

Contents

I	Introduction	1
1	Rappels de probabilités	4
1.1	Loi binomiale et loi de Poisson	4
1.2	Loi normale et exponentielle	9
2	Chaînes de Markov sur un espace fini	13
2.1	Définition et premières propriétés	13
2.2	Exemples de chaîne de Markov	15
2.2.1	La chaîne à deux états	15
2.2.2	Jeu de pile ou face	16
2.2.3	Modèle d'Ehrenfest	17
2.3	La relation de Chapman-Kolmogorov	19
2.4	Chaînes de Markov absorbantes	21
2.5	Chaînes de Markov irréductibles	26
2.6	Etats récurrents et transients	38
2.7	Périodicité	48
3	Chaîne de Markov sur un ensemble dénombrable	51
3.1	Marches aléatoires	51
3.2	Distributions stationnaires	56
3.3	Convergence vers la distribution stationnaire	62
3.4	Chaînes de Markov réversibles	64
4	Le processus ponctuel de Poisson	66
4.1	Définition et premières propriétés	66
4.2	Construction par la fonction de comptage	67
4.3	Construction par les temps d'attente	72
4.4	Exemples de processus de Poisson	76
4.5	Généralisations 0	80

Première partie

Introduction

Introduction : Bon nombre de phénomènes physiques se décrivent par l'évolution d'une ou de plusieurs grandeurs au cours du temps. À un instant donné, ces grandeurs présentent souvent un caractère imprévisible, aléatoire, et il est alors naturel de les représenter par une variable aléatoire. L'évolution du phénomène est alors décrite par l'ensemble des variables aléatoires modélisant le phénomène à chaque instant. Cet ensemble de variables aléatoires forme un processus stochastique ou aléatoire.

Un processus stochastique est donc une collection de variables aléatoires indexées par un paramètre. Celui-ci peut représenter le temps, discret ou continu, ou une variable d'espace. Présenté comme ceci, cet objet mathématique ne présente que peu d'intérêt et est difficilement exploitable. En revanche, la connaissance des relations entre ces variables aléatoires lorsque le paramètre varie permet d'obtenir des propriétés intéressantes qui caractérisent l'évolution du phénomène.

Les applications des processus stochastiques sont très nombreuses. Ceux-ci sont notamment utilisés par l'ingénieur pour la construction de modèles mathématiques de nombreux phénomènes. On peut par exemple citer :

- *La théorie économique et l'économétrie* dont l'objectif est de rendre compte des mécanismes qui régissent les faits économiques (souvent aléatoires). *La théorie de la prévision*, qui regroupe l'ensemble des méthodes permettant de donner une estimation (probabiliste) de l'évolution d'une variable économique à partir de données sur ses valeurs passées, utilise les processus stochastiques. On parle dans ce cas de statistique des processus stochastiques.
- *Les transports et le trafic*, qu'il s'agisse de transport de personnes, de biens, ou de trafic dans les réseaux (téléphoniques, mobiles, Internet, etc.).
- *La fiabilité* des systèmes ou d'un matériel, c'est-à-dire l'évolution dans le temps de ses défaillances.
- *L'ingénierie financière*, où les modèles financiers font intervenir des notions complexes de processus et de calcul stochastique.

- *La théorie de l'information et du filtrage.*
- *Les sciences de l'environnement.*

Chapitre 1

Rappels de probabilités

1.1 Loi binomiale et loi de Poisson

Nous commençons par rappeler quelques lois de probabilités usuelles qui joueront un rôle important dans la suite. Une expérience de Bernoulli de longueur n et probabilité de succès $p \in [0, 1]$ consiste à répéter n fois, de manière indépendante, une expérience élémentaire qui n'admet que deux issues possibles : le succès, qui se produit avec probabilité p , et l'échec, qui se produit avec probabilité $1 - p$. Si par exemple l'expérience consiste à jeter un dé équilibré, et que l'on considère comme succès uniquement l'obtention de 6 *points*, on aura $p = \frac{1}{6}$.

Soit X la variable aléatoire donnant le nombre de succès de l'expérience de longueur n . Elle pourra prendre les valeurs $0, 1, \dots, n$, la valeur k étant obtenue avec probabilité :

$$\mathbb{P}(X = k) = b_{n,p}(k) = C_n^k p^k (1 - p)^{n-k}. \quad (1.1.1)$$

avec $C_n^k = \frac{n!}{k!(n-k)!}$. En effet, il y a C_n^k manières de choisir les k succès et $n - k$ échecs parmi les n expériences, et chacun de ces choix se produit avec probabilité $p^k (1 - p)^{n-k}$.

Signalons que lors-que'on parle de variable aléatoire on sousentend qu'elle est

definie sur un espace probabilisé (Ω, \mathcal{F}, p) , dans le cas ou Ω est finie ou dénombrable on prend $\mathcal{F} = \mathcal{P}(\Omega)$ l'ensemble de toutes les parties de Ω

Définition 1.1.1. (*Loi binomiale*). Soit X une variable aléatoire prenant ses valeurs dans $\{0, 1, \dots, n\}$ et satisfaisant (1.1.1). On dit alors que X suit une loi binomiale de paramètres (n, p) et on note $X \sim b_{n,p}$.

On peut représenter X comme somme de n variables aléatoires Y_i indépendantes et identiquement distribuées (*i.i.d*), de loi de Bernoulli de paramètres p , c'est-à-dire telles que $\mathbb{P}(Y_i = 1) = p = 1 - \mathbb{P}(Y_i = 0)$. Alors comme l'espérance de chaque Y_i vaut $\mathbb{E}(Y_i) = 0 \cdot \mathbb{P}(Y_i = 0) + 1 \cdot \mathbb{P}(Y_i = 1) = p$, on obtient pour l'espérance de X

$$\mathbb{E}(X) = \sum_{k=0}^n k \mathbb{P}\{X = k\} = \sum_{i=1}^n \mathbb{E}(Y_i) = np.$$

De plus, comme la variance de chaque Y_i vaut $\text{var}(Y_i) = \mathbb{E}(Y_i^2) - \mathbb{E}(Y_i)^2 = p(1 - p)$, on voit que la variance de X est donnée par :

$$\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = np(1 - p)$$

Définition 1.1.2. (*Loi de Poisson*). On dit que la variable aléatoire X suit une loi de Poisson de paramètre $\lambda > 0$, et noterons $X \sim \pi_\lambda$ si elle prend des valeurs entières non-négatives, avec probabilités

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} = \pi_\lambda(k).$$

Avant de discuter sa signification, mentionnons quelques propriétés de base de cette loi.

Proposition 1.1.3.

1. Si X suit une loi de Poisson de paramètre λ , alors :

$$\mathbb{E}(X) = \text{var}(X) = \lambda.$$

2. Si X et Y sont indépendantes, et suivent des lois de Poisson de paramètres λ et μ respectivement, alors $X + Y$ suit une loi de Poisson de paramètre $\lambda + \mu$.

L'importance de la loi de Poisson π_λ vient du fait qu'elle donne une bonne approximation de la loi binomiale $b_{n,p}$ lorsque la longueur n de l'expérience est grande et que la probabilité de succès p est faible, avec $np = \lambda$.

En effet, nous avons le résultat de la convergence suivant :

Proposition 1.1.4. Soit $\{p_n\}_{n \geq 0}$ une suite telle que $\lim_{n \rightarrow \infty} np_n = \lambda > 0$. Alors, pour tout $k \in \mathbb{N}$,

$$\lim_{n \rightarrow \infty} b_{n,p_n}(k) = \pi_\lambda(k) \quad (1.1.2)$$

Démonstration. On a pour tout $k = 0, 1, \dots, n$

$$\begin{aligned} b_{n,p_n}(k) &= C_n^k p_n^k (1 - p_n)^{n-k} \\ &= \frac{n!}{k!(n-k)!} p_n^k (1 - p_n)^{n-k} \end{aligned}$$

En posant $\lambda_n = np_n$. On obtient

$$\begin{aligned} b_{n,p_n}(k) &= \frac{n(n-p) \dots (n-k+1)}{k!(1 - \frac{\lambda_n}{n})^k} \cdot \frac{(\lambda_n)^k}{n^k} \cdot (1 - \frac{\lambda_n}{n})^n \\ &= \frac{(\lambda_n)^k}{k!} (1 - \frac{1}{n})(1 - \frac{2}{n}) \dots (1 - \frac{k-1}{n}) \frac{1}{(1 - \frac{\lambda_n}{n})^k} (1 - \frac{\lambda_n}{n})^n. \end{aligned}$$

On a :

$$\text{i)} \quad \lim_{n \rightarrow \infty} (1 - \frac{j}{n}) = 1, \forall j = 1, \dots, k-1.$$

$$\text{ii)} \quad \lim_{n \rightarrow \infty} \frac{\lambda_n}{n} = \lim_n \frac{np_n}{n} = \lim_{n \rightarrow \infty} p_n = 0.$$

$$\begin{aligned} \text{iii)} \quad \lim_{n \rightarrow \infty} (1 - \frac{\lambda_n}{n})^n &= \lim_{n \rightarrow \infty} \left(\left(1 - \frac{1}{\frac{n}{\lambda_n}}\right)^{\frac{n}{\lambda_n}} \right)^{\lambda_n} = \lim_{n \rightarrow \infty} \left(\left(1 - \frac{1}{\frac{n}{\lambda_n}}\right)^{\frac{n}{\lambda_n}} \right)^{\lim_{n \rightarrow \infty} \lambda_n} \\ &= (e^{-1})^\lambda = e^{-\lambda}. \end{aligned}$$

On déduit que $\lim_{n \rightarrow \infty} b_{n,p_n}(k) = \pi_\lambda(k)$ ■

Nous donnons maintenant un résultat bien plus fort, à savoir la convergence dans L^1 de la loi de Bernoulli vers la loi de Poisson.

Théorème 1.1.5. *On a*

$$\sum_{k=0}^{\infty} |b_{n,p}(k) - \pi_{n,p}(k)| \leq 2np^2. \quad (1.1.3)$$

Démonstration. Nous commençons par introduire les espaces probabilisés $(\Omega_i, \mathcal{P}(\Omega_i), p_i)$, pour $i = 1, \dots, n$, donnés par $\Omega_i = \{-1, 0, 1, 2, \dots\}$ et

$$p_i(k) = \begin{cases} e^{-p} - (1-p) & \text{si } k = -1, \\ 1-p & \text{si } k = 0, \\ e^{-p} \frac{p^k}{k!} & \text{si } k \geq 1. \end{cases}$$

On vérifiera que les p_i définissent bien une distribution de probabilité. Sur chaque Ω_i nous introduisons les deux variables aléatoires

$$X_i(w_i) = \begin{cases} 0 & \text{si } w_i = 0, \\ 1 & \text{sinon,} \end{cases} \quad Y_i(w_i) = \begin{cases} w_i & \text{si } w_i \geq 1, \\ 0 & \text{sinon.} \end{cases}$$

De cette manière, on a $\mathbb{P}\{X_i = 0\} = 1-p$, $\mathbb{P}\{X_i = 1\} = p$, et $\mathbb{P}\{Y_i = k\} = \pi_p(k)$ pour tout $k \geq 0$. De plus,

$$\begin{aligned} \mathbb{P}(X_i = Y_i) &= \mathbb{P}(X_i = 0, Y_i = 0) + \mathbb{P}(X_i = 1, Y_i = 1) \\ &= p_i(0) + p_i(1) = 1-p + pe^{-p}, \end{aligned}$$

donc

$$\mathbb{P}\{X_i \neq Y_i\} = p(1 - e^{-p}) \leq p^2. \quad (1.1.4)$$

Soit (Ω, p) l'espace produit des (Ω_i, p_i) . Alors

- $X = X_1 + \dots + X_n$ suit la loi binomiale $\mathbb{P}\{X = k\} = b_{n,p}(k)$;
- $Y = Y_1 + \dots + Y_n$ suit la loi de Poisson $\mathbb{P}\{Y=k\} = \pi_{np}(k)$, en vertu de la proposition(1.1.3).

Comme $X \neq Y$ implique que $X_i \neq Y_i$ pour un i au moins, il suit de 1.1.4 que

$$\mathbb{P}(X \neq Y) \leq \sum_{i=1}^n \mathbb{P}(X_i \neq Y_i) \leq np^2.$$

Montrons maintenant que :

$$\sum_{k=0}^{\infty} |b_{n,p}(k) - \pi_{np}(k)| \leq 2\mathbb{P}(X \neq Y).$$

Nous posons, pour abréger l'écriture, $f(k) = \mathbb{P}\{X = k\}$, $g(k) = \mathbb{P}\{Y = k\}$ et $A = \{k : f(k) > g(k)\}$. Alors

$$\begin{aligned} \sum_{k=0}^{\infty} |b_{n,p}(k) - \pi_{np}(k)| &= \sum_{k=0}^{\infty} |f(k) - g(k)| \\ &= \sum_{k \in A} (f(k) - g(k)) - \sum_{k \notin A} (f(k) - g(k)) \\ &= 2 \sum_{k \in A} (f(k) - g(k)) - \underbrace{\sum_{k \in \mathbb{N}} (f(k) - g(k))}_{= 1 - 1 = 0} \end{aligned}$$

Or nous pouvons écrire

$$\begin{aligned} \sum_{k \in A} (f(k) - g(k)) &= \mathbb{P}(X \in A) - \mathbb{P}(Y \in A) \\ &= \mathbb{P}(X \in A, Y \in A) + \mathbb{P}(X \in A, Y \neq X) - \mathbb{P}(Y \in A) \\ &\leq \mathbb{P}(X \in A, Y \in A) + \mathbb{P}(X \neq Y) - \mathbb{P}(Y \in A) \\ &\leq \mathbb{P}(X \neq Y), \end{aligned}$$

ce qui conclut la démonstration. Si nous prenons par exemple $p = \lambda/n$, la borne (1.1.3) nous fournit

$$\sum_{k=0}^{\infty} |b_{n,\lambda/n}(k) - \pi_{\lambda}(k)| \leq 2 \frac{\lambda^2}{n}.$$

■

1.2 Loi normale et exponentielle

Nous aurons besoins de certaines variables aléatoires réelles continues. Pour les définir, le plus simple est de passer par la notion de fonction de répartition.

Définition 1.2.6. (Fonction de répartition). Une fonction $F : \mathbb{R} \rightarrow [0, 1]$ est une fonction de répartition si

- F est croissante : $x \leq y \implies F(x) \leq F(y)$.
- F est continue à droite : $\lim_{y \rightarrow x^+} F(y) = F(x) \quad \forall x \in \mathbb{R}$.
- $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$.

Une fonction de répartition F est dite absolument continue s'il existe une fonction mesurable non négatif f telle que :

$$F(x) = \int_{-\infty}^x f(y) dy \quad \forall x \in \mathbb{R}.$$

f est appelée fonction densité de F .

Le lien entre la notion de fonction de répartition et les variables aléatoires vient du fait que pour toute variable aléatoire réelle, $\mathbb{P}(X \leq t)$ est une fonction de répartition.

En effet :

- Si $s \leq t$, alors $\{X \leq s\} \subset \{X \leq t\}$, et donc $\mathbb{P}(X \leq s) \leq \mathbb{P}(X \leq t)$;
- $\lim_{s \rightarrow t^+} \mathbb{P}(X \leq s) - \mathbb{P}(X \leq t) = \lim_{s \rightarrow t^+} \mathbb{P}(t < X \leq s) = 0$;
- $\lim_{t \rightarrow -\infty} \mathbb{P}(X \leq t) = 0$ et $\lim_{t \rightarrow +\infty} \mathbb{P}(X \leq t) = 1$.

Ceci motive la définition suivante.

Définition 1.2.7. (Variable aléatoire à densité). Si X est une variable aléatoire,

$$F_X(t) = \mathbb{P}\{X \leq t\}$$

est appelée fonction de répartition de X . Si F_X est absolument continue de densité

f , on dit que X admet la densité f et on a les relations :

$$\begin{aligned}\mathbb{P}\{X \leq t\} &= \int_{-\infty}^t f(s) \, ds. \\ \mathbb{P}\{a < X \leq b\} &= \mathbb{P}\{X \leq b\} - \mathbb{P}\{X \leq a\} = \int_a^b f(s) \, ds.\end{aligned}$$

Dans ce cas, on peut remplacer $<$ par \leq et inversement.

Une première expérience importante de variable aléatoire réelle à densité est celle des variables gaussiennes ou normales.

Définition 1.2.8. (Loi normale). On dit que la variable aléatoire X suit une loi normale de moyenne μ et d'écart-type σ , et on note $X \sim N(\mu, \sigma^2)$ si elle admet la densité

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}.$$

Si $X \sim N(0, 1)$, on dit qu'elle suit une loi normale centrée réduite, ou standard.

Si $X \sim N(\mu, \sigma^2)$, alors son espérance vaut $\mathbb{E}(X) = \mu$ et sa variance vaut $\text{var}(X) = \sigma^2$.

Théorème 1.2.9. (Théorème de la limite centrale). Soit X_1, X_2, \dots, X_n une suite de variables aléatoires indépendantes, identiquement distribuées, d'espérance finie μ et de variance finie σ^2 alors la variable aléatoire $S_n = \sum_{i=1}^n X_i$ satisfait

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ a \leq \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq b \right\} = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} \, dx.$$

C'est-à-dire que $(S_n - n\mu)/\sqrt{n\sigma^2}$ converge en loi vers une variable normale standard.

Un second exemple de loi à densité, particulièrement important dans notre travail,

est celui des variables exponentielles.

Définition 1.2.10. (Variable exponentielle). *On dit que la variable aléatoire X suit une loi exponentielle de paramètre $\lambda > 0$, et on note $X \sim \exp(\lambda)$, si elle satisfait*

$$\mathbb{P}\{X > t\} = e^{-\lambda t}$$

pour tout $t \geq 0$. Sa fonction de répartition est donc $F_X(t) = 1 - e^{-\lambda t}$ pour $t > 0$, et sa densité est $\lambda e^{-\lambda t}$, toujours pour $t > 0$.

On vérifie qu'une variable de loi exponentielle a espérance $1/\lambda$ et variance $1/\lambda^2$. Une propriété remarquable de loi exponentielle est la propriété de Markov : pour $t > s \geq 0$,

$$\mathbb{P}\{X > t \mid X > s\} = e^{-\lambda(t-s)} = \mathbb{P}\{X > t - s\}.$$

Nous aurons parfois affaire à des couples, ou des n-uplets de variables aléatoires à densité, aussi appelés vecteurs aléatoires. Leur densité conjointe est définie comme la fonction de n variables f telle que

$$\mathbb{P}\{X_1 \leq t_1, X_2 \leq t_2, \dots, X_n \leq t_n\} = \int_{-\infty}^{t_1} \int_{-\infty}^{t_2} \dots \int_{-\infty}^{t_n} f(x_1, x_2, \dots, x_n) dx_n \dots dx_2 dx_1$$

pour tout choix de (t_1, t_2, \dots, t_n) . Autrement dit, on a

$$f(t_1, t_2, \dots, t_n) = \frac{\partial^n}{\partial t_1 \partial t_2 \dots \partial t_n} \mathbb{P}\{X_1 \leq t_1, X_2 \leq t_2, \dots, X_n \leq t_n\}.$$

Les variables aléatoires X_1, X_2, \dots, X_n sont dites indépendantes si on a

$$\mathbb{P}\{X_1 \leq t_1, X_2 \leq t_2, \dots, X_n \leq t_n\} = \mathbb{P}\{X_1 \leq t_1\} \mathbb{P}\{X_2 \leq t_2\} \dots \mathbb{P}\{X_n \leq t_n\}$$

pour tout choix de t_1, t_2, \dots, t_n . On montre que c'est équivalent à ce que la densité

conjointe s'écrit sous la forme

$$f(t_1, t_2, \dots, t_n) = f_1(t_1)f_2(t_2)\dots f_n(t_n)$$

pour des densités f_1, f_2, \dots, f_n (appelées densités marginales X_1, X_2, \dots, X_n)

Chapitre 2

Chaînes de Markov sur un espace fini

2.1 Définition et premières propriétés

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires à valeurs dans l'ensemble E des états, supposé une partie de \mathbb{N} . On dit que cette suite est une chaîne de Markov, si pour tout $n \geq 1$ et toute suite $(i_0, \dots, i_{n-1}, i, j)$ d'éléments de E , pour laquelle les probabilités conditionnelles suivantes existent, on a :

$$\mathbb{P}(X_{n+1} = j \mid X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) = \mathbb{P}(X_{n+1} = j \mid X_n = i). \quad (2.1.1)$$

Autrement dit, l'état du processus à l'instant $(n + 1)$ ne dépend que de celui à l'instant n précédent, mais non de ses états antérieurs. (On dit que le processus est sans mémoire ou non héréditaire).

Définition 2.1.11. *La chaîne de Markov est dite homogène (dans le temps), si la*

probabilité précédente ne dépend pas de n . Soit

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = p_{i,j} \quad \forall n \in \mathbb{N}$$

cette probabilité s'appelle la probabilité de passage ou de transition de l'état i à l'état j en une étape.

Définition 2.1.12. La matrice

$$P = \begin{pmatrix} p_{0,0} & p_{0,1} & p_{0,2} & \cdots \\ p_{1,0} & p_{1,1} & p_{1,2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

dont les coefficients sont les probabilités de transition $p_{i,j}$ est appelée matrice de passage (ou matrice de transition) de la chaîne. C'est une matrice finie ou dénombrable, selon que l'ensemble des états est fini ou dénombrable.

Proposition 2.1.13. Toute matrice de transition $P = (p_{i,j}) \ ((i,j) \in E^2)$ vérifie les propriétés suivantes :

- (1) pour tout couple (i,j) , on a : $p_{i,j} \geq 0$;
- (2) pour tout $i \in E$, on a $\sum_{j \in E} p_{i,j} = 1$.

Démonstration.

1. Les nombres $p_{i,j}$ sont des probabilités, donc des nombres non négatifs.
2. Pour chaque $i \in E$, l'application $B \mapsto \sum_{j \in B} p_{i,j}$ définit une mesure de probabilité sur E ($B \subset E$).

■

Remarque 2.1.14. Une matrice P qui vérifie les conditions 1) et 2) de la proposition précédente, est appelée stochastique.

Proposition 2.1.15. *Soit P une matrice de transition. Alors :*

- (1) P admet la valeur propre 1.
- (2) Le vecteur V ayant toutes ses composantes égales à 1 est un vecteur propre associé à la valeur propre 1.

Démonstration. En effet, en considérant V comme un vecteur -colonne, on a :

$PV = V$ si et seulement si, pour tout $i \in E$, la relation suivante est satisfaite :

$$\sum_{j \in E} p_{i,j} v_j = v_i. \text{ Il suffit donc, pour tout } i \in E, \text{ de prendre } v_i = 1. \quad \blacksquare$$

Définition 2.1.16. **(graphe associé à une matrice de transition).** *A toute matrice de transition, on peut associer son graphe. Il y a une flèche, étiquetée $p_{i,j}$, entre le sommet étiqueté i et le sommet étiqueté j si et seulement si la probabilité de transition de l'état i à l'état j est strictement positive : $p_{i,j} > 0$.*

Lorsque l'ensemble des états est fini, cette présentation de la matrice de transition par son graphe est particulièrement utile et parlante.

2.2 Exemples de chaîne de Markov

Il y a des exemples classiques de chaîne de Markov homogène.

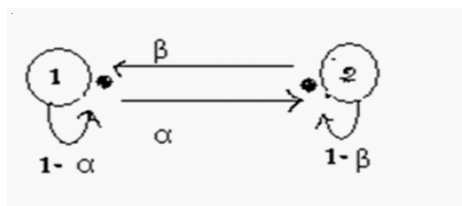
2.2.1 La chaîne à deux états

En excluant le cas trivial de la matrice-unité, la matrice de transition correspondante est de la forme :

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix} \quad (0 < \alpha, \beta < 1).$$

Les calculs sont explicites. Pour tout $n \geq 0$, on peut évaluer la $n^{\text{ième}}$ puissance P^n , ainsi que la valeur limite $\lim_n P^n$.

Le graphe associé est très simple :



2.2.2 Jeu de pile ou face

Deux joueurs A et B jouent à la variante suivante de Pile ou Face. Ils jettent une pièce de monnaie (parfaitement équilibrée) de manière répétée. A gagne dès que la pièce tombe trois fois de suite sur Face, alors que B gagne dès que la suite Pile-Face-Pile apparaît.

On se pose les questions suivantes :

1. avec quelle probabilité est-ce A qui gagne le jeu ?
2. Au bout de combien de jets de la pièce l'un des deux joueurs gagne-t-il ?

On peut alors décrire le jeu par une chaîne de Markov sur l'ensemble

$$E = \{pp, pf, fp, ff, A \text{ gagne}, B \text{ gagne}\},$$

où par exemple pp signifie que la pièce est tombée sur Piles lors des deux derniers jets. La matrice de transition vaut

$$\begin{array}{c}
pp \\
pf \\
fp \\
ff \\
A \\
B
\end{array}
\begin{pmatrix}
pp & pf & fp & ff & A & B \\
1/2 & 1/2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1/2 & 0 & 1/2 \\
1/2 & 1/2 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/2 & 0 & 1/2 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

2.2.3 Modèle d'Ehrenfest

C'est un système motivé par la physique, qui a été introduit pour modéliser de manière simple la répartition d'un gaz entre deux récipients.

N boules, numérotées de 1 à N , sont réparties sur deux urnes. De manière répétée, on tire au hasard, de façon équiprobable un numéro entre 1 et N , et on change d'urne la boule correspondante.

On peut décrire le système par une chaîne de Markov, sur l'espace des états $E = \{0, 1, \dots, N\}$, où le numéro d'état correspond au nombre de boules dans l'urne de gauche, par exemple.

La matrice vaut

$$\begin{array}{c}
0 \\
1 \\
2 \\
3 \\
4 \\
\vdots \\
N
\end{array}
\begin{pmatrix}
0 & 1 & 2 & 3 & 4 & 5 & \dots & N-1 & N \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{N} & 0 & \frac{N-1}{N} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{2}{N} & 0 & \frac{N-2}{N} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{3}{N} & 0 & \frac{N-3}{N} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{4}{N} & 0 & \frac{N-4}{N} & 0 & 0 & 0 \\
\vdots & & & & & & & & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}$$

Définition 2.2.17. Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov. La loi de la variable aléatoire X_0 s'appelle la loi initiale de la chaîne.

Remarque 2.2.18. La loi d'une chaîne de Markov est définie à partir de sa loi initiale et sa matrice de transition. Comme il est indiqué dans le résultat suivant.

Théorème 2.2.19. Soient (X_n) une suite de variables aléatoires à valeurs dans E , ν une mesure de probabilité sur E et P une matrice stochastique. Alors (X_n) est une chaîne de Markov de matrice de transition P et de distribution initiale ν si et seulement si pour tout $n \geq 0$, et pour tout choix de i_0, i_1, \dots, i_n d'éléments de E , on a

$$\mathbb{P}(X_{[0,n]} = i_{[0,n]}) = \nu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n}. \quad (2.2.2)$$

Démonstration.

1. Supposons que (X_n) est une chaîne de Markov de matrice de transition P et de loi initiale ν . Procédons par récurrence :

- 1) Pour $n = 0$ on a par définition

$$\mathbb{P}(X_0 = i) = \nu_i, \forall i \in E \quad (2.2.3)$$

- 2) Supposons que l'expression (2.2.2) est vérifiée pour n .

$$\begin{aligned} \text{On a pour } n+1, \mathbb{P}(X_{n+1} = i_{n+1}, X_n = i_n, \dots, X_0 = i_0) &= \\ \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n, \dots, X_0 = i_0) \cdot \mathbb{P}(X_n = i_n, \dots, X_0 = i_0) &= \\ = \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n) \nu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n} &= \\ = \nu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n} \cdot p_{i_n i_{n+1}} \end{aligned}$$

2. Inversement : Supposons que l'expression (2.2.2) est satisfaite pour la suite de variables aléatoires (X_n) montrons que la (X_n) est chaîne de Markov. Soit $n \in \mathbb{N}$, et $i_0, i_1, \dots, i_n \in E$.

$$\begin{aligned} \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n, \dots, X_0 = i_0) &= \frac{\mathbb{P}(X_{n+1} = i_{n+1}, X_n = i_n, \dots, X_0 = i_0)}{\mathbb{P}(X_n = i_n, \dots, X_0 = i_0)} \\ &= \frac{\nu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n} \cdot p_{i_n i_{n+1}}}{\nu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n}} = p_{i_n i_{n+1}} = \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n). \end{aligned}$$

■

2.3 La relation de Chapman-Kolmogorov

Pour $n > 0$ et $i, j \in E$, on désigne par $p_{i,j}^{(n)}$ la probabilité, partant de l'état i à l'instant 0, d'être dans l'état j à l'instant n . En d'autres termes, on pose :

$$p_{i,j}^{(n)} = \mathbb{P}(X_n = j | X_0 = i). \quad (2.3.4)$$

On désigne également par $P^{(n)}$ la matrice dont les éléments sont $p_{i,j}^{(n)}$ $((i, j) \in E^2)$.

Théorème 2.3.20. (*Relation de Chapman-Kolmogorov*). *Pour tout $n \geq 0$, la matrice de transition en n étapes est égale à la puissance $n^{\text{ième}}$ de la matrice de transition en une étape :*

$$P^{(n)} = (P)^n \quad (2.3.5)$$

Démonstration. Procédons par récurrence. Le résultat est vrai pour $n = 0$, puisque $P^0 = I$ (la matrice identité) et pour $n = 1$, puisque $P^{(1)} = P$. Prenons $n \geq 2$; on a, $P^n = P^{n-1}P = P^{(n-1)}P^{(1)}$. Par conséquent, si l'on désigne par $p_{i,j}^n$ le coefficient en (i, j) de la matrice P^n , on a :

$$\begin{aligned} P_{i,j}^n &= \sum_{k \in E} p_{i,k}^{(n-1)} p_{k,j} \\ &= \sum_{k \in E} \mathbb{P}(X_{n-1} = k | X_0 = i) \mathbb{P}(X_n = j | X_{n-1} = k) \\ &= \sum_{k \in E} \mathbb{P}(X_{n-1} = k | X_0 = i) \mathbb{P}(X_n = j | X_{n-1} = k) \quad [\text{car la chaîne est homogène}] \end{aligned}$$

Considérons l'évènement $A(i_1, \dots, i_{n-2}) = \{X_1 = i_1, \dots, X_{n-2} = i_{n-2}\}$.

Alors

$\mathbb{P}(X_n = j | X_{n-1} = k) = \mathbb{P}(X_n = j | X_{n-1} = k, A(i_1, \dots, i_{n-2}), X_0 = i_0)$, car la chaîne est de Markov et on a :

$$\mathbb{P}(X_{n-1} = k | X_0 = i) = \sum_{i_1, \dots, i_{n-2}} \mathbb{P}(X_{n-1} = k, A(i_1, \dots, i_{n-2}) | X_0 = i),$$

puisque les événements $A(i_1, \dots, i_{n-2})$ forment un système complet d'événements. On en tire :

$$\begin{aligned}
p_{i,j}^n &= \sum_{k \in E} \sum_{i_1, \dots, i_{n-2}} \mathbb{P}(X_{n-1} = k, A(i_1, \dots, i_{n-2}) | X_0 = i) \mathbb{P}(X_n = j | X_{n-1} = k) \\
&= \sum_{k \in E} \sum_{i_1, \dots, i_{n-2}} \mathbb{P}(X_{n-1} = k, A(i_1, \dots, i_{n-2}) | X_0 = i) \mathbb{P}(X_n = j | X_{n-1} = k, A(i_1, \dots, i_{n-2}), X_0 = i) \\
&= \sum_{k \in E} \sum_{i_1, \dots, i_{n-2}} \mathbb{P}(X_n = j, X_{n-1} = k, A(i_1, \dots, i_{n-2}) | X_0 = i) \\
&= \mathbb{P}(X_n = j | X_0 = i) = p_{i,j}^{(n)}. \quad \blacksquare
\end{aligned}$$

Corollaire 2.3.21. *Pour tout $n \geq 0$, la matrice $P^{(n)}$ est une matrice stochastique.*

Démonstration. En effet, pour tout $i \in E$, on a :

$$\sum_{j \in E} p_{i,j}^{(n)} = \sum_{j \in E} \mathbb{P}(X_n = j | X_0 = i) = 1. \quad \blacksquare$$

Corollaire 2.3.22. *Pour tout $(i, j) \in E^2$ et tout couple (m, n) d'entiers positifs, on a l'identité :*

$$\mathbb{P}(X_{m+n} = j | X_0 = i) = \sum_{k \in E} \mathbb{P}(X_m = k | X_0 = i) \mathbb{P}(X_n = j | X_0 = k) \quad (2.3.6)$$

ou encore

$$p_{i,j}^{(m+n)} = \sum_{k \in E} p_{i,k}^{(m)} p_{k,j}^{(n)}$$

Démonstration. Cette identité résulte immédiatement de l'associativité du produit matriciel : $P^{(m+n)} = P^{m+n} = P^m P^n = P^{(m)} P^{(n)}$. ■

Proposition 2.3.23. *Soient $n \geq 0$, $r \geq 1$ deux entiers. Alors*

$$\begin{aligned}
\mathbb{P}(X_{n+1} = j_{n+1}, \dots, X_{n+r} = j_{n+r} | X_0 = i_0, \dots, X_n = i_n) & \quad (2.3.7) \\
= p_{i_n j_{n+1}} p_{j_{n+1} j_{n+2}} \cdots p_{j_{n+r-1} j_{n+r}}.
\end{aligned}$$

Démonstration. Lorsque $r = 1$, l'identité (2.3.7) se réduit à (2.1.1). Il suffit donc

de procéder par récurrence sur r . Pour $r \geq 2$, posons :

$$\begin{aligned} A_1 &= \{X_0 = i_0, \dots, X_n = i_n\}; \\ A_2 &= \{X_{n+1} = j_{n+1}, \dots, X_{n+r-1} = j_{n+r-1}\}; \\ A_3 &= \{X_{n+r} = j_{n+r}\}. \end{aligned}$$

D'après (2.1.1) on a $\mathbb{P}(A_3|A_2 \cap A_1) = p_{j_{n+r-1}j_{n+r}}$. De plus, $\mathbb{P}(A_2|A_1) = p_{i_n j_{n+1}} p_{j_{n+1} j_{n+2}} \dots p_{j_{n+r-2} j_{n+r-1}}$, par récurrence sur r . On conclut alors, en utilisant l'identité $\mathbb{P}(A_3 \cap A_2|A_1) = \mathbb{P}(A_3|A_2 \cap A_1) \mathbb{P}(A_2|A_1)$. ■

2.4 Chaînes de Markov absorbantes

Définition 2.4.24. Soit (X_n) une chaîne de Markov de matrice de transition P . On dit que l'état j est accessible ou atteignable à partir de l'état i , s'il existe un entier $n \geq 0$ tel que $p_{i,j}^{(n)} > 0$. On écrit :

$$i \rightsquigarrow j$$

Proposition 2.4.25. La relation d'accessibilité entre états est réflexive et transitive.

Démonstration. Comme $p_{i,i}^{(0)} = \mathbb{P}(X_0 = i|X_0 = i) = 1$ pour tout état i , on a bien $i \rightsquigarrow i$. Ensuite, si $i \rightsquigarrow l$ et $l \rightsquigarrow j$, alors $p_{i,l}^{(m)} > 0$ et $p_{l,j}^{(n)} > 0$ pour certains entiers $m, n \geq 0$. D'après ce qui précède, on en tire :

$$p_{i,j}^{(m+n)} = \sum_{k \in E} p_{i,k}^{(m)} p_{k,j}^{(n)} \geq p_{i,l}^{(m)} p_{l,j}^{(n)} > 0, \text{ d'où } i \rightsquigarrow j \quad \blacksquare$$

Définition 2.4.26. On dit que deux états i et j communiquent et l'on écrit $i \longleftrightarrow j$, si on a à la fois $i \rightsquigarrow j$ et $j \rightsquigarrow i$.

Proposition 2.4.27. La relation de communication entre états est une relation

d'équivalence.

Démonstration. Les propriétés de réflexivité et de transivité déjà vérifiées pour la relation d'accessibilité restent naturellement encore valable pour la relation de communication. Enfin, cette dernière relation est symétrique par définition-même. ■

Définition 2.4.28. Soit (X_n) une chaîne de Markov sur E . On dit d'un état $i \in E$ qu'il est absorbant si $p_{i,i} = 1$ (i.e $\forall j \neq i \ p_{i,j} = 0$)

Définition 2.4.29. On dit qu'une chaîne de Markov (X_n) est absorbante si $\forall i \in E$, il existe un état absorbant $j \in E$ telle que j est atteignable partant de i

Dans le cas où $(X_n)_{n \in \mathbb{N}}$ est chaîne de Markov absorbante. On ordonne les états de façon que les éléments non absorbants soient classés les premiers et les absorbants les derniers,

En désignons les non absorbants par $\{1, 2, \dots, q\}$ et les absorbants par $\{q+1, \dots, M\}$, on obtient la matrice de transition

$$P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}.$$

appelée la forme canonique. Avec

Q : matrice carré d'ordre q . Les éléments $Q_{i,j}$ sont définis par $Q_{i,j} = p_{i,j}$ pour $1 \leq i, j \leq q$

R : une matrice de q lignes et r colonnes ($r = M - q$). $R_{i,k} = p_{i,k}$ avec $1 \leq i \leq q$ et $q + 1 \leq k \leq M$.

I : la matrice identité d'ordre r

0 : la matrice nulle de taille $r \times q$

Proposition 2.4.30. On a $P^n = \begin{pmatrix} Q^n & (I + Q + \dots + Q^{n-1})R \\ 0 & I_r \end{pmatrix}, \forall n \in \mathbb{N}$.

Démonstration. On procède par récurrence :

a) Pour $n = 1$, on a $P = \begin{pmatrix} Q & I_q.R \\ 0 & I_r \end{pmatrix}$.

b) Supposons que la proposition est vérifiée pour n . On a

$$\begin{aligned} P^{n+1} &= \begin{pmatrix} Q^n & (I_q + Q + \dots + Q^{n-1})R \\ 0 & I_r \end{pmatrix} \times \begin{pmatrix} Q & R \\ 0 & I_r \end{pmatrix} \\ &= \begin{pmatrix} Q^{n+1} & Q^n R + (I_q + Q + \dots + Q^{n-1})R \\ 0 & I_r \end{pmatrix} \\ &= \begin{pmatrix} Q^{n+1} & (I_q + Q + \dots + Q^n)R \\ 0 & I_r \end{pmatrix} \end{aligned}$$

■

Théorème 2.4.31. Soit (X_n) une chaîne de Markov absorbante. En considérant les mêmes notations précédentes. On a :

1. $\lim_n Q^n = 0$ (la matrice nulle).
2. La matrice $[I - Q]$ est inversible, son inverse est $[I - Q]^{-1} = \sum_{n \geq 0} Q^n$.

Démonstration.

1) le nombre $Q_{i,j}^{(n)}$ est la probabilité pour que la chaîne se trouve dans j en n pas partant de i avec $i, j \leq q$. Cette valeur est inférieure à la probabilité de ne pas atteindre un état absorbant en n pas partant de i . C'est à dire on a :

$$Q_{i,j}^{(n)} = \mathbb{P}(X_n = j \mid X_0 = i) = \mathbb{P}_i(X_n = j) \leq \mathbb{P}_i(X_n \leq q).$$

Pour tout $i \in \{1, \dots, q\}$, posons $m_i = \min\{n \in \mathbb{N}^* \mid p_{i,k}^{(n)} > 0\}$. m_i est le nombre de pas minimal pour atteindre un absorbant partant de i avec une probabilité non nulle.

Posons $p_i = \mathbb{P}(X_{m_i} \leq q)$, p_i est la probabilité de ne pas atteindre un état absorbant partant de i en m_i pas. On a $p_i < 1$ car $\mathbb{P}(X_{m_i} > q) > 0$.

Posons $M = \max_{1 \leq i \leq q} \{m_i\}$ et $p = \max_{1 \leq i \leq q} \{p_i\}$. La probabilité de ne pas atteindre un état absorbant en M pas partant de i est inférieure à p . En effet, $\{w \in W/X_{m_i}(w) > q\} \supset \{w \in W/X_M(w) > q\}$ et $m_i \leq M$, donc $\mathbb{P}(X_{m_i} > q) > \mathbb{P}(X_M > q)$. D'où $\mathbb{P}_i(X_M \leq q) \leq \mathbb{P}_i(X_{m_i} \leq q)$.

Or $\mathbb{P}_i(X_{m_i} \leq q) = p_i \leq p$. Donc la probabilité de ne pas atteindre un état absorbant partant de i en nM , pas est inférieure à p^n . Or la probabilité d'atteindre j en M pas partant de i , est inférieure à la probabilité d'atteindre un état absorbant en M pas partant de i , c'est à dire $Q_{i,j}^{(M)} \leq \mathbb{P}_i(X_M \leq q) \leq p$ ($1 \leq i, j \leq q$). On en déduit que $Q_{i,j}^{(nM)} \leq \mathbb{P}_i(X_{nM} \leq q) \leq p^n$.

Or $\lim_n Q_{i,j}^{(nM)} = \lim_n p^n = 0$ (car $p < 1$).

Maintenant, montrons que $\lim_n Q_{i,j}^{(n)} = 0$:

La suite de terme générale $U_n = \mathbb{P}_i(X_n \leq q)$ est décroissante. En effet ; $U_{n+1} = \mathbb{P}_i(X_{n+1} \leq q) \leq U_n = \mathbb{P}_i(X_n \leq q)$ (car $\mathbb{P}_i(X_n > q) \leq \mathbb{P}_i(X_{n+1} > q)$). On a $\lim_n U_{nM} \leq \lim_n p^n = 0$ (une suite extraite qui converge vers 0) et (U_n) décroissante donc $\lim_n U_n = 0$.

D'autre part on a : $Q_{i,j}^{(n)} = \mathbb{P}_i(X_n = j) \leq \mathbb{P}_i(X_n \leq q)$. Donc $\lim_n Q^n = 0$

2) Supposons qu'il existe $x \in \mathbb{R}^N$ telle que : $Qx = x$, donc $Q^2x = Qx = x$ d'où $\forall n \in \mathbb{N}, Q^n x = x$. Or $\lim_n Q^n = 0$ donc $x = 0$.

En posant $f(x) = (I - Q)x$. On a $\ker f = \{0\}$. En effet ; $\ker f = \{x \in \mathbb{R}^N | f(x) = 0\}$ d'où f est injective, or $\dim \mathbb{R}^N = \dim \text{Im } f + \dim \ker f$ donc $\dim \text{Im } f = \dim \mathbb{R}^N$ d'où f est surjective. Par conséquent f est bijective

On déduit donc que la matrice $(I - Q)$ est inversible.

On a $\forall n \in \mathbb{N}, (I - Q) \sum_{k=0}^n Q^k = \sum_{k=0}^n Q^k - \sum_{k=0}^{n+1} Q^k = I - Q^{n+1}$.

Par suite $(I - Q) \sum_{k=0}^{\infty} Q^k = \lim_n [(I - Q) \sum_{k=0}^n Q^k] = I - \lim_n Q^{n+1} = I$.

■

Définition 2.4.32. La matrice $F = \sum_{n \geq 0} Q^n = [I - Q]^{-1}$ s'appelle la matrice fonda-

mentale de la chaîne de Markov.

On a :

$$\lim_{n \rightarrow \infty} P^n = \begin{pmatrix} 0 & FR \\ 0 & I_r \end{pmatrix}$$

L'expression $\lim_n Q^n = 0$ signifie qu'à l'infini la probabilité d'atteindre un état absorbant partant d'un état non absorbant est égale à 0. Donc la matrice $B = F.R$ devrait représenter à la limites des temps à l'infini les probabilités de transition entre états non absorbants et états absorbants.

Théorème 2.4.33. *Soit F une matrice fondamentale d'une chaîne de Markov absorbante.*

1. *L'élément $f_{i,j}$ de la matrice F est l'esperance du nombre de passages en j partant de i :*

$$\mathbb{E}_i\left(\sum_{n=0}^{\infty} 1_{\{X_n=j\}}\right) = f_{i,j} = (F)_{i,j} = \sum_{n \geq 0} (Q^n)_{i,j} \quad 1 \leq i, j \leq q$$

2. *Soit $\tau = \min\{n \in N^*, X_n > q\}$, donnant le temps jusqu'à l'absorption de la chaîne. Alors*

$$\mathbb{E}_i(\tau) = \sum_{j=1}^q f_{i,j}.$$

3. *Posons $B = F.R$. Alors pour tout état non absorbant et tout état absorbant la probabilité que la chaîne soit absorbée par k partant de i est :*

$$\mathbb{P}_i(X_\tau = k) = b_{i,k}.$$

où $b_{i,k}$ est l'élément de la matrice B .

Démonstration.

1. On a d'après le théorème de Beppo-Liville :

$$\begin{aligned}
\mathbb{E}_i\left(\sum_{n=0}^{\infty} 1_{\{X_n=j\}}\right) &= \sum_{n=0}^{\infty} \mathbb{E}_i(1_{\{X_n=j\}}) = \sum_{n=0}^{\infty} [0 \cdot \mathbb{P}_i(X_n \neq j) + 1 \cdot \mathbb{P}_i(X_n = j)] \\
&= \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = j) = \sum_{n=0}^{\infty} P_{i,j}^{(n)} = \sum_{n=0}^{\infty} Q_{i,j}^{(n)} = f_{i,j} = (F)_{i,j} \\
2. \sum_{j=1}^q f_{i,j} &= \sum_{j=1}^q \left(\sum_{n \geq 0} Q_{i,j}^{(n)} \right) = \sum_{n \geq 0} \left(\sum_{j=1}^q P_i(X_n = j) \right) = \sum_{n \geq 0} \mathbb{P}_i(X_n \leq q) = \sum_{n \geq 0} \mathbb{P}_i(\tau > n) \\
&= \mathbb{P}_i(\tau > 1) + \mathbb{P}_i(\tau > 2) + \mathbb{P}_i(\tau > 3) + \dots \\
&= \mathbb{P}_i(\tau = 1) + \mathbb{P}_i(\tau = 2) + \mathbb{P}_i(\tau = 3) + \dots \\
&\quad + \mathbb{P}_i(\tau = 2) + \mathbb{P}_i(\tau = 3) + \dots \\
&\quad + \mathbb{P}_i(\tau = 3) + \dots \\
&= \mathbb{P}_i(\tau = 1) + 2\mathbb{P}_i(\tau = 2) + 3\mathbb{P}_i(\tau = 3) + \dots + n\mathbb{P}_i(\tau = n) + \dots \\
&= \sum_{n \geq 1} n\mathbb{P}_i(\tau = n) = \mathbb{E}_i(\tau) \\
3. \text{ Soit } k \in \{q+1, \dots, M\}. \text{ On a pour } i = \{1, \dots, q\} \\
\mathbb{P}_i(X_\tau = k) &= \sum_{n=0}^{\infty} \mathbb{P}_i(X_{n+1} = k, X_n \leq q) = \sum_{n=0}^{\infty} \sum_{j=1}^q \mathbb{P}_i(X_{n+1} = k, X_n = j) \\
&= \sum_{n=0}^{\infty} \sum_{j=1}^q \mathbb{P}(X_{n+1} = k | X_n = j) \mathbb{P}_i(X_n = j) \\
&= \sum_{n=0}^{\infty} \sum_{j=1}^q Q_{i,j}^{(n)} P_{j,k} = \sum_{j=1}^q \sum_{n=0}^{\infty} (Q_{i,j}^{(n)}) \cdot R_{j,k} = \sum_{j=1}^q f_{i,j} R_{j,k} = (B)_{i,k} \\
&= (FR)_{i,k}
\end{aligned}$$

■

2.5 Chaînes de Markov irréductibles

Définition 2.5.34. Soit (X_n) une chaîne de Markov de matrice de transition P . On dit que la chaîne $(X_n)_{n \in \mathbb{N}}$ est irréductible si $\forall i, j \in E$, j est atteignable partant de i , et inversement.

Définition 2.5.35. Soit (X_n) une chaîne de Markov, sur un espace fini $E = \{1, 2, \dots, N\}$, de matrice de transition $P = p_{(i,j)}$. On dit que la chaîne est régulière s'il existe $m \in \mathbb{N}^*$ tel que la matrice P^m a tous ses éléments non nuls (i.e $P_{i,j}^{(m)} > 0$ $\forall i, j = 1, \dots, M$)

Remarque 2.5.36. Si P est une matrice stochastique telle que $\exists m \in \mathbb{N}^*$ pour le quel $P_{i,j}^{(m)} > 0, \forall i, j = 1, \dots, M$. Alors pour tout $n \geq m$ on a $P_{i,j}^n > 0, \forall i, j = 1, \dots, M$

Remarque 2.5.37. Toute chaîne régulière est irréductible

En effet : Si (X_n) est régulier, alors $\exists m \in \mathbb{N}^*$ tq : $P_{i,j}^m > 0, \forall i, j = 1, \dots, M$. Donc tout état est atteignable en m pas partant d'un autre état.

Remarque 2.5.38. On peu avoir une chaîne qui est irréductible mais non régulière.

En effet : si P est une matrice qui contient des *zéros*, en élevant P par des puissance en peut avoir des *zéros* qui change de position ce qui rend la chaîne irréductible mais non régulière puisque toutes les puissances de P contiennent des *zéros*.

Exemple 2.5.39. $P = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ ($\forall m \in \mathbb{N}^*, \exists i, j$ tels que $p_{ij}^{(m)} = 0$)

Remarque 2.5.40. La chaîne décrivant le modèle d'Ehrenfest est irréductible. En effet, quelque soit le nombre de boules dans l'urne de gauche, on peut atteindre tout autre état en déplaçant au plus N boules d'une urne à l'autre. Cependant, la chaîne n'est pas régulière. En effet, comme à chaque pas de temps on déplace exactement une boule, le nombre de boules dans l'urne de gauche sera alternativement pair et impair. Par conséquent, chaque élément de matrice des puissances P^n sera nul pour un n sur deux.

Proposition 2.5.41. Soit (X_n) une chaîne de Markov irréductible sur un espace fini $E = \{1, 2, \dots, N\}$ et P sa matrice de transition et soit $A \subset E$. On pose

$$\tau_A = \inf\{n \in \mathbb{N}^*, X_n \in A\}. \quad (2.5.8)$$

Alors :

$$P(\tau_A < \infty) = 1 \quad (2.5.9)$$

Remarque 2.5.42. τ_A est une variable aléatoire qui représente le temps du premier passage de la chaîne par A (on dit aussi le temps d'entrée dans A)

Démonstration. Soit (X_n) une chaîne irréductible de matrice de transition $P = (p_{ij})_{1 \leq i, j \leq M}$. Considérons la chaîne (Y_n) définie à partir de sa matrice de transition \tilde{P}

$$\text{donnée par } \tilde{p}_{i,j} = \begin{cases} p_{ij} & \text{si } i \notin A \\ \delta_{ij} & \text{si } i \in A \end{cases}$$

$$\text{avec } \delta_{i,j} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

La chaîne (Y_n) est une chaîne absorbante car : pour tout $i \in A$, $\tilde{p}_{ii} = 1$ et $\tilde{p}_{ij} = 0$ si $i \neq j$.

Les deux chaînes (X_n) et (Y_n) ont le même comportement jusqu'au temps du premier passage par A . En effet : $\forall i, j \notin A$ $\tilde{p}_{ij} = \mathbb{P}_i(Y = j) = p_{ij} = \mathbb{P}_i(X = j)$.

La matrice de la chaîne absorbante (Y_n) s'écrit sous la forme canonique

$$\tilde{P} = \begin{pmatrix} Q & R \\ 0 & I_r \end{pmatrix}$$

Le temps du premier passage par A de la chaîne (X_n) , représente le temps jusqu'à absorption pour la chaîne (Y_n) . Or la probabilité de ne pas atteindre un état absorbant partant de $i \notin A$ est $\mathbb{P}_i(Y_n \leq q) = \sum_{j=1}^q \mathbb{P}_i(Y_n = j) = \sum_{j=1}^q Q_{ij}^n$.

Cependant $\lim_n Q_{ij}^n = 0$, donc $\lim_n \sum_{j=1}^q Q_{ij}^n = 0$. Par conséquent, $\lim_n \mathbb{P}_i(Y_n \leq q) = 0$.

Par ailleurs, $\mathbb{P}_i(Y_n < q) = \mathbb{P}_i(X_n < q) = \mathbb{P}_i(\tau_A > n) = 0$.

Donc $\lim_n \mathbb{P}_i(\tau_A > n) = 0$, par conséquent $\lim_n \mathbb{P}_i(\tau_A \leq n) = 1$

d'où $\mathbb{P}(\tau_A < \infty) = 1$. ■

Remarque 2.5.43. Ce résultat n'est pas vrai lorsque l'espace des états E est infini comme nous allons voir par la suite.

Théorème 2.5.44. Soit (X_n) une chaîne de Markov régulière de matrice de transition P d'ordre N . Alors il existe une matrice stochastique dont toutes les lignes sont égales

$$\Pi = \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_N \\ \pi_1 & \pi_2 & \dots & \pi_N \\ \vdots & & \dots & \vdots \\ \pi_1 & \pi_2 & \dots & \pi_N \end{pmatrix} \quad (2.5.10)$$

telle que $\lim_n P^n = \Pi$.

En plus, $0 < \pi_j < 1 \forall j = 1, \dots, N$.

Démonstration. Puisque (X_n) est une chaîne de Markov régulière, $\exists m \in N^*$ telle que : $P_{i,j}^{(m)} > 0, \forall i, j = 1 \dots N$.

a) Pour s'implifier la preuve nous supposons que $m = 1$. On a donc $P_{ij} > 0, \forall i, j = 1, \dots, N$. Posons $d = \min_{1 \leq i, j \leq N} p_{ij}$.

Le résultat du théorème est évidemment vrai pour $n = 1$. La preuve sera faite pour $n \geq 2$. On a $Nd \leq 1$, car la somme des éléments d'une ligne quelconque est égale à 1. Donc : $d \leq \frac{1}{N} \leq \frac{1}{2}$.

Dans le cas où $N = 2$, on peut avoir $d = \frac{1}{2}$, dans ce cas la matrice de transition est $P = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$, par conséquent $P^{(n)} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}, \forall n \in N^*$. D'où $\lim_n P^{(n)} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} = \Pi$.

Il nous reste à considérer les cas avec $d < \frac{1}{2}$, et $N \geq 3$ et par conséquent $d < \frac{1}{2}$.

Soit $y = (y_1, \dots, y_N)$ un vecteur quelconque de \mathbb{R}^N et soit m_0 et M_0 telle que $m_0 \leq y_i \leq M_0, \forall i = 1, \dots, N$.

Posons $z = P.y = (P_{i,j}). \begin{pmatrix} y_1 & \dots & y_N \end{pmatrix}^T = \begin{pmatrix} z_1 & \dots & z_N \end{pmatrix}^T$.

La plus grande valeur que les composantes de $z = (z_1, \dots, z_N)$ peuvent prendre est

celle pour la quelle toutes les composantes de y prennent la valeur M_0 sauf un qui prend la valeur m_0 . On a alors $z_i \leq P_{i1}m_0 + P_{i2}M_0 + \dots + P_{iN}M_0$, d'où

$$z_i \leq P_{i1}m_0 + (1 - P_{i1})M_0 \leq dm_0 + (1 - d)M_0.$$

La dernière inégalité est vraie car la fonction $g(x) = xm_0 + (1 - x)M_0 = (m_0 - M_0)x + M_0$ est décroissante, ($g'(x) = (m_0 - M_0) \leq 0$)

On a $P_{i,j} \geq d$, $\forall i, j = 1, \dots, N$. Donc $g(P_{i,j}) \leq g(d)$.

De la même façon on a $P_{i1}M_0 + P_{i2}m_0 + \dots + P_{iN}m_0 \leq z_i \implies P_{i1}M_0 + (1 - P_{i1})m_0 \leq z_i$. Or $dM_0 + (1 - d)m_0 \leq P_{i1}M_0 + (1 - P_{i1})m_0$, car la fonction $h(x) = xM_0 + (1 - x)m_0 = (M_0 - m_0)x + m_0 > 0$ est croissante d'où $h(d) \leq h(P_{i,j}) \quad \forall i, j = 1, \dots, N$.

On a ainsi,

$$dM_0 + (1 - d)m_0 \leq z_i \leq dm_0 + (1 - d)M_0$$

Notons $m_1 = dM_0 + (1 - d)m_0$ et $M_1 = dm_0 + (1 - d)M_0$.

On a donc $m_1 \leq z_i \leq M_1$, $\forall i = 1, \dots, N$.

$$\text{En plus } \begin{cases} m_0 = dm_0 + (1 - d)m_0 \leq dM_0 + (1 - d)m_0 = m_1 \\ M_0 = dM_0 + (1 - d)M_0 \geq dm_0 + (1 - d)M_0 = M_1 \end{cases}$$

D'autre part $M_1 - m_1 = (M_0 - m_0)(1 - 2d)$. Maintenant, en posons $h = P.z$, on trouve de la même façon :

$$m_2 = dM_1 + (1 - d)m_1 \leq h_i \leq dm_1 + (1 - d)M_1 = M_2$$

avec h_i les composantes du vecteur h .

On montre facilement que $m_0 \leq m_1 \leq m_2$, $M_2 \leq M_1 \leq M_0$ et $(M_2 - m_2) = (M_1 - m_1)(1 - 2d) = (M_0 - m_0)(1 - 2d)^2$.

En poursuivant ce processus on trouve deux suites adjacentes (m_n) et (M_n) telles que $m_n \leq (P_y^m)_i \leq M_n \quad \forall i = 1, \dots, N$ et $\forall n \in \mathbb{N}^*$, avec $M_n - m_n = (M_0 - m_0)(1 - 2d)^n$.

Puisque $d < \frac{1}{2}$ alors $\lim_n (M_n - m_n) = 0$. On déduit que les deux suites sont convergentes et tendent vers la même limite u . Donc $\lim_n P^n y = (u, \dots, u)^T$.

Cherchons maintenant la limite de chaque colonne de la matrice P .

Considérons la base canonique (e_1, \dots, e_N) c'est-à-dire $e_1 = (1, 0, \dots, 0)^T$, $e_2 = (0, 1, \dots, 0)^T$, ..., $e_N = (0, 0, \dots, 1)^T$, le produit $P.e_j$ nous donne la colonne j de la matrice P .

On a $\lim_n (P^n e_j) = \lim_n \begin{pmatrix} P_{1,j}^{(n)} & \dots & P_{N,j}^{(n)} \end{pmatrix}^T = \begin{pmatrix} \pi_j & \dots & \pi_j \end{pmatrix}^T$. Donc pour $j = 1, \dots, N$, on obtient $\lim_n P^n = \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_N \\ \pi_1 & \pi_2 & \dots & \pi_N \\ \vdots & & & \vdots \\ \pi_1 & \pi_2 & \dots & \pi_N \end{pmatrix}$.

b) Montrons que $\pi_j > 0$, $\forall j = 1, \dots, N$.

On a $p_{ij}^{(m)} > 0$, $\forall i, j = 1, \dots, N$. Posons $d = \min\{p_{ij}^{(m)}, i, j = 1, \dots, N\}$. Evidement on a $d > 0$. Puisque $P^{m+1} = P.P^m$

$$\text{alors } P_{ij}^{(m+1)} = \sum_{k=1}^N p_{ik} p_{kj}^{(m)} \geq \sum_{k=1}^N p_{ik} d = d.$$

On déduit que $p_{ij}^{(m+n)} \geq d \forall n \in \mathbb{N}^*$. Donc $\lim_n p_{ij}^{(n)} = \pi_j > 0$. ■

Corollaire 2.5.45. On a :

$$\pi.P = \pi \tag{2.5.11}$$

avec $\pi = (\pi_1, \pi_2, \dots, \pi_N)$, la première ligne de la matrice Π .

Démonstration. On a $P^{n+1} = P^n.P$ donc $\lim_n P^{n+1} = \lim_n P^n.P$

$$\implies \Pi = \Pi.P.$$

On particulier en considérant la première ligne de la matrice Π on obtient :

$$(\pi_1, \pi_2, \dots, \pi_N) = (\pi_1, \pi_2, \dots, \pi_N)P$$

■

Définition 2.5.46. Soit (X_n) une chaîne de Markov irréductible de matrice de transition P et π une loi de probabilité sur E . On dit que π est une loi stationnaire ou invariante de (X_n) si $\pi P = \pi$.

Remarque 2.5.47.

1. Le corollaire précédant montre que la chaîne de Markov régulière possède toujours une loi stationnaire.
2. L'équation (2.5.11) implique qu'à chaque instant n , on a

$$\mathbb{P}_\pi(X_n = j) = \sum_{i \in E} \pi_i (P^n)_{i,j} = (\pi P^n)_j = \pi_j, \quad \forall j = 1, \dots, N \quad (2.5.12)$$

Cela signifie que si la loi initiale de la chaîne est sa loi stationnaire, alors la loi de la chaîne à n'importe quel instant est la même (égale à la loi stationnaire)

3. Par (2.5.10) on a, $\forall i, j \in E$,

$$\lim_{n \rightarrow \infty} \mathbb{P}_i(X_n = j) = \lim_{n \rightarrow \infty} (P^n)_{i,j} = \pi_j \quad (2.5.13)$$

π décrit donc la loi de probabilité asymptotique de la chaîne, qui est indépendante de l'état initial.

Théorème 2.5.48. *Pour toute loi initiale ν d'une chaîne de Markov régulière on a :*

$$\lim_n \nu P^n = \pi = (\pi_1, \pi_2, \dots, \pi_N). \quad (2.5.14)$$

Ce qui est équivalent à $\lim_n \mathbb{P}_\nu(X_n = j) = \pi_j, \quad \forall j = 1, \dots, N$.

Démonstration. La preuve très simple :

$$\lim_n \nu P^n = \nu \Pi. \text{ Or } \nu \Pi = (\nu_1, \nu_2, \dots, \nu_N) \cdot \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_N \\ \pi_1 & \pi_2 & \dots & \pi_N \\ \vdots & & \dots & \vdots \\ \pi_1 & \pi_2 & \dots & \pi_N \end{pmatrix} =$$

$$(\underbrace{(\nu_1 + \nu_2 + \dots + \nu_N)}_{=1} \pi_1, \dots, \underbrace{(\nu_1 + \nu_2 + \dots + \nu_N)}_{=1} \pi_N) = (\pi_1, \dots, \pi_N) \text{ d'où la preuve.}$$

Il est plus intéressant de présenter une autre preuve, due à **Doeblin**.

Considérons une autre Chaîne de Markov, définie sur l'espace $E \times E$. Ses probabilités de transition P^* sont données par

$$P_{((i,j),(k,l))}^* = p_{ik}p_{jl} \quad (2.5.15)$$

Nous supposons que la loi initiale de cette chaîne est une mesure produit $\rho = \nu \otimes \pi$, c'est-à-dire que

$$\rho((i,j)) = \nu_i \pi_j, \quad \forall (i,j) \in E \times E \quad (2.5.16)$$

Nous dénotons cette chaîne par $((X_n, Y_n))_{n \geq 0}$. Par construction les variables aléatoires X_0 et Y_0 sont indépendantes. Il suit donc de la définition (2.5.15) des probabilités de transition que $(X_n)_{n \geq 0}$ et $(Y_n)_{n \geq 0}$ sont en fait deux chaînes de Markov sur E de matrice de transition P , et de distributions initiales respectivement données par ν et π .

La matrice de transition P^* est régulière car les éléments de la matrice puissance $(P^*)^n$ sont donnés par les produits $p_{ik}^{(n)} p_{jl}^{(n)}$.

Considérons alors l'ensemble $A = \{(i,i) : i \in E\} \subset E \times E$. Le temps de premier passage τ_A peut s'écrire

$$\tau_A = \inf\{n > 0, X_n = Y_n\}. \quad (2.5.17)$$

Alors les deux chaînes ont la même loi pour $n \geq \tau_A$. Plus précisément

$$\mathbb{P}_\rho(X_n = j, \tau_A \leq n) = \mathbb{P}_\rho(Y_n = j, \tau_A \leq n), \quad \forall j \in E, \quad \forall n \geq 0. \quad (2.5.18)$$

Pour montrer cela, nous introduisons un nouveau processus $\{Z_n\}_{n \in \mathbb{N}}$ définie par

$$Z_n = \begin{cases} X_n & \text{pour } n \leq \tau_A, \\ Y_n & \text{pour } n > \tau_A. \end{cases} \quad (2.5.19)$$

Donc on a bien $\mathbb{P}_\rho(Z_{[0,n]} = i_{[0,n]}) = \nu_{i_0} \prod_{m=1}^n p_{i_{m-1}i_m}$ pour tout $n \geq 0$ et tout choix de $i_{[0,n]} \in E^{n+1}$. Par le théorème (2.2.19), il suit que $\{Z_n\}_{n \in \mathbb{N}}$ est une chaîne de

Markov de loi initiale ν et matrice de transition P , et donc égale en loi à $\{X_n\}_{n \in \mathbb{N}}$.

Ceci prouve (2.5.18).

Cependant, on a

$$\mathbb{P}_\nu(X_n = j) = \mathbb{P}_\rho(X_n = j, \tau_A \leq n) + \mathbb{P}_\rho(X_n = j, \tau_A > n), \quad (2.5.20)$$

$$\pi_j = \mathbb{P}_\pi(Y_n = j) = \mathbb{P}_\rho(Y_n = j, \tau_A \leq n) + \mathbb{P}_\rho(Y_n = j, \tau_A > n).$$

Donc d'après (2.5.18), on a :

$$\begin{aligned} |\mathbb{P}_\nu(X_n = j) - \pi_j| &\leq |\mathbb{P}_\rho(X_n = j, \tau_A > n) - \mathbb{P}_\rho(Y_n = j, \tau_A > n)| \quad (2.5.21) \\ &\leq 2\mathbb{P}_\rho(\tau_A > n). \end{aligned}$$

Or cette dernière quantité tend vers zéro lorsque $n \rightarrow \infty$, car τ_A est fini presque sûrement. ■

Théorème 2.5.49. *Toute chaîne de Markov régulière possède une loi stationnaire unique.*

Démonstration. On a déjà montré l'existence. Montrons maintenant l'unicité. Supposons qu'il existe une loi $\nu = (\nu_1, \dots, \nu_N)$ telle que $\nu.P = \nu$. On a alors $\forall n \in \mathbb{N}^*$ $\nu.P^n = \nu$.

Donc $\lim_n \nu.P^n = \nu$. par conséquent, $\nu\Pi = \nu$.

$$\begin{aligned} \text{On a ainsi, } (\nu_1, \dots, \nu_N) \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_N \\ \pi_1 & \pi_2 & \dots & \pi_N \\ \vdots & \vdots & & \vdots \\ \pi_1 & \pi_2 & \dots & \pi_N \end{pmatrix} &= ((\nu_1 + \dots + \nu_N)\pi_1, \dots, (\nu_1 + \dots + \nu_N)\pi_N) = \\ &(\nu_1, \dots, \nu_N). \end{aligned}$$

D'où $(\pi_1, \dots, \pi_N) = (\nu_1, \dots, \nu_N)$. ■

Remarque 2.5.50. Si la chaîne est irréductible la loi de (X_n) ne converge pas

nécessairement vers une loi π . Mais elle possède une loi stationnaire

Proposition 2.5.51. *Soit P la matrice de transition d'une chaîne de Markov irréductible. Alors il existe une mesure de probabilité unique $\pi = (\pi_1, \dots, \pi_N)$ sur E telle que $\Pi.P = \Pi$.*

Démonstration. Considérons la matrice stochastique $Q = \frac{1}{2}[P + I]$. Soit

$$m = \max_{i,j \in E} \{ \min\{n \geq 1 : p_{ij}^{(n)} > 0\} \}.$$

Considérons la matrice

$$Q^m = (I + C_m^1 P + C_m^2 P^2 + \dots + C_m^{m-1} P^{m-1} + P^m) / 2^m.$$

Pour tout couple (i, j) , il existe un terme de cette somme dont l'élément de la matrice se trouvant dans la i -ième ligne et la j -ième colonne soit strictement positif. Comme tous les autres éléments de la même matrice sont non-négatifs, on conclut que $(Q^m)_{ij} > 0$. Par conséquent, Q est la matrice de transition d'une chaîne régulière. Donc il existe une unique mesure de probabilité π telle que $\pi.Q = \pi$, ce qui implique $\frac{1}{2}[\pi + \pi P] = \pi$, donc $\pi P = \pi$. ■

Exemple 2.5.52. On a déjà noté que le modèle d'Erenfest est une chaîne de Markov irréductible. Elle a comme distribution stationnaire la loi binomiale de paramètre $p = 1/2$: $\nu_i = C_N^i \left(\frac{1}{2}\right)^N$ pour $i = 1, \dots, N$.

L'interprétation de la distribution stationnaire nous indique que si la variable aléatoire X_n suit la loi π à un temps n , alors X_m suivra la même loi π à tous les temps ultérieurs $m > n$. Cependant, les théorèmes (2.5.44) et (2.5.48) ne sont plus nécessairement vrais dans le cas d'une chaîne irréductible : Il suffit de considérer l'exemple du modèle d'Ehrenfest. Toutefois, on a encore convergence vers la distribution stationnaire dans le sens de la moyenne ergodique (ou moyenne de Cesaro)

Théorème 2.5.53. *Pour une chaîne de Markov irréductible, et pour toute distribution initiale ν , la fréquence moyenne de passage en tout état j converge vers π_j :*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_\nu \left(\sum_{m=0}^{n-1} 1_{\{X_m=j\}} \right) = \pi_j \quad \forall j \in E. \quad (2.5.22)$$

Démonstration. Soit Π la matrice carrée dont toutes les lignes sont égales à π . Alors on a $\Pi P = \Pi$, du fait que $P.1 = 1$ on a $P\Pi = \Pi$. Donc :

$$(I + P + \dots + P^{n-1})(I - P + \Pi) = I - P^n + n\Pi. \quad (2.5.23)$$

Montrons que la matrice $I - P + \Pi$ est inversible. Soit x un vecteur colonne tel que $(I - P + \Pi)x = 0$. Alors on a

$$0 = \pi(I - P + \Pi)x = \underbrace{\pi(I - P)x}_{=0} + \pi\Pi x = \pi x. \quad (2.5.24)$$

Il suit que $\pi x = 0$. Or tous les π_i sont strictement positifs donc $x = 0$. On déduit que la matrice $I - P + \Pi$ est inversible. Soit $Z = (I - P + \Pi)^{-1}$. Comme $\pi(I - P + \Pi) = \pi$, on a aussi $\pi = \pi Z$ et $\Pi = \Pi Z$. En multipliant (3.5.24) à droite par Z , il vient

$$I + P + \dots + P^{n-1} = (I - P^n)Z + n\Pi Z = (I - P^n)Z + n\Pi.$$

Or nous avons, pour tout état initial i ,

$$\frac{1}{n} \mathbb{E}_i \left(\sum_{m=0}^{n-1} 1_{\{X_m=j\}} \right) = \frac{1}{n} \sum_{m=0}^{n-1} (P^m)_{ij} = \left[\frac{1}{n} (I - P^n)Z + \Pi \right]_{ij}.$$

Comme les éléments de la matrice P^n sont bornés par 1, cette quantité converge vers $(\Pi)_{ij} = \pi_j$, lorsque $n \rightarrow \infty$. ■

Théorème 2.5.54. *Soit (X_n) une chaîne de Markov irréductible de distribution*

stationnaire π . Alors le temps de récurrence moyen en un état $i \in E$ est :

$$\mathbb{E}_i(\tau_i) = \frac{1}{\pi_i}. \quad (2.5.25)$$

Démonstration. Pour $i, j \in E$, on a

$$\begin{aligned} \mathbb{E}_i(\tau_j) &= \mathbb{P}_i(\tau_j = 1) + \sum_{n \geq 2} n \mathbb{P}_i(\tau_j = n) \\ &= P_{ij} + \sum_{n \geq 2} n \sum_{k \neq j} \mathbb{P}_i(\tau_j = n, X_1 = k) \\ &= P_{ij} + \sum_{n \geq 2} n \sum_{k \neq j} p_{ik} \mathbb{P}_k(\tau_j = n - 1) \\ &= P_{ij} + \sum_{k \neq j} p_{ik} \sum_{m \geq 1} (m + 1) \mathbb{P}_k(\tau_j = m) \\ &= P_{ij} + \sum_{k \neq j} p_{ik} \left[\mathbb{E}_k(\tau_j) + \underbrace{\sum_{m \geq 1} \mathbb{P}_k(\tau_j = m)}_{=1} \right] \\ &= 1 + \sum_{k \neq j} p_{ik} \mathbb{E}_k(\tau_j). \end{aligned}$$

Par conséquent,

$$1 - \mathbb{E}_i(\tau_j) = - \sum_{k \neq j} p_{ik} \mathbb{E}_k(\tau_j) = - \sum_{k \in E} (1 - \delta_{kj}) p_{ik} \mathbb{E}_k(\tau_j).$$

Il suit que

$$\begin{aligned}
1 - \pi_j \mathbb{E}_j(\tau_j) &= \sum_{i \in E} \pi_i [1 - \delta_{ij} \mathbb{E}_i(\tau_j)] \\
&= \sum_{i \in E} \pi_i [1 - \mathbb{E}_i(\tau_j) + (1 - \delta_{ij}) \mathbb{E}_i(\tau_j)] \\
&= \sum_{i \in E} \pi_i \sum_{k \in E} (\delta_{kj} - 1) p_{ik} \mathbb{E}_k(\tau_j) + (1 - \delta_{ij}) \mathbb{E}_i(\tau_j) \\
&= \sum_{k \in E} \sum_{i \in E} \pi_i (\delta_{ik} - p_{ik}) (1 - \delta_{kj}) \mathbb{E}_k(\tau_j) = 0
\end{aligned}$$

La somme sur i s'annule, car $\pi_k = \sum_i \pi_i p_{ik}$. ■

Exemple 2.5.55. Dans le cas du modèle d'Ehrenfest avec N boules, le temps de récurrence moyen vers l'état à i boules est donné par

$$\mathbb{E}_i(\tau_i) = \frac{1}{\nu_i} = 2^N \frac{i!(N-i)!}{N!}.$$

En particulier, le temps moyen entre configurations où toutes les boules sont dans l'urne de gauche est de 2^N . Ce temps devient gigantesque pour des nombres de boules de l'ordre du nombre d'Avogadro, c'est-à-dire du nombre de molécules dans un échantillon d'une mol de gaz. Ce modèle simple peut donc justifier pourquoi, lorsque deux récipients contenant des gaz sont mis en contact, on n'observe jamais toutes les molécules dans le même récipient.

2.6 Etats récurrents et transients

On se donne une chaîne de Markov (X_n) à valeurs dans E , définie sur l'espace de probabilité (Ω, \mathcal{F}, P) . On notera \mathcal{F}_n la tribu engendrée par (X_0, X_1, \dots, X_n) c'est-à-

dire :

$$\begin{aligned}\mathcal{F}_n &= \{ \{w \in \Omega : (X_0(w), X_1(w), \dots, X_n(w)) \in B_n\}, B_n \in \mathcal{P}(E^{n+1}) \} \\ &= \sigma(X_0^{-1}(\mathcal{P}(E)) \cup X_1^{-1}(\mathcal{P}(E)) \cup \dots \cup X_n^{-1}(\mathcal{P}(E))).\end{aligned}$$

c'est l'ensemble des événements se produisant jusqu'à l'instant n .

Remarque 2.6.56. $(\mathcal{F}_n)_{n \in \mathbb{N}}$ est famille croissante au sens de l'inclusion

Définition 2.6.57. (*temps d'arrêt*). Une variable aléatoire T à valeurs dans $\mathbb{N} \cup +\infty$ est appelée temps d'arrêt pour la chaîne de Markov si pour tout $n \in \mathbb{N}$

$$\{T = n\} \in \mathcal{F}_n$$

ou de manière équivalente $\{T \leq n\} \in \mathcal{F}_n$.

Cela signifie qu'en observant la chaîne jusqu'à l'instant n , on peut décider si

$\{T = n\}$ a lieu ou non.

Autrement dit, cet événement ne dépend que des variables aléatoires X_0, X_1, \dots, X_n .

Exemple 2.6.58. On définit la variable S_x par :

$$S_x = \inf\{n \in \mathbb{N} : X_n = x\}$$

avec la convention $\inf \emptyset = +\infty$.

S_x représente le temps de premier passage à l'état x de la chaîne. Il est clair que S_x est un temps d'arrêt :

$$\{S_x = n\} = \{X_0 \neq x\} \cap \{X_1 \neq x\} \cap \dots \cap \{X_{n-1} \neq x\} \cap \{X_n = x\} \in \mathcal{F}_n$$

Exemple 2.6.59. On définit la variable T_x par :

$$T_x = \inf\{n > S_x : X_n = x\}.$$

T_x représente le temps du premier retour à l'état x de la chaîne. Alors T_x est un temps d'arrêt.

Exemple 2.6.60. On définit la variable L_x par :

$$L_x = \sup\{n \in \mathbb{N} : X_n = x\}.$$

L_x représente le temps du dernier passage à l'état x de la chaîne. L_x n'est pas un temps d'arrêt.

Désignons par τ_j le temps d'atteindre l'état j de la chaîne à partir de l'instant 1. Autrement dit :

$$\tau_j = \inf\{n \geq 1 : X_n = j\}.$$

Ce temps d'atteinte est un temps d'arrêt de la chaîne. Rappelons que cela signifie que pour tout $n \geq 1$, l'événement $\{\tau_j = n\}$, qui est égal à $\{X_1 \neq j, \dots, X_{n-1} \neq j, X_n = j\}$, ne dépend que de X_1, \dots, X_n .

Formules de conditionnement

Dans les énoncés suivants, A désigne un évènement appartenant à la tribu

$\mathcal{F}_n (n \geq 0)$, une tribu qui est engendrée par le vecteur (X_0, \dots, X_n) . L'évènement A est une réunion dénombrable d'évènements, disjoints deux à deux, de la forme $\{X_0 = i_0, \dots, X_n = i_n\}$.

L'évènement $\{A, X_n = i\} = A \cap \{X_n = i\}$ est alors un évènement de la tribu \mathcal{F}_n . Le présent $\{X_n = i\}$ est donc fixé.

Définition 2.6.61. (loi de probabilité conditionnelle du temps d'atteindre)

Pour tout couple (i, j) d'états et tout $n \geq 1$, on pose :

$$f_{ij}^{(n)} = \mathbb{P}(\tau_j = n \mid X_0 = i). \quad (2.6.26)$$

Ainsi, $f_{i,j}^{(n)}$ ($n \geq 1$) est la probabilité pour que le processus, partant de l'état i , atteigne l'état j , pour la première fois, à l'instant n . Pour tout couple d'états (i, j) , on pose, par convention, $f_{i,j}^{(0)} = 0$.

Théorème 2.6.62. *Pour tout entier $n \geq 1$, on a l'identité :*

$$p_{ij}^{(n)} = \sum_{k=0}^n f_{ij}^{(k)} p_{jj}^{(n-k)}. \quad (2.6.27)$$

Démonstration. Le processus passe de i à j en n étapes, si et seulement s'il passe de i à j pour la première fois en k étapes ($0 \leq k \leq n$) et s'il passe ensuite de j à j en les $(n - k)$ étapes suivantes. Ces chemins, pour des k distincts, sont disjoints et la probabilité pour un k fixé est $f_{ij}^{(k)} p_{jj}^{(n-k)}$.

Ce raisonnement intuitif peut être rendu rigoureux de la façon suivante.

Comme, pour $n \geq 1$

$$\{X_n = j\} = \sum_{k=1}^{n-1} \{\tau_j = k, X_n = j\} + \{\tau_j = n\},$$

On en déduit

$$\begin{aligned} \mathbb{P}(X_n = j \mid X_0 = i) &= \sum_{k=1}^{n-1} \mathbb{P}(\tau_j = k, X_n = j \mid X_0 = i) + f_{ij}^{(n)} \\ &= \sum_{k=1}^{n-1} \mathbb{P}(\tau_j = k \mid X_0 = i) \mathbb{P}(X_n = j \mid \tau_j = k, X_0 = i) + f_{ij}^{(n)}. \end{aligned}$$

Or, pour $1 \leq k \leq n - 1$, l'événement $\{\tau_j = k, X_0 = i\}$ est de la forme $\{A, X_k = j\}$,

où A appartient à $\mathcal{F}_{n-1} \subset \mathcal{F}_n$. Par conséquent,

$$\mathbb{P}(X_n = j | \tau_j = k, X_0 = i) = \mathbb{P}(X_n = j | A, X_k = j) = p_{jj}^{(n-k)},$$

Comme $\mathbb{P}\{X_n = j | X_0 = i\} = p_{ij}^{(n)}$ et $\mathbb{P}\{\tau_j = k | X_0 = i\} = f_{ij}^{(k)}$, il résulte

$$p_{ij}^{(n)} = \sum_{k=1}^{n-1} f_{ij}^{(k)} p_{jj}^{(n-k)} + f_{ij}^{(n)} = \sum_{k=0}^n f_{ij}^{(k)} p_{jj}^{(n-k)}.$$

■

D'après (3.6.27) on peut déterminer les $f_{ij}^{(n)}$ par récurrence à partir des $p_{ij}^{(n)}$:

$$\begin{aligned} f_{ij}^{(1)} &= p_{ij}^{(1)}; \\ f_{ij}^{(n)} &= p_{ij}^{(n)} - \sum_{k=1}^{n-1} f_{ij}^{(k)} p_{jj}^{(n-k)} \quad (n \geq 2). \end{aligned} \tag{2.6.28}$$

Posons

$$f_{ij} = \mathbb{P}\{\tau_j < +\infty | X_0 = i\} = \sum_{n \geq 1} f_{ij}^{(n)}. \tag{2.6.29}$$

C'est la probabilité pour que le processus, partant de i , passe par j au moins une fois au cours du temps ; si $i = j$, le nombre $f_{j,j}$ est la probabilité pour que le processus, partant de j retourne en j , au moins une fois au cours du temps.

Définition 2.6.63. On dit que l'état j est récurrent, si $f_{jj} = 1$. On dit qu'il est transient ou transitoire, si $f_{j,j} < 1$.

Théorème 2.6.64. (*Critère de récurrence*). Soit $j \in E$ Alors

$$j \text{ est recurrent} \iff \sum_{n \geq 0} p_{jj}^{(n)} = +\infty \tag{2.6.30}$$

et que :

$$j \text{ est transient} \iff \sum_{n \geq 0} p_{jj}^{(n)} < +\infty \quad (2.6.31)$$

Remarque 2.6.65. Ces formules ont une interprétation intuitive :

Notons $N_j = \sum_{n \geq 0} 1_{\{X_n=j\}}$ le nombre de retours dans l'état j après l'instant 0. Alors le nombre moyen $\mathbb{E}[N_j] = \sum_{n \geq 0} p_{jj}^{(n)}$ est infini si et seulement si j est récurrent.

Il y a plusieurs techniques pour démontrer ce théorème. Nous utilisons le lemme d'Abel.

Lemme 2.6.66. (Lemme d'Abel)

(1) Si la série de terme général α_n converge et a pour somme α , alors

$$\lim_{\substack{s \rightarrow 1 \\ s \leq 1}} \sum_{n \geq 0} \alpha_n s^n = \alpha.$$

(2) Si les α_n sont positifs et si $\lim_{\substack{s \rightarrow 1 \\ s \leq 1}} \sum_{n \geq 0} \alpha_n s^n = \alpha \leq +\infty$, alors la série de termes général α_n a pour somme α .

Afin de démontrer ce théorème, nous considérons les fonctions génératrices

$$P_{i,j}(s) = \sum_{n \geq 0} p_{ij}^{(n)} s^n, \quad F_{i,j}(s) = \sum_{n \geq 0} f_{ij}^{(n)} s^n, \quad (2.6.32)$$

et à l'aide de la relation de récurrence (3.6.28), établissons une relation fonctionnelle entre elles.

On a :

$$\begin{aligned} P_{jj}(s) &= 1 + \sum_{n \geq 1} p_{jj}^{(n)} s^n = 1 + \sum_{n \geq 1} s^n \sum_{0 \leq k \leq n} f_{jj}^{(k)} p_{jj}^{(n-k)} \\ &= 1 + \sum_{n \geq 0} s^n \sum_{0 \leq k \leq n} f_{jj}^{(k)} p_{jj}^{(n-k)} \quad (\text{car } f_{jj}^{(0)} = 0) \\ &= 1 + F_{jj}(s) P_{jj}(s). \end{aligned} \quad (2.6.33)$$

Lorsque $i \neq j$, il vient de même : $P_{ij}(s) = F_{ij}(s)P_{jj}(s)$. Se rappelant que $p_{ij}^{(0)} = \delta_{ij}$.
Donc le lemme suivant est démontré.

Lemme 2.6.67. *On a les identités*

$$P_{jj}(s) = \frac{1}{1 - F_{jj}(s)}, \quad P_{ij}(s) = F_{ij}(s)P_{jj}(s) \quad (i \neq j) \quad (2.6.34)$$

Proposition 2.6.68. *que l'on peut réunir en une seule formule :*

$$P_{ij}(s) = \delta_{i,j} + F_{ij}(s)P_{jj}(s) . \quad (2.6.35)$$

Le théorème (2.6.64) se démontre alors ainsi : supposons que j est récurrent, donc $\sum_{n \geq 0} f_{jj}^{(n)} = 1$. D'après le Lemme d'Abel partie(1), $\lim_{s \rightarrow 1, s \leq 1} \sum_{n \geq 0} f_{jj}^{(n)} s^n = 1$ et donc $\lim_{s \rightarrow 1, s \leq 1} F_{jj}(s) = 1$. Donc d'après (3.6.34), il en résulte $\lim_{s \rightarrow 1, s \leq 1} P_{jj}(s) = +\infty$ et donc $\lim_{s \rightarrow 1, s \leq 1} \sum_{n \geq 0} p_{jj}^{(n)} s^n = +\infty$. On peut appliquer alors la partie (2) du Lemme d'Abel et conclure que $\sum_{n \geq 0} p_{jj}^{(n)} = +\infty$.

Réciproquement, en supposant que j est transient, on aura $\sum_{n \geq 0} f_{jj}^{(n)} < 1$, Suivant la même technique, on a $\lim_{s \rightarrow 1, s \leq 1} P_{jj}(s) < +\infty$, d'où , par le Lemme d'Abel partie(2), $\sum_{n \geq 0} p_{jj}^{(n)} < +\infty$.

Exemple 2.6.69. Considérons la chaîne de Markov, dont l'ensemble des états est $E = \{0, 1, 2, 3\}$, et de matrice de transition

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{pmatrix} .$$

En calculons les coefficients diagonaux des matrices P^n pour tout $n \geq 0$, on trouve :

$p_{00}^{(n)} = p_{11}^{(n)} = 1/2, p_{22}^{(n)} = (1/4)^n$ et $p_{33}^{(n)} = 1$. Les séries $\sum p_{00}^{(n)}, \sum p_{11}^{(n)}$ et $\sum p_{33}^{(n)}$ divergent. Les états 0, 1, 3 sont récurrents. La série $\sum p_{22}^{(n)}$ converge, l'état 2 est transient.

Remarque 2.6.70. De tels états sont appelés états de **non-retour** s'il existe des états i tels que pour tout $n \geq 1$ on ait $p_{ii}^{(n)} = 0$.

Exemple 2.6.71. L'état 1 dans la chaîne de Markov, telle que $E = \{0, 1\}$ et $P = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ est un état de non-retour.

Proposition 2.6.72. *Tout état de **non-retour** est **transient**; tout état **absorbant** est **récurrent***

Démonstration. Pour un état j de non-retour, on a $p_{j,j}^{(n)} = 1$ si $n = 0$ et 0 autrement. La série de terme général $p_{j,j}^{(n)}$ est convergente. L'état est donc transient. Pour un état absorbant, tout les termes $p_{j,j}^n$ de la série vaut 1. La série est divergente et l'état est récurrent. ■

Proposition 2.6.73. *Si $i \leftrightarrow j$ et si i est récurrent, alors j est aussi récurrent.*

Démonstration. Comme $i \leftrightarrow j$. On a $p_{i,j}^{(n_1)} > 0$ et $p_{j,i}^{(n_2)} > 0$ pour certains entiers n_1, n_2 . De là,

$$\sum_n p_{jj}^{(n_2+n+n_1)} \geq \sum_n p_{ji}^{(n_2)} p_{ij}^{(n_1)} p_{ii}^{(n)} = p_{ji}^{(n_2)} p_{ij}^{(n_1)} \sum_n p_{ii}^{(n)} = +\infty$$

■

Proposition 2.6.74. *Soit j un état transient. Alors, pour tout état i , on a :*

$$\sum_{n \geq 0} p_{ij}^{(n)} = \delta_{i,j} + f_{ij} \sum_{n \geq 0} p_{jj}^{(n)}; \quad (2.6.36)$$

$$\sum_{n \geq 0} p_{jj}^{(n)} = \frac{1}{1 - f_{jj}}; \quad (2.6.37)$$

$$\sum_{n \geq 1} p_{ij}^{(n)} = \frac{f_{i,j}}{1 - f_{jj}} \quad (i \neq j) \quad (2.6.38)$$

En particulier, la série de terme général $p_{ij}^{(n)}$ est convergente et $p_{ij}^{(n)} \rightarrow 0$, lorsque n tend vers l'infini.

Démonstration. L'identité (2.6.36) découle de l'identité (2.6.35) appliquée pour $s \xrightarrow{\leq} 1$ et du théorème (2.6.64). Comme la série de terme général $p_{jj}^{(n)}$ converge vers $\beta = \sum_{n \geq 0} p_{jj}^{(n)}$, on a $\lim_{s \xrightarrow{\leq} 1} P_{jj}(s) = \beta$, par le Lemme d'Abel.

La série de terme général $f_{ij}^{(n)}$ a pour somme $f_{ij} \leq 1$. On a donc aussi $\lim_{s \xrightarrow{\leq} 1} F_{ij}(s) = f_{ij}$. D'où $\lim_{s \xrightarrow{\leq} 1} P_{ij}^{(s)} = \delta_{ij} + f_{ij}\beta$ et $\sum_{n \geq 0} p_{ij}^{(n)} = \delta_{ij} + f_{ij}\beta < \infty$.

Les identités (2.6.37) et (2.6.38) sont des conséquences de la première formule (2.6.36)

■

Proposition 2.6.75. *Si une chaîne de Markov à un nombre fini d'états, elle a au moins un état récurrent.*

Démonstration. Soit N le nombre d'états de la chaîne. Pour tout entier $n \geq 0$, on a $\sum_{1 \leq i, j \leq N} p_{ij}^{(n)} = N$, d'où $\sum_{n \geq 0} \sum_{1 \leq i, j \leq N} p_{ij}^{(n)} = +\infty$, ce qui, d'après la proposition (2.6.73) serait une contradiction, si tous les états étaient transients.

■

Proposition 2.6.76. *Soit j un état récurrent et $k \neq j$ tel que $j \rightsquigarrow k$, alors $k \rightsquigarrow j$, de sorte que k est aussi récurrent et dans la même classe que j . En particulier, une*

chaîne de Markov ne peut aller d'un état récurrent vers un état transient.

Démonstration. Supposons que pour $n \geq 1$ on ait $p_{jk}^{(n)} > 0$. Il s'agit de montrer que l'on a $p_{kj}^{(m)} > 0$ pour un certain $m \geq 1$. Il suffit de montrer la proposition pour $n = 1$, soit $p_{jk} > 0$. Si l'on avait $\mathbb{P}(X_m = j | X_0 = k) = p_{kj}^{(m)} = 0$ pour tout $m \geq 1$, on aurait $\mathbb{P}(N_j = \infty | X_0 = k) \leq \sum_{m \geq 1} p_{kj}^{(m)} = 0$. De là $\mathbb{P}(N_j = \infty | X_0 = j) = \sum_{l \neq k} p_{jl} \mathbb{P}(N_j = \infty | X_0 = l) \leq \sum_{l \neq k} p_{jl} = 1 - p_{jk} < 1$ et j ne serait pas récurrent, contradiction avec l'hypothèse. ■

Proposition 2.6.77. *Si i et j sont dans la même classe récurrente, alors*

$$\mathbb{P}_i(\tau_j < \infty) = \mathbb{P}_j(\tau_i < \infty) = 1. \quad (2.6.39)$$

Démonstration. Soit $A_M = \cup_{m=1}^M \{X_m = j\}$ l'événement “la chaîne visite l'état j lors des M premiers pas”. Alors

$$\lim_{M \rightarrow \infty} \mathbb{P}_j(A_M) = \sum_{m=1}^{\infty} P_j(\tau_j = m) = 1. \quad (2.6.40)$$

Soit n_0 le plus petit entier tel que $\mathbb{P}_j\{X_{n_0} = i\} > 0$. Alors pour tout $M > n_0$,

$$\mathbb{P}_j(A_M \cap \{X_{n_0} = i\}) = \sum_{n=1}^{M-n_0} \mathbb{P}_j(X_{n_0} = i, \tau_j = n_0 + n) \quad (2.6.41)$$

$$\begin{aligned} &= \sum_{n=1}^{M-n_0} \mathbb{P}_j(X_{n_0} = i, j \notin X_{[1, n_0]}) \mathbb{P}_i(\tau_j = n) \\ &\leq \mathbb{P}_j(X_{n_0} = i) \sum_{n=1}^{M-n_0} \mathbb{P}_i(\tau_j = n). \end{aligned} \quad (2.6.42)$$

La première égalité suit du fait que la chaîne ne peut pas retourner en j avant n_0 et visiter i au temps n_0 , par définition de n_0 . En tendant M vers l'infinie des deux

côtés de l'inégalité, on obtient

$$\mathbb{P}_j(X_{n_0} = i) \leq \mathbb{P}_j(X_{n_0} = i)\mathbb{P}_i(\tau_j < \infty). \quad (2.6.43)$$

Comme $\mathbb{P}_j\{X_{n_0} = i\} \neq 0$ et $\mathbb{P}_i\{\tau_j < \infty\} \leq 1$, on a nécessairement $\mathbb{P}_i\{\tau_j < \infty\} = 1$. ■

2.7 Périodicité

Il s'agit d'étudier dans quelles conditions le temps qui sépare deux retours au même état j est ou n'est pas multiple d'un temps minimum. On introduit pour ce faire la notion de **période**.

Définition 2.7.78. Soit j un état de retour ; on appelle **période** de j , le *p.g.c.d* de tous les entiers $n \geq 1$ pour lesquels $p_{j,j}^{(n)} > 0$. On note $d(j)$ la période de j . Si $d(j) = d \geq 2$, on dit que j est périodique de période d ; si $d = 1$, on dit que j est apériodique. Si j est un état de non-retour, à savoir que, pour tout $n \geq 1$, on a $p_{j,j}^{(n)} = 0$, on pose $d(j) = +\infty$.

Théorème 2.7.79. Si i est périodique de période d finie et si $i \leftrightarrow j$, $j \neq i$, alors j est aussi périodique de période d . La propriété de périodicité est une propriété de classe.

Démonstration. Si $i \leftrightarrow j$, alors il existe deux entiers n et m tels que $p_{i,j}^{(n)} > 0$ et $p_{j,i}^{(m)} > 0$. Comme i est de période $d(i) = d$, il existe aussi un entier $s \geq 1$ tel que $p_{i,i}^{(s)} > 0$. On a donc $p_{j,j}^{(m+s+n)} \geq p_{j,i}^{(m)} p_{i,i}^{(s)} p_{i,j}^{(n)} > 0$. Comme $p_{i,i}^{(s)} > 0 \implies p_{i,i}^{(2s)} > 0$, on a aussi $p_{j,j}^{(m+2s+n)} > 0$. La période $d(j)$ divise donc à la fois $m + 2s + n$ et $m + s + n$, donc aussi leur différence s , et en particulier la période $d(i)$ de i . De la même façon,

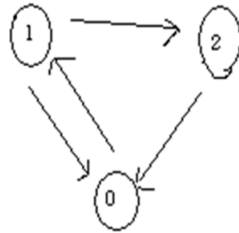
on montre que $d(i)$ divise $d(j)$. Ainsi $d(j) = d(i) = d$. ■

Remarque 2.7.80. toute chaîne régulière est apériodique. En effet, pour tout état i , il existe un état j tel que $p_{i,j} > 0$. Par définition, il existe un temps n tel que $\mathbb{P}_k(X_n = l) > 0$ pour tout $k, l \in E$. Par conséquent, on a $\mathbb{P}_i\{X_n = i\} > 0$ et aussi

$$\mathbb{P}_i(X_{n+1} = i) \geq \mathbb{P}_i(X_1 = j, X_{n+1} = i) = p_{ij}\mathbb{P}_j(X_n = i) > 0$$

Ceci implique que $d_i = \text{pgcd}\{n, n+1\} = 1$.

Exemple 2.7.81. Considérons la chaîne de Markov, à trois états 0, 1, 2, dont le graphe associé est donné dans la figure ci-dessous, où toutes les flèches présentes correspondent à des probabilités de transition strictement positives.



L'état 0 est de retour ; les lacets $0 \rightarrow 1 \rightarrow 0$ et $0 \rightarrow 1 \rightarrow 2 \rightarrow 0$ ont pour longueur 2 et 3, respectivement ; leur p.g.c.d est $d = 1$; l'état 0 est donc apériodique. Maintenant, la chaîne est irréductible. Les deux autres états 1 et 2 sont aussi apériodiques, ce que l'on peut aussi vérifier directement.

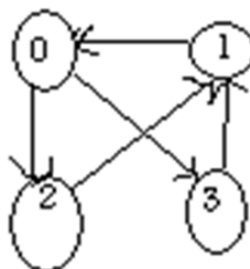
Exemple 2.7.82. Nous allons voir dans le chapitre qui va suivre que dans la premenade sur \mathbb{Z} , tous les états sont périodiques, de période 2.

Exemple 2.7.83. Considérons la chaîne de Markov, dont l'ensemble des états est

$E = \{0, 1, 2, 3\}$ et de matrice de transition :

$$P = \begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

on obtient le graphe



Tous les états communiquent. Il y a donc une seule classe(récurrente). Il y a exactement deux lacets issus de 0 : $0 \rightarrow 2 \rightarrow 1 \rightarrow 0$ et $0 \rightarrow 3 \rightarrow 1 \rightarrow 0$, tous deux de longueur 3. La classe est donc périodique de période 3.

Chapitre 3

Chaîne de Markov sur un ensemble dénombrable

3.1 Marches aléatoires

Les marches aléatoires constituent un exemple relativement simple, et néanmoins très important de chaîne de Markov sur un ensemble dénombrable infini. Dans ce cas en effet, $E = \mathbb{Z}^d$ est un réseau infini, de dimension $d \in \mathbb{N}^*$. D'habitude, on considère que la chaîne démarre en $X_0 = 0$. Ensuite, elle choisit à chaque instant l'un des $2d$ sites voisins, selon une loi fixée d'avance.

Définition 3.1.84. Une marche aléatoire sur \mathbb{Z}^d est une chaîne de Markov à valeurs dans \mathbb{Z}^d , de distribution initiale $\nu = \delta_0$, et de probabilités de transition satisfaisant

$$p_{ij} = 0 \quad \text{si } i = j \quad \text{ou} \quad \|i - j\| > 1 \quad (3.1.1)$$

La marche est dite symétrique si

$$p_{ij} = \frac{1}{2d} \quad \text{pour } \|i - j\| = 1 \quad (3.1.2)$$

Les trajectoires de la marche aléatoire sont des suites de points de \mathbb{Z}^d à distance 1.

Dans le cas symétrique, chaque trajectoire $X_{[0,n]}$, qui est de longueur n , a pour probabilité $(2d)^{-n}$.

Nous allons d'abord déterminer quelques propriétés élémentaires de la loi de (X_n)

Proposition 3.1.85. *Pour la marche aléatoire symétrique sur \mathbb{Z}^d , les variables aléatoires X_n satisfont :*

1. $\mathbb{E}(X_n) = 0$

2. $COV(X_n) = \frac{n}{d} I_d$ (avec I_d la matrice unité d'ordre d)

On plus on a $\frac{X_n}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \frac{1}{d} I_d)$

Démonstration.

(a) Considérons la suite de variables aléatoires $Y_n = X_n - X_{n-1}$ pour $n \in \mathbb{N}^*$.

D'après la définition de la marche aléatoire (X_n) , $Y_n \in F = \{\pm e_j / j = 1, \dots, d\}$, avec e_j le vecteur unité dont la j -ième composante est égale à 1 et toutes les autres composantes soient nulles.

Puisque $\text{card } F = 2d$, la loi de Y_n est $\mathbb{P}_{Y_n}(y) = \frac{1}{2d}$, $\forall y \in F$.

On déduit que les variables aléatoires de Y_n possèdent la même loi. Or toute chaîne de Markov est à accroissement indépendants. Et puisque Y_n sont des accroissements de la chaîne (X_n) , les Y_n sont des variables indépendantes. On déduit que les Y_n sont *i.i.d*.

On a pour tout $n \in \mathbb{N}^*$: $\mathbb{E}(Y_n) = \frac{1}{d} \sum_{j=1}^d \pm e_j = 0$

Par conséquent $\mathbb{E}(Y_n) = 0$.

(b) La covariance de (Y_n) est $cov(Y_n) = (\mathbb{E}(Y_n^i - \mathbb{E}(Y_n^i))(Y_n^j - \mathbb{E}(Y_n^j)))_{1 \leq i, j \leq d}$

avec $Y_n^1, Y_n^2, \dots, Y_n^d$ les composantes de Y_n .

Or on a $\mathbb{E}(Y_n^i) = 0 \quad \forall i = 1, \dots, d$.

D'où $\mathbb{E}(Y_n^i \cdot Y_n^j) = 0 \quad \text{si } i \neq j$.

Si $i = j$ on a : $\mathbb{E}(Y_n^i \cdot Y_n^i) = \mathbb{E}((Y_n^i)^2) = 1^2 \frac{1}{2d} + (-1)^2 \frac{1}{2d} + 0(\frac{2d-2}{2d}) = \frac{1}{d}$

On deduit que $COV(Y_n) = \frac{1}{d} \cdot I_d =$

Calculons maintenant $\mathbb{E}(X_n)$.

On a $X_n = Y_1 + Y_2 + \dots + Y_n$ donc $\mathbb{E}(X_n) = \sum_{k=1}^n \mathbb{E}(Y_k) = 0$.

D'autre part les Y_n sont indépendantes donc

$$cov(X_n) = cov(Y_1 + Y_2 + \dots + Y_n) = \sum_{k=1}^n cov(Y_k) = \frac{n}{d} I_d.$$

Le resultat $\frac{X_n}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \frac{1}{d} I_d)$ est le théorème de la limite centrale.

En effet $X_n = Y_1 + Y_2 + \dots + Y_n$, et Y_1, Y_2, \dots, Y_n sont des variables aléatoires indépendantes et de même loi d'esperance nulle et de covariance égale à $\frac{1}{d} I_d$

alors $\frac{Y_1 + \dots + Y_n}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \frac{1}{d} I_d)$.

■

En conséquence, la position de la marche aléatoire au temps n se trouvera en grande probabilité dans une boule de rayon d'ordre \sqrt{n} autour de l'origine. On dit que la marche aléatoire a un comportement **diffusif** (par opposition à **ballistique**, où la distance à l'origine croîtrait proportionnellement à n).

Considérons le cas unidimensionnel, c'est à dire la marche aléatoire sur \mathbb{Z} .

Proposition 3.1.86. *Soit $(X_n)_{n \in \mathbb{N}}$ une marche aléatoire sur \mathbb{Z} . Alors pour tout $n \in \mathbb{N}$, la loi de X_n est une loi binomiale centrée. C'est à dire définie par*

$$\mathbb{P}(X_n = k) = C_n^{\frac{n+k}{2}} \left(\frac{1}{2}\right)^n \quad \text{pour } k = -n, -n+2, \dots, n-2, n \quad (3.1.3)$$

Démonstration. Calculons $\mathbb{P}(X_n = k)$. Le fait que $X_n = k$, X_n aurait avancé p pas et reculé $n-p$ pas de façon à ce que $(n-p)(-1) + p(1) = k$. Or la probabilité d'avancer p pas et de reculer $(n-p)$ pas est : $C_n^p (\frac{1}{2})^p (\frac{1}{2})^{n-p} = C_n^p (\frac{1}{2})^n$.

On a $(n-p)(-1) + p(1) = k \implies -n + p + p = k \implies 2p = k + n$.

On déduit que : $\mathbb{P}(X_n = k) = \binom{n}{\frac{k+n}{2}} \left(\frac{1}{2}\right)^n$.

On a $k = -n + 2p$ et $-n \leq k \leq n$. Or p prend les valeurs $0, 1, \dots, n$, le nombre k prend les valeurs $-n, -n + 2, \dots, n - 2, n$.

On a

$$\mathbb{E}(X_n) = \sum_{p=0}^n [(-n + 2p)C_n^p + (n - 2p)C_n^{n-p}] = 0.$$

■

Remarque 3.1.87. En particulier, la probabilité que le processus se trouve en 0 au $n^{\text{ème}}$ pas est donnée par :

$$\mathbb{P}(X_n = 0) = \begin{cases} 0 & \text{si } n \text{ est impair} \\ \frac{(2m)!}{2^{2m}(m!)^2} & \text{si } n = 2m \text{ est pair.} \end{cases} \quad (3.1.4)$$

En effet : on a $k = -n, -n + 2, \dots, n - 2, n$; si n est impair k ne peut pas être égal à 0 donc $\mathbb{P}(X_n = 0) = 0$.

Dans le cas où n est pair, $n = 2m$, on a :

$$\mathbb{P}(X_n = 0) = \binom{n}{\frac{n}{2}} \left(\frac{1}{2}\right)^n = \binom{2m}{m} \left(\frac{1}{2}\right)^{2m} = \frac{(2m)!}{(m!)^2} \cdot \frac{1}{2^{2m}}$$

Pour n assez grand. En utilisant la formule de Stirling $n! \simeq \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$, on obtient

$$\mathbb{P}(X_n = 0) \simeq \sqrt{4\pi m} \left(\frac{2m}{e}\right)^{2m} \left(\frac{e}{m}\right)^{2m} \frac{1}{\sqrt{2\pi m}^2} \frac{1}{2^{2m}} \simeq \frac{1}{\sqrt{\pi m}} \simeq \sqrt{\frac{2}{\pi n}}. \quad (3.1.5)$$

En tout temps pair, L'origine est l'endroit le plus probable où trouver la marche aléatoire, mais cette probabilité décroît avec le temps.

Définition 3.1.88. Soit X_n une marche aléatoire sur Z . On pose

$$\tau_o = \inf\{n \in \mathbb{N}^*, \quad X_n = 0\} \quad (3.1.6)$$

τ_0 est le temps du premier retour au point 0 (on a $X_0 = 0$).

Donc τ_0 ne peut prendre que des valeurs paires, de plus, si $\tau_0 = n$ alors $X_n = 0$, donc $\mathbb{P}(\tau_0 = n) \leq \mathbb{P}(X_n = 0)$. En effet, on a :

$$\mathbb{P}(\tau_0 = n) = \mathbb{P}(X_1 \neq 0, X_2 \neq 0, \dots, X_{n-1} \neq 0, X_n = 0) \quad (3.1.7)$$

Théorème 3.1.89. Soit (X_n) une marche aléatoire sur \mathbb{Z} . Alors

$$\mathbb{P}\{\tau_0 = n\} = \begin{cases} \frac{1}{n}P(X_{n-2} = 0) & \text{si } n \text{ est pair} \\ 0 & \text{si } n \text{ est impair.} \end{cases} \quad (3.1.8)$$

Démonstration. Supposons que n est pair :

$$\begin{aligned} & \text{On a } \mathbb{P}(\tau_0 = n) = \mathbb{P}(X_1 \neq 0, X_2 \neq 0, \dots, X_{n-1} \neq 0, X_n = 0) \\ &= \mathbb{P}(X_1 > 0, \dots, X_{n-1} > 0, X_n = 0) + \mathbb{P}(X_1 < 0, \dots, X_{n-1} < 0, X_n = 0) \\ &= 2\mathbb{P}(X_1 > 0, X_2 > 0, \dots, X_{n-1} > 0, X_n = 0) \\ &= 2\mathbb{P}(X_n = 0 / X_1 = 1, X_2 > 0, \dots, X_{n-2} > 0, X_{n-1} = 1) \mathbb{P}(X_1 = 1, X_2 > 0, \dots, X_{n-2} > 0, X_{n-1} = 1) \\ &= 2\mathbb{P}(X_n = 0 / X_{n-1} = 1) \cdot \mathbb{P}(X_1 = 1, X_2 > 0, \dots, X_{n-2} > 0, X_{n-1} = 1) \\ &= 2\mathbb{P}(X_1 = 0 / X_0 = 1) \cdot \mathbb{P}(X_1 = 1, X_2 > 0, \dots, X_{n-2} > 0, X_{n-1} = 1) \\ &= 2 \cdot \frac{1}{2} \mathbb{P}(X_1 = 1, X_2 > 0, \dots, X_{n-2} > 0, X_{n-1} = 1) \\ &= \mathbb{P}(X_1 = 1, X_2 > 0, \dots, X_{n-2} > 0, X_{n-1} = 1) \\ &= \mathbb{P}(X_1 = 1, X_{n-1} = 1) - \mathbb{P}(X_1 = 1, X_{n-1} = 1, \exists m \in \{2, \dots, n-2\} \text{ tel que } X_m = 0) \\ &= \mathbb{P}(X_1 = 1, X_{n-1} = 1) - \mathbb{P}(X_1 = -1, X_{n-1} = 1) \quad (\text{par le principe de réflexion}). \\ &= \mathbb{P}(X_{n-1} = 1 / X_1 = 1) \mathbb{P}(X_1 = 1) - \mathbb{P}(X_{n-1} = 1 / X_1 = -1) \mathbb{P}(X_1 = -1) \\ &= \frac{1}{2} [\mathbb{P}(X_{n-2} = 1 / X_0 = 1) - \mathbb{P}(X_{n-2} = 1 / X_0 = -1)] \\ &= \frac{1}{2} [\mathbb{P}(X_{n-2} = 0 / X_0 = 0) - \mathbb{P}(X_{n-2} = 2 / X_0 = 0)] = \frac{1}{2} [\mathbb{P}(X_{n-2} = 0) - \mathbb{P}(X_{n-2} = 2)]. \end{aligned}$$

D'autre part on a :

$$\begin{aligned} \frac{\mathbb{P}(X_{n-2}=2)}{\mathbb{P}(X_{n-2}=0)} &= \frac{C_{\frac{n-2}{2}}^{\frac{n-2}{2}+2} (\frac{1}{2})^{n-2}}{C_{\frac{n-2}{2}}^{\frac{n-2}{2}} (\frac{1}{2})^{n-2}} = \frac{((\frac{n}{2}-1)!)^2}{(\frac{n}{2}-2)! (\frac{n}{2})!} = \frac{\frac{n}{2}-1}{\frac{n}{2}} = 1 - \frac{2}{n}. \\ \implies \frac{\mathbb{P}(X_{n-2}=2)}{\mathbb{P}(X_{n-2}=0)} &= (1 - \frac{2}{n}) \implies \mathbb{P}(X_{n-2} = 2) = \mathbb{P}(X_{n-2} = 0) (1 - \frac{2}{n}) \end{aligned}$$

$$\text{Donc } \mathbb{P}\{\tau_0 = n\} = \frac{1}{2}[\mathbb{P}(X_{n-2} = 0) - \mathbb{P}(X_{n-2} = 0)(1 - \frac{2}{n})] = \frac{1}{n}\mathbb{P}\{X_{n-2} = 0\} = \frac{C_{\frac{n-2}{2}}^{\frac{n}{2}}}{C_{n-2}^{\frac{n-2}{2}}}.$$

Or on a dit que n est pair donc posons $n = 2m$. Il s'ensuit que :

$$\frac{C_{\frac{2m-2}{2}}^m}{C_{2m-2}^{\frac{2m-2}{2}}} = \frac{m-1}{m} = 1 - \frac{1}{m} = 1 - \frac{2}{n}$$

D'où :

$\mathbb{P}\{\tau_0 = n\} = \frac{1}{2}\mathbb{P}\{X_{n-2} = 0\}[1 - 1 + \frac{2}{n}] = \frac{1}{n}\mathbb{P}\{X_{n-2} = 0\}$ ce qui conclut la démonstration. ■

Corollaire 3.1.90. $\mathbb{E}(\tau_0) = +\infty$

Démonstration. $\mathbb{E}(\tau_0) = \sum_{n \in \mathbb{N}} n\mathbb{P}(\tau_0 = n) = \sum_{m \in \mathbb{N}} 2m\mathbb{P}(\tau_0 = 2m)$

$$= \sum_{m \in \mathbb{N}} 2m \frac{1}{2m} \mathbb{P}(X_{2m-2} = 0) = \sum_{m \in \mathbb{N}} \mathbb{P}(X_{2m-2} = 0) = \sum_{m \in \mathbb{N}} \mathbb{P}(X_m = 0) \quad (*)$$

Or pour m assez grand on a $\mathbb{P}(X_{2m} = 0) \simeq \frac{1}{\sqrt{\pi m}}$. Donc la série $(*)$ est de même nature que la série $\sum_{m \in \mathbb{N}} \frac{1}{\sqrt{\pi m}} = \frac{1}{\sqrt{\pi}} \sum_{m=1}^{\infty} \frac{1}{(m)^{\frac{1}{2}}}$ qui est divergente, car c'est série de Riemann avec $\alpha = 1/2$. Donc $\sum_{m \in \mathbb{N}} \frac{1}{\sqrt{\pi m}} = \infty$. ■

En d'autres termes, la marche aléatoire finit toujours par revenir en 0, mais la loi de τ_0 décroît trop lentement pour que son espérance soit finie. Ceci est dû au fait que si la marche aléatoire s'éloigne beaucoup de 0, elle met longtemps pour y revenir.

Corollaire 3.1.91. *La marche aléatoire symétrique sur \mathbb{Z}^d est récurrente pour $d = 1$ et $d = 2$ et transiente pour $d > 3$.*

3.2 Distributions stationnaires

Nous considérons une chaîne de Markov irréductible sur un ensemble dénombrable E , de matrice de transition $P = (p_{ij})_{i,j \in E}$.

Définition 3.2.92. Une distribution de probabilité π sur E est dite stationnaire si

elle satisfait

$$\pi_j = \sum_{i \in E} \pi_i p_{ij} \quad \forall j \in E \quad (3.2.9)$$

Plus généralement, une mesure μ sur E (qui n'est pas nécessairement une probabilité) satisfaisant $\mu_j = \sum_{i \in E} \mu_i p_{ij}$ pour tout $j \in E$ est appelée une mesure invariante de la chaîne.

Dans le cas où E est fini nous avons vu qu'une chaîne irréductible admettait toujours une distribution stationnaire. Dans le cas infini, ce n'est plus nécessairement le cas.

Nous allons maintenant dériver une condition nécessaire et suffisante pour qu'une chaîne de Markov irréductible admette une distribution stationnaire, qui sera toujours unique dans ce cas. Un rôle important est joué par la quantité

$$\gamma_i^{(k)} = \mathbb{E}_k \left(\sum_{n=1}^{\tau_k} 1_{\{X_n=i\}} \right), \quad (3.2.10)$$

C'est-à-dire le nombre moyen de passage en i entre deux passages en k . Intuitivement, si k est récurrent alors la chaîne revient infiniment souvent en k , et donc $\gamma_i^{(k)}$ devrait mesurer le temps moyen passé en i .

Proposition 3.2.93. *Supposons que la chaîne est irréductible et récurrente. Alors on a $\forall k \in E$:*

1. $\gamma_k^{(k)} = 1$;
2. $\gamma^{(k)}$ est une mesure invariante ;
3. pour tout $i \in E$, on a $0 < \gamma_i^{(k)} < \infty$;
4. $\gamma^{(k)}$ est l'unique mesure invariante telle que $\gamma_k^{(k)} = 1$.

Démonstration.

1) Evidente, puisque $X_{\tau_k} = k$ et $X_n \neq k$ pour $1 \leq n < \tau_k$.

2) Nous avons

$$\gamma_i^{(k)} = \mathbb{E}_k\left(\sum_{n=1}^{\infty} 1_{\{X_n=i, n \leq \tau_k\}}\right) = \sum_{n=1}^{\infty} \mathbb{P}_k(X_n = i, n \leq \tau_k) \quad (3.2.11)$$

$$\begin{aligned} &= \sum_{j \in E} \sum_{n=1}^{\infty} \mathbb{P}_k(X_{n-1} = j, n \leq \tau_k) p_{ji} \\ &= \sum_{j \in E} p_{ji} \sum_{m=0}^{\infty} \mathbb{P}_k(X_m = j, m \leq \tau_k - 1). \end{aligned} \quad (3.2.12)$$

Or la seconde somme dans cette expression peut s'écrire $\mathbb{E}_k\left(\sum_{m=0}^{\tau_k-1} 1_{\{X_m=i\}}\right) = \mathbb{E}_k\left(\sum_{m=1}^{\tau_k} 1_{\{X_m=i\}}\right) = \gamma_j^{(k)}$, vu que $\mathbb{P}_k(X_0 = j) = \delta_{kj} = \mathbb{P}_k(X_{\tau_k} = j)$. Ceci prouve l'invariance de mesure $\gamma^{(k)}$.

3) L'invariance de la mesure implique que pour tout $n \geq 0$,

$$\gamma_i^{(k)} = \sum_{j \in E} \gamma_j^{(k)} P_j\{X_n = i\}. \quad (3.2.13)$$

En particulier, $1 = \gamma_k^{(k)} \geq \gamma_j^{(k)} \mathbb{P}_j(X_n = k)$ pour tout j . Comme par l'irréductibilité, il existe un n tel que $\mathbb{P}_j(X_n = k) > 0$. On déduit donc $\gamma_j^{(k)} < \infty$ pour tout j . D'autre part, on a aussi $\gamma_i^{(k)} \geq \mathbb{P}_k(X_n = i)$, qui est strictement positif pour au moins un n .

4) Soit λ une mesure invariante telle que $\lambda_k = 1$. Alors pour tout j on a

$$\lambda_j = \sum_{i \neq k} \lambda_i p_{ij} + p_{kj} \geq p_{kj}. \quad (3.2.14)$$

Il vient alors, en minorant λ_i par p_{ki} dans l'expression ci-dessus,

$$\lambda_j \geq \sum_{i \neq k} p_{ki} p_{ij} + p_{kj} \quad (3.2.15)$$

$$= P_k\{X_2 = j, \tau_k \geq 2\} + P_k\{X_1 = j, \tau_k \geq 1\}. \quad (3.2.16)$$

Par récurrence, on trouve donc pour tout $n \geq 1$ ($a \wedge b$ désigne le minimum de a et b)

$$\lambda_j \geq \sum_{m=1}^{n+1} P_k\{X_m = j, \tau_k \geq m\} = \mathbb{E}_k\left(\sum_{m=1}^{(n+1) \wedge \tau_k} 1_{\{X_m=j\}}\right)$$

Lorsque n tend vers l'infini, le membre de droite tend vers $\gamma_j^{(k)}$. On a donc $\lambda_j \geq \gamma_j^{(k)}$ pour tout j . Par conséquent $\mu = \lambda - \gamma^{(k)}$ est une mesure invariante, satisfaisant $\mu_k = 0$. Comme $\mu_k = \sum_j \mu_j P_j\{X_n = k\}$ pour tout n , l'irréductibilité implique $\mu_j = 0 \quad \forall j$, donc nécessairement $\lambda = \gamma^{(k)}$. ■

Théorème 3.2.94. *Pour une chaîne de Markov irréductible, les propriétés suivantes sont équivalentes :*

1. *Il existe une distribution stationnaire.*
2. *Il existe un état $k \in E$ tel que*

$$\mu_k = \mathbb{E}_k(\tau_k) < \infty. \tag{3.2.17}$$

3. *La relation (3.2.17) est vérifiée pour tout $k \in E$.*

De plus, si ces propriétés sont vérifiées, alors la distribution stationnaire est unique, et donnée par

$$\pi_i = \frac{1}{\mu_i} \quad \forall i \in E \tag{3.2.18}$$

Démonstration.

2 \Rightarrow 1) Si $\mu_k < \infty$, alors k est récurrent, donc la chaîne, étant irréductible, est récurrente.

Par la proposition précédente, $\gamma^{(k)}$ est l'unique mesure invariante prenant la valeur

1 en k . Or nous avons

$$\sum_{j \in E} \gamma_j^{(k)} = \mathbb{E}_k \left(\underbrace{\sum_{n=1}^{\tau_k} \sum_{j \in E} 1_{\{X_n=j\}}}_{=1} \right) = \mathbb{E}_k(\tau_k) = \mu_k < \infty \quad (3.2.19)$$

Par conséquent, La mesure π définie par $\pi_j = \gamma_j^{(k)} / \mu_k < \infty$ est une mesure de probabilité invariante, c'est-à-dire une loi stationnaire de la chaîne.

1 \Rightarrow 3) Soit π une loi stationnaire, et $k \in E$. Alors $\hat{\gamma} = \pi_j / \pi_k$ est une mesure invariante telle que $\hat{\gamma}_k = 1$. Par la proposition précédente, on a nécessairement $\gamma^{(k)} = \hat{\gamma}$. Il suit par le même calcul que ci-dessus,

$$\mathbb{E}_k(\tau_k) = \sum_{j \in E} \hat{\gamma}_j = \frac{\sum_j \pi_j}{\pi_k} = \frac{1}{\pi_k} < \infty \quad (3.2.20)$$

3 \Rightarrow 2) Evidente.

Dans ce cas, l'unicité de la mesure suit de celle de $\gamma^{(k)}$, et la relation (3.2.18) suit de (3.4.20).

■

Définition 3.2.95. On dit qu'un état i est récurrent positif si $\mathbb{E}_i(\tau_i) < \infty$. Sinon il est dit récurrent nul. La chaîne est dite récurrente positive si tous ses états le sont.

Ainsi un état est récurrent positif lorsque le temps d'attente moyen pour un retour en i est fini.

Remarque 3.2.96. Une chaîne irréductible admet donc une loi stationnaire si et seulement si elle est récurrente positive.

Par conséquent, on distingue dans le cas d'une chaîne irréductible récurrente, deux cas de figures :

- Le cas récurrent positif : tous les états sont récurrents positifs et il existe une

unique probabilité invariante.

- Le cas récurrent nul : tous les états sont récurrents nuls et toutes les mesures invariantes sont de masse infinie.

Il est clair que lorsque $\text{card } E < \infty$, le cas récurrent nul disparaît et tout état $i \in E$ récurrent est récurrent positif.

Remarque 3.2.97. Pour $d \geq 3$ on sait que tout état de \mathbb{Z}^d est un état transient.

Donc dans ce cas la promenade aléatoire ne possède pas de loi stationnaire.

Dans le cas où $n = 1, 2$ on a le résultat suivant :

Théorème 3.2.98. *La marche aléatoire symétrique est récurrente nulle en dimensions $d = 1$ et $d = 2$.*

Démonstration. Supposons que la chaîne est récurrente positive donc elle possède une loi stationnaire unique.

Soit π cette loi. On a alors $\forall j \in \mathbb{Z}; \sum_{i \in \mathbb{Z}} \pi_i p_{ij} = \pi_j$.

Notons $T_s(\pi)$ le translaté de π avec $s \in \mathbb{Z}$.

On a $\forall j \in \mathbb{Z}$:

$$T_s(\pi_j) = \pi_{j+s}.$$

Posons $\tilde{\pi} = T_s(\pi)$. Montrons que $\tilde{\pi}$ est aussi une loi stationnaire.

On a $\forall j \in \mathbb{Z}$

$$\sum_{i \in \mathbb{Z}} \tilde{\pi}_i p_{ij} = \sum_{i \in \mathbb{Z}} \pi_{i+s} p_{ij}$$

or $p_{ij} = p_{i+s, j+s}$, donc $\sum_{i \in \mathbb{Z}} \tilde{\pi}_i p_{ij} = \sum_{i \in \mathbb{Z}} \pi_{i+s} p_{i+s, j+s} = \pi_{j+s} = \pi_j$.

Or la chaîne possède une loi stationnaire unique et que

$$\forall j \in \mathbb{Z}, \tilde{\pi}_j = \pi_j.$$

Donc $\pi_{j+s} = \pi_j$. On a alors

$$\pi_{j+s} = \pi_j \quad \forall s, j \in \mathbb{Z}.$$

On déduit que π est une loi uniforme sur \mathbb{Z} . Ce qui est impossible, car \mathbb{Z} est un ensemble infini dénombrable.

On déduit que la chaîne ne peut pas être récurrente positive pour $n = 1, 2$. ■

3.3 Convergence vers la distribution stationnaire

Dans le cas fini, nous avons montré que si la chaîne était régulière, alors la loi de X_n convergeait vers la distribution stationnaire. Dans le cas d'un espace infini, une chaîne de Markov ne peut jamais être régulière : les probabilités de transition étant sommables, elles ne peuvent être minorées par une quantité strictement positive. Il s'avère toutefois que la récurrence positive et l'apériodicité suffisent à garantir la convergence vers la distribution stationnaire.

Théorème 3.3.99. *Soit $(X_n)_{n \geq 0}$ une chaîne de Markov irréductible, apériodique et récurrente positive, et soit π son unique distribution stationnaire. Alors pour toute distribution ν initiale, on a*

$$\lim_{n \rightarrow \infty} \mathbb{P}_\nu(X_n = j) = \pi_j, \quad \forall j \in E.$$

Démonstration. Nous allons généraliser la preuve de Doeblin, déjà vue dans le cas fini.

Nous introduisons une chaîne de Markov $(X_n, Y_n)_{n \geq 0}$ sur $E \times E$, de probabilités de transition

$$p_{(i,j),(k,l)}^* = p_{ik}p_{jl}, \quad (3.3.21)$$

et de distribution initiale $\rho = \nu \otimes \pi$. Dans ce cas, (X_n) et (Y_n) sont deux chaînes indépendantes de matrice de transition P , et de distributions initiales ν et π .

Le seul point non trivial de la généralisation est de montrer que P^* est irréductible et apériodique. Pour cela, fixons un état $k \in E$. Considérons d'abord l'ensemble

$$\Gamma_k = \{n \in \mathbb{N} : \mathbb{P}_k(X_n = k) > 0\} \quad (3.3.22)$$

La propriété de Markov implique que si $n, m \in \Gamma_k$, alors $n + m \in \Gamma_k$. D'autre part, par définition d'apériodicité, $p \text{ gcd } \Gamma_k = 1$. Alors il existe un n_0 tel que pour tout $t \geq n_0$ appartienne à Γ_k . En effet pour cela, supposons d'abord qu'il existe $n, m \in \Gamma_k$ premiers entre eux. Par le théorème de Bezout, il existe des entiers $p, q \geq 1$ tels que $pm - qn = \pm 1$. Quitte à intervertir n et m , on peut supposer que $pn - qm = 1$. Soit $n_0 = qnm$. Alors pour $1 \leq r \leq n$, on a $n_0 + r = qnm + r(pn - qm) = qm(n - r) + rpn \in \Gamma_k$. Il suffit alors d'écrire tout $t > n_0$ comme $t = n_0 + r + ns$ avec $1 \leq r \leq n$ pour conclure que $t \in \Gamma_k$.

Il se pourrait que $p \text{ gcd } \Gamma_k = 1$ sans que cet ensemble ne contienne deux entiers premiers entre eux. Mais par le théorème de Bezout, il existe forcément un ensemble d'éléments de Γ_k dont une combinaison linéaire vaut 1, et le raisonnement ci-dessus s'adapte facilement à ce cas.

Fixons des états $i, j, k, l \in E$. P étant supposée irréductible, il existe $r \in \mathbb{N}$ tel que $\mathbb{P}_i(X_r = k) > 0$. Comme pour tout $n \geq n_0$,

$$\mathbb{P}_i(X_{r+n} = k) \geq \mathbb{P}_i(X_r = k) \mathbb{P}_k(X_n = k) > 0, \quad (3.3.23)$$

il suit que $\mathbb{P}_i\{X_n = k\} > 0$ pour tous les $n \geq n_0 + r$. Pour des raisons similaires, il existe $m_0, s \in \mathbb{N}$ tels que $\mathbb{P}_j(X_m = l) > 0$ pour tous les $m > m_0 + s$. Par conséquent, il existe un temps M tel que $\mathbb{P}_{(i,j)}^*((X_t, Y_t) = (k, l)) > 0$ pour tous les $t \geq M$. Ceci implique que la chaîne composée est irréductible et apériodique.

Comme la chaîne composée admet manifestement la loi stationnaire $\pi \otimes \pi$, le Thé-

rème (3.2.93) implique qu'elle est récurrente positive.

Le reste de la preuve est identique au cas fini. On introduit le temps τ_A de premier passage sur la diagonale $A = \{(i, j) : i \in E\}$, et on montre comme dans le cas fini que

$$|\mathbb{P}_\nu\{X_n = j\} - \pi_j| \leq 2\mathbb{P}_\rho\{\tau_A > n\}. \quad (3.3.24)$$

La Proposition (2.6.76) implique que A est fini presque sûrement, et donc que la différence ci-dessus tend vers zéro pour $n \rightarrow \infty$. ■

3.4 Chaînes de Markov réversibles

Définition 3.4.100. Soit P une matrice stochastique. Un vecteur $\alpha = \{\alpha_i\}_{i \in E} \in [0, \infty)^E$, $\alpha \neq 0$, est dit réversible par rapport à P si

$$\alpha_i p_{ij} = \alpha_j p_{ji} \quad (3.4.25)$$

Une chaîne de Markov est dite réversible si sa matrice admet un vecteur réversible.

La condition (3.4.25) est appelée condition d'équilibre détaillé en physique. Elle signifie que si les états i et j sont occupés avec probabilités proportionnelles i et j respectivement, alors les taux de transition de i à j et de j à i sont égaux.

Théorème 3.4.101. Soit P une matrice stochastique et $\alpha \in [0, \infty)^E$ un vecteur non nul.

1. Si α est réversible par rapport P , alors α est une mesure invariante.
2. Si α est réversible par rapport P , et $\sum_{j \in E} \alpha_j < \infty$, alors la mesure π définie par $\pi_i = \alpha_i / \sum_{j \in E} \alpha_j$ est une distribution stationnaire.

3. Si π est une distribution stationnaire, alors

$$\mathbb{P}_\pi(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \mathbb{P}_\pi(X_0 = i_n, X_1 = i_{n-1}, \dots, X_n = i_0) \quad (3.4.26)$$

pour tout $n \in \mathbb{N}$ et tout choix de $i_0, \dots, i_n \in E$.

Démonstration.

1. On a

$$\sum_{i \in E} \alpha_i p_{ij} = \alpha_j \sum_{i \in E} p_{ji} = \alpha_j$$

2. Suit immédiatement de 1.

3. Par le théorème (2.2.19)

$$\begin{aligned} \mathbb{P}_\pi(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) &= \pi_{i_0} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n} \\ &= p_{i_1 i_0} \pi_{i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n} \\ &\quad \vdots \end{aligned} \quad (3.4.27)$$

$$= p_{i_1 i_0} p_{i_2 i_1} \dots p_{i_n i_{n-1}} \pi_{i_n} \quad (3.4.28)$$

■

La relation (3.4.26) signifie qu'une trajectoire a la même probabilité que la trajectoire renversée dans le temps. C'est ce qui justifie le terme de réversibilité.

Remarque 3.4.102. Il est clair que si (3.4.25) a lieu, alors α est une mesure invariante. Par contre, la réciproque est fausse. En effet, considérons une chaîne irréductible récurrente positive de mesure invariante α . Alors s'il existe $i, j \in E$ tels que $P_{ij} > 0$ et $P_{ji} = 0$, la relation (3.4.25) n'est pas vérifiée. Alors P n'est pas réversible par rapport à α .

Chapitre 4

Le processus ponctuel de Poisson

Le processus de Poisson sur la droite est un processus à temps continu et à valeurs entières positives. On dit encore que c'est un processus de comptage, que l'on note $\{N(t) : t \geq 0\}$. Il s'agit d'étudier le nombre aléatoire $N(t)$ de certains évènements qui se produisent dans un intervalle de temps $[0, t]$ donné. Sa grande popularité dans les applications vient notamment du fait que beaucoup de calculs le concernant sont explicites.

4.1 Définition et premières propriétés

On se propose d'étudier la répartition dans le temps d'instantanés aléatoires, appelés instantanés d'arrivée. Dans les applications, ce sont les instantanés où se produisent certains évènements spécifiques, comme, par exemple : les arrivées de clients devant un guichet, les émissions de particules radioactives, les appels dans un centrale téléphonique, etc. Par analogie avec le premier exemple cité, on appelle ces évènements des tops, qui se produisent donc aux dits instantanés d'arrivée.

Le processus peut être caractérisé de plusieurs manières différentes. Une réalisation

peut être spécifiée par une suite croissante de nombres réels positifs

$$X_0 = 0 < X_1 < X_2 < X_3 < \dots,$$

désignant les points dans \mathbb{R}_+ . Alternativement, on peut décrire une réalisation en donnant le nombre de points $N_I(\omega)$ contenus dans chaque intervalle I de la forme : $I =]t, t + s]$. Si nous abrégeons $N_{]0,t]}$ par N_t (communément appelé *fonction de comptage*), nous aurons $N_{]t,t+s]} = N_{]0,t+s]} - N_{]0,t]}$, et les N_t sont donnés en fonction des X_n par

$$N_t(\omega) = \sup\{n \geq 0 : X_n(\omega) \leq t\}.$$

Inversement, les X_n se déduisent des N_t par la relation

$$X_n(\omega) = \inf\{t \geq 0 : N_t(\omega) \geq n\}.$$

Nous allons voir deux constructions équivalentes du processus de Poisson. La première construction part de la distribution des N_t .

4.2 Construction par la fonction de comptage

Définition 4.2.103. (processus de poisson). *Le processus ponctuel de Poisson satisfait les conditions suivantes :*

- a) N_I ne dépend que de la longueur de I , i.e. $N_{]t,t+s]}$ a la même loi que N_s .
- b) Si I_1, \dots, I_k sont deux à deux disjoints, N_{I_1}, \dots, N_{I_k} sont indépendants.
- c) $E(N_I)$ existe pour tout I (de longueur finie).
- d) Il existe un intervalle I tel que $\mathbb{P}(N_I > 0) > 0$.
- e) Absence de points doubles : $\lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} P(N_\varepsilon \geq 2) = 0$.

Supposant qu'un tel processus existe, nous pouvons dériver quelques-unes de ses propriétés

Proposition 4.2.104.

1. Pour tout intervalle borné $I \subset \mathbb{R}_+$, on a $\mathbb{P}(N_I > 0) \leq E(N_t)$.
2. Soit $\alpha(t) = E(N_t)$. Alors il existe $\lambda > 0$ tel que $\alpha(t) = \lambda t$.

Démonstration.

1) Soit I intervalle de \mathbb{R}_+ . $\mathbb{E}(N_I) = \sum_{j=0}^{\infty} j \mathbb{P}(N_I = j) = \sum_{j=1}^{\infty} j \mathbb{P}(N_I = j) \geq \sum_{j=1}^{\infty} \mathbb{P}(N_I = j) = \mathbb{P}(N_I \geq 1)$.

2) Posons $f(x) = \mathbb{E}(N_s)$ pour $s \in \mathbb{R}_+$. f est une application linéaire. En effet :

$$\begin{aligned} \text{a) } \forall s, t \in \overline{\mathbb{R}_+} \quad f(s+t) &= \mathbb{E}(N_{s+t}) = \mathbb{E}(N_{[0, s+t]}) = \mathbb{E}(N_{[0, s]} + N_{[s, s+t]}) \\ &= \mathbb{E}(N_{[0, s]}) + \mathbb{E}(N_{[s, s+t]}) = f(s) + f(t). \end{aligned}$$

b) $\forall a \in \mathbb{R}_+$ et $\forall s \in \mathbb{R}_+$ on a :

$$f(as) = \mathbb{E}(N_{as}) = \mathbb{E}(N_{[0, as]}) = \mathbb{E}(aN_{[0, s]}) = a \mathbb{E}(N_{[0, s]}) = a f(s).$$

On déduit que f est linéaire sur \mathbb{R}_+ . Donc $\exists \lambda \geq 0$ telle que $\forall t \in \mathbb{R}_+ \quad f(t) = \lambda t$.

■

La propriété remarquable du processus de Poisson est alors que les variables aléatoires $N_{[s, s+t]}$ suivent nécessairement une loi de Poisson de paramètre λs .

Théorème 4.2.105. *Si le processus satisfait les 5 conditions de la définition (4.2.102), alors les variables aléatoires $N_{[s, s+t]}$ suivent des lois de Poisson de paramètre λs :*

$$\mathbb{P}(N_{[s, s+t]} = k) = \pi_{\lambda s}(k) = e^{-\lambda s} \frac{(\lambda s)^k}{k!}.$$

Démonstration. Par la propriété a), il suffit de montrer le résultat pour $t = 0$, c'est-à-dire pour N_s . Partageons $]0, s]$ en k intervalles de longueur égale, de la forme

$$]s_{j-1}, s_j] \quad \text{où } s_0 = 0 \text{ et } s_j = \frac{js}{k} \text{ pour } 1 \leq j \leq k.$$

L'idée de la preuve est que pour k suffisamment grand, il est peu probable d'avoir plus d'un point par intervalle, donc la loi de $Y_j^{(k)} = N_{]s_{j-1}, s_j]}$ est à peu près une loi de Bernoulli.

La loi de N_s est donc proche d'une loi binomiale, que l'on peut approximer par la loi de Poisson pour k grand.

Il suit donc des conditions que les $Y_j^{(k)}$ sont i.i.d, de mêmes loi que $N_{s_1} = N_{s/k}$ et on a

$$N_s = \sum_{j=1}^k Y_j^{(k)}.$$

Introduisons alors des variables aléatoires

$$\overline{Y_j^{(k)}} = \begin{cases} 0 & \text{si } Y_j^{(k)} = 0, \\ 1 & \text{si } Y_j^{(k)} \geq 1. \end{cases}$$

les $\overline{Y_j^{(k)}}$ sont également i.i.d, et suivent une loi de Bernoulli. La variable aléatoire

$$\overline{N_s^{(k)}} = \sum_{j=1}^k \overline{Y_j^{(k)}}$$

satisfaisant $\overline{N_s^{(k)}} \leq N_s$, pour tout k , on a

$$\mathbb{P}\{\overline{N_s^{(k)}} \geq m\} \leq \mathbb{P}\{N_s \geq m\}$$

pour tout k et tout m . De plus, $\overline{N_s^{(k)}}$ suit une loi binomiale de paramètre

$$p_k = \mathbb{P}(\overline{Y_j^{(k)}} = 1) = \mathbb{P}(Y_j^{(k)} \geq 1) = \mathbb{P}(N_{s/k} \geq 1).$$

Estimons maintenant la différence entre les lois de $\overline{N_s^{(k)}}$ et N_s . Nous avons

$$\begin{aligned} \mathbb{P}(\overline{N_s^{(k)}} \neq N_s) &= \mathbb{P}(\exists j \in \{1, \dots, k\} : Y_j^{(k)} \geq 2) \\ &\leq \sum_{j=1}^k \mathbb{P}(Y_j^{(k)} \geq 2) \\ &= k\mathbb{P}(Y_1^{(k)} \geq 2) = k\mathbb{P}(N_{s/k} \geq 2). \end{aligned}$$

La condition 5 avec $\varepsilon=s/k$ implique alors

$$\lim_{k \rightarrow \infty} \mathbb{P}(\overline{N_s^{(k)}} \neq N_s) = 0.$$

Comme on a d'une part la minoration

$$\mathbb{P}(N_s = m) = \mathbb{P}(N_s = \overline{N_s^{(k)}} = m) \geq \mathbb{P}(\overline{N_s^{(k)}} = m) - \mathbb{P}(\overline{N_s^{(k)}} \neq N_s),$$

et d'autre part la majoration

$$\begin{aligned} \mathbb{P}(N_s = m) &= \mathbb{P}(N_s = \overline{N_s^{(k)}} = m) + \mathbb{P}(\overline{N_s^{(k)}} \neq N_s = m) \\ &\leq \mathbb{P}(\overline{N_s^{(k)}} = m) + \mathbb{P}(\overline{N_s^{(k)}} \neq N_s), \end{aligned}$$

il suit que

$$\lim_{k \rightarrow \infty} \mathbb{P}(\overline{N_s^{(k)}} = m) = \mathbb{P}(N_s = m).$$

Il reste à montrer que kP_k tend vers λs pour $k \rightarrow \infty$. Si c'est le cas, alors la proposition 1.1.4 montre que N_s suit une loi de Poisson de paramètre λs . Or nous

avons

$$\begin{aligned}
 kp_k &= \mathbb{E}(\overline{N_s^{(k)}}) = \sum_{j=1}^{\infty} j \mathbb{P}(\overline{N_s^{(k)}} = j) = \mathbb{P}(\overline{N_s^{(k)}} = 1) + 2\mathbb{P}(\overline{N_s^{(k)}} = 2) + 3\mathbb{P}(\overline{N_s^{(k)}} = 3) + \dots \\
 &= \sum_{l=1}^{\infty} \sum_{j \geq l} \mathbb{P}(\overline{N_s^{(k)}} = j) = \sum_{l=1}^{\infty} \mathbb{P}(\overline{N_s^{(k)}} \geq l).
 \end{aligned}$$

Or

$$\lambda s = E(N_s) = \sum_{j=0}^{\infty} j P\{N_s = j\} = \sum_{l=1}^{\infty} P(N_s \geq l).$$

On a

$$\lim_{k \rightarrow \infty} kp_k = \lim_{k \rightarrow \infty} \sum_{l=1}^{\infty} P(\overline{N_s^{(k)}} \geq l) = \sum_{l=1}^{\infty} \lim_{k \rightarrow \infty} P(\overline{N_s^{(k)}} \geq l) = \sum_{l=1}^{\infty} P\{N_s \geq l\} = \lambda s \quad (4.2.1)$$

Puisque $\lim_{k \rightarrow \infty} kp_k = \lambda s$ et kp_k sont des paramètres d'une loi binomiale, d'après la proposition (1.1.4), les lois de $\overline{N_s^{(k)}}$ convergent vers la loi de Poisson de paramètre λs . On déduit aussi que la loi de N_s est de Poisson de paramètre λs (car $\overline{N_s^{(k)}}$ converge en loi vers N_s)

Montrons l'égalité suivante :

$$\lim_{k \rightarrow \infty} \sum_{l=1}^{\infty} \mathbb{P}(\overline{N_s^{(k)}} \geq l) = \sum_{l=1}^{\infty} \lim_{k \rightarrow \infty} \mathbb{P}(\overline{N_s^{(k)}} \geq l).$$

En effet :

On a $\overline{N_s^{(k)}} \leq N_s \quad \forall k \in N$, par conséquent : $\{\omega \in \Omega \mid \overline{N_s^{(k)}} \geq l\} \subset \{\omega \in \Omega \mid N_s \geq l\}$.

Donc

$$\chi_{\overline{N_s^{(k)}} \geq l} \leq \chi_{N_s \geq l}; \text{ avec } \chi \text{ fonction indicatrice.}$$

Posons $f_k = \sum_{l=1}^{\infty} \chi_{\overline{N_s^{(k)}} \geq l}$ et $f = \sum_{l=1}^{\infty} \chi_{N_s \geq l}$.

On a $\int f \, dp = \sum_{l=1}^{\infty} P\{N_s \geq l\} = \lambda s$ et $f_k \leq f \quad \forall k \in N$

D'où $\int f_k \, dp < \infty \quad \forall k \in N$. Par conséquent $\lim_k \int f_k \, dp = \int \lim_k f_k \, dp$, d'après le théorème de la convergence dominée de Lebesgue.

On en déduit $\lim_k \sum_{l=1}^{\infty} P(\overline{N_s^{(k)}} \geq l) = \sum_{l=1}^{\infty} \lim_k P(\overline{N_s^{(k)}} \geq l)$ ■

4.3 Construction par les temps d'attente

La seconde construction du processus ponctuel de Poisson se base sur la distribution des différences de position $Z_n = X_n - X_{n-1}$. Celles-ci caractérisent également de manière univoque le processus, via la relation

$$X_n(\omega) = \sum_{j=1}^n Z_j(\omega).$$

Le résultat remarquable est alors que les Z_j sont i.i.d et suivent une loi bien particulière, à savoir une loi exponentielle de paramètre λ .

Théorème 4.3.106. *Pour tout n , les variables aléatoires Z_1, \dots, Z_n sont indépendantes, et suivent la même loi exponentielle $\exp(\lambda)$.*

Démonstration. *Fixons des instants*

$$t_0 = 0 < s_1 < t_1 < s_2 < t_2 < \dots < s_n < t_n.$$

$$\begin{aligned} & \mathbb{P}(X_1 \in]s_1, t_1], X_2 \in]s_2, t_2], \dots, X_n \in]s_n, t_n]) \\ &= \mathbb{P}(N_{]0, s_1]} = 0, N_{]s_1, t_1]} = 1, N_{]t_1, s_2]} = 0, \dots, N_{]t_{n-1}, s_n]} = 0, N_{]s_n, t_n]} \geq 1) \\ &= \prod_{k=1}^n \mathbb{P}(N_{]t_{k-1}, s_k]} = 0) \prod_{k=1}^{n-1} \mathbb{P}(N_{]s_k, t_k]} = 1) \mathbb{P}(N_{]s_n, t_n]} \geq 1) \\ &= \prod_{k=1}^n e^{-\lambda(s_k - t_{k-1})} \prod_{k=1}^{n-1} \lambda(t_k - s_k) e^{-\lambda(t_k - s_k)} [1 - e^{-\lambda(t_n - s_n)}] \\ &= \lambda^{n-1} \prod_{k=1}^{n-1} (t_k - s_k) [e^{-\lambda s_n} - e^{-\lambda t_n}] \end{aligned}$$

$$= \int_{s_1}^{t_1} \int_{s_2}^{t_2} \dots \int_{s_n}^{t_n} \lambda^n e^{-\lambda x_n} dx_n dx_{n-1} \dots dx_2 dx_1.$$

La loi conjointe de (X_1, \dots, X_n) admet donc la densité

$$f(x_1, \dots, x_n) = \begin{cases} \lambda^n e^{-\lambda x_n} & \text{si } 0 < x_1 < \dots < x_n \\ 0 & \text{sinon.} \end{cases}$$

Déterminons maintenant la loi de $Z_k, k \in N^*$

Posons $Z_1 = X_1 - X_0, Z_2 = X_2 - X_1, \dots, Z_n = X_n - X_{n-1}$.

La fonction de répartition de Z_1, Z_2, \dots, Z_n est :

$$\begin{aligned} \forall z_1; \dots, z_n \in R_+, F(z_1, z_2, \dots, z_n) &= P\{Z_1 \leq z_1, Z_2 \leq z_2, \dots, Z_n \leq z_n\} \\ &= P\{X_1 \leq z_1, X_2 - X_1 \leq z_2, \dots, X_n - X_{n-1} \leq z_n\} \\ &= \int_0^{z_1} \int_0^{z_2+x_1} \int_0^{z_3+x_2} \dots \int_0^{z_n+x_{n-1}} f(x_1, x_2, \dots, x_n) dx_n dx_{n-1} \dots dx_2 dx_1 \end{aligned}$$

On a :

$$\frac{\partial F(z_1, \dots, z_n)}{\partial z_1} = \int_0^{z_1+z_2} \int_0^{z_3+x_2} \dots \int_0^{z_n+x_{n-1}} f(z_1, x_2, \dots, x_n) dx_n \dots dx_2$$

et par conséquent :

$$\frac{\partial^2 F(z_1, z_2, \dots, z_n)}{\partial z_2 \partial z_1} = \int_0^{z_1+z_2+z_3} \dots \int_0^{z_n+x_{n-1}} f(z_1, z_2 + z_1, x_3, \dots, x_n) dx_n \dots dx_3$$

Ainsi de suite...

On obtient finalement :

$$\begin{aligned} \frac{\partial^n F(z_1, z_2, \dots, z_n)}{\partial z_n \dots \partial z_1} &= f(z_1, z_1 + z_2, z_1 + z_2 + z_3, \dots, z_1 + z_2 + \dots + z_n) \\ &= \lambda^n e^{-\lambda(z_1 + \dots + z_n)} \\ &= (\lambda e^{-\lambda z_1})(\lambda e^{-\lambda z_2}) \dots (\lambda e^{-\lambda z_n}). \end{aligned}$$

Or cette densité f représente la densité d'une famille de variables aléatoires indépendantes et identiquement distribuées gouvernées par la loi $\lambda e^{-\lambda x}$. ■

Proposition 4.3.107. (Superposition de deux processus de Poisson indépendants).

Soient $\{N_t^i\}_{t \in \mathbb{R}}$, $i = 1, 2$ deux processus de Poisson indépendants de taux respectivement λ_1, λ_2 . Alors le processus $N = N_t^1 + N_t^2$ est encore un processus de Poisson de taux $\lambda = \lambda_1 + \lambda_2$.

Démonstration. N_t est la variable aléatoire représentant le nombre d'arrivées dans l'intervalle $]0, t]$.

On a pour $k \in \mathbb{N}$,

$$\begin{aligned}
 \mathbb{P}(N_t = k) &= \sum_{j=0}^k \mathbb{P}(N_t^1 = j \text{ et } N_t^2 = k - j) \\
 &= \sum_{j=0}^k \mathbb{P}(N_t^1 = j) \cdot \mathbb{P}(N_t^2 = k - j) \\
 &= \sum_{j=0}^k \frac{(t\lambda_1)^j}{j!} e^{-\lambda_1 t} \cdot \frac{(t\lambda_2)^{k-j}}{(k-j)!} e^{-\lambda_2 t} \\
 &= \frac{t^k}{k!} e^{-(\lambda_1 + \lambda_2)t} \sum_{j=0}^k \frac{k! \lambda_1^j \lambda_2^{k-j}}{j!(k-j)!} \\
 &= \frac{1}{k!} ((\lambda_1 + \lambda_2)t)^k e^{-(\lambda_1 + \lambda_2)t}.
 \end{aligned}$$

On déduit que $\forall t > 0$, N_t est une variable aléatoire gouvernée par la loi de Poisson de paramètre $(\lambda_1 + \lambda_2)t$.

On vérifie facilement que (N_t) est à accroissement indépendants et stationnaire par rapport au temps.

On en déduit que (N_t) est un processus de Poisson ■

Proposition 4.3.108. Soient $\{X(t) : t \geq 0\}$ et $\{Y(t) : t \geq 0\}$ deux processus de Poisson indépendants, de taux d'arrivées respectifs λ_X et λ_Y . Alors le nombre d'arrivées du processus $\{Y(t) : t \geq 0\}$ se produisant entre deux arrivées successives du processus $\{X(t) : t \geq 0\}$ suit une distribution géométrique modifiée de paramètre $\frac{\lambda_X}{\lambda_X + \lambda_Y}$.

Remarque 4.3.109. une variable aléatoire Z suit une loi géométrique modifiée de paramètre p ($0 < p < 1$) si $\forall i \in \mathbb{N}$, $P(Z = i) = p(1 - p)^i$.

Démonstration. Soit τ_X la variable aléatoire qui représente le temp d'attente entre deux arrivées de X . Calculons $P(N_{\tau_X}^Y = k)$ pour $k \in \mathbb{N}$.

C'est la loi du nombre d'arrivées de Y entre deux arrivées de X .

$$\begin{aligned}
 \mathbb{P}(N_{\tau_X}^Y = k) &= \int_0^{+\infty} \mathbb{P}(N_t^Y = k) \lambda_X e^{-\lambda_X t} dt = \frac{\lambda_X \lambda_Y^k}{k!} \int_0^{+\infty} t^k e^{-(\lambda_X + \lambda_Y)t} dt \\
 &= \frac{\lambda_X \lambda_Y^k}{k!} \frac{1}{(\lambda_X + \lambda_Y)^{k+1}} \int_0^{+\infty} s^k e^{-s} ds \quad \text{avec } s = (\lambda_X + \lambda_Y)t. \\
 &= \frac{\lambda_X \lambda_Y^k}{k!} \frac{1}{(\lambda_X + \lambda_Y)^{k+1}} \Gamma(k+1) \quad (\text{avec } \Gamma(\alpha) \text{ est fonction d'Euler}) \\
 &= \frac{\lambda_X \lambda_Y^k}{k!} \frac{k!}{(\lambda_X + \lambda_Y)^{k+1}} = \frac{\lambda_X}{(\lambda_X + \lambda_Y)} \left(\frac{\lambda_Y}{\lambda_X + \lambda_Y} \right)^k \\
 &= p(1 - p)^k \quad \text{avec } p = \frac{\lambda_X}{\lambda_X + \lambda_Y}. \quad \blacksquare
 \end{aligned}$$

Proposition 4.3.110. Soit $\{T_n\}_{n \in \mathbb{N}}$, $T_0 = 0$, les temps d'arrivée associés à un processus de Poisson $\{N_t\}_{t \in \mathbb{R}}$ de paramètre λ . et soit B un borélien borné de \mathbb{R} . On introduit la variable aléatoire :

$$N_B = \sum_{n=1}^{+\infty} \mathbf{1}_{T_n \in B}.$$

Alors N_B suit une loi de Poisson de paramètre $\lambda m(B)$, où $m(B)$ est la mesure de Lebesgue de B

Démonstration. B est une réunion de ses composantes connexes

$$B = \bigcup_{j \in I} I_j, \quad I \subset \mathbb{N}.$$

Les composantes connexes de B sont des intervalles. Donc I_j est un intervalle de \mathbb{R} $\forall j \in I$.

a) Supposons que I est finie, donc $B = \bigcup_{j=1}^n I_j$, on a $m(B) = \sum_{j=1}^n l(I_j)$.

D'autre par N_{I_j} suit la loi de Poisson de paramètre $\lambda l(I_j) \forall j \in I$ donc $\sum_{j=1}^n N_{I_j}$ suit

la loi Poisson de paramètre $\lambda \sum_{j=1}^n l(I_j)$.

Or $N_B = \sum_{j=1}^n N_{I_j}$. Donc N_B suit la loi Poisson de paramètre $\lambda \sum_{j=1}^n l(I_j)$

mais $m(B) = \sum_{j=1}^n l(I_j)$ donc N_B suit la loi de Poisson de paramètre $\lambda m(B)$.

■

4.4 Exemples de processus de Poisson

Exemple 4.4.111. (Décomposition d'un processus de Poisson).

Le nombre de chocs $N(t)$ affectant une composante d'un système au cours de l'intervalle de temps $(0, t)$ est gouverné par un processus de Poisson de taux λ . Lorsque la composante subit un choc, elle continue néanmoins à fonctionner avec une probabilité p ; le choc sera par contre fatal avec une probabilité $q = 1 - p$, auquel cas la composante est instantanément remplacée par une nouvelle composante identique. Soit $N_F(t)$ le nombre de chocs fatals survenus au cours de l'intervalle $(0, t)$. Alors $N_F = \{N_F(t); t \geq 0\}$ est un processus de Poisson de taux λq .

Démonstration. Il est clair que le support de $N_F(t)$ est \mathbb{N}

Pour $k \in \mathbb{N}$ on a :

$$\begin{aligned} \mathbb{P}(N_F(t) = k) &= \sum_{l=0}^{\infty} \mathbb{P}(\{N_F(t) = k, N(t) = l\}) = \sum_{l=k}^{\infty} \mathbb{P}(\{N_F(t) = k, N(t) = l\}) \\ &= \sum_{l=k}^{\infty} \mathbb{P}(N_F(t) = k | N(t) = l) \mathbb{P}(N(t) = l) = \sum_{l=k}^{\infty} C_l^k q^k (1-q)^{l-k} \left(\frac{(\lambda t)^l}{l!} e^{-\lambda t} \right) \\ &= (\lambda t)^k q^k \sum_{l=k}^{\infty} \frac{l!}{k!(l-k)!} \frac{(1-q)^{l-k} (\lambda t)^{l-k} e^{-\lambda t}}{l!} = \frac{(\lambda t q)^k}{k!} e^{-\lambda t} \sum_{l \geq k} \frac{((1-q) \lambda t)^{l-k}}{l-k!} \end{aligned}$$

$$= \frac{(\lambda tq)^k}{k!} e^{-\lambda t} \sum_{j \geq 0} \frac{((1-q)\lambda t)^j}{j!} = \frac{(\lambda tq)^k}{k!} e^{-\lambda t} e^{(1-q)\lambda t} = \frac{(\lambda tq)^k}{k!} e^{-q\lambda t}.$$

Puisque $N(t)$ est à accroissements indépendants et stationnaire par rapport au temps on montre facilement que $N_F(t)$ l'est aussi. ■

Exemple 4.4.112. On considère une suite de variables aléatoires positives X_i , $i \in \mathbb{N}^*$, définies sur un même espace $(\Omega, \mathcal{A}, \mathbb{P})$, identiquement distribuées selon une loi de fonction de répartition continue F .

Les résultats de l'exemple pourront être interprétés, par exemple, du point de vue d'un auto-stoppeur qui, placé en un point d'une route, observe qu'il s'écoule un temps X_i entre le passage de la $(i-1)^{i\text{ème}}$ voiture et de la $i^{\text{ème}}$. Si la circulation est fluide, on peut considérer en première approximation que les $\{X_i\}_{i \in \mathbb{N}^*}$ sont indépendantes et de loi exponentielle.

1) On pose :

$$T_0 = 0, \quad T_n = \sum_{t=1}^n X_t, \quad N_t = \sum_{n=1}^{+\infty} 1_{\{T_n \leq t\}}.$$

L'interprétation de ces variables indique que :

- a) T_n : est la variable aléatoire qui représente le temp de la $n^{\text{ième}}$ arrivée.
- b) N_t : est la variable aléatoire qui représente le nombre d'arrivées avant l'instant t .
- c)

$$N_t \geq n \Leftrightarrow T_n \leq t.$$

$N_t \geq n \Rightarrow$ le nombre d'arrivées avant l'instant t a dépassé n donc $T_n \leq t$. (c'est à dire le n arrivée est realise avant t)

$T_n \leq t \Rightarrow$ le n arrivée est réalisé avant t , donc le nombre d'arrivée avant l'instant t est surement plus grand que n .

2) Considerons les variables aléatoires $W_t = T_{N_t+1} - t$ et $Z_t = t - T_{N_t}$. W_t est le temps que l'auto-stoppeur doit attendre à partir de l'instant t pour voir passer

une voiture et Z_t est le temps écoulé entre le dernier passage d'une voiture et l'instant d'arrivée de l'auto-stoppeur qui est t .

a) $W_t + Z_t = (T_{N_t+1} - t) + (t - T_{N_t}) = T_{N_t+1} - T_{N_t}.$

$W_t + Z_t$ est le temp d'attente entre deux arrivées. C'est une variable aléatoire qui possède la même loi que X_1 .

Donc $W_t + Z_t$ est gouvernée par la loi exponentielle de paramètre λ . C'est à dire de densité $f(x) = \lambda e^{-\lambda x}$.

b) Donnons la loi du couple (W_t, Z_t) .

1. Montrons d'abord que W_t et Z_t sont indépendant :

La fonction de repartition de W_t est:

1^{ere} cas : $s > t$

$$\begin{aligned} F_{Z_t}(s) &= \mathbb{P}(Z_t \leq s) \\ &= \mathbb{P}(t - T_{N_t} \leq s) = 1 \end{aligned}$$

D'autre part la fonction de répartition de Z_t est :

2^{eme} cas : $s \leq t$

$$\begin{aligned} F_{Z_t}(s) &= \mathbb{P}(Z_t \leq s) \\ &= \mathbb{P}(t - T_{N_t} \leq s) \\ &= \mathbb{P}(N_{]t-s, t]} \geq 1) \end{aligned}$$

Or $N_{]t-s, t]}$ et $N_{]t, t+s]}$ sont des variables aléatoires indépendantes donc W_t et Z_t sont indépendantes.

2. Déterminons la loi de W_t :

$$\begin{aligned} F_{W_t}(s) &= \mathbb{P}(W_t \leq s) = \mathbb{P}(T_{N_t+1} - t \leq s) \\ &= \mathbb{P}(T_{N_t+1} \leq t + s) = \mathbb{P}(N_{]t, t+s]} \geq 1) \end{aligned}$$

$$= 1 - \mathbb{P}(N_{]t, t+s]} = 0) = 1 - \mathbb{P}(N_s = 0)$$

$$= 1 - e^{-\lambda s}.$$

Donc la densité de la loi de W_t est $\frac{dF_{W_t}(s)}{ds} = \lambda e^{-\lambda s} = f(s)$.

3. Déterminons la loi de Z_t : On a pour $s \geq 0$

1^{ère} cas : $s > t$

$$\begin{aligned} F_{Z_t}(s) &= \mathbb{P}(Z_t \leq s) \\ &= \mathbb{P}(t - T_{N_t} \leq s) = 1 \\ &= \mathbb{P}(\min(X_1, t) \leq s). \end{aligned}$$

2^{ème} cas : $s \leq t$

$$\begin{aligned} F_{Z_t}(s) &= \mathbb{P}(Z_t \leq s) \\ &= \mathbb{P}(t - T_{N_t} \leq s) \\ &= \mathbb{P}(N_{]t-s, t]} \geq 1) \\ &= \mathbb{P}(N_s \geq 1) = 1 - e^{-\lambda s} \\ &= \mathbb{P}(X_1 \leq s) = \mathbb{P}(\min(X_1, t) \leq s) \end{aligned}$$

On déduit que dans les deux cas, on a :

$$\mathbb{P}(Z_t \leq s) = \mathbb{P}(\min(X_1, t) \leq s)$$

Donc Z_t possède la même loi que $\min(X_1, t)$.

Puisque W_t, Z_t sont indépendantes alors $\forall A, B \in \mathcal{B}(\mathbb{R}_+)$ on a :

$$\begin{aligned} \mathbb{P}((W_t, Z_t) \in A \times B) &= \mathbb{P}((W_t \in A) \cap (Z_t \in B)) \\ &= \mathbb{P}(W_t \in A) \mathbb{P}(Z_t \in B). \\ &= \mathbb{P}_{W_t} \otimes \mathbb{P}_{Z_t}(A \times B). \end{aligned}$$

c) Montrons que Z_t converge en loi vers X_1 quand t tend vers l'infini.

En effet, soit $s \in \mathbb{R}_+$

$$\lim_{t \rightarrow +\infty} \mathbb{P}(Z_t \leq s) = \lim_{t \rightarrow +\infty} \mathbb{P}(\min(X_1, t) \leq s) = \mathbb{P}(X_1 \leq s).$$

car t tend vers $+\infty$ donc $t > s$ donc

$$\mathbb{P}(\min(X_1, t) \leq s) = \mathbb{P}(X_1 \leq s).$$

4.5 Généralisations

Il existe plusieurs généralisations du processus ponctuel de Poisson discuté précédemment :

1. **Le processus de Poisson inhomogène** : Dans ce cas le nombre de points $N]t, t + s]$ suit une loi de Poisson de paramètre

$$\int_t^{t+s} \lambda(u) du$$

où $\lambda(u)$ est une fonction positive, donnant le taux au temps u . Ce processus permet de décrire des situations où les points apparaissent avec une intensité variable, par exemple si l'on veut tenir compte des variations journalières du trafic influençant les horaires de passage de bus. On retrouve le processus de Poisson homogène si $\lambda(u)$ est constant.

2. **Le processus de Poisson de dimension $n \geq 2$** : Ce processus peut être défini par sa fonction de comptage, en remplaçant les intervalles I par des sous-ensembles (mesurables) de \mathbb{R}^n . Les nombres de points dans deux ensembles disjoints sont à nouveau indépendants, et le nombre de points dans un ensemble

est proportionnel à son volume. Ce processus peut par exemple modéliser la distribution des étoiles dans une région de l'espace ou du ciel.

3. **Le processus de naissance et de mort :** Le processus ponctuel de Poisson peut être considéré comme un processus de naissance pur : Si N_t est interprété comme le nombre d'individus dans une population au temps t , ce nombre augmente avec un taux constant λ . Plus généralement, dans un processus de naissance et de mort, de nouveaux individus naissent avec un taux λ et meurent avec un taux μ ; éventuellement, $\lambda = \lambda(N_t)$ et $\mu = \mu(N_t)$ peuvent dépendre de la taille actuelle de la population.

Bibliographie

- [1] I. Adan, J. Resing, *Queueing Theory*, notes de cours, Eindhoven University of Technology, 2002.
- [2] C. Bélisle, *Processus de Poisson*, notes de cours, Université Laval
- [3] N. Berglund, *Processus aléatoires et applications*, notes de cours, Université d'Orléans, 2010.
- [4] P. Bougerol, *Processus de Sauts et Files d'Attente*, notes de cours, Université Pierre et Marie Curie, 2002.
- [5] D. Foata, A. Fuchs, *Processus stochastiques*, Dunod, 2006.
- [6] R. Durrett, *Essentials of Stochastic Processes*, Springer, 1999.
- [7] O. Garet, *Probabilités et Modélisation Stochastique*, notes de cours, Université de Nancy 1, 2009.
- [8] C. Graham, *Chaînes de Markov*, Dunod, 2006.
- [9] C. M. Grinstead, J. L. Snell, *Introduction to Probability*, web book, <http://math.dartmouth.edu/~doyle/docs/prob/prob.pdf>
- [10] J. Lacroix, *Chaînes de Markov et Processus de Poisson*, notes de cours, Université Pierre et Marie Curie, 2002
- [11] S. Loustau, *Chaînes de Markov et Processus markoviens de sauts. Application aux files d'attentes*, notes de cours, Ecole Centrale de Marseille, 2009