov et martingales. Dunod 2004.

[10] L. MAZLIAK, P. PRIOURET et P. BALDI : Martingales et chaînes de Markov. Hermann 1998.

[11] E. SENETA : Non-negative matrices and Markov chains. Springer Verlag 1981.

[12] B. YCART : Modèles et algorithmes markoviens. Springer 2002.