

الجمهورية الجزائرية الديمقراطية الشعبية  
République Algérienne Démocratique et Populaire

وزارة التعليم العالي و البحث العلمي  
Ministère de l'Enseignement Supérieur et de la Recherche  
Scientifique

Centre Universitaire Salhi Ahmed- Naama  
Institut des sciences et technologies  
Département de Mathématiques et Informatique



## *Mémoire de fin d'études*

En vue de l'obtention du diplôme de Master  
En : Mathématiques

Spécialité : Probabilités, Statistique et Application

## Intitulé

---

### CHAÎNE DE MARKOV ET SIMULATION

---

Présenté par :  
BOUDIA Nesrine

Soutenu : Juillet 2022

Devant le jury composé de :

Dr.LATTI Fethi	C-Univ Naâma	Président
Dr.BELGUERNA Abderrahmane	C-Univ Naâma	Examineur
Dr.MOULAI KHATIR Smain	C-Univ Naâma	Encadreur

Année universitaire 2021/2022

# Dédicace

A mes parents : Vous me rendez heureux et me donnez de la force.

Vous serez toujours celui à suivre.

Papa, j'admire ton courage, ta force et ton honnêteté.

Maman, merci d'être gentille, patiente et dévouée envers nous.

Merci pour tout ce que vous avez fait pour aider vos enfants,

Grandissent et prospèrent

Merci de travailler dur, malgré l'existence de hauts et de bas dans la vie.

Pour que vos enfants soient en sécurité,

Merci d'être mes parents en général.

Tu es la raison pour laquelle j'ai bien fait.

Je suis heureux de vous le donner.

Sabrina, Redouane et Asiya, mes frères et soeur

Comme preuve de la proximité, de l'amour et combien je tiens à toi.

Parce que je veux que tu sois heureux, en bonne santé et que tu réussisses, je te confie cette responsabilité.

À mes amies proches Ayda, Zineb et Maryem, qui sont aussi mes collègues :

Je ne sais pas comment te dire à quel point je tiens à toi.

Dis-moi à quel point je tiens à toi et ce que je pense de toi.

J'ai des frères et sœurs et des amis sur qui je peux compter.

En l'honneur de l'amitié qui nous unit et en l'honneur de

Je vous raconte tous les moments passés ensemble.

Ce travail, et Je vous souhaite longévité et bonne santé.

Bonheur.

A toutes les personnes à qui je tiens...

# Remerciement

Louange à Allah, Le Tout Puissant, Le plus Miséricordieu, pour son aide et sa bénédiction.

Tout d'abord, nous tenons à remercier notre encadrant, le Dr MOULAI KHATIR Smain, pour son aide, son soutien, ses encouragements, sa gentillesse et ses larmes. Les vérifications nécessaires nous ont beaucoup aidés dans ce travail.

Je tiens à remercier les membres du jury d'être là et d'avoir lu attentivement ma thèse. Je tiens également à les remercier pour leurs commentaires lors de la soutenance, qui m'ont permis d'améliorer mon travail.

Mes remerciements vont à notre professeur, le Dr LAALA Zeyneb, qui a été mon meilleur guide et sans l'aide de qui je n'aurais pas pu terminer.

Merci beaucoup, mes collègues. J'espère que vos vies personnelles et professionnelles sont remplies de bonheur et de prospérité.

Tout le monde devrait avoir la meilleure des chances et des espoirs!

# Table des Matières

<b>Introduction</b>	<b>1</b>
<b>Préliminaires</b>	<b>3</b>
0.1 Espace mesuré . . . . .	3
0.2 Espace de probabilité . . . . .	5
0.2.1 Probabilité sur un ensemble fini . . . . .	5
0.2.2 Espace probabilisé général . . . . .	5
0.3 Variables aléatoires discrètes . . . . .	5
0.3.1 Couple de variables aléatoires-indépendantes . . . . .	6
0.4 Le théorème de Bayes . . . . .	7
0.4.1 Probabilités conditionnelles . . . . .	8
0.4.2 Probabilités totales . . . . .	8
0.4.3 Formule de Bayes . . . . .	9
0.5 Espérance Conditionnelle . . . . .	9
0.6 Marche aléatoire sur $\mathbb{Z}^d$ . . . . .	9
<b>1 Chaînes de Markov</b>	<b>10</b>
1.1 Définitions et propriétés générales . . . . .	10
1.1.1 Chaîne de Markov homogène . . . . .	10
1.1.2 Probabilité de transition . . . . .	11
1.1.3 Matrice de transition . . . . .	11
1.1.4 Graphe de transition . . . . .	12
1.1.5 Chaîne réversible . . . . .	13
1.1.6 Distribution invariante . . . . .	13
1.1.7 Unicité de la distribution invariante . . . . .	13
1.1.8 Irréductibilité . . . . .	14
1.1.9 Périodicité . . . . .	17
1.1.10 Equation de Chapman-Kolmogorov . . . . .	19
1.1.11 Interprétation . . . . .	20
1.1.12 Loi de probabilité de $X_n$ . . . . .	21
1.1.13 Propriété de Markov faible . . . . .	21
1.1.14 Propriété de Markov forte . . . . .	23

## TABLE DES MATIÈRES

---

1.1.15	Chaîne de Markov à deux états . . . . .	26
1.1.16	Existence et unicité des distributions stationnaire . . . . .	28
1.2	Classification des états de la chaîne de Markov . . . . .	30
1.3	Chaîne de Markov ergodique . . . . .	32
1.4	Exercice sur les chaînes de Markov : . . . . .	33
<b>2</b>	<b>Simulation</b>	<b>37</b>
2.1	Simulation . . . . .	37
2.1.1	Générateur de nombres aléatoires . . . . .	37
2.1.2	Méthode d'inversion de la fonction de répartition . . . . .	38
2.1.3	La méthode du rejet . . . . .	39
2.1.4	Méthode de Monte Carlo . . . . .	41
2.1.5	Histogramme . . . . .	46
2.2	Simulation par chaîne de Markov . . . . .	47
2.2.1	Méthode de MCMC . . . . .	47
2.2.2	Algorithme de Metropolis . . . . .	47
2.2.3	Simulation des lois de Gibbs . . . . .	51
2.2.4	Optimisation globale et recuit simulé . . . . .	51
2.2.5	Simulation exacte : l'algorithme de Propp-Wilson . . . . .	54
<b>3</b>	<b>Application à la simulation à l'aide de R</b>	<b>56</b>
3.1	Qu'est ce que R ? . . . . .	56
3.2	Application . . . . .	56
	<b>Conclusion</b>	<b>60</b>
	<b>Bibliographie</b>	<b>61</b>

# Liste des figures

1.1	Graphe de transition non values d'une chaîne de Markov irréductible (à gauche) et d'une chaîne de Markov non irréductible (à droite). . . . .	14
1.2	Graphe de transition non valeur d'une chaîne de Markov non irréductible avec deux ensembles absorbants fortement connexes: $\{4, 5\}$ et $\{6, 7, 8\}$ . .	17
1.3	Graphe de transition d'une chaîne de Markov irréductible de période 2. . .	19
2.1	Génération d'une variable aléatoire de loi exponentielle de paramètre 1. .	39
2.2	Estimation de $\pi/4$ . . . . .	41

# Résumé

Ce mémoire présente une étude sur les chaînes de Markov et la simulation, les chaînes de Markov sont un moyen assez courant et relativement simple de modéliser statistiquement des processus aléatoires. Elles ont été utilisées dans de nombreux domaines différents, allant de la génération de textes à la modélisation financière.

La simulation est l'expérimentation d'un modèle. Le comportement de modèle imite certains aspects saillants du comportement du système étudié et l'utilisation expérimenté avec le modèle pour déduire ce comportement. Ce cadre général s'est avéré être un adjuvant puissant pour l'apprentissage, la résolution de problèmes et la conception.

**Mots-clés :** chaînes de Markov, simulation, processus aléatoires, modélisation financière.

# Abstract

This thesis presents a study on Markov chains and simulation, Markov chains are a common and easy way to use statistics to model random processes. They have been used in many different areas, from making text to modeling finances.

Simulation is the experimentation of a model. The model behavior mimics some salient aspects of the behavior of the studied system and experienced use with the model to deduce this behavior. This general framework has proven to be a powerful aid for learning, problem solving and design.

**Keywords :** Markov chains, simulation, random processes, financial modelling.

# Introduction

Les chaînes de Markov sont exceptionnellement utiles pour modéliser des processus stochastiques en temps et en espace discrets dans divers domaines tels que la physique, la biologie et aussi l'informatique...etc.

Afin de comprendre ce qu'est une chaîne de Markov, voyons d'abord ce qu'est un processus stochastique, puisque la chaîne de Markov est une sorte unique de processus stochastique.

Un processus stochastique est défini comme collection de variable aléatoire  $X = \{X_t : t \in T\}$  défini sur un espace de probabilité commun, prenant des valeurs dans un ensemble commun  $S$  (l'espace d'état) et indexé par un ensemble  $T$  souvent soit  $N$  soit  $[0, \infty[$  et pensé comme le temps(respectivement discret au continu). Cela signifie que nous avons des observations à un certain moment et le résultat est une variable aléatoire. En terme simple, un processus stochastique est tout processus qui explique l'évolution d'un phénomène aléatoire dans le temps.

Maintenant que nous avons une intuition de base d'un processus stochastique, que nous allons nous atteler à comprendre l'un des concepts mathématiques les plus utiles pour la science des données: les chaînes de Markov!

Une chaîne de Markov est un processus stochastique à temps discret qui progresse d'un état à un autre avec certaines probabilités qui peuvent être présentées par une matrice. Cette matrice est appelée matrice de transition d'état ou matrice de probabilité de transition généralement représentée par  $P$ . Un graphe de transition d'état présente généralement une chaîne de Markov.

La simulation est utilisée pour modéliser la probabilité de différents résultats dans un processus difficile à prévoir car des variables aléatoires sont impliquées. De plus, c'est une méthode pour comprendre comment le risque et l'incertitude affectent les modèles de prédiction et de prévision. La simulation peut être utilisée pour résoudre un grand nombre de problèmes dans presque tous les domaines, tels que la finance, l'ingénierie,...

Décrivons maintenant les éléments et l'organisation de notre mémoire:

Premièrement, commençons par mentionner quelques préliminaires essentielles utilisés dans ce mémoire.

Dans le premier chapitre 1 nous avons présenté les chaînes de Markov et ces propriétés.

On passe au deuxième chapitre 2 qui traite de la simulation, les différentes méthodes de simulation.

On termine par une application de simulation à l'aide de programme R dans le dernier chapitre 3.

# Préliminaires

## 0.1 Espace mesuré

### Définition 1.

Soit  $X$  un ensemble. Nous appelons une tribu ou  $\sigma$ -algèbre sur  $X$  une famille  $\mathcal{M}$  de parties de  $X$  ayant les caractéristiques suivantes :

- i.  $X \in \mathcal{M}$
- ii. Si  $A_n \in \mathcal{M}$  donc  $A^c \in \mathcal{M}$  (où  $A^c = X \setminus A$  est le complémentaire de  $A$  dans  $X$ ).
- iii. Si  $A_n \in \mathcal{M}$ ,  $\forall n \in \mathbb{N}$ , donc  $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{M}$ .

Les éléments de  $\mathcal{M}$  sont appelés les parties mesurables de  $X$ . Alors que  $(X, \mathcal{M})$  est un espace mesurable.

### Remarque 1.

Si nous demandons seulement au lieu de (iii) :  
Si  $A, B \in \mathcal{M}$ , donc  $A \cup B \in \mathcal{M}$

On obtient un anneau booléen si  $\mathcal{M}$  est stable sous intersection finie.

### Conséquences :

- $\phi \in \mathcal{M}$  car  $X \in \mathcal{M}$ .
- Si  $A_n \in \mathcal{M}, \forall n \in \mathbb{N}$  alors  $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{M}$  (car  $\left(\bigcap_{n \in \mathbb{N}} A_n\right)^c = \bigcup_{n \in \mathbb{N}} A_n^c$ ).
- $\mathcal{M}$  est stable sous l'intersection ou l'union finie.
- Si  $A$  et  $B$  peuvent tous deux mesurables, la différence non symétrique

$$A \setminus B = A \cap B^c \in \mathcal{M}.$$

De toute évidence, tout ensemble  $X$  a des tribus. Comme exemple,

- $\mathcal{M} = \{\emptyset, X\}$  la plus petite.
- $\mathcal{M} = P(X)$  plus grande.

On utilise souvent le résultat suivant pour construire des tribus "intéressantes" sur  $X$  :

**Lemme 1.**

Soit  $\{\mathcal{M}_i\}_{i \in I}$  une famille quelconque de tribus sur  $X$ . Donc  $\mathcal{M} = \bigcap_{i \in I} \mathcal{M}_i$  est encore une tribu sur  $X$ .

**Définition 2.**

Disons que  $F$  est une famille de parties de  $X$ . Notons

$$\sigma(F) = \bigcap_{\mathcal{M} \text{ tribu sur } X, \mathcal{M} \supset F} \mathcal{M},$$

donc,  $\sigma(F)$  est une tribu sur  $X$  appelée tribu engendrée par  $F$ . C'est la plus petite tribu contenant  $F$  sur  $X$ .

**Définition 3.**

Une topologie sur  $X$  est une famille  $\mathcal{T}$  de parties de  $X$  ayant les propriétés suivantes :

- $\emptyset \in \mathcal{T}$ ,  $X \in \mathcal{T}$ .
- Si  $O_1, \dots, O_n \in \mathcal{T}$ , alors  $\bigcap_{i=1}^n O_i \in \mathcal{T}$ .
- Si  $\{O_i\}_{i \in I}$  est une famille quelconque d'éléments de  $\mathcal{T}$  d'où  $\bigcup_{i \in I} O_i \in \mathcal{T}$ .

Les éléments de  $\mathcal{T}$  sont appelés les ouverts de  $X$ . Nous disons que  $(X, \mathcal{T})$  est un espace topologique.

**Définition 4** (Tribu de Borel).

Soit  $(X, \mathcal{T})$  un espace topologique. Nous appelons tribu de Borel sur  $X$  la tribu engendrée par les ouverts de  $X$  :  $\mathcal{M} = \sigma(\mathcal{T})$ .

Considérant plus en détail le cas de la tribu de Borel sur  $\mathbb{R}$  notée  $\mathcal{B}(\mathbb{R})$ .

- $\mathcal{B}(\mathbb{R})$  comprend à la fois les ouverts et les fermés de  $\mathbb{R}$ .
- $\mathcal{B}(\mathbb{R})$  contient les unions dénombrables de fermés (ensemble  $F_\sigma$ ).
- $\mathcal{B}(\mathbb{R})$  contient les intersections dénombrables d'ouverts (ensemble  $G_\delta$ ).

Nous pouvons prouver que la tribu  $\mathcal{B}(\mathbb{R})$  a la puissance du continu. Par conséquent,  $\mathcal{B}(\mathbb{R}) \neq \mathcal{P}$ .

**Proposition 1.** La tribu  $\mathcal{B}(\mathbb{R})$  est engendrée par les intervalles  $]a, +\infty[$  pour  $a \in \mathbb{R}$ .

## 0.2 Espace de probabilité

### 0.2.1 Probabilité sur un ensemble fini

**Définition 5.**

Soit  $\Omega$  un ensemble fini et soit  $\mathcal{P}(\Omega)$  l'ensemble de ses parties. Un espace probabilisable est le couple  $(\Omega, \mathcal{P}(\Omega))$ .

**Remarque 2.**

L'ensemble  $\mathcal{P}(\Omega)$  vérifie plusieurs propriétés fondamentales : On dit que c'est une tribu. En effet :

- i.  $\Omega \in \mathcal{P}(\Omega)$ .
- ii.  $\forall A \in \mathcal{P}(\Omega), \bar{A} \in \mathcal{P}(\Omega)$ .
- iii.  $\forall A_1, \dots, A_k \in \mathcal{P}(\Omega), \bigcup_{i=1}^k A_i \in \mathcal{P}(\Omega)$ .

( $\mathcal{P}(\Omega)$  est stable par passage au complémentaire, et aussi par réunion).

### 0.2.2 Espace probabilisé général

**Définition 6.**

Soit  $(\Omega, A)$  un espace probabilisable. La probabilité définie sur cet espace est toute application :  $\mathcal{P} : A \rightarrow [0, 1]$  qui vérifie :

- i.  $\mathcal{P}(\Omega) = 1$ .
- ii. Si  $(A_n)$  est une suite d'évènements incompatible (disjoints) deux à deux donc :

$$\mathcal{P}\left(\bigcup_{n=0}^{+\infty} A_n\right) = \sum_{n=0}^{+\infty} \mathcal{P}(A_n).$$

Le triplet  $(\Omega, A, \mathcal{P})$  c'est un espace probabilisé.

## 0.3 Variables aléatoires discrètes

$(\Omega, \mathcal{T}, \mathcal{P})$  est un espace de probabilité et  $E$  un ensemble.

**Définition 7.**

- Nous appelons *variable aléatoire discrète* une application  $X$  de  $\Omega$  dans  $E$  telle que  $X(\Omega)$  est dénombrable ou fini et,  $\forall x \in E, X^{-1}(\{x\}) \in \mathcal{T}$ . On dit que  $X$  est une *variable aléatoire discrète réelle* si  $E = \mathbb{R}$ .
- Soit  $X$  une variable aléatoire discrète et notons  $X(\Omega) = \{x_n, n \in I\}$  où  $I$  est dénombrable ou fini. La loi de probabilité de  $X$  est la suite  $(P_n)_{n \in I}$ , où pour tout  $n \in I, P_n = P(X = x_n)$ .
- Soit  $(\Omega_1, \mathcal{T}_1, \mathcal{P}_1)$  et  $(\Omega_2, \mathcal{T}_2, \mathcal{P}_2)$  deux espaces de probabilité. Soit  $X$  (resp.  $Y$ ) une variable aléatoire discrète définie sur  $\Omega_1$  (resp.  $\Omega_2$ ). On dit que  $X$  et  $Y$  ont même loi si  $X(\Omega_1) = Y(\Omega_2)$  et si, pour tout  $x \in X(\Omega_1), P_1(X = x) = P_2(Y = x)$ . On note  $X \sim Y$ .

**0.3.1 Couple de variables aléatoires-indépendantes**

- On appelle couple de variables aléatoires discrètes un couple  $(X, Y)$  où  $X$  et  $Y$  sont deux variables aléatoires discrètes. La loi conjointe du couple  $(X, Y)$  est la loi de  $(X, Y)$  vue comme variable aléatoire. Autrement dit, la loi conjointe est la donnée de toutes les valeurs de  $P(X = x, Y = y)$  pour  $(x, y) \in X(\Omega) \times Y(\Omega)$ . Les lois de  $X$  et de  $Y$  sont les lois marginales de  $X$  et de  $Y$ .
- Soit  $x$  un élément de  $X(\Omega)$  tel que  $P(X = x) > 0$ . La probabilité  $P_x$  définie sur  $Y(\Omega)$  est donnée par :

$$\forall y \in Y(\Omega), P_x(\{y\}) = P(Y = y | X = x) = \frac{P(X = x, Y = y)}{P(X = x)},$$

est la loi conditionnelle de  $Y$  sachant que  $(X = x)$ .

- Ces définitions se généralisent à des n-uplets de variables aléatoires discrètes. Si  $X_1, \dots, X_n$  sont  $n$  variables aléatoires discrètes,  $(X_1, \dots, X_n)$  s'appelle un vecteur aléatoire discret.
- On dit que les variables aléatoires discrètes  $X$  et  $Y$  sont indépendantes si,  $\forall x \in X(\Omega)$  et  $\forall y \in Y(\Omega)$ , on a :

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

**Proposition 2.**

- Deux variables aléatoires discrètes  $X$  et  $Y$  sont indépendantes  $\Leftrightarrow$  pour tout  $A \subset X(\Omega)$  et tout  $B \subset Y(\Omega)$  on a :

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

- Soit  $(X_n)_{n \in I}$  une famille de variables aléatoires, où  $I$  est fini ou dénombrable. On dit que les variables aléatoires  $(X_n)_{n \in I}$  sont mutuellement indépendantes lorsque, pour toute partie finie  $J = \{i_1, \dots, i_p\} \subset I$ , pour tout  $(x_{i_1}, \dots, x_{i_p}) \in x_{i_1}(\Omega) \times \dots \times X_{i_p}(\Omega)$  on a :

$$P(X_{i_1} = x_{i_1}, \dots, X_{i_p} = x_{i_p}) = P(X_{i_1} = x_{i_1}) \dots P(X_{i_p} = x_{i_p}).$$

**Proposition 3.**

Si  $X_1, X_2, \dots, X_n$  sont des variables aléatoires mutuellement indépendantes alors pour tout  $m$  compris entre 1 et  $n-1$ , et pour toutes fonction  $f$  et  $g$ , les variables  $f(X_1, \dots, X_m)$  et  $g(X_{m+1}, \dots, X_n)$  sont indépendantes.

**Probabilités conditionnelles :****Définition 8.**

$A$  et  $B$  étant deux évènements lesquels  $P(A) \neq 0$ , la probabilité de  $B$  sachant  $A$  est le réel :

$$P_A(B) = \frac{P(A \cap B)}{P(A)}.$$

**Remarque 3.**

- Au lieu de  $P_A(B)$  on note parfois  $P(B/A)$ .
- Le symbole  $\cap$  (intersection) correspond à "et".

De même, si  $P(B) \neq 0$ , la probabilité de  $A$  sachant  $B$  est

$$P_B(A) = \frac{P(A \cap B)}{P(B)}.$$

**0.4 Le théorème de Bayes**

Le théorème de Bayes est une conséquence directe des probabilités totales et des probabilités conditionnelles.

### 0.4.1 Probabilités conditionnelles

Il y a 100 livres dans une bibliothèque. 40 d'entre eux sont écrits en anglais, et huit d'entre eux portent sur la biologie. Considérons les événements suivants :

$A =$  "L'anglais est la langue utilisée pour écrire le livre.",  $P(A) = \frac{40}{100}$ ,

$B =$  "la biologie est le sujet du livre.",

$A \cap B =$  "le livre est écrit en anglais et parle de biologie",  $P(A \cap B) = \frac{8}{100}$ .

**Probabilité conditionnelle :**

$B|A =$  "le livre parle de biologie sachant qu'il est écrit en anglais",  $P(B|A) = \frac{8}{40}$ .

C'est le nombre de livres anglais qui traitent de la biologie.

On a les relations 
$$P(B|A) = \frac{8}{40} = \frac{\frac{8}{100}}{\frac{40}{100}} = \frac{P(A \cap B)}{P(A)}.$$

Retenons

$$P(B|A) = \frac{P(A \cap B)}{P(A)},$$

ou encore

$$P(A \cap B) = P(A).P(B|A).$$

### 0.4.2 Probabilités totales

Considérons une partition  $A_1, A_2, \dots, A_n$  de l'ensemble des événements  $E$ , c'est-à-dire  $P(E) = 1$ ,  $A_1 \cup A_2 \cup \dots \cup A_n = E$  et  $A_i \cap A_j = \phi$  pour  $i \neq j$ . Alors :

$$P(B) = P(A_1).P(B|A_1) + P(A_2).P(B|A_2) + \dots + P(A_n).P(B|A_n).$$

**Démonstration :**

$$\begin{aligned} P(B) &= P(B \cap E) \\ &= P(B \cap A_1) + P(B \cap A_2) + \dots + P(B \cap A_n) \\ &= P(A_1).P(B|A_1) + P(A_2).P(B|A_2) + \dots + P(A_n).P(B|A_n). \end{aligned}$$

□

### 0.4.3 Formule de Bayes

Considérons une partitions  $A_1, A_2, \dots, A_n$  de l'ensemble des évènements  $E$ , Alors :

$$P(B) = P(A_1).P(B|A_1) + P(A_2).P(B|A_2) + \dots + P(A_n).P(B|A_n)$$

$$P(A_1|B) = \frac{P(A_1).P(B|A_1)}{P(B)}$$

$$P(A_2|B) = \frac{P(A_2).P(B|A_2)}{P(B)}$$

$$\dots$$

$$P(A_n|B) = \frac{P(A_n).P(B|A_n)}{P(B)}.$$

## 0.5 Espérance Conditionnelle

Munie de la loi conditionnelle, on peut calculer les espérances dites « conditionnelles » de  $Y$  sachant  $X$ .

**Définition 9** (Espérance conditionnelle).

Soit  $X$  et  $Y$  deux variables aléatoires et  $P_{Y|X=x}$  la loi conditionnelle de  $Y$  sachant  $X$ . Pour tout  $x \in \text{supp}X$ , on appelle "espérance conditionnelle de  $Y$  sachant  $X = x$ ", que l'on note :  $E[Y|X = x]$ , la quantité :

$$E[Y|X = x] = \int y dP_{Y|X=x}(y).$$

**Corollaire 1.**

Soit  $X$  et  $Y$  deux variables aléatoires discrètes. Pour tout  $x \in \text{supp}X$ , l'espérance conditionnelle de  $Y$  sachant  $X = x$  est donnée par :

$$E[Y|X = x] = \sum_{y \in \text{supp}P_y} y P(Y = y|X = x).$$

## 0.6 Marche aléatoire sur $\mathbb{Z}^d$

**Théorème 1.**

Soit  $d \geq 3$ . On définit une marche aléatoire sur  $\mathbb{Z}^d$  issue de 0 par :

$$\begin{cases} X_0 = 0 \\ X_{n+1} = X_n + \mathcal{E}_n, \end{cases}$$

avec les  $(\mathcal{E}_n)_n$  des variables aléatoires indépendantes à valeurs dans  $\mathbb{Z}^d$  suivant la loi uniforme sur  $\{\pm e_1, \dots, \pm e_d\}$  où  $(\{e_1, \dots, e_d\})$  est la base canonique de  $\mathbb{R}^d$ , alors :

$$\mathbb{P}(\|X_n\| \xrightarrow[n \rightarrow \infty]{} \infty) = 1.$$

# Chapitre 1

## Chaînes de Markov

Pour mieux comprendre les chaînes de Markov et leurs propriétés, nous les aborderons plus en détail dans ce chapitre.

### 1.1 Définitions et propriétés générales

**Définition 10** (Chaîne de Markov).

Soit  $\{X_n, n \geq 0\}$  une suite de variables aléatoires à valeurs dans l'ensemble des états  $E$  supposé égal à  $N$ . On dit que cette suite est une chaîne de Markov si, pour tout  $n \geq 0$  et toute suite  $(i_0, i_1, \dots, i_{n-1}, i, j)$ , on a :

$$P\left(\underbrace{X_{n+1} = j}_{\text{Le futur}} \mid \underbrace{X_n = i, X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \dots, X_0 = i_0}_{\text{Le passé et le présent}}\right) = P\left(\underbrace{X_{n+1} = j}_{\text{le futur}} \mid \underbrace{X_n = i}_{\text{le présent}}\right).$$

**Remarque 4.**

- L'état du processus au temps  $(n + 1)$  ne dépend que de son état au temps  $n$ , mais pas de ses états antérieurs.
- Nous dirons que ce processus est sans mémoire.

#### 1.1.1 Chaîne de Markov homogène

**Définition 11.**

Une chaîne de Markov est dite homogène si, pour tout  $i, j \in E$ , la probabilité  $P(X_n = i) | (X_{n+1} = j)$  ne dépend pas de  $n$ . On la note alors  $P_{ij}$ ,

$$P_{ij}(n) = P(X_{n+1} = j | X_n = i) = P_{ij} = P(X_1 = j | X_0 = i), (n \geq 0).$$

### 1.1.2 Probabilité de transition

**Définition 12.**

On définit la probabilité de transition de l'état  $i$  à l'état  $j$  entre les instants  $n$  et  $n + 1$  par la quantité.

$$P_{ij}(n) = P(X_{n+1} = j | X_n = i) \quad \forall i, j \in E.$$

Où  $P_{ij}$ :  $P$  (pour que le système soit dans l'état  $j$  à l'instant  $n + 1$  sachant à l'instant  $n$  il se trouvait à l'état  $i$ ).

### 1.1.3 Matrice de transition

**Définition 13.**

L'évolution de la chaîne de Markov est uniquement déterminée par sa matrice de transition  $P$ , qui est définie comme suit :

$$\forall i, j = 1, \dots, N, P_{ij} = P(X_1 = j | X_0 = i).$$

C'est une matrice stochastique, au sens où :

$$\forall i = 1, \dots, N, \sum_j P_{ij} = 1,$$

Soit  $P_1 = 1$  en notation vectorielle, avec  $1 = (1, \dots, 1)^t$ .

**Exemple 1.**

La matrice suivante est stochastique et définit donc une chaîne de Markov :

$$P = \begin{pmatrix} 1/4 & 1/2 & 1/4 \\ 1/3 & 0 & 2/3 \\ 0 & 1/5 & 4/5 \end{pmatrix}.$$

En particulier, la chaîne passe de l'état 1 à l'état 2 avec probabilité  $1/2$ .

Par la loi des probabilités totales, on a pour tout  $n \in N$  :

$$\forall i = 1, \dots, N, P(X_{n+1} = i) = \sum_{j=1}^N P(X_n = j)P(X_{n+1} = i | X_n = j).$$

Nous désignons la loi de  $X_n$  comme un vecteur ligne  $\pi(n)$  à  $N$  dimensions, avec  $\pi(n)1 = 1$ , cela s'écrit :

$$\forall i = 1, \dots, N, \pi(n+1)_i = \sum_{j=1}^N \pi(n)_j P_j.$$

Soit  $\pi(n+1) = \pi(n)P$  en notation vectorielle, donc on déduit :

$$\forall n \in \mathbb{N}, \pi(n) = \pi(0)P^n.$$

En prenant  $\pi(0)$  le vecteur nul sauf la  $i$ .ème composante égale à 1. On obtient :

$$\pi(n) = \left( P(X_n = j | X_0 = i) \right), 1 \leq j \leq N = \left( (P^n)_{ij}, 1 \leq j \leq N \right),$$

donc :

$$\forall n \in \mathbb{N}, (p^n)_{ij} = P(X_n = j | X_0 = i).$$

### 1.1.4 Graphe de transition

#### Définition 14.

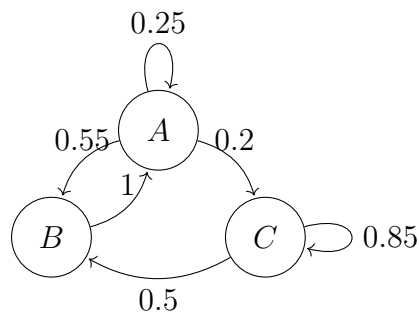
- Le graphe de transition est un graphe orienté pondéré dans lequel :
- Tous les poids sont compris entre 0 et 1.
- La somme des poids des chemins provenant d'un sommet est égale à 1.
- La matrice de transition d'un graphe probabiliste est une matrice stochastique.

#### Remarque 5.

Le terme "stochastique" est équivalent à "aléatoire".

On a la matrice de transition suivante :

$$M = \begin{pmatrix} 0.25 & 0.55 & 0.2 \\ 1 & 0 & 0 \\ 0 & 0.15 & 0.85 \end{pmatrix}$$



La somme des coefficients de chaque ligne vaut 1.  
C'est une matrice stochastique.

### 1.1.5 Chaîne réversible

**Définition 15.**

Soit  $\pi$  une probabilité sur  $M$ . La matrice de transition  $P$  (respectivement la chaîne  $X$ ) est réversible par rapport à  $\pi$  si :

$$\pi(x)P(x, y) = \pi(y)P(y, x) \quad \text{pour tout } x, y \in M.$$

### 1.1.6 Distribution invariante

**Définition 16.**

Soit une chaîne de Markov homogène  $(X_n)$  de matrice de transition  $P$ .  $\pi$  est une distribution invariante si la matrice ligne  $\pi$  vérifie :  $\pi P = \pi$ .

**Remarque 6.**

$\pi$  n'existe pas si  $(P - I_N)$  est inversible car une distribution ne peut correspondre au vecteur nul.

### 1.1.7 Unicité de la distribution invariante

**Théorème 2.**

Soit une chaîne de Markov homogène  $(X_n)$  de matrice de transition  $P$  d'ordre  $N$ . Si  $P$  ne possède aucun coefficient nul autre que sur sa diagonale principale, alors  $(X_n)$  admet une unique distribution invariante.

**Exemple 2.**

Si  $P = \begin{pmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{pmatrix}$  alors la distribution invariante est  $\pi = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \end{pmatrix}$ .

**Théorème 3.**

Soit une chaîne homogène de Markov  $(X_n)$  et soit  $(\pi_n)$  la suite de ses distributions.

- Si  $(\pi_n)$  est convergente donc même  $(\pi_n)$  est convergente vers une distribution invariante  $\pi$ .
- Si  $(X_n)$  à deux états a une distribution invariante  $\pi$ , alors la suite  $(\pi_n)$  converge vers  $\pi$ , quelque soit la distribution initial  $\pi_0$ .

### 1.1.8 Irréductibilité

#### Définition 17.

Une chaîne de Markov est dite irréductible si son graphe de transition est fortement connexe, c'est-à-dire si pour tous sommets distincts  $i, j$ , il existe un chemin de  $i$  vers  $j$  dans le graphe de transition. De ce fait, la propriété d'irréductibilité ne dépend que de la structure du graphe de transition et non de ses poids.

Une chaîne de Markov irréductible « visite » tous ses états infiniment souvent tandis qu'une chaîne de Markov non irréductible finit par ne prendre qu'un sous-ensemble strict des états possibles. Par exemple, pour les chaînes de Markov de la figure 1.1, la chaîne de gauche prendra les états 1, 2, 3 infiniment souvent tandis que la chaîne de droite finira par ne prendre que les états 2, 3 (on dit que l'ensemble 2, 3 est absorbant).

Le résultat suivant montre comment la matrice de transition peut être utilisée pour décrire une chaîne de Markov irréductible.

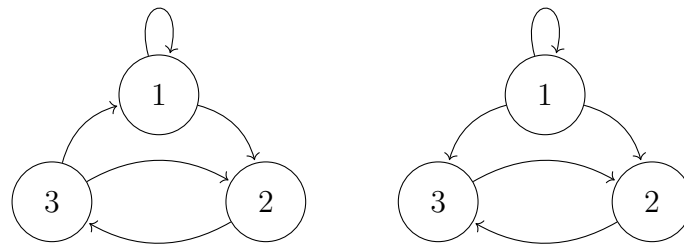


Figure 1.1: Graphe de transition non values d'une chaîne de Markov irréductible (à gauche) et d'une chaîne de Markov non irréductible (à droite).

#### Proposition 4.

Une chaîne de Markov de matrice de transition  $P$  est irréductible si et seulement si il existe un entier  $n \geq 1$  tel que la matrice  $(I + P)^n$  est strictement positive.

#### Démonstration :

On commence par la condition suffisante. Soit  $n \geq 1$  tel que la matrice  $(I + P)^n$  est strictement positive. Comme :

$$(I + P)^n = \sum_{k=0}^n \binom{n}{k} P^k,$$

il existe pour toute paire d'états distincts  $i \neq j$  un entier  $k$  tel que  $(P^k)_{ij} > 0$ , et donc un chemin de longueur  $k$  de  $i$  à  $j$  dans le graphe de transition.

Pour la condition nécessaire. Il suffit de prendre  $n = N$  car il existe pour toute paire

d'états distincts  $i \neq j$  un chemin de longueur inférieure à  $N$  de  $i$  à  $j$  dans le graphe de transition.  $\square$

**Théorème 4** (Perron-Frobenius (partiel)).

*Une chaîne de Markov irréductible admet une unique loi stationnaire.*

**Démonstration :**

La matrice  $P$  étant stochastique admet  $1 = (1, \dots, 1)^T$  comme vecteur propre droit associé à la valeur propre 1. Soit  $P1 = 1$ . La matrice transposée  $P^T$  ayant les mêmes valeurs propres que  $P$ , il existe un vecteur propre gauche  $v$  de  $P$  correspondant à la valeur propre 1, soit  $vP = v$ .

On montre tout d'abord que le vecteur ligne  $|v|$  satisfait également l'égalité  $|v|P = |v|$ .

On a en effet pour tout  $j$ ,

$$|v_j| = \left| \sum_i v_i P_{ij} \right| \leq \sum_i |v_i| P_{ij}.$$

En sommant ces inégalités, on trouve :

$$\sum_j |v_j| \leq \sum_j \sum_i |v_i| P_{ij} = \sum_i |v_i| \sum_j P_{ij} = \sum_i |v_i|,$$

de sorte que toutes les inégalités précédentes sont en fait des égalités.

On vérifie ensuite que le vecteur  $|v|$  est strictement positif, par irréductibilité, il existe en effet d'après la proposition précédente un entier  $n \geq 1$  tel que la matrice  $(I + P)^n$  est strictement positive. Le résultat découle alors du fait que :

$$|v|(I + P)^n = |v| \sum_{k=0}^n \binom{n}{k} P^k = 2^n |v|.$$

On sait donc que tout vecteur propre  $v$  de  $P$  pour la valeur propre 1 n'a aucune coordonnée nulle. Il reste à montrer que ce vecteur est de signe constant, soit  $v = |v|$  ou  $v = -|v|$ . Ceci vient du fait que si le vecteur  $v + |v|$  est non nul, c'est un vecteur propre gauche de  $P$  pour la valeur propre 1; comme un tel vecteur aucune coordonnée nulle, cela implique que toutes les coordonnées du vecteur  $v$  sont positives, c'est-à-dire que  $v = |v|$ .

Enfin, on définit le vecteur ligne  $\pi$  par renormalisation de  $|v|$  :

$$\pi_j = \frac{|v_j|}{\sum_{i=1}^N |v_i|}.$$

C'est un vecteur strictement positif qui satisfait  $\pi P = \pi$  et  $\sum_j \pi_j = 1$ . Par ailleurs, s'il existe un autre vecteur  $\pi'$  qui satisfait ces égalités, alors le vecteur  $\pi - \pi'$  est nul ou vecteur propre gauche de  $P$  pour la valeur propre 1. Mais si le vecteur  $\pi - \pi'$  était vecteur propre gauche de  $P$  pour la valeur propre 1, il serait de signe constant, ce qui est impossible car les vecteurs  $\pi$  et  $\pi'$  sont tous deux positifs et de somme unitaire.

On en déduit que  $\pi' = \pi$ , et donc que la loi stationnaire de la chaîne de Markov est unique.  $\square$

La preuve du théorème nous permet également de savoir que la loi stationnaire d'une chaîne de Markov irréductible est supportée par l'ensemble des états.

**Corollaire 2.**

*La loi stationnaire  $\pi$  d'une chaîne de Markov irréductible vérifie  $\pi_i > 0$  pour tout  $i = 1, \dots, N$ .*

**Exemple 3.** *Pensez à une chaîne de Markov avec une matrice de transition :*

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

*Cette chaîne de Markov est irréductible et, par conséquent, n'admet qu'une seule loi stationnaire. Voici comment s'écrivent les équations d'équilibre :*

$$\begin{aligned} \pi_1 &= \frac{1}{3}\pi_1 + \frac{1}{2}\pi_3 \\ \pi_2 &= \frac{2}{3}\pi_1 + \frac{1}{2}\pi_3 \\ \pi_3 &= \pi_2. \end{aligned}$$

Ces équations définissent un vecteur propre gauche de  $P$  pour la valeur propre 1 et ne sont donc pas indépendantes (la solution est unique à une constante multiplicative près). La dernière équation définissant la loi stationnaire  $\pi$  est uniquement la condition de normalisation :

$$\pi_1 + \pi_2 + \pi_3 = 1.$$

On trouve :  $\pi = (3/11, 4/11, 4/11)$ .

L'irréductibilité n'est pas une condition nécessaire d'existence et d'unicité d'une loi stationnaire. La chaîne de Markov associée au graphe de transition de droite de la figure 1.1 a pour unique loi stationnaire  $\pi = (0, 1/2, 1/2)$  par exemple. Par contre, l'irréductibilité est

une condition essentielle pour que la loi stationnaire ait pour support l'ensemble des états.

Une chaîne de Markov admet en fait toujours une loi stationnaire (dans le cas d'un nombre fini d'états considéré ici). Il suffit de décomposer le graphe de transition en ensembles absorbants fortement connexes et de considérer la restriction de la chaîne de Markov à chacun de ces ensembles : chacune de ces chaînes de Markov « réduites » définit une chaîne de Markov irréductible, qui admet donc une loi stationnaire. Toute combinaison convexe de ces lois résulte en une loi stationnaire pour la chaîne de Markov initiale ; il n'y a ainsi pas d'unicité en général.

**Exemple 4.** Le graphe de transition de la figure 1.2 a deux ensembles absorbants fortement connexes :  $\{4, 5\}$  et  $\{6, 7, 8\}$ . Les lois stationnaires associées sont :

$\pi^{(1)} = (0, 0, 0, 1/2, 1/2, 0, 0, 0)$  et  $\pi^{(2)} = (0, 0, 0, 0, 0, 1/3, 1/3, 1/3)$  ; on a ainsi :  $\pi^{(1)}P = \pi^{(1)}$  et  $\pi^{(2)}P = \pi^{(2)}$ . Toute combinaison convexe de ces deux lois est une loi stationnaire de la chaîne de Markov .

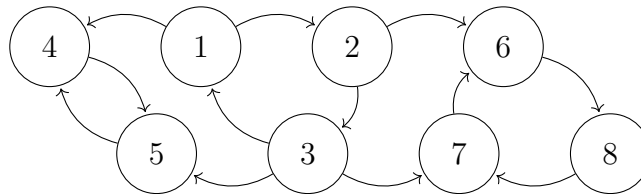


Figure 1.2: Graphe de transition non valeur d'une chaîne de Markov non irréductible avec deux ensembles absorbants fortement connexes:  $\{4, 5\}$  et  $\{6, 7, 8\}$

### 1.1.9 Périodicité

Certaines chaînes de Markov sont périodiques, ce qui signifie que les mêmes sous-ensembles d'états sont visités de manière cyclique.

**Définition 18.**

On appelle période d'un état le plus grand commun diviseur des longueurs des cycles du graphe transition passant par cet état. Un état est apériodique lorsqu'il est de période 1.

**Proposition 5.**

Tous les états d'une chaîne de Markov irréductible ont la même période.

**Démonstration :**

Soit  $i, j$  deux états distincts de périodes respectives  $d_i, d_j$ . La chaîne de Markov étant irréductible, il existe  $n, m$  tel que  $(P^n)_{ij} > 0$  et  $(P^m)_{ji} > 0$ . Comme :

$$(P^{n+m})_{ii} > (P^n)_{ij}(P^m)_{ji} > 0,$$

il existe un cycle de longueur  $n + m$  passant par  $i$ , donc  $d_i$  divise  $n + m$ . Par ailleurs, pour tout cycle de longueur  $k$  passant par  $j$ ,

$$(P^{n+k+m})_{ii} \geq (P^n)_{ij} (P^k)_{jj} (P^m)_{ji} > 0,$$

de sorte qu'il existe un cycle de longueur  $n+k+m$  passant par  $i$ . Ainsi  $d_i$  divise  $n+k+m$ , et comme  $d_i$  divise  $n+m$ ,  $d_i$  divise  $k$ . Finalement comme  $d_i$  divise toutes les longueurs de cycle passant par  $j$ ,  $d_i$  divise  $d_j$ . Par symétrie,  $d_j$  divise  $d_i$  et  $d_i = d_j$ .

On peut dire avec certitude que la période d'une chaîne de Markov irréductible est la somme des périodes de tous ses états.

**Théorème 5.**

Une chaîne de Markov  $(X_n)_{n \in \mathbb{N}}$  irréductible de période  $d$  admet une partition  $C_0, \dots, C_{d-1}$  de ses états telle que :

$$\forall k = 0, \dots, d-1, \forall n \in \mathbb{N}, P(X_n \in C_{k+n \bmod d} | X_0 \in C_k) = 1.$$

Les chaînes de Markov  $(X_{nd})_{n \in \mathbb{N}}$  restreintes aux ensembles  $C_0, \dots, C_{d-1}$  sont irréductibles, de lois stationnaires respectives  $\pi^{(0)}, \dots, \pi^{(d-1)}$  telles que :

$$\pi^{(1)} = \pi^{(0)}P, \dots, \pi^{(d)} = \pi^{(d-1)}P \text{ et } \pi^{(0)} = \pi^{(d)}P.$$

La loi stationnaire de la chaîne de Markov  $(X_n)_{n \in \mathbb{N}}$  est donnée par :

$$\pi = \frac{1}{d}(\pi^{(0)} + \dots + \pi^{(d-1)}).$$

**Exemple 5.**

La chaîne de Markov dont le graphe de transition est représenté en figure 1.3, est irréductible de période 2. Elle visite les états  $\{1, 3\}$  et  $\{2, 4\}$  de manière cyclique .

Sa matrice de transition est :

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

On vérifie que la matrice :

$$P^2 = \begin{pmatrix} 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 2/3 & 0 & 1/3 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

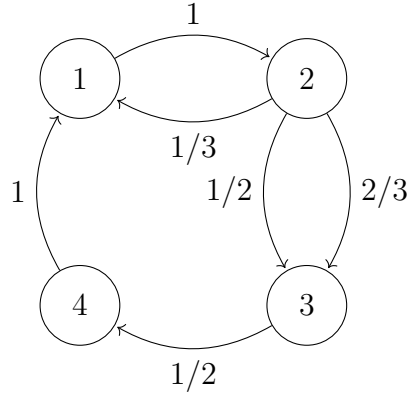


Figure 1.3: Graphe de transition d’une chaîne de Markov irréductible de période 2.

définit bien deux chaînes de Markov irréductibles, restreintes à  $\{1, 3\}$  et  $\{2, 4\}$ , respectivement. Les lois stationnaires associées sont :

$$\pi^{(0)} = (1/2, 0, 1/2, 0) \text{ et } \pi^{(1)} = (0, 3/4, 0, 1/4).$$

On vérifie que  $\pi^{(1)} = \pi^{(0)}P$  et  $\pi^{(0)} = \pi^{(1)}P$ . De plus, la chaîne de Markov initiale a bien pour unique loi stationnaire  $\pi = (\pi^{(0)} + \pi^{(1)})/2$ , soit  $\pi = (1/4, 3/8, 1/4, 1/8)$ .

Une chaîne de Markov irréductible est dite apériodique si sa période est 1. Le résultat suivant en donne une caractérisation à partir de la matrice de transition.

### 1.1.10 Equation de Chapman-Kolmogorov

Soit :

$$P_{i,j}^{(n)} = P[X_n = j | X_0 = i],$$

la probabilité de passer de  $i$  à  $j$  en exactement  $n$  étapes.

L’équation de Chapman-Kolmogorov est la suivante :

$$P_{ij}^{(n+m)} = \sum_{k=0}^{\infty} P[X_{n+m} = j | X_n = k] P[X_n = k | X_0 = i] = \sum_{k=0}^{\infty} P_{i,k}^{(n)} P_{k,j}^{(m)}.$$

En notation matricielle, si  $P^{(n)}$  est la matrice contenant les  $P_{i,j}^{(n)}$ , cela donne :

$$P^{(n+m)} = P^{(n)} \cdot P^{(m)}.$$

En particulier,  $P^{(2)} = P \cdot P$  et par induction sur  $n$ , on a :

$$P^{(n)} = P^{(n-1)} \cdot P = P^n.$$

Semblable aux matrices qui donnent le nombre de chemins de longueur  $n$  entre chaque paire de sommets dans un graphe.

**Exemple 6.** Dans l'exemple de "pluie" vs "non pluie", supposons que  $\alpha = 0,7$  et  $\beta = 0,4$ , de sorte que :

$$P = \begin{pmatrix} \alpha & 1 - \alpha \\ \beta & 1 - \beta \end{pmatrix} = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}.$$

Les probabilités de transition en deux jours et en 4 jours sont :

$$P^2 = \begin{pmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{pmatrix}, P^4 = \begin{pmatrix} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{pmatrix}.$$

Que se passe-t-il avec  $P^n$  quand  $n \rightarrow \infty$  ? On a :

$$\lim_{n \rightarrow \infty} P^n = \begin{pmatrix} 4/7 & 3/7 \\ 4/7 & 3/7 \end{pmatrix}.$$

Les lignes sont identiques et donnent les probabilités limites des deux états.

### 1.1.11 Interprétation

#### Exemple 7.

On dispose de balles rouges et de balles bleues, et d'une urne qui contient 2 balles. À chaque étape, on tire une balle de l'urne et on la remplace par une balle de la même couleur avec probabilité 0.8 et de l'autre couleur avec probabilité 0.2. On définit une chaîne où  $X_n$  est le nombre de balles rouges dans l'urne à l'étape  $n$ . L'espace d'états est  $0, 1, 2$  et la matrice des probabilités de transition est :

$$P = \begin{pmatrix} 0.8 & 0.2 & 0 \\ 0.1 & 0.8 & 0.1 \\ 0 & 0.2 & 0.8 \end{pmatrix}.$$

Si  $X_0 = 2$  (2 balles rouges au départ), alors la première balle tirée est certainement rouge.

Soit  $b_n$  la probabilité que la  $(n + 1)$  ième balle tirée soit rouge. On aura :

$$b_n = 1 \cdot P[X_n = 2 | X_0 = 2] + (1/2)P[X_n = 1 | X_0 = 2] = P_{2,2}^{(n)} + P_{2,1}^{(n)}/2.$$

Par exemple, en calculant  $P^4$  et on trouve :  $P_{2,2}^{(4)} = 0.4872$  et  $P_{2,1}^{(4)} = 0.4352$ , ce qui donne  $b_4 = 0.4872 + 0.4352/2 = 0.704$ .

Que se passe-t-il quand  $n \rightarrow \infty$  ? On peut montrer que  $P^n$  converge vers une matrice dont les 3 lignes sont  $(1/4, 1/2, 1/4)$ , et que  $b_n \rightarrow 1/2$  ce qui correspond à l'intuition.

### 1.1.12 Loi de probabilité de $X_n$

L'espace des états est noté :  $E = \{1; 2; \dots; N\}$ .

Pour tout  $n \in \mathbb{N}$ , la variable aléatoire  $X_n$  est définie par les  $N$  probabilités :

$$P(X_n = 1), P(X_n = 2), \dots, P(X_n = N).$$

Soit  $\pi_n$  la matrice ligne :  $\pi_n = (P(X_n = 1) \ P(X_n = 2) \dots P(X_n = N))$ , on a :

$$\forall n \in \mathbb{N}, \pi_{n+1} = \pi_n P \Rightarrow \pi_n = \pi_0 P^n.$$

**Remarque 7.** La loi d'une chaîne de Markov homogène est entièrement déterminée par sa distribution initiale  $\pi_0$  et sa matrice de transition  $P$ .

**Preuve :**

Établissons la relation de récurrence :  $\forall n \in \mathbb{N}, \pi_{n+1} = \pi_n P$ .

$(X_n = i)_{i \in E}$  forme une partition de l'univers  $\Omega$ .

On rappelle que  $P_{ij} = P_{(X_n=i)}(X_{n+1} = j) = \frac{P[(X_n = i) \cap (X_{n+1} = j)]}{P(X_n = i)}$ .

Posons :  $\pi_n P = (q_{1j})$  matrice ligne.

$$q_{1j} = \sum_{i=1}^n P(X_n = i) P_{ij} = \sum_{i=1}^n P(X_n = i) P_{(X_n=i)}(X_{n+1} = j) = \sum_{i=1}^n \underbrace{P[(X_n = i) \cap (X_{n+1} = j)]}_{\text{Proba totales} = P(X_{n+1} = j)}.$$

On a donc :  $\forall n \in \mathbb{N}, (q_{1j}) = (P(X_{n+1} = j)) = \pi_{n+1}$ . □

### 1.1.13 Propriété de Markov faible

**Notation 1.** Par la suite, on note  $\mathcal{F}_n$  la tribu engendrée par les variables aléatoires  $X_0, \dots, X_n$ . Plus précisément,  $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$ .

En outre, nous noterons  $P_x$  la probabilité conditionnelle sachant  $\{X_0 = x\}, x \in E$ .

**Théorème 6** (Propriété de Markov faible).

Soit  $(X_n)_{n \in \mathbb{N}}$  une chaîne de Markov. Donc pour tout  $n \in \mathbb{N}, x \in E$  et conditionnellement à  $\{X_n = x\}, (X_{n+p})_{p \in \mathbb{N}}$  est une chaîne de Markov de loi initiale  $\delta_x$ .

De plus elle est indépendante de  $\mathcal{F}_n$  : c'est-à-dire, pour tout  $A \in \mathcal{F}_n$ , pour tout  $m > 0$ ,  $x_0, x_1, \dots, x_m \in E$  :

$$\begin{aligned} & P(A \cap \{X_{n+1} = x_1, \dots, X_{n+m} = x_m\} | X_n = x) \\ &= P(X_{n+1} = x_1, \dots, X_{n+m} = x_m | X_n = x) P(A | X_n = x) \\ &= P_x(X_1 = x_1, \dots, X_m = x_m) P(A | X_n = x). \end{aligned}$$

Avant la preuve, deux lemmes seront établis.

**Lemme 2.** Soit  $(\Omega, \mathcal{F}, P)$  un espace probabilisé et soient  $A, B, C \in \mathcal{F}$  tels que  $P(B \cap C) > 0$ , donc :

$$P(A|B|C) = P(A|B \cap C).$$

**Démonstration :**

Soit  $Q(\cdot) := P(\cdot|C)$ .  $Q$  est une probabilité. Par la définition de la probabilité conditionnelle, nous avons :

$$\begin{aligned} P(A|B|C) &= Q(A|B) = \frac{Q(A \cap B)}{Q(B)} = \frac{P(A \cap B|C)}{P(B|C)} \\ &= \frac{P(A \cap B \cap C)}{P(B \cap C)} = P(A|B \cap C). \end{aligned}$$

□

**Lemme 3.** Soit  $(X_n)_{n \geq 0}$  une chaîne de Markov. Soit  $n \geq 0$  et  $x_0, x_n, x_{n+1} \in E$  tels que  $P(\{X_n = x_n\} \cap \{X_0 = x_0\}) > 0$ . Alors :

$$P(X_{n+1} = x_{n+1} | X_n = x_n, X_0 = x_0) = P(X_{n+1} = x_{n+1} | X_n = x_n).$$

**Démonstration:**

En utilisant la notion de chaîne de Markov et la somme de toutes les trajectoires potentielles, on détermine ce qui suit :

$$\begin{aligned} &P(X_{n+1} = x_{n+1}, X_n = x_n, X_0 = x_0) \\ &= \sum_{z_1, z_2, \dots, z_{n-1} \in E} P(X_{n+1} = x_{n+1}, X_n = x_n, X_{n-1} = z_{n-1}, \dots, X_1 = z_1, X_0 = x_0) \\ &= \sum_{z_1, z_2, \dots, z_{n-1} \in E} P(X_{n+1} = x_{n+1} | X_n = x_n, X_{n-1} = z_{n-1}, \dots, X_1 = z_1, X_0 = x_0) \\ &\quad \times P(X_n = x_n, X_{n-1} = z_{n-1}, \dots, X_1 = z_1, X_0 = x_0) \\ &= P(X_{n+1} = x_{n+1} | X_n = x_n) \sum_{z_1, z_2, \dots, z_{n-1} \in E} P(X_n = x_n, X_{n-1} = z_{n-1}, \dots, X_1 = z_1, X_0 = x_0) \\ &= P(X_{n+1} = x_{n+1} | X_n = x_n) P(X_n = x_n | X_0 = x_0). \end{aligned}$$

D'où :

$$P(X_{n+1} = x_{n+1} | X_n = x_n, X_0 = x_0) = P(X_{n+1} = x_{n+1} | X_n = x_n).$$

□

Revenons à montrer la propriété de Markov faible.

**Démonstration :** Prouvons tout d'abord que  $(X_{n+p})_{p \in \mathbb{N}}$  est une chaîne de Markov. Soit,  $m \geq 1$  et  $x_1, \dots, x_m \in E$ , on a en utilisant le lemme ci-dessus :

$$\begin{aligned}
& P_x(X_{m+n} = x_m | X_{n+m-1} = x_{m-1}, \dots, X_{n+1} = x_1) \\
&= P(X_{m+n} = x_m | X_{n+m-1} = x_{m-1}, \dots, X_{n+1} = x_1, X_n = x) \\
&= P(X_{m+n} = x_m | X_{n+m-1} = x_{m-1}) \\
&= P(X_{m+n} = x_m | X_{n+m-1} = x_{m-1}, X_n = x) \\
&= P_x(X_{m+n} = x_m | X_{n+m-1} = x_{m-1}).
\end{aligned}$$

Pour montrer l'indépendance conditionnelle, il suffit de la prouver pour  $A = \{X_0 = y_0, \dots, X_n = y_n\}$ , En effet par  $\sigma$ -additivité de  $P$ , nous pouvons l'acquérir pour tout  $A \in \mathcal{F}_n$ . De plus, il suffit d'examiner l'instance  $y_n = x$  (sinon les deux membres sont nuls). On tient :

$$\begin{aligned}
& P(A \cap \{X_{n+1} = x_1, \dots, X_{n+m} = x_m\} | X_n = x) \\
&= P(X_0 = y_0, \dots, X_n = x, X_{n+1} = x_1, \dots, X_{n+m} = x_m | X_n = x), \\
&= \frac{P(X_0 = y_0, \dots, X_n = x, X_{n+1} = x_1, \dots, X_{n+m} = x_m)}{P(X_n = x)}, \\
&= \frac{P(X_0 = y_0, \dots, X_n = x)}{P(X_n = x)} Q_{n+1}(x, x_1) Q_{n+2}(x_1, x_2) \dots Q_{n+m}(x_{m-1}, x_m), \\
&= \frac{P(A)}{P(X_n = x)} P(X_{n+1} = x_1, \dots, X_{n+m} = x_m | X_n = x), \\
&= P(A | X_n = x) P(X_{n+1} = x_1, \dots, X_{n+m} = x_m | X_n = x).
\end{aligned}$$

□

### 1.1.14 Propriété de Markov forte

#### Définition 19.

- Une filtration est une suite croissante de tribus, c'est à dire une suite de tribus  $(\mathcal{G}_n)_{n \in \mathbb{N}}$  vérifiant  $\mathcal{G}_0 \subset \mathcal{G}_1 \subset \dots \subset \mathcal{G}_n \subset \dots$
- $(\sigma(X_0, \dots, X_n))_{n \in \mathbb{N}} = (\mathcal{F}_n)_{n \in \mathbb{N}}$  est appelée filtration engendrée par  $(X_n)_{n \in \mathbb{N}}$ .

#### Définition 20.

Soient  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires définies sur  $(\Omega, \mathcal{F}, P)$ , et  $\tau$  une variable aléatoire définie sur le même espace probabilisé.  $\tau$  est dite un temps d'arrêt adapté à la suite  $(X_n)_{n \in \mathbb{N}}$  (ou temps d'arrêt adapté à la filtration  $(\mathcal{F}_n)_{n \in \mathbb{N}}$ ) si :

- $\tau$  prend ses valeur dans  $\mathbb{N} \cup \{\infty\}$ ,

b.  $\forall n \geq 0$ , on a  $\{\tau \leq n\} \in \mathcal{F}_n$ .

**Remarque 8.** On peut remplacer la deuxième condition par la condition :  
 $\forall n \geq 0$ , on a  $\{\tau = n\} \in \mathcal{F}_n$ .

**Proposition-définition :**

Soit  $\tau$  un temps d'arrêt adapté à la filtration  $(\mathcal{F}_n)_n$ .

Posons :  $\mathcal{F}_\infty = \sigma(\mathcal{F}_n, n \geq 0)$ . On définit :

$$\mathcal{F}_\tau \subset \mathcal{F}_\infty \text{ par } \mathcal{F}_\tau = \{A \in \mathcal{F}_\infty, \forall n \geq 0, A \cap \{\tau \leq n\} \in \mathcal{F}_n\}.$$

est une tribu appelée tribu des événements antérieurs à  $\tau$ .

**Preuve :**

- $\forall n \geq 0, \Omega \cap \{\tau = n\} = \{\tau = n\} \in \mathcal{F}_n$  donc  $\Omega \in \mathcal{F}_\tau$ .
- Soit  $(A_k)_{k \in \mathbb{N}} \in \mathcal{F}_\tau$ , montrons que  $\bigcup_{k \in \mathbb{N}} A_k \in \mathcal{F}_\tau$ . On a :

$$\left( \bigcup_{k \in \mathbb{N}} A_k \right) \cap \{\tau \leq n\} = \bigcup_{k \in \mathbb{N}} (A_k \cap \{\tau \leq n\}) \in \mathcal{F}_n.$$

Donc  $\bigcup_{k \in \mathbb{N}} A_k \in \mathcal{F}_\tau$ .

- Pour  $A \in \mathcal{F}_\tau$ , prouvons que  $\bar{A} \in \mathcal{F}_\tau$ . Soit  $B = \bar{A} \cap \{\tau \leq n\}$ , on a :

$$\begin{aligned} \bar{A} \cap \{\tau \leq n\} &= (\bar{A} \cup \overline{\{\tau \leq n\}}) \cap \{\tau \leq n\} \\ &= (\overline{A \cap \{\tau \leq n\}}) \cap \{\tau \leq n\} \in \mathcal{F}_n, \end{aligned}$$

car  $A \cap \{\tau \leq n\}$  et  $\{\tau \leq n\}$  sont dans  $\mathcal{F}_n$ . Donc  $\bar{A} \in \mathcal{F}_\tau$ .

□

**Théorème 7** (Propriété de Markov forte).

Soit  $(X_n)_{n \in \mathbb{N}}$  une chaîne de Markov homogène de paramètres  $(\mu, P)$  et  $\tau$  un temps d'arrêt adapté à  $(X_n)_{n \in \mathbb{N}}$ . Pour tout  $x \in E$  conditionnellement à  $\{\tau < \infty, X_\tau = x\}$ , la suite  $(X_{\tau+n})_{n \in \mathbb{N}}$  est une chaîne de Markov de paramètres  $(\delta_x, P)$ . De plus elle est indépendante de  $\mathcal{F}_\tau$ ; i.e., pour tout  $A \in \mathcal{F}_\tau, x_1, \dots, x_m \in E$  :

$$\begin{aligned} &P(\{X_{\tau+n} = x_n, \dots, X_\tau = x_0\} \cap A | \{\tau < +\infty\} \cap X_\tau = x_0) \\ &= P_{x_0}(\{X_n = x_n, \dots, X_1 = x_1\})P(A | \{\tau < +\infty\} \cap X_\tau = x_0). \end{aligned}$$

**Remarque 9.** Nous rétablissons la propriété de Markov faible en fixant  $\tau = k$ .

**Remarque 10.** Particulièrement, nous avons ce qui suit :

$$\begin{aligned} &P(\{X_{\tau+n} = x_n, \dots, X_\tau = x_0\} | \{\tau < +\infty\} \cap X_\tau = x_0) \\ &= P_{x_0}(\{X_n = x_n, \dots, X_1 = x_1\}). \end{aligned}$$

**Démonstration :**

Soit  $A \in \mathcal{F}_\tau$ , en écrivant  $A \cap \{\tau < \infty\}$  comme l'union disjointe

$$A \cap \{\tau < \infty\} = \bigcup_{m \geq 0} A \cap \{\tau = m\} \text{ on a :}$$

$$\begin{aligned} & P(\{X_{\tau+n} = x_n, \dots, X_\tau = x_0\} \cap A \cap \{\tau < +\infty\}) \\ &= \sum_{m \geq 0} P(\{X_{\tau+n} = x_n, \dots, X_\tau = x_0\} \cap A \cap \{\tau = m\} \cap \{X_\tau = x_0\}) \\ &= \sum_{m \geq 0} P(\{X_{m+n} = x_n, \dots, X_m = x_0\} \cap A \cap \{\tau = m\} \cap \{X_m = x\}) \\ &= \sum_{m \geq 0} P(\{X_{m+n} = x_n, \dots, X_m = x_0\} | X_m = x) P(A \cap \{\tau = m\} | \{X_m = x\}) P(\{X_m = x_0\}) \\ &= P_{x_0}(\{X_n = x_n, \dots, X_1 = x\}) \sum_{m \geq 0} P(A \cap \{\tau = m\} \cap \{X_m = x_0\}) \\ &= P_{x_0}(\{X_n = x_n, \dots, X_1 = x\}) P(A \cap \{\tau < +\infty\} \cap \{X_\tau = x_0\}); \end{aligned}$$

où nous avons employé la propriété de Markov faible pour la 3<sup>ème</sup> égalité, on note que :  $A \cap \{\tau = m\} \in \mathcal{F}_m$ . Alors, pour  $A \in \mathcal{F}_\tau$ , par division sur  $P(\{\tau < +\infty\} \cap X_\tau = x_0)$ , on obtient :

$$\begin{aligned} & P(\{X_{\tau+n} = x_n, \dots, X_\tau = x_0\} \cap A | \{\tau < +\infty\} \cap \{X_\tau = x_0\}) \\ &= P_{x_0}(\{X_n = x_n, \dots, X_1 = x\}) P(A | \{\tau < +\infty\} \cap \{X_\tau = x_0\}). \end{aligned}$$

En particulier, on observe que :

$$\begin{aligned} & P(\{X_{\tau+n} = x_n, \dots, X_\tau = x_0\} | \{\tau < +\infty\} \cap \{X_\tau = x_0\}) \\ &= P_{x_0}(\{X_n = x_n, \dots, X_1 = x\}), \end{aligned}$$

et donc, on remarque que :

$$\begin{aligned} & P_{\{\tau < +\infty \cap X_\tau = x_0\}}(X_{\tau+n+1} = x_{n+1} | X_{\tau+n} = x_n, \dots, X_\tau = x_0) \\ &= P_{x_0}(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0, \dots, X_1 = x) \\ &= P_{x_0}(X_{n+1} = x_{n+1} | X_n = x_n) \\ &= P_{\{\tau < +\infty \cap X_\tau = x_0\}}(X_{\tau+n+1} = x_{n+1} | X_{\tau+n} = x_n). \end{aligned}$$

Cette égalité finale transfère précisément la propriété de Markov de la suite  $(X_{\tau+n})_{n \in \mathbb{N}}$  conditionnellement à l'évènement  $\{\tau < \infty, X_\tau = x\}$ .  $\square$

**Exemple :**

Soit  $(X_n)_n$  une marche aléatoire sur  $Z$  partant de 0. Soit  $\tau$  le temps pour atteindre 10 :

$$\tau := \inf\{k \geq 0, X_k = 10\}.$$

Alors conditionnellement à  $\{\tau < +\infty\}$ ,  $(X_{\tau+n})_n \geq 0$  est aussi une marche aléatoire partant de 10. De plus elle est indépendante de  $\mathcal{F}_\tau$ , c'est-à-dire qu'elle est indépendante des trajectoires précédant  $\tau$ .

### 1.1.15 Chaîne de Markov à deux états

En général, pour une chaîne de Markov à deux états, 0 et 1, la matrice de transition ressemble à ceci :

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix},$$

avec :  $0 \leq \alpha, \beta \leq 1$ , supposons ici que :  $0 < \alpha, \beta < 1$  de sorte que les probabilités de chaque transition sont toutes strictement positives.

Voici le résultat de la détermination de la matrice de transition en  $n$  étapes :

**Proposition 6.** *Nous avons pour tout  $n \in \mathbb{N}$ ,*

$$P^n = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta + \alpha(1 - \alpha - \beta)^n & \alpha(1 - (1 - \alpha - \beta)^n) \\ \beta(1 - (1 - \alpha - \beta)^n) & \alpha + \beta(1 - \alpha - \beta)^n \end{pmatrix}.$$

**Démonstration :**

Il suffit de diagonaliser la matrice  $P$  pour montrer ce résultat.

On a  $(1, 1)^t$  et  $(-\alpha, \beta)^t$  les vecteurs propres correspondant respectivement aux valeurs propres  $\lambda_1 = 1$  et  $\lambda_2 = 1 - \alpha - \beta$  de la matrice  $P$ , D'où :

$$P = M.D.M^{-1}. \quad (1.1)$$

Alors la relation 1.1 devient :

$$P = \begin{pmatrix} 1 & -\alpha \\ 1 & \beta \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \cdot \begin{pmatrix} \frac{\beta}{\alpha + \beta} & \frac{\alpha}{\alpha + \beta} \\ -\frac{1}{\alpha + \beta} & \frac{1}{\alpha + \beta} \end{pmatrix},$$

où :

$$M = \begin{pmatrix} 1 & -\alpha \\ 1 & \beta \end{pmatrix}, D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \quad \text{et} \quad M^{-1} = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ -1 & 1 \end{pmatrix}.$$

Comme conséquence de la relation 1.1, nous avons :

$$\begin{aligned} P^n &= (M.D.M^{-1})^n = (M.D.M^{-1}) \dots (M.D.M^{-1}) \\ &= M.D \dots D.M^{-1} \\ &= M.D^n.M^{-1}, \end{aligned}$$

avec :

$$D^n = \begin{pmatrix} 1 & 0 \\ 0 & \lambda_2^n \end{pmatrix}, \quad n \in \mathbb{N},$$

d'où :

$$\begin{aligned} P^n &= \begin{pmatrix} 1 & -\alpha \\ 1 & \beta \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & \lambda_2^n \end{pmatrix} \cdot \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ -\frac{1}{\alpha+\beta} & \frac{1}{\alpha+\beta} \end{pmatrix}, \\ &= \frac{1}{\alpha+\beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{\lambda_2^n}{\alpha+\beta} \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}, \\ &= \frac{1}{\alpha+\beta} \begin{pmatrix} \beta + \alpha\lambda_2^n & \alpha(1 - \lambda_2^n) \\ \beta(1 - \lambda_2^n) & \alpha + \beta\lambda_2^n \end{pmatrix}. \end{aligned}$$

□

## Distributions stationnaires et limites pour les chaînes de Markov homogènes

**Définition 21** (Loi stationnaire).

Une distribution de probabilité discrète  $\pi = (\pi_1, \dots, \pi_k, \dots)$  est une distribution stationnaire par rapport à la matrice de transition  $P$  si et ssi :

$$\pi = \pi.P,$$

où  $\pi$  un vecteur propre à gauche de la matrice  $P$  associé à la valeur propre 1, avec :

$$\sum_{j=0}^{\infty} \pi_j = 1, \pi_j \geq 0, \forall j;$$

ou bien :

$$\pi_j = \sum_{i \in E} \pi_i.P_{ij}.$$

**Définition 22** (Loi limite).

Une chaîne de Markov converge vers  $\pi$  ou possède une distribution limite  $\pi$  si :

$$\lim_{n \rightarrow \infty} \pi^{(n)} = \pi,$$

indépendamment de la loi initiale de  $\pi^{(0)}$  et, si  $\pi$  est une loi de probabilité. La convergence d'une chaîne de Markov est une propriété qui dépend uniquement de la matrice de transition  $P$ .

**Définition 23** (Distribution ergodique).

Une distribution est appelé distribution ergodique si  $\pi_j > 0, \forall j$ .

### 1.1.16 Existence et unicité des distributions stationnaire

#### **Théorème 8.**

*Il existe toujours au moins une distribution fixe dans une chaîne de Markov finie. Cependant, les distributions stationnaires ne sont pas nécessairement admissibles dans une chaîne de Markov finie.*

*Nous allons énoncer le critère d'unicité d'une distribution stationnaire.*

#### **Théorème 9.**

*Une chaîne de Markov finie admet une unique distribution stationnaire si et seulement si elle contient une classe récurrente unique.*

**Exemple 8.** *Considérons la matrice de transition de la chaîne de Markov suivante :*

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

*Cette chaîne de Markov permet un nombre infini de distributions stationnaires définies comme suit :*

$$(0, 1 - 2\alpha, \alpha, \alpha), \quad \text{avec } 0 \leq \alpha < \frac{1}{2}.$$

#### **Remarque 11.**

*Une chaîne de Markov infinie n'admet pas toujours de distribution stationnaire.*

*Par exemple, la chaîne de Markov à valeurs dans  $\mathbb{N}$  qui à  $i$  associe  $i + 1$  avec probabilité 1, i.e. qui possède la matrice de transition :*

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \dots & \dots \\ 0 & 0 & 1 & 0 & \dots & \dots \\ 0 & 0 & 0 & 1 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix},$$

*n'admet pas de loi stationnaire.*

### Recherche des distributions stationnaires

Pour calculer les composantes du vecteur  $\pi = (\pi_0, \pi_1, \dots, \pi_n)$  d'une chaîne de Markov finie, on résout le système d'équations linéaires suivant :

$$\begin{cases} \pi = \pi.P \\ \sum_{k \in E} \pi_k = 1 \end{cases}.$$

**Exemple 9.** *Considérons la matrice de transition :*

$$P = \begin{pmatrix} 1 - \alpha & 0 & \alpha & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 1/3 & 0 & 2/3 \\ 1/3 & 0 & 2/3 & 0 \end{pmatrix},$$

*d'une chaîne de Markov prenant ses valeurs dans l'ensemble  $\{0, 1, 2, 3\}$  avec  $\alpha \in [0, 1]$ .*

**Recherchons les solutions stationnaires :** Pour trouver ces solutions stationnaires, il faut résoudre le système suivant :

$$\left\{ \begin{array}{l} \pi = \pi.P \\ \sum_{k=0}^3 \pi_k = 1 \end{array} \right. \iff \left\{ \begin{array}{l} \pi_0 = (1 - \alpha)\pi_0 + \frac{1}{3}\pi_1 + \frac{1}{3}\pi_3 \\ \pi_1 = \frac{2}{3}\pi_2 \\ \pi_2 = \alpha\pi_0 + \frac{2}{3}\pi_1 + \frac{2}{3}\pi_3 \\ \pi_3 = \frac{2}{3}\pi_2 \\ \sum_{k=0}^3 \pi_k = 1 \end{array} \right. .$$

Les équations 1) et 3) sont identiques d'où :

$$\left\{ \begin{array}{l} \pi_1 = \frac{1}{3}\pi_2 \\ \pi_3 = \frac{2}{3}\pi_2 \\ \alpha\pi_0 = \pi_2 - \frac{2}{3}\pi_1 - \frac{2}{3}\pi_3 = \frac{1}{3}\pi_2 \\ \sum_{k=0}^3 \pi_k = 1 \end{array} \right. .$$

On en déduit :

$$\pi_2 = 3\alpha\pi_0, \quad \pi_1 = \alpha\pi_0, \quad \pi_3 = 2\alpha\pi_0.$$

Cette dernière relation donne :

$$\pi_0(1 + \alpha + 3\alpha + 2\alpha) = 1 \Rightarrow \pi_0(1 + 6\alpha) = 1.$$

Par conséquent, la distribution stationnaire est exprimée comme suit :

$$\pi = \left( \frac{1}{1 + 6\alpha}, \frac{\alpha}{1 + 6\alpha}, \frac{3\alpha}{1 + 6\alpha}, \frac{2\alpha}{1 + 6\alpha} \right).$$

On notera que  $\alpha \in [0, 1]$  ce qui implique que  $\alpha \neq \frac{-1}{6}$ .

**Théorème(Existence lois limites)**

**Hypothèse 1.** Il existe des entiers  $k_0, n_0$  et un réel  $\delta > 0$  tels que :

$$P_{ik_0}^{(n_0)} > \delta > 0, \quad i = 0, \dots, l-1,$$

(la condition signifie que les éléments de la colonne  $k_0$  de  $P^{n_0}$  sont bornés inférieurement par  $\delta$ ).

**Conclusion :**

Il existe  $q_0, \dots, q_{l-1}$  tel que :

$$\lim_{n \rightarrow \infty} P_{jk}^{(n)}, \quad j = 0, \dots, l-1,$$

avec :

$$\lim_{n \rightarrow \infty} P^n = \begin{pmatrix} \pi_0 & \pi_1 & \cdot & \cdot & \cdot & \cdot & \pi_{l-1} \\ \pi_0 & \pi_1 & \cdot & \cdot & \cdot & \cdot & \pi_{l-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \pi_0 & \pi_1 & \cdot & \cdot & \cdot & \cdot & \pi_{l-1} \end{pmatrix},$$

et :

$$\lim_{n \rightarrow \infty} q(n) = \lim_{n \rightarrow \infty} q(0)P^n = \pi = (\pi_0, \dots, \pi_{l-1}).$$

Il est à noter que la limite  $\pi$  ne dépend pas de  $q(0)$ .

**Détermination de la loi limite :**  $\pi$  est l'unique solution de :

$$\pi = \pi P \text{ avec } \sum_{k=0}^{l-1} \pi_k = 1.$$

**1.2 Classification des états de la chaîne de Markov****État accessible**

Un état  $j$  est accessible à partir de l'état  $i$  : si  $P_{ij}^{(n)} > 0$  quelque soit  $n \geq 0$ .

**Proposition 7.** La relation d'accessibilité entre états est réflexive et transitive.

**Preuve :**

- a. Réflexivité :** Comme  $P_{ii}^{(0)} = P(X_0 = i | X_0 = i) = 1$ , pour tout état : on a bien  $i \rightarrow i$ .
- b. Transitivité :** Supposons que pour les états  $i, j$  et  $k$  on ait :  $i \rightarrow j$  et  $j \rightarrow k$ . Alors, il existe des entiers positifs  $m$  et  $n$  tels que :  $P_{ij}^{(m)} > 0$  et  $P_{jk}^{(n)} > 0$ . D'où on aura

$$P_{ik}^{(m+n)} = \sum_{l \in E} P_{il}^{(m)} P_{lk}^{(n)} \geq P_{ij}^{(m)} P_{jk}^{(n)} > 0,$$

Ce qui montre bien que l'on a  $i \rightarrow k$  une relation transitive.

□

## États communicants

On dit que les états  $i$  et  $j$  communiquent, si :  $i \leftrightarrow j$  si on a à la fois  $i \rightarrow j$  et  $j \rightarrow i$ .

**Proposition 8.** *La relation de "communication" est bien une relation d'équivalence.*

- a. (Réflexivité).** *Tout état  $i$  de la chaîne communique avec lui même, ie.  $i \rightarrow i$ .*
- b. (Symétrie).** *Si un état  $i$  communique avec un état  $j$ , donc la réciproque est vraie, ie.  $i \rightarrow j \Leftrightarrow j \rightarrow i$ .*
- c. (Transitivité).** *Si un état  $i$  communique avec un état  $j$  qui lui même communique avec un état  $k$ , donc l'état  $i$  communique avec l'état  $k$ , ie. si  $i \rightarrow j$  et  $j \rightarrow k$  alors  $i \rightarrow k$ .*

## Remarque 12.

- Il est clair que tout état d'une chaîne de Markov communique avec lui même, puisque l'on a  $P_{ii}^{(0)} = 1$ . Un état est appelé état de retour, s'il existe  $n \geq 1$  tel que  $P_{ii}^{(n)} > 0$ . Il existe des états  $i$  tels que pour tout  $n \geq 1$  on ait  $P_{ii}^{(n)} = 0$ . De tels états sont appelés états de non retour.*
- L'ensemble des états  $E$  se partitionne en classes d'équivalence, disjointes et non vides, dites classes indécomposables. Si  $C_1$  et  $C_2$  sont deux classes distinctes, il est possible de passer de  $C_1$  à  $C_2$ , mais pas l'inverse (on peut pas retourner de  $C_2$  à  $C_1$ ). En revanche, tous les états d'une même classe communiquent. Certaines classes peuvent ne contenir qu'un seul élément; ce sont les singletons. A titre d'exemple :*
  - Un état de non retour  $i$  :  $P_{ii}^{(0)} = 1, P_{ii}^{(n)} = 0$  pour  $n \geq 1$ ,*
  - Un état absorbant  $i$  :  $P_{ii}^{(0)} = 1, P_{ii}^{(n)} = 1$  pour  $n \geq 1$ .*

## Irréductibilité

S'il n'y a qu'une seule classe pour la relation de communication, ou, si tous les états communiquent entre eux, la chaîne est considérée comme irréductible.

**Exemple 10.** *La marche aléatoire sur  $Z$  est irréductible (tous les états communiquent).*

### Exemple 11.

*Considérons la chaîne de Markov, dont l'ensemble des états est représenté par  $E = \{0, 1, 2\}$ , et sa matrice de transition est :*

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1/3 & 2/3 \end{pmatrix}.$$

*Tous les états communiquent, même si  $P_{0,2} = P_{2,0} = 0$ . Alors, cette chaîne est irréductible.*

### Exemple 12.

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

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

*La chaîne comporte trois classes  $\{0, 1\}$ ,  $\{2\}$  et  $\{3\}$ . L'état 3 absorbant, l'état 2 n'est plus visité après un certain temps. On va voir que les classes  $\{0, 1\}$  et  $\{3\}$  sont récurrentes alors que la classe  $\{2\}$  est transiente.*

## État absorbant

On affirme qu'un état est absorbant si le processus ne peut pas quitter l'état. En d'autres termes, l'état  $i$  est absorbant si et seulement si  $P_{ii} = 1$ .

### Définition 24.

*Si une chaîne de Markov contient au moins un état absorbant et peut passer de n'importe quel état à un état absorbant, on dit qu'elle est absorbante.*

## 1.3 Chaîne de Markov ergodique

### Définition 25.

*Une chaîne de Markov est ergodique si  $(\pi_n)$  converge, quelle que soit la valeur de  $\pi_0$ .*

**Remarque 13.**

*On montre qu' alors l'espace propre associé à la valeur propre 1 est de dimension 1.  $\pi_1$  converge vers  $\pi^*$ , appelée distribution stationnaire.*

**Théorème 10.**

*Les chaînes de Markov irréductibles et apériodiques sont ergodiques.*

**1.4 Exercice sur les chaînes de Markov :**

Avec une probabilité de 0,05, une maladie est contractée. Une fois infecté, un individu peut soit guérir, soit acquérir des séquelles irréversibles. Ces séquelles sont associées à une immunité totale par la suite. Si on guérit, en revanche, on n'est immunisé que dans 50% des cas. Par ailleurs, 1/5 de la population est naturellement immunisé.

1. Modéliser l'état d'un individu dans la période de temps  $(n, n + 1)$  par une chaîne de Markov à 4 états [immunisé = 1; malade = 2; soin(mais pas immunisé) = 3; séquelles irréversibles = 4].
2. Tracer son graphe, sa loi initiale et sa matrice de transition.
3. Classifier les états, on précisera les classes récurrentes, transitoires, absorbantes ainsi que la période.
4. Quelle est la probabilité qu'une personne s'attrape la maladie deux fois de suite et de s'en sortir sans séquelle mais non immunisé ?
5. Y a-t-il une probabilité invariante ? Est-elle unique ? Pourquoi ?

**Correction :**

1. Soit pour  $n \in \mathbb{N}$ , soit  $X_n$  la variable aléatoire donnant l'état de santé d'un individu à la date  $t = n$ .

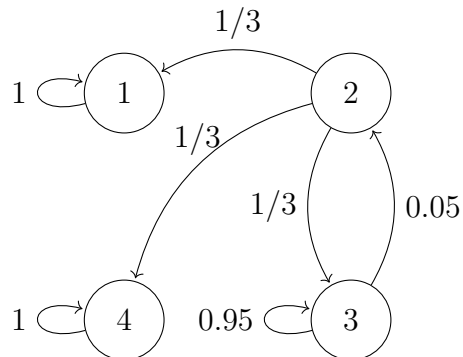
On pose :  $E = 1, 2, 3, 4$  l'ensemble des états.

L'état d'un futur  $t = (n + 1)$  d'un individu est conditionné uniquement par son état présent ( $t=n$ ) :

$$P(X_{n+1} = j | X_n = i, X_{n-1} = \dots, \dots, X_0 = \dots) = P(X_{n+1} = j | X_n = i), \quad \forall i, j \in E.$$

$(X_n)_{n \in \mathbb{N}}$  est donc une chaîne de Markov.

2. Le graphe :



Loi initiale :

$$\mu_0 = \left( \frac{1}{5}, 0, \frac{4}{5}, 0 \right).$$

Matrice de transition :

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 0.05 & 0.95 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

3. On a les classes d'équivalence suivantes :

$$C_1 = 1, C_2 = 2, 3, C_3 = 4.$$

**Proposition :**

Un critère simple pour montrer la périodicité d'un état  $i$  est  $p_{ii} > 0$

**Proposition :**

Soit deux états  $i, j$  tels que  $i \leftrightarrow j$ , alors  $i$  et  $j$  sont de même périodicité.

On a :

- $P_{11} = 1 > 0$  donc l'état est apériodique.
- $P_{33} = 0.95 > 0$  donc l'état 3 est apériodique, et comme  $2 \leftrightarrow 3$ , l'état 2 est aussi apériodique.
- $P_{44} = 1 > 0$  donc l'état 4 est apériodique.
- L'état 1 est absorbant et apériodique.

- L'état 2 est absorbant et apériodique.
- L'état 3 est absorbant et apériodique.
- L'état 4 est absorbant et apériodique.

4. On calcule :  $P(X_4 = 3, X_3 = 2, X_2 = 3, X_1 = 2)$  :

Comme  $(X_n)$  est une chaîne de Markov, on a :

$$\begin{aligned} P(X_4 = 3, X_3 = 2, X_2 = 3, X_1 = 2) &= P(X_4 = 3 | X_3 = 2) \times \\ &P(X_3 = 2 | X_2 = 3) \times P(X_2 = 3 | X_1 = 2) \times P(X_1 = 2) \\ &= P_{23} \times P_{32} \times P_{23} \times P(X_1 = 2). \end{aligned}$$

Or :

$$P(X_1 = 2) = \frac{1}{0} \times 0 + 0 \times 0 + \frac{4}{5} \times 0.05 + 0 \times 0 = \frac{1}{25}.$$

D'où :

$$P(X_4 = 3, X_3 = 2, X_2 = 3, X_1 = 2) = \frac{1}{3} \times 0.05 \times \frac{1}{3} \times \frac{1}{25} = \frac{1}{4500}.$$

5.

$$P_I = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1/3 & -1 & 1/3 & 1/3 \\ 0 & 0.05 & -0.05 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

On a :  $\det(P_I) = 0$ , donc il ya l'existence d'une probabilité invariante.

Unicité de la loi invariante ?

$\mu(a, b, c, d)$  est une loi invariante si  $\mu P = \mu$ , c-à-d :  $\mu(P_I) = 0$ , et  $a + b + c + d = 1$ ,

$$\mu P_I = 0.$$

On a :

$$(a, b, c, d) \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1/3 & -1 & 1/3 & 1/3 \\ 0 & 0.05 & -0.05 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = 0,$$

$$\begin{cases} \frac{b}{3} = 0 \\ -b + 0.05c = 0 \\ \frac{b}{3} - 0.05c = 0 \\ \frac{b}{3} = 0 \end{cases}$$

On trouve :  $b = c = 0$ ,

$$\mu = (a, 0, 0, d)$$

De plus :  $a + B + C + d = 1$ , donc :  $a + d = 1$ , d'où :

$$\mu = (a, 0, 0, 1 - a),$$

où :  $a \in [0, 1]$ .

Donc, il n'y a pas d'unicité de loi invariante.

# Chapitre 2

## Simulation

Dans ce chapitre nous discutons du concept de simulation et de certaines de ses méthodes, ainsi que la simulation par chaîne de Markov.

### 2.1 Simulation

La simulation est un outil mathématique et informatique utile pour observer et analyser des phénomènes réels extrêmement complexes, ce qui n'est pas toujours possible par l'expérimentation. Cela vaut pour la chimie, la physique, la biologie, l'économie et la médecine. Utilisant des modèles mathématiques qui décrivent le fonctionnement d'un système de transport sur de longues périodes de temps réel, cette méthode est utilisée pour exécuter des expériences sur un ordinateur. Ces expériences peuvent porter à la fois sur de petites et de grandes échelles.

#### 2.1.1 Générateur de nombres aléatoires

Les ordinateurs usuels n'étant pas équipés d'une source physique (mesurant l'état d'un photon par exemple), une technique courante pour simuler des variables aléatoires consiste à générer une suite de nombres  $u_1, u_2, \dots$  à valeur dans l'ensemble  $\{0, 1, \dots, m - 1\}$ , pour un entier  $m$  grand, à partir d'une valeur initiale  $u_0$  et d'une opération arithmétique simple du type  $u_{n+1} = au_n + b \pmod{m}$ . Pour des entiers  $a$  et  $b$  bien choisis, la suite  $u_1, u_2, \dots$  ressemble à une suite de variables aléatoires indépendantes de loi uniforme sur:  $\{0, 1, \dots, m - 1\}$ ; elle est cependant parfaitement déterministe et, de ce fait, plutôt qualifiée de suite de nombres *pseudo-aléatoires*. On notera en particulier que cette suite est périodique, de période au plus  $m$ ; il faut ainsi veiller à ne pas faire appel à plus de  $m$  fois consécutives au même générateur aléatoire, sous peine de biaiser les résultats de la simulation.

**Remarque 14.**

*Savoir si la suite générée « ressemble » à une suite parfaitement aléatoire est un problème difficile. Idéalement, il faudrait qu'aucun test statistique ne permette de distinguer lune de l'autre avec une probabilité supérieure à 1/2. La formalisation de cette propriété pour des suites grandes mais finies fait appel à la théorie de la complexité.*

La valeur initiale  $u_0$  de la suite est déterminante (aux deux sens du terme). Considérons un programme informatique faisant appel 10 fois de suite au générateur aléatoire. Lorsque la valeur de  $u_0$  est fixée, deux appels successifs du même programme donneront deux suites de nombres  $u_1, u_2, \dots, u_{10}$  strictement identiques. Pour que deux appels du même programme donnent deux suites de nombres  $u_1, u_2, \dots, u_{10}$  distinctes, il faut changer la valeur initiale  $u_0$ .

Une technique classique consiste à initialiser la valeur de  $u_0$  à partir de l'horloge de l'ordinateur au moment de l'appel du programme. Comme deux appels successifs du même programme ne commencent jamais exactement en même temps, les deux suites de nombres générés  $u_1, u_2, \dots, u_{10}$  seront distinctes. C'est un aspect important lorsque l'on souhaite faire plusieurs expériences aléatoires indépendantes. La valeur initiale  $u_0$  s'appelle la graine du générateur de nombres aléatoires.

Après normalisation de la suite  $u_1, u_2, \dots$  par  $m$ , et pourvu que  $m$  soit assez grand, on obtient une suite de nombres sur  $[0, 1]$  ayant les caractéristiques d'une suite de variables aléatoires indépendantes de loi uniforme sur  $[0, 1]$ . On peut à partir de cette suite générer des variables aléatoires réelles de loi quelconque. Nous verrons deux méthodes génériques : l'inversion de la fonction de répartition et la méthode du rejet. Des méthodes ad-hoc existent également pour certaines lois, comme les lois gaussiennes.

**2.1.2 Méthode d'inversion de la fonction de répartition**

Cette méthode permet de générer une variable aléatoire réelle  $X$  de fonction de répartition  $F$  quelconque à partir d'une variable aléatoire  $U$  de loi uniforme sur  $[0, 1]$ . On suppose tout d'abord que la fonction  $F$  est continue et strictement croissante sur  $[a, +\infty[$ , avec  $a \in \mathbf{R}$  et  $F(a) = 0$ ; elle définit alors une bijection de  $[a, +\infty[$  vers  $]0, 1[$  et on définit la variable aléatoire  $X = F^{-1}(U)$ .

Vérifions que  $X$  a bien pour fonction de répartition  $F$  :

$$\begin{aligned} \forall x > a, \mathbf{P}(X \leq x) &= \mathbf{P}(F^{-1}(U) \leq x) \\ &= \mathbf{P}(U \leq F(x)) \\ &= F(x). \end{aligned}$$

**Exemple 13.** Pour une loi exponentielle de paramètre  $\lambda$ , on a  $F(x) = 1 - e^{-\lambda x}$  sur

$[0, +\infty[$  de sorte que :

$$X = -\frac{1}{\lambda} \ln(1 - U).$$

Graphiquement, cela revient à générer une variable aléatoire uniforme sur  $[0, 1[$  sur l'axe des ordonnées et à prendre son antécédent par la fonction  $F$  sur l'axe des abscisses, comme illustré par la figure 2.1.

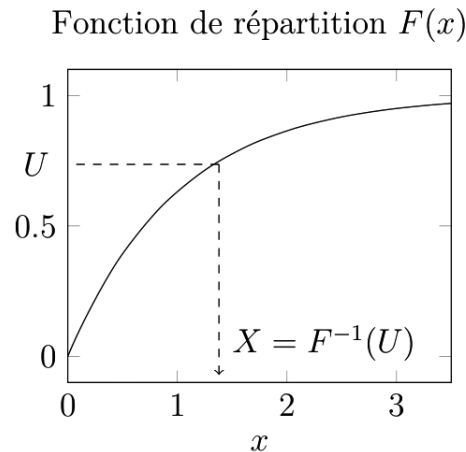


Figure 2.1: Génération d'une variable aléatoire de loi exponentielle de paramètre 1.

Le principe est le même pour une fonction de répartition  $F$  quelconque (croissante sur  $\mathbf{R}$ , continue à droite), en prenant pour fonction inverse la fonction  $F^{-1}$  définie par :

$$\forall y \in ]0, 1[, F^{-1}(y) \equiv \inf\{x \in \mathbf{R} : y \leq F(x)\},$$

puisque l'on a toujours  $F^{-1}(y) \leq x$  si et seulement si  $y \leq F(x)$ .

### 2.1.3 La méthode du rejet

La méthode d'inversion nécessitait la connaissance explicite de  $F_x^{-1}$ , ce qui n'est pas toujours le cas (surtout pour les lois normales). Elle est d'abord utilisée pour simuler des lois conditionnelles ou des lois uniformes sur des domaines de  $\mathbb{R}^d$  en utilisant le résultat suivant.

**Proposition 9.** Soit  $(Y_n)$  une suite de variables aléatoires indépendantes de loi  $\mu$  et  $B$  un borelien tel que  $\mu(B) > 0$ . Soit  $T$  le plus petit entier  $n \geq 1$  tel que  $Y_n \in B$ . Donc :

- (i)  $T$  est une variable aléatoire de loi géométrique de paramètre  $\mu(B)$ ,
- (ii)  $Y_T$  est une variable aléatoire ayant pour loi la loi conditionnelle  $\mu(\cdot|B)$ . En particulier lorsque  $\mu$  est la loi uniforme sur un borélien  $C$  contenant  $B$  et tel que  $0 < \lambda(B) < \lambda(C)$ , cette loi conditionnelle est la loi uniforme sur  $B$ .

On a noté  $\lambda$  la mesure de Lebesgue. En pratique on peut donc simuler une variable  $X$  de loi uniforme sur  $B$  de la façon suivante: On tire  $Y_1$  uniforme sur  $C$ , si  $Y_1 \in B$ , on pose  $X = Y_1$ , sinon on retire une variable  $Y_2$  uniforme sur  $C$  et on recommence.

Cette méthode permet également de simuler des lois à densités en considérant le domaine sous le graphe de la densité.

Le facteur crucial est le "théorème" suivant :

**Théorème 11. Théorème fondamental de la simulation :**

Soit  $f$  une densité sur  $\mathbb{R}^d$ . Alors simuler  $X$  de densité  $f$  est équivalent à simuler  $(X, U)$  de loi uniforme (ie. Lebesgue) sur :

$$\{(x, u) : (x, u) \in \mathbb{R}^{d+1}, 0 < u < f(x)\}.$$

En effet, la loi marginale du couple  $(X, U)$  admet bien la densité  $f$ , car par le théorème de Fubini-Tonelli, pour tout  $g$  mesurable positive,

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}^{d+1}} g(x) \mathbf{1}_{0 < u < f(x)} dx du = \int_{\mathbb{R}^d} \left( \int \mathbf{1}_{0 < u < f(x)} du \right) g(x) dx = \int_{\mathbb{R}^d} f(x) g(x) dx.$$

Inversement, conditionnellement à  $X = x$ , la variable aléatoire  $U$  suit la loi uniforme dans l'intervalle  $[0, f(x)]$  qui peut facilement être simulée.

**Méthode du rejet comparatif :**

On souhaite simuler  $X$  de densité  $f$  sur  $\mathbb{R}^d$ . On suppose qu'il existe une densité de probabilité  $g$  telle que :

- Nous savons comment simuler une variable de densité  $g$ ,
- Il existe une constante  $K (K \geq 1!)$  telle que  $f \leq Kg$ .

Ensuite, nous utilisons la méthode suivante, qui est presque certaine de se terminer et qui simule une variable  $X$  de densité  $f$  :

1. On simule une variable  $U$  de densité  $g$  et une variable  $W$  indépendante de  $X$  de loi uniforme sur  $[0, 1]$ .

Posons :  $V = KWg(U)$ .

2. Si  $V \leq f(U)$  on pose  $X = U$ , sinon, il faut revenir à (1).

### 2.1.4 Méthode de Monte Carlo

La méthode de Monte Carlo fait référence à une famille de techniques qui estiment une valeur numérique en simulant un système aléatoire. Nous pouvons, par exemple, estimer le temps de service moyen dans un problème de file d'attente en développant un algorithme de simulation pour ce système. En mathématiques, statistiques, physique, chimie et ingénierie, on rencontre fréquemment des intégrales qui doivent être résolues numériquement.

En règle générale, ces intégrales peuvent être exprimées sous forme de probabilité et estimées à l'aide de la méthode de Monte Carlo. C'est une méthode probabiliste car il faut générer des variables aléatoires. Prenons un exemple classique pour illustrer cette méthode.

**Exemple 14.** Une technique pour estimer le nombre  $\pi/4$  consiste à calculer la probabilité que la variable aléatoire  $X \sim \text{Unif}([0, 1] \times [0, 1])$  soit de norme inférieure ou égale à un,

$$\mathbb{P}[\|X\| \leq 1] = \mathbb{E}[\mathbb{I}(\|X\| \leq 1)] = \int \int_{[0,1] \times [0,1]} (\mathbb{I}\|x\| \leq 1) dx.$$

Nous pouvons estimer cette intégrale en deux étapes avec la méthode de Monte Carlo :

(i) produire  $n$  variables indépendantes  $X_i \sim \text{Unif}([0, 1] \times [0, 1])$  pour  $i = 0, \dots, n - 1$ ,

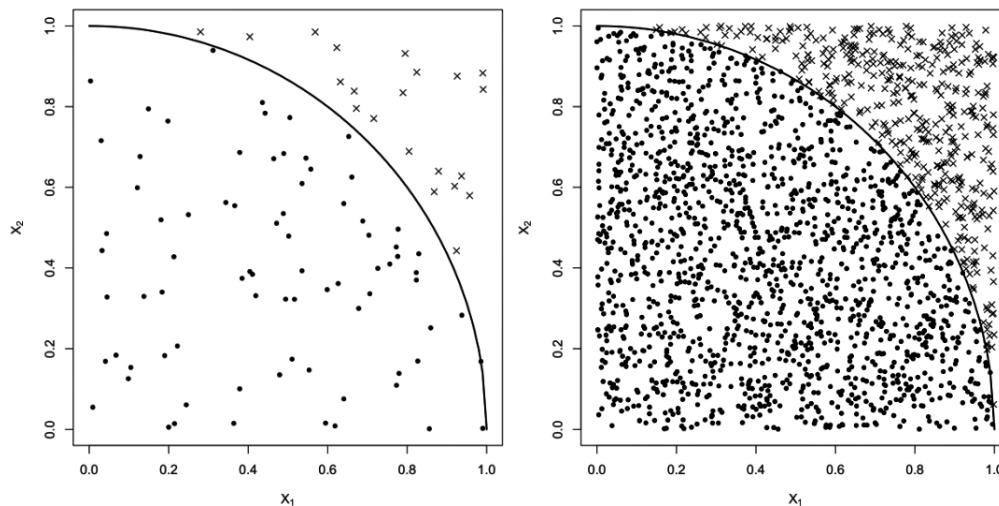


Figure 2.2: Estimation de  $\pi/4$ .

Pour  $n = 100$  (gauche) et  $n = 2000$  (droite) nous notons d'un point quand la variable  $X$  est de norme inférieure ou égale à un, sinon nous notons  $X$  par une croix.

(ii) L'intégrale  $\hat{I}_n$  a comme estimateur :

$$\hat{I}_n := \sum_{i=0}^{n-1} \frac{\mathbb{I}(\|X_i\| \leq 1)}{n}.$$

À la figure 2.2, nous illustrons cette estimation pour  $n = 100$  ( $\hat{I}_n = 3/4$ ) et  $n = 2000$  ( $\hat{I}_n = 3,11/4$ ).

De façon générale, pour une variable aléatoire  $X \in \mathcal{X}$  de mesure de probabilité  $\pi$  et une fonction  $f$  définie sur  $\mathcal{X}$  à valeur dans  $\mathbb{R}$ , la difficulté réside dans le calcul de l'intégrale:

$$\mathbb{E}_\pi[f(X)] = \int_{\mathcal{X}} f(x)\pi(dx), \quad (2.1)$$

supposons qu'elle existe. Comme l'exemple précédent, on va estimer de cette intégrale par la méthode de Monte Carlo en deux étapes :

1. générer  $n$  variables indépendantes  $X_i$  selon la mesure de probabilité  $\pi$ ;
2. l'estimateur de l'intégrale est :

$$\bar{f}_n := \sum_{i=0}^{n-1} \frac{f(X_i)}{n}.$$

L'estimateur de Monte Carlo est simplement la moyenne échantillonnale. Il est sans biais, comme les variables aléatoires  $X_i$  sont générées selon  $\pi$ . De plus, l'estimateur converge vers l'espérance avec probabilité un. En effet, comme les variables sont indépendantes et identiquement distribuées (i.i.d.), nous savons, grâce à la loi des grands nombres, que :

$$\bar{f}_n \xrightarrow[n \rightarrow +\infty]{p.s.} \mathbb{E}_\pi[f(X)],$$

où p.s. fait référence à la convergence presque sûre. Le point crucial pour pouvoir utiliser la méthode de Monte Carlo afin d'estimer  $\mathbb{E}_\pi[f(X)]$  est de pouvoir générer des variables i.i.d. . Pour les problèmes à haute dimension et/ou les statistiques bayésiennes, il n'est pas toujours viable de construire un échantillon i.i.d. selon  $\pi$ .

L'approche Monte Carlo via les chaînes de Markov apporte une solution à ce problème. Avant de la présenter, nous rappelons brièvement le modèle des chaînes de Markov. La variable aléatoire  $X$  de mesure  $\pi$  sur  $(\mathcal{X}, \mathcal{X})$  sera toujours la variable d'intérêt dans le calcul de l'intégrale 2.1. pour des définitions plus générales, nous utilisons la variable  $Z$  et la mesure  $\mu$  sur  $(E, \mathcal{E})$ .

**Exemple 15.** En utilisant l'approche de Monte Carlo, pour le calcul de l'intégrale :  
Afin de calculer l'intégral :

$$I = \int_2^{12} \exp(-(x^2 + x)/3) \sin(x) dx.$$

On pose :

$$\psi(x) = \exp(-(x^2 + x)/3) \sin(x).$$

Pour l'écriture de la fonction, on utilise la syntaxe "function" du langage R comme suit :

```
1 le nom de la fonction <- function(liste des arguments) {le
2 corps de la fonction}.
```

Le code de R ci-dessous :

1. # L'écriture de la fonction :  $\psi(x)$ .

```
1 psi=function(x) {exp(-(x^2+x)/3)*sin(x)}
2
3 # Est ce qu'elle fonctionne ?
4
5 psi(0);psi(-3);psi(pi/2)
6
```

2. Exécution dans la console R :

```
1 > psi=function(x) {exp(-(x^2+x)/3)*sin(x)}
2
3 > # Est ce qu'elle fonctionne ?
4
5 > psi(0);psi(-3);psi(pi/2)
6 [1] 0
7 [1] -0.01909852
8 [1] 0.2602622
9
```

Il est nécessaire de charger le package contenant la syntaxe de deux manières différentes. La première étape consiste à cliquer sur "package" ensuite "charger package" puis sélectionne le nom de package après clique sur "ok". La seconde est plus simple ; nous écrivons Library(le nom de package) dans la console R.

**Utilisation de syntaxe "integrate" :**

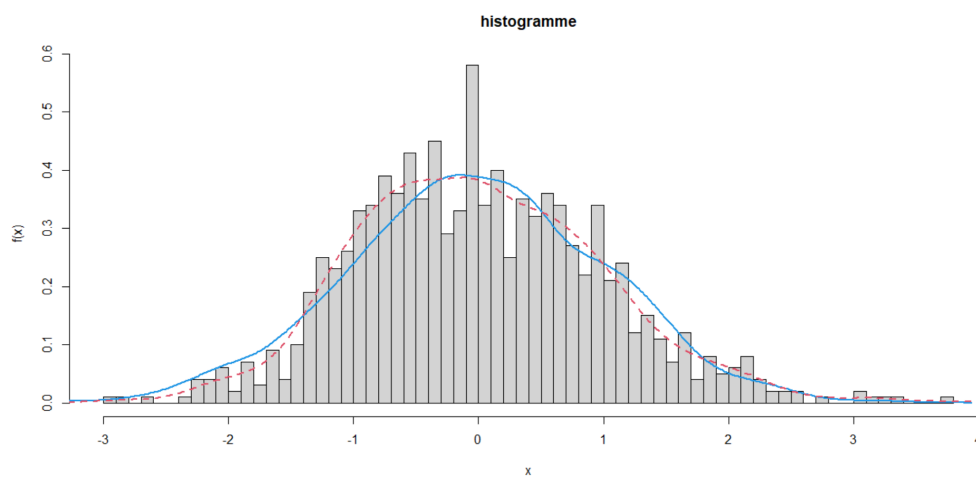
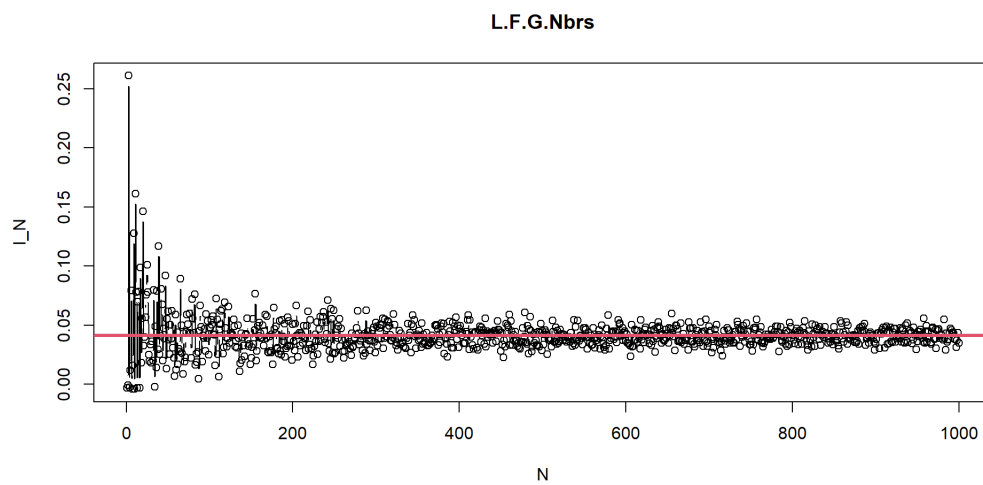
Pour utiliser la syntaxe "integrate", vous devez charger le package "stats". On utilise la syntaxe "integrate" du langage de programmation R sous la forme :

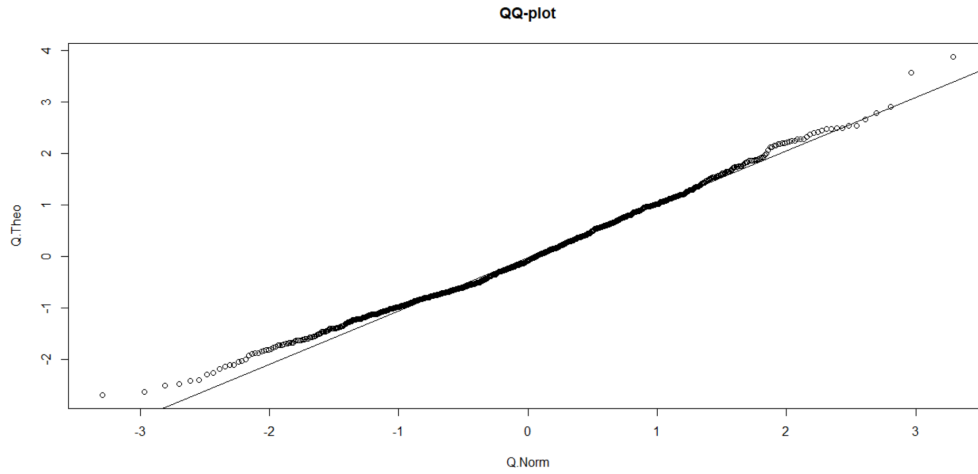
```
1 le nom d'integrale <- integrate(le nom de la fonction,
2 la borne infereur d'integrale, la borne superieur d'integrale)
3
```

**Le programme R :**

```
1 ### Simulation par methode de Monte-Carlo
2 # Integration par simulation de Monte-Carlo
3
4 psi=function(x){exp(-(x^2+x)/3)*sin(x)}
5
6 library(stats) #charger le package "stats"
7 I=integrate(psi,2,12)$v
8
9 pis2=function(x){(exp(-(x^2+x)/3)*sin(x))^2}
10
11 J=integrate(pis2,2,12)$v
12
13 var=10*J-I^2
14
15 ##Loi Forte des Gr Nbr
16 N=1000
17
18 In=numeric(N)
19
20 Zn=numeric(N)
21
22 for(K in 1:N){
23 V=runif(K,2,12)
24 In[K]=10*mean(psi(V))
25 Zn[K]=(In[K]-I)/sqrt(var/K)
26 }
27
28 N=1:1000
29
30 plot(N, In, typ='b', xlab="N", ylab=expression(I_N), main="L.F.G.Nbrs")
31
32 abline(h=I, lwd=3, col=2)
33 ## T C L
34 for(j in 1:N){}
35
36 x11()
37
38 qqnorm(Zn, xlab="Q.Norm", ylab="Q.Theo", main="QQ-plot")
39
40 qqline(Zn)
41
42 x11()
43
44 hist(Zn, 50, prob=TRUE, xlab="x", ylab="f(x)", main="histogramme")
45
46 x=rnorm(length(Zn))
47
48 lines(density(x), col=4, lwd=2)
```

```
49  
50 lines(density(Zn), col=2, lwd=2, lty=2)  
51  
52 legend(1, -4, c("densite theorique", "densite Empirique"),  
53  
54 col=c(4, 2), lty=c(1, 2), lwd=c(2, 2), bty="n", cex= 1.3)
```





### 2.1.5 Histogramme

Pour vérifier si la simulation produit bien la loi de probabilité recherchée, on peut tracer l'histogramme obtenu à partir d'un nombre  $N$  d'échantillons  $x_1, \dots, x_N$ , censés correspondre à  $N$  réalisations d'une variable aléatoire réelle  $X$ . On prend pour simplifier des intervalles de même largeur  $\delta > 0$  et on note  $g_k$  la fraction des échantillons se trouvant dans l'intervalle  $[k\delta, (k+1)\delta[$  :

$$\forall k \in \mathbf{Z}, g_k = \frac{1}{N} \sum_{j=1}^N \mathbf{1}_{\{k\delta \leq x_j < (k+1)\delta\}}.$$

Lorsque le nombre d'échantillons  $N$  est grand,  $g_k$  doit être proche de la probabilité que  $X$  soit dans l'intervalle  $[k\delta, (k+1)\delta[$ . Par la loi forte des grands nombres, on sait en effet que pour toute suite de variables aléatoires indépendantes  $X_1, X_2, \dots$  de même loi que  $X$  :

$$\frac{1}{N} \sum_{j=1}^N \mathbf{1}_{\{k\delta \leq X_j < (k+1)\delta\}} \xrightarrow{p.s.} \mathbf{E} \left[ \mathbf{1}_{\{k\delta \leq X < (k+1)\delta\}} \right] = \mathbf{P}(k\delta \leq X < (k+1)\delta).$$

Lorsque la variable aléatoire  $X$  a pour densité  $f$ , et pour une largeur d'intervalle  $\delta$  suffisamment faible, on peut comparer  $f$  à la fonction de densité  $g$  constante par morceaux définie par :

$$\forall x \in \mathbf{R}, g(x) = \frac{1}{\delta} \sum_{k \in \mathbf{Z}} g_k \mathbf{1}_{[k\delta, (k+1)\delta[}(x).$$

## 2.2 Simulation par chaîne de Markov

Les méthodes de simulation des chaînes de Markov ont parcouru un long chemin ces dernières années. Ils sont aujourd'hui utilisés dans un large éventail de domaines, du traitement d'images à la physique statistique.

Si  $M$  est un ensemble fini (en pratique très grand) et  $\pi$  une loi sur  $M$  que l'on cherche à simuler; l'idée de ces méthodes consiste à exprimer  $\pi$  comme la mesure invariante d'une chaîne de Markov ergodique sur  $M$ .

### 2.2.1 Méthode de MCMC

Les méthodes MCMC nous facilitent la vie en nous fournissant des Algorithmes qui pourraient créer une chaîne de Markov dont la distribution bêta est sa distribution stationnaire étant donnée que nous pouvons échantillonner à partir d'une distribution uniforme.

Soit  $M$  un espace d'états fini et  $\pi$  la mesure invariante d'un noyau  $P$  irréductible et apériodique sur  $M$ . Supposons que l'on cherche à calculer la valeur moyenne d'une fonction  $f : M \rightarrow \mathbb{R}$  sous  $\pi$

$$\mathbb{E}_\pi(f) = \sum_{x \in M} f(x)\pi(x);$$

ou bien que l'on veuille produire un échantillon de loi  $\pi$ .

La méthode MCMC consiste à générer les états successifs  $X_0, \dots, X_N$  d'une chaîne de Markov de noyau  $P$  à partir d'un état initial  $X_0$  de loi quelconque (ex: la loi uniforme ou une masse de Dirac). D'après le théorème ergodique, la fréquence empirique :

$$S_N(f) = \frac{1}{N} \sum_{k=1}^N f(X_k),$$

est (pour  $N$  assez grand!) une bonne approximation de  $E_\pi(f)$ , et la loi de  $X_N$  est proche de  $\pi$ .

### 2.2.2 Algorithme de Metropolis

L'algorithme de Metropolis-Hastings est l'un de ces algorithmes MCMC.

Dans de nombreux cas, la loi  $\pi$  que l'on cherche à simuler n'est pas donnée a priori comme la mesure invariante d'une chaîne de Markov. L'algorithme de Metropolis produit une chaîne de Markov réversible par rapport à  $\pi$ .

On se donne une matrice de transition markovienne  $Q$  sur  $M$ , appelée matrice de sélection, telle que tout couple  $(x, y) \in M$ ,

$$Q(x, y) > 0 \Rightarrow Q(y, x) > 0.$$

Soit:  $h : ]0, \infty[ \rightarrow ]0, 1]$  une fonction vérifiant :

$$h(u) = uh\left(\frac{1}{u}\right).$$

Par exemple :

$$h(u) = \inf(u, 1),$$

ou bien :

$$h(u) = \frac{u}{1+u}.$$

Pour  $x \neq y$ , posons :

$$R(x, y) = \begin{cases} h\left(\frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)}\right) & \text{si } Q(x, y) \neq 0, \\ 0 & \text{sinon.} \end{cases} \quad (2.2)$$

Ensuite, on construit une matrice de transition  $P$ , qui est donnée par :

$$\begin{aligned} P(x, y) &= Q(x, y)R(x, y) \\ \text{pour } x \neq y, \text{ et} & \\ P(x, x) &= 1 - \sum_{y \neq x} P(x, y). \end{aligned} \quad (2.3)$$

La proposition suivante est immédiate.

**Proposition 10.** *Supposons que  $\pi$  charge tous les points de  $M$ . La matrice  $P$  définie par 2.2 et 2.3 est réversible par rapport à  $\pi$ . Elle est irréductible si  $Q$  est irréductible. Si de plus  $h(u) < 1$  (par exemple:  $h(u) = \frac{u}{1+u}$ ), elle est apériodique.*

L'algorithme suivant, correspondant à la simulation de la chaîne de noyau  $P$ , s'appelle l'algorithme de Metropolis.

**Étape 0.** Initialisation de  $X_0$ .

**Étape  $n+1$ .**

(Sélection) Choisir  $y$  avec la loi  $Q(X_n, y)$ .

Choisissez un nombre  $U$  au hasard dans  $[0, 1]$ .

Si  $U < R(X_n, y)$  accepter la sélection:

$X_{n+1} = y$ ;

Sinon, la sélection :  $X_{n+1} = X_n$  est refusée.

**Exemple 16. (Le modèle des sphères dures).**

Soit  $S$  un ensemble fini et  $G = (S, E)$  un graphe symétrique sur  $S$ . Une configuration de sphères dures sur  $G$  est une suite finie,

$$x = (x(s))_{s \in S} \in \{0, 1\}^S,$$

telle que :

$$x(s)x(v) = 0.$$

Si  $s$  et  $v$  sont deux sites adjacents. La condition  $x(s) = 1$  (respectivement  $x(s) = 0$ ) s'interprète comme la présence (respectivement l'absence) d'une particule au site  $s$ .

Les particules ont un diamètre non négligeable de sorte que deux particules ne peuvent pas occuper des positions voisines.

Soit  $M \subset \{0, 1\}^S$  l'ensemble des configurations de sphères dures et soit  $\pi$  la probabilité uniforme sur  $M$ . On peut simuler  $\pi$  par l'algorithme suivant :

**Étape 0.** (Initialisation)  $X_0(s) = 0$  pour tout  $s \in S$ .

**Étape  $n+1$ .**

Choisir un sommet  $s$  au hasard dans  $S$ .

Si tous les voisins de  $s$  sont libres, choisir  $U$  au hasard dans (Pile, Face).

Si  $U = \text{Pile}$  :

$$X_{n+1}(s) = 1;$$

Si  $U = \text{Face}$ :

$$X_{n+1}(s) = 0.$$

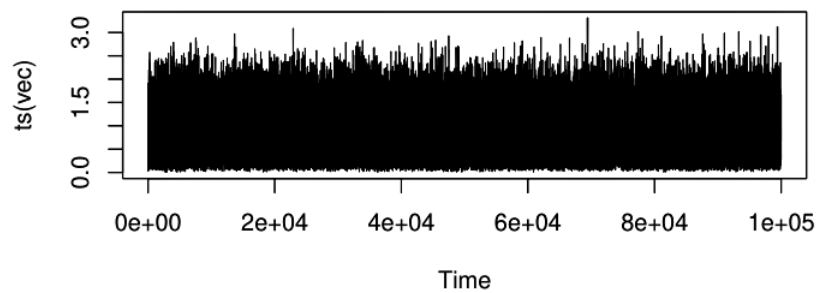
**Exemple 17.** Ceci est un exemple de l'algorithme de Metropolis simulant la distribution Gamma  $\text{Gam}(2,3 - 2,7)$

En utilisant 100000 échantillons de la distribution de proposition normale  $N(0.85, 0.30)$ . Les graphiques ci-dessous montrent que la chaîne se mélange bien et qu'il existe une indépendance entre les échantillons générés. Également un histogramme des échantillons générés ressemblant à la courbe de densité gamma.

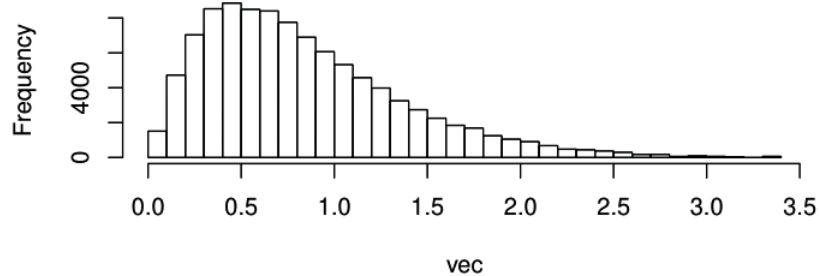
**Code R :**

```
1 > gamm<-function (n, a, b)
2 + {
3 + mu<-a/b
4 + sig<-sqrt (a/(b*b))
5 + vec<-vector("numeric", n)
```

```
6 + x<-a/b
7 + vec[1]<-x
8 + for(i in 2:n){
9 + candidate<-rnorm(1, mu, sig)
10 + accept<-min(1, (dgamma(candidate, a, b)/dgamma(x,
11 + mu, sig)))
12 + u<-runif(1)
13 + if(u<accept)
14 + x<-candidate
15 + vec[i]<-x
16 + }
17 + vec
18 + }
19 > vec<-gamm(100000, 2.3, 2.7)
20 > c(mean(vec), var(vec))
21 [1] 0.8401326 0.2851852
22 > par(mfrow=c(2,1))
23 > plot(ts(vec))
24 > hist(vec, 30)
25 > par(mfrow=c(1,1))
26 >
```



**Histogram of vec**



### 2.2.3 Simulation des lois de Gibbs

Soit  $X = (X_1, \dots, X_d)$  un vecteur aléatoire de probabilité  $\pi(X)$  que l'on veut simuler. On reste pour simplifier dans le cas où  $X$  prend un nombre fini de valeurs bien que ce ne soit pas nécessaire.

L'échantillonneur de Gibbs consiste à créer directement une chaîne de Markov réversible indécomposable apériodique de mesure invariante  $\pi$  en tirant chaque coordonnée à son tour selon sa loi conditionnelle aux autres. On note :

$$X^{k,x} = (X_1, \dots, X_{k-1}, x, X_{k+1}, \dots, X_d).$$

#### Algorithme :

Tirage de  $X'$  (échantillon suivant) après  $X$  :

1. Tirer  $k$  uniformément dans  $\{1, \dots, d\}$ .
2. Tirer  $x$  selon la loi de  $X_k$  conditionnelle aux autres :

$$Q_k(x|X) = \pi(X_k = x | X_j, j \neq k) = \frac{\pi(X^{k,x})}{\sum_y \pi(X^{k,y})},$$

puis  $X' = X^{k,x}$ .

### 2.2.4 Optimisation globale et recuit simulé

Le recuit simulé est un algorithme d'optimisation qui permet le calcul des minima globaux d'une fonction définie sur un ensemble fini. Avant de parler de l'algorithme, précisons quelques points :

#### La composition énergie-entropie :

Revenons sur l'algorithme de Metropolis. Nous savons, que l'entropie de Kullback de la loi de  $X_n$  par rapport à  $\pi_T$  décroît au cours de temps. Ici cette entropie s'écrit :

$$H(\mu, \pi_T) = \sum_{x \in M} \mu(x) \ln \left( \frac{\mu(x)}{\pi_T(x)} \right) = -S(\mu) + \frac{1}{T} E_\mu(V) + \ln(Z_T),$$

où :

$$S(\mu) = - \sum_{x \in M} \mu(x) \ln(\mu(x)),$$

et l'entropie de  $\mu$ , et :

$$E_\mu(V) = \sum_{x \in M} \mu(x) V(x),$$

est l'énergie moyenne pour la loi  $\mu$ .

Si  $\mu_n$  désigne la loi de  $X_n$ , l'algorithme de Metropolis est donc tel que la suite :

$$\left( S(\mu_n) - \frac{1}{T} E_{\mu_n}(V) \right)_{n \geq 0},$$

décroît au cours du temps.

À haute température, le terme d'entropie est dominant et les variables  $X_n$  se comportent approximativement comme des variables i.i.d. de loi uniforme sur  $M$ . À basse température, c'est le terme d'énergie qui domine, et l'algorithme de Metropolis vise essentiellement à minimiser  $V$ .

Cette compétition entre énergie et entropie se retrouve à l'équilibre, comme le montre le lemme suivant :

**Lemme 4.**

(a)  $\lim_{T \rightarrow \infty} \pi_T(x) = \frac{1}{|M|}.$

(b) Soit  $V_{\min}$  l'ensemble des minima globaux de  $V$ . Alors :

$$\lim_{T \rightarrow 0} \pi_T(x) = \frac{\mathbf{1}_{\{x \in V_{\min}\}}}{|V_{\min}|}.$$

**Démonstration :**

Pour  $x \neq y$  :

$$\frac{\pi_T(x)}{\pi_T(y)} = \exp\left(-\frac{1}{T}(V(x) - V(y))\right).$$

D'où le résultat. □

**Le recuit simulé :**

En métallurgie, le recuit est un procédé qui consiste à faire fondre plusieurs fois un métal puis à le laisser refroidir lentement afin d'améliorer ses propriétés mécaniques. Cette

méthode est la base de l'algorithme de recuit simulé. On se donne une suite décroissante de températures  $\{T(n)\}_{n \geq 0}$  telle que  $\lim_{n \rightarrow \infty} T(n) = 0$ . L'algorithme est alors semblable à l'algorithme de Metropolis, sauf que la température change avec le temps.

**Étape 0.** Initialiser  $X_0$

**Étape n+1.** (Sélection) Choisir au hasard  $y$  parmi les voisins de  $X_n$ .

Choisissez un nombre  $U$  au hasard parmi  $[0, 1]$ .

Si  $U < h \left( \exp\left(\frac{1}{T(n)}(V(X_n) - V(y))\frac{|N(y)|}{|N(x_n)|}\right) \right)$  la sélection est acceptée :

$$X_{n+1} = y;$$

Si ce n'est pas le cas, n'accepter pas la sélection :

$$X_{n+1} = X_n.$$

**Théorème 12.** Il existe une constante  $C(V) > 0$  (dépendante de  $V$ ) telle que, si

$$\lim_{n \rightarrow \infty} T(n) = 0 \text{ et } \sum_n \exp\left(-\frac{C(V)}{T(n)}\right) = \infty,$$

l'algorithme du recuit simulé sélectionne les minima globaux de  $V$ , c'est-à-dire :

$$\lim_{n \rightarrow \infty} P(X_n \in V_{\min}) = 1.$$

Pour utiliser l'algorithme, vous pouvez choisir l'un des schémas de température ci-dessous :

(i) (Décroissance logarithmique)

$$T(n) = \frac{C}{\ln(n)}.$$

(ii) (Recuit par palier)

$$T(n) = \frac{1}{k} \text{ pour } e^{(k-1)C} \leq n < e^{kC}.$$

Si  $C > C(V)$ , les conditions du théorème sont vérifiées.

## 2.2.5 Simulation exacte : l'algorithme de Propp-Wilson

### Présentation de l'algorithme de Propp-Wilson :

Soit  $E$  un ensemble fini et  $\pi$  une loi de probabilités sur  $E$ .

L'algorithme de Propp-Wilson est basé sur la construction d'une chaîne de Markov  $(X_n^x)_n$  interactive de loi stationnaire  $\pi$ , et du processus  $(H_n^x)_n$  associé appelé "backward process", plus précisément :

Soit  $(Y_n)_n$  une suite de variables aléatoires i.i.d. à valeurs dans  $E$ , et de loi  $\mu$ .

Pour  $x, y \in E$ , soit la fonction  $x \rightarrow f_Y(x)$  définie de  $E$  dans  $E$ .

La chaîne de Markov  $(X_n^x)_n$  d'espace des états  $E$  définie par :

$$\begin{cases} X_0^x = x; \\ X_n^x = f_{Y_n}(X_{n-1}^x), \quad \forall n > 0. \end{cases}$$

Supposons que la chaîne de Markov  $(X_n^x)_n$  est de loi stationnaire  $\pi$ . De plus on suppose qu'elle est irréductible et apériodique.

#### Définition 26.

On appelle "backward process" associé à  $(X_n^x)_n$ , le processus  $(H_n^x)_n$  définie par :

$$\begin{cases} H_0^x = x; \\ H_n^x = f_{Y_1} \circ f_{Y_2} \circ \dots \circ f_{Y_n}(x), \quad \forall n > 0 \end{cases}$$

Dans les conditions (H) que nous allons préciser, l'algorithme est basé sur le fait que :

1.  $\lim_n H_n^x$  est constante.
2. Si  $(X_n)_n$  admet la loi stationnaire  $\pi$ , alors  $\lim_n H_n^x$  est une variable aléatoire de loi  $\pi$ .
3. Si  $T$  est la variable aléatoire définie par :

$$T = \inf_n \{n \in \mathbb{N}, H_n^x = cte\},$$

alors  $T$  est fini, p.s.

L'algorithme consiste donc à construire un échantillon  $(Z_1, Z_2, \dots, Z_n)$  de loi stationnaire  $\pi$ , en considérant :

$$Z_1 = \lim_n H_n^{x_1},$$

$$Z_2 = \lim_n H_n^{x_2},$$

$$\vdots$$

$$Z_n = \lim_n H_n^{x_n},$$

avec les  $Z_n$  distribués suivant la loi  $\pi$ , car si  $\pi$  est une loi stationnaire pour  $X_n^x$ , c'est aussi la loi de  $H_n^x$ , car les processus  $(X_n^x)_n$  et  $(H_n^x)_n$  sont de même loi. Il en résulte que  $\lim_n H_n^x$  est une variable aléatoire de loi  $\pi$ .

## Application à la simulation à l'aide de R

Il existe plusieurs façons d'atteindre l'exécution informatique de la simulation, mais certains sont plus efficaces que l'autres. Dans ce chapitre, nous utiliserons un exemple simple mais bien détaillé pour montrer comment utiliser la simulation avec R.

### 3.1 Qu'est ce que R ?

R est un langage de programmation avec des fonctions statistiques intégrées et des modules complémentaires qui aident les ingénieurs et les scientifiques à résoudre des problèmes statistiques avec moins de lignes de code que les langages de programmation traditionnels comme Java ou C ou C++.

Il a également des graphismes de haute qualité qui peuvent être modifiés. R fonctionne sous Unix/Linux, Windows et Mac. R possède toutes ces fonctionnalités et est également gratuit.

### 3.2 Application

Les pourcentages de tirs d'une équipe de basket-ball seront désormais modélisés comme une distribution normale à l'aide de l'algorithme Metropolis. Le pourcentage de tirs d'une équipe de basket-ball est le nombre de tirs à 3 points et à 2 points effectués par chaque joueur divisé par le nombre de tirs effectués par l'équipe dans son ensemble.

Considérons maintenant le pourcentage de tirs d'une équipe de basket-ball sur  $n$  matchs :

$$x = (x_1, \dots, x_n).$$

La distribution bêta peut être utilisée pour modéliser ces pourcentages,

$$x_i \sim \text{Beta}(\theta, 2) \text{ pour } \theta > 0$$

$$f(x_i|\theta) = \theta(1 + \theta)x_i^{\theta-1}(1 - x_i)$$

et on utilise  $\text{Gamma}(a,b)$  comme distribution antérieure pour  $\theta$ ,

$$p(\theta|x) \sim \theta^{n+a-1}(\theta + 1)^n e^{-b\theta} \left( \prod_{i=1}^n x_i - i \right)^\theta := h(\theta).$$

Utilisons les pourcentages de tirs sur les  $n = 20$  matchs avec :

$$\sum_{i=1}^{20} \log x_i = -9.89.$$

Choisissez  $a = b = 1$  pour la distribution antérieure.

Postérieure approximée antérieurement on utilise le théorème de la centrale limite bayésienne :

$$p(\theta|x) \approx N(3.24, 0.33).$$

Utiliser l'algorithme de Metropolis pour échantillonner  $p(\theta|x)$ .

#### Les étapes de l'algorithme de Metropolis :

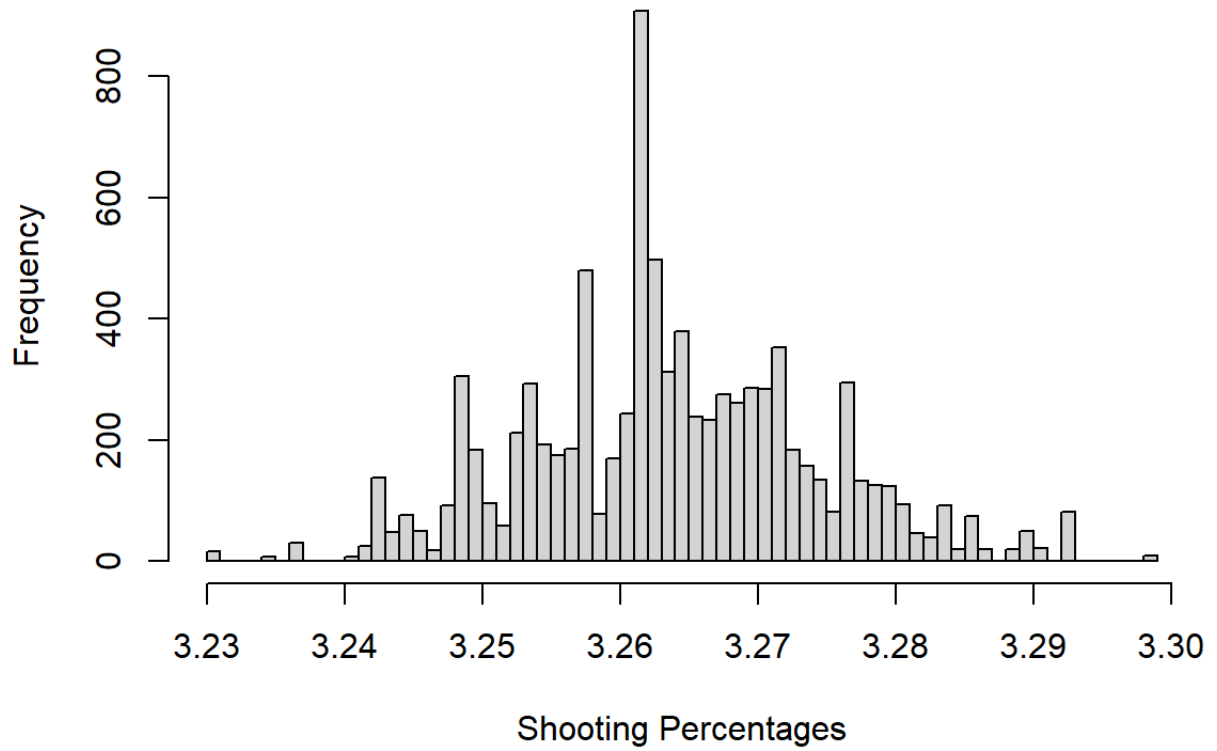
- La valeur initiale  $\theta^{(0)} = 3.24$ .
- L'échantillon de  $p(\cdot|\theta) = N(\theta^{t-1}, 0.33)$ .
- L'échantillon  $\mu$  de l'uniforme  $(0,1)$ .
- Accepter ou rejeter la proposition basée sur si  $\mu < \frac{N(\theta^t, 0.33)}{N(\theta^{t-1}, 0.33)}$  alors  $\theta^{i+1} = \theta$  sinon  $\theta^{i+1} = \theta$ .

Nous ferons cette itération 1000 fois, et les 50 premières seront traitées comme des "burnins".

**Code R :**

```
1
2
3 x=rnorm(1000,3.24,0.33);
4
5 T<-10000;B<-1000;u.mu<-runif(T);u.sig<-runif(T)
6
7 mu<-sig<-numeric(T);mu[1]<-3;sig[1]<-5
8
9 REJmu<-0;REJsig<-0
10
11 logpost=function(x,mu,sig){
12   loglike=sum(dnorm(x,mu,sig,log=TRUE))
13   return(loglike-log(sig))}
14
15
16
17 for(t in 2:T){mut<-mu[t-1];sigt<-sig[t-1]
18
19 mucand<-mut+runif(1,-0.5,0.5)
20
21 sigcand<-abs(sigt+runif(1,-0.5,0.5))
22
23 log.alpha.mu=logpost(x,mucand,sigt)-logpost(x,mut,sigt)
24
25 if(log(u.mu[t])<log.alpha.mu)mu[t]<-mucand
26
27 else{mu[t]<-mut;REJmu<-REJmu+1}
28
29 log.alpha.sig=logpost(x,mu[t],sigcand)-logpost(x,mu[t],sigt)
30
31 if(log(u.sig[t])<log.alpha.sig)sigt[t]<-sigcand
32
33 else{sigt[t]<-sigt;REJsig<-REJsig+1}}
34
35 REJratemu<-REJmu/T;REJratesig<-REJsig/T;
36
37 samp<-1001:10000;musamp<-mu[samp];sigsamp<-sig[samp]
38
39 hist(musamp,50,xlab="Shooting Percentages",main="Histogram of Shooting
40   Percentage")
41
42 mu.den<-density(musamp)
43
44 plot(mu.den)
45
46 plot(musamp,sigsamp)
47
48 00
```

### Histogram of Shooting Percentage



# Conclusion

Pour conclure, soulignons une fois de plus à quel point les chaînes de Markov sont puissantes pour modéliser des problèmes lorsqu'il s'agit de dynamique aléatoire. En raison de leurs bonnes propriétés, elles sont utilisées dans divers domaines. Les méthodes de Monte-Carlo par chaîne de Markov sont très répandues aujourd'hui, elles sont nécessaires pour la simulation. Dans l'avenir il sera intéressant de poursuivre l'étude de différentes méthodes de Monte-Carlo à chaîne de Markov.

# Bibliographie

- [1] Aude Rondepierre et Adeline Rouchon, *Notes du cours de Mathématiques de l'UF Outils mathématiques pour l'ingénieur*. INSA Département STPI Année 2016/2017.
- [2] Bernard Delyon, *Simulation et modélisation*. IRMAR, Université Rennes I, Campus de Beaulieu, 35042 Rennes cédex, France.
- [3] Marcel Déleze, *Probabilités conditionnelles-Le théorème de Bayes*, EDP sciences17.
- [4] Michel Benaïm & Nicole El Karoui, *Promenade aléatoire : Chaînes de Markov et simulations; martingales et stratégies*. École Polytechnique, 2005.
- [5] MICHEL BONNEFONT, *CHAÎNES DE MARKOV*. Master MIMSE Bordeaux.
- [6] Mohamed El Merouani, *Simulation par méthode de Monte-Carlo*. Université Abdelmalek Essaâdi Faculté des Sciences de Tétouan. 2018/2019.
- [7] Nicholas Burke, *Metropolis, Metropolis-Hastings and Gibbs Sampling Algorithms*. Lakehead University Thunder Bay, Ontario, 2018.
- [8] Pascal Bianchi, T. Bonald, L. Decreusefond, G. Fort, J. Najim, *Cours de Probabilités - MDI 104–114*, 2017–2018.
- [9] Paul Sabatier. *Simulation de variables et vecteurs aléatoires*. Université Toulouse III. M1 ESR 2019/2020.
- [10] Pierre L'Ecuyer, *IFT-3655, Modèles Stochastiques Chaînes de Markov en temps discret*. DIRO, Université de Montréal, 2012-2013.
- [11] THIERRY GALLAY, *THÉORIE DE LA MESURE ET DE L'INTÉGRATION.. UNIVERSITÉ JOSEPH FOURIER, GRENOBLE*, 2009.

- [12] TOUCHE Nassim, *Chaînes de Markov à temps discret*. Faculté des Sciences Exactes -Département de Recherche Opérationnelle, Université Abderrahmane Mira de Béjaia, 2017 – 2018.