

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



# *Mémoire de fin d'études*

pour l'obtention du diplôme

Master en Mathématique

Option :Analyse fonctionnelle et EDP

## Thème

---

Méthodes de Monte Carlo

---

Présenté par :  
Zidoun Youcef

Soutenu le : 2021

Devant le jury composé de :

**Encadreur :** Dr. Laala Zeyneb MCB, C- Univ. Salhi ahmed-Naama.  
**Président :** Mekki Slimane MAA, C- Univ. Salhi ahmed-Naama.  
**Examineur :** Kanouza Djamel MCB, C- Univ. Salhi ahmed-Naama.

Année universitaire 2020/2021

# Table des matières

<b>1</b>	<b>Modes de convergences et théorèmes limites</b>	<b>5</b>
1.1	Les modes de convergences . . . . .	5
1.1.1	Convergence presque sûre . . . . .	5
1.1.2	Convergence en probabilités . . . . .	6
1.1.3	Convergence en moyenne quadratique . . . . .	6
1.1.4	Convergence en loi . . . . .	7
1.2	Loi des grands nombres . . . . .	7
1.2.1	Loi faible de grand nombre . . . . .	7
1.2.2	Loi forte de grand nombre . . . . .	8
1.3	Théorème limite central . . . . .	8
<b>2</b>	<b>Simulation de Variables aléatoires</b>	<b>10</b>
2.1	Simulation de la loi uniforme . . . . .	10
2.1.1	Loi de probabilité discrète . . . . .	10
2.1.2	Simulation de la loi uniforme sur intervalle [a,b] . . . . .	11
2.2	Simulation de la lois non uniformes . . . . .	12
2.2.1	Simulation des variables aléatoires discrètes . . . . .	12
2.2.2	Loi de Bernoulli de paramètre $p$ . . . . .	12
2.2.3	Loi Binomiale de paramètres $(n, p)$ . . . . .	13
2.3	Simulation des variables aléatoires continue . . . . .	13
2.3.1	La Méthode d'inversion . . . . .	14
2.3.2	Loi exponentielle de paramètre $\lambda$ . . . . .	15
2.4	Simulation par la méthode de rejet . . . . .	15
2.4.1	Algorithme de rejet . . . . .	15
2.5	Simulation par la méthode de Box-Muller . . . . .	16
<b>3</b>	<b>Calcul d'intégrale par La méthode de Monte-Carlo</b>	<b>17</b>
3.1	Intégration Unidimensionnelle . . . . .	17
3.2	Intégration Multidimensionnelles . . . . .	18
<b>4</b>	<b>Méthode de Monte Carlo par chaîne de markov</b>	<b>20</b>
4.1	Chaîne de Markov . . . . .	20
4.1.1	Chaîne de Markov et processus stochastique . . . . .	20
4.2	Propriété de Markov . . . . .	22
4.2.1	Matrice de transition et loi initiales . . . . .	22
4.2.2	Récurrence . . . . .	22

4.2.3	Irréductibilité . . . . .	24
4.2.4	Mesure Invariante . . . . .	24
4.3	Simulation d'une Chaîne de Markov . . . . .	27
4.4	Loi stationnaire . . . . .	28
4.5	Théorème Ergodique . . . . .	30
4.6	Algorithme de Metropolis-Hastings . . . . .	31
4.7	Algorithme de Metropolis . . . . .	34
4.8	Algorithme de Gibbs . . . . .	36
<b>5</b>	<b>Conclusion</b>	<b>37</b>

# Remerciement

*Avant tout, nous remercions "Allah", le tout Puissant de nous avoir accordé la volonté et la patience pour accomplir ce travail.*

Je tiens à remercier infiniment Dr. **Laala zeyneb** de m'avoir encadrés et je tiens Dr. **Tahri Kamel** d'avoir pu bénéficier aussi bien de leurs conseils et ses compétences scientifiques, de nous avoir témoigné une grande confiance en nous laissant une large part d'initiative. Son soutien incessant aussi bien moral, ses conseils et ses encouragements furent déterminants.

J'adresse également mes remerciements à **Mr. Mekki** de l'honneur qu'elle nous a fait de présider le jury de notre master, et à **Dr. Kanouza** d'avoir accepté la lourde tâche d'être examinateurs de ce travail .

Mes sincères remerciements ont aussi adressé à tous les enseignants de Mathématiques au centre universitaire **salhi-ahmed** de Nâama et surtout **Dr. latti** , ses conseils, critiques et remarques vont significativement améliorer la qualité de ce mémoire.

Je souhaite qu'ils trouvent ici l'assurance de notre gratitude, et aussi que nous n'oublions pas **Dr D.Bouchiha** et **Dr. Benaissa**.

Enfin, je ne trouve pas de mots pour exprimer ma gratitude et ma reconnaissance à mes parents, mes frères, mes soeurs, mes amis et surtout le groupe *Naami.Y* et *Zaoui.M* et *Otmane.N* et *Boussag.M* et *Belhadje.K* et toutes les personnes qui ont contribué d'une manière directe ou indirecte à ce travail.

# Introduction

La modélisation aléatoire (stochastique) a le but de représenter un comportement par des modèles probabilistes. Nous allons concentrer sur une partie de la modélisation stochastique qui est les méthodes de Monte-Carlo.

Les méthodes de monte-carlo sont des techniques stochastiques repose essentiellement sur la loi des grands nombres, c'est-à-dire une répétition d'expérience aléatoire, permettant d'évaluer une quantité donnée mais inconnue. Des exemples de monte carlo incluent l'intégration stochastique, où nous utilisons une méthode basé sur la simulation pour évaluer une intégrale, Monte-carlo par chaine de markov, où nous construisons une chaine de Markov qui converge vers la distribution d'inérêt.

Les deux outils clés de la méthode de Monte-carlo sont d'une part la loi des grands nombres, où on utilise les modes de convergences et d'autre part le théorème central limite. Ce qui explique que ces deux thèmes constituent le premier chapitre que nous le faisons sous forme d'un rappel.

Ensuite, nous nous intéressons à la simulation des variables aléatoires qui joue un rôle essentiel car elle permet de justifier la méthode de Monte-Carlo.

Dans ce sujet, nous présentons des méthodes pour la simulation des variables aléatoires discrètes et continues. Pour ces derniers, nous proposons la méthode de la fonction de répartition, mais nous remarquons qu'on ne peut pas l'appliquer si la fonction inverse n'a pas d'expression analytique simple, et c'est le cas pour la loi normale. C'est pour ça qu'on propose d'autre méthodes telle que la méthode de rejet, Box-muller. Elles seront abordées dans le deuxième chapitre.

Dans le troisième chapitre, nous expliquons brièvement comment utiliser la méthode de Monte-Carlo pour le calcul d'intégrales.

Il existe des cas où on ne peut pas simuler la loi de probabilité par les méthodes précédentes, alors que la méthode de Monte-Carlo par Chaines de Markov est utilisé pour résoudre tels problèmes qui est le sujet principale du quatrième chapitre.

Dans ce chapitre, on a commencer par un rappel sur les chaînes de Markov, puis on a introduit l'algorithme de Metropolis-Hasting et les deux algorithme qui sont des cas particulier de ce dernier, se sont l'algorithme de Metropolis et l'algorithme de Gibbs. On a donné des exemples d'application en utilisant le logiciel R qui est un outils de programmation très utile.

# Chapitre 1

## Modes de convergences et théorèmes limites

Dans ce chapitre on s'intéresse à introduire les modes de convergences et les théorèmes essentiels pour la suite.

Nous considérons des suites de variables aléatoires notées  $X_n$  définies sur un espace de probabilités  $(\Omega, \mathbb{A}, \mathbb{P})$ .

### 1.1 Les modes de convergences

Nous considérons des suites  $(X_n)_{n \geq 1}$  de variable aléatoire définies sur même espace probabilisé  $(\Omega, \mathbb{A}, \mathbb{P})$ , et nous étudions le comportement asymptotique de telle suite lorsque  $n$  tend vers l'infini.

Plusieurs types de convergence se sont imposés (convergence en loi, convergence en probabilités, presque sûre, en moyenne d'ordre  $r > 0$ ). Nous passons en revue les plus importants de ce type.

#### 1.1.1 Convergence presque sûre

Il existe une notation très forte de convergence : la convergence presque sûre. Il s'agit du mode le plus puissant.

**Définition 1.1.1** Soit  $(X_n)_{n \geq 1}$  une suite de variable aléatoire définies sur même espace probabilisé  $(\Omega, \mathbb{A}, \mathbb{P})$ .

1. On dit que la suite  $(X_n)_{n \geq 1}$  converge vers 0 presque sûrement, et l'on écrit :

$$X_n \xrightarrow[n \rightarrow \infty]{p.s.} 0$$

s'il existe un ensemble  $P$  négligeable  $A \in \mathbb{A}$  tel que pour tout  $\omega \in \Omega/A$   $X_n(\omega) \rightarrow 0$  lorsque  $n \rightarrow \infty$ .

2. On dit que la suite  $(X_n)_{n \geq 1}$  converge vers  $X$  presque sûrement et l'on écrit :

$$X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$$

si la suite  $(X_n - X)_{n \geq 1}$  converge vers 0 presque sûrement.

**Proposition 1.1.1** *Il y a une manière équivalente de définir la convergence presque sûre :*

$$\mathbb{P}(\lim_{n \rightarrow \infty} X_n = X) = 1 \Leftrightarrow X_n \xrightarrow[n \rightarrow \infty]{p.s.} X.$$

### 1.1.2 Convergence en probabilités

**Définition 1.1.2** *Une suite de variable aléatoire  $(X_n)_{n \in \mathbb{N}}$  converge en probabilité vers un nombre réel  $X$  si :*

*Pour tout  $\epsilon > 0$  on a :*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\omega / |X_n(\omega) - X| > \epsilon) = 0.$$

*Et en note :*

$$\mathbb{P} - \lim_{n \rightarrow \infty} X_n = X \iff X_n \xrightarrow[n \rightarrow \infty]{p} X.$$

**Proposition 1.1.2** *La convergence presque sûre implique la convergence en probabilité :*

$$X_n \xrightarrow[n \rightarrow \infty]{p.s.} X \Rightarrow X_n \xrightarrow[n \rightarrow \infty]{p} X.$$

### 1.1.3 Convergence en moyenne quadratique

**Définition 1.1.3** *Une suite de variables aléatoires réelles  $(X_n)_{n \in \mathbb{N}}$  converge en moyenne quadratique vers une variable aléatoire réelle  $X$  si :*

$$\lim_{n \rightarrow \infty} (\mathbb{E}|X_n - X|^2) = 0.$$

**Proposition 1.1.3** *Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires réelles d'espérances et de variances finies si :*

1.  $\mathbb{E}(X_n) \rightarrow \mu < \infty$ .
2.  $\text{Var}(X_n) \rightarrow 0$ .

*Alors :*

$$X_n \xrightarrow{m.q.} \mu.$$

### 1.1.4 Convergence en loi

Soient  $(F_n)$  la suite de fonctions de répartition associées aux variables aléatoires réelles  $(X_n)$  et  $F$  la fonction de répartition de variables aléatoires réelles  $X$ .

**Définition 1.1.4** *De manière équivalente, la suite  $(X_n)$  converge en loi vers  $X$  si et seulement si pour toute fonction continue bornée :*

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)].$$

**Théorème 1.1.1 (Lévy)** *Soit  $\varphi_n(t)$  la fonction caractéristique de  $X_n$  est  $\varphi(t)$  celle de  $X$ . Alors :*

$$\forall t \in \mathbb{R} : (\varphi_n(t) \rightarrow \varphi(t)) \Leftrightarrow (X_n \xrightarrow{\ell} X).$$

## 1.2 Loi des grands nombres

### Définition 1.2.1

- La suite  $(X_n)_{n \geq 1}$  satisfait la loi faible des grands nombres si la suite de terme générale  $M_n = \frac{1}{n} \sum_{k=1}^n X_k$  converge vers  $\mathbb{E}(X)$  en probabilité.
- La suite  $(X_n)_{n \geq 1}$  satisfait la loi forte des grands nombres si la suite de terme général  $M_n = \frac{1}{n} \sum_{k=1}^n X_k$  converge vers  $\mathbb{E}(X)$  presque sûrement.

### 1.2.1 Loi faible de grand nombre

Dans cette section  $(X_n)_{n \geq 1}$  désigne une suite de variables aléatoires réelles indépendantes, on s'intéresse au comportement asymptotique de la suite  $(S_n)_{n \geq 1}$  avec :

$$S_n = X_1 + X_2 + \dots + X_n.$$

Lorsque les variables aléatoires  $(X_n)_{n \geq 1}$  ont la même loi on notera :

$\mathbb{E}(X)$  : l'espérance commune des  $X_n$ .

$\text{var}(X) = \sigma(X)^2$  : leur variance commune (lorsque bien sûr ces quantités existent).

**Théorème 1.2.1** *On suppose que la suite  $(X_n)_{n \geq 1}$  est formée de variables aléatoires indépendantes, équidistribuées et possédant un moment d'ordre 2*

1- La suite  $(\frac{S_n}{n})_{n \geq 1}$  converge en probabilité vers  $\mathbb{E}(X)$ .

2- Pour tout  $\varepsilon \geq 0$  et  $n \geq 1$ , on a :

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}(X)\right| > \varepsilon\right) \leq \frac{\sigma(X)^2}{\varepsilon^2} \frac{1}{n}$$

**Preuve :** Posons :  $M_n = \frac{S_n}{n}$ , puisque  $\mathbb{E}(X_i) = \mathbb{E}(x)$  alors  $\mathbb{E}(M_n) = \mathbb{E}(X)$ .

D'après l'inégalité de *Bienaymie Tchebycheff* :

$$\mathbb{P}(|M_n - \mathbb{E}(M_n)| > \varepsilon) = \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}(X)\right| > \varepsilon\right) \leq \frac{\text{var}(M_n)}{\varepsilon^2}$$

Les variables aléatoires  $X_1, X_2, \dots, X_n$  étant indépendantes et de même loi

$$\text{var}(M_n) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma(X)^2 = \frac{\sigma(X)^2}{n}.$$

### 1.2.2 Loi forte de grand nombre

La loi forte des grands nombres joue un rôle clé dans la simulation stochastique, c'est ce résultat qui permet en particulier de justifier la méthode de Monte Carlo.

**Théorème 1.2.2** *Soit  $(X_n)_n \geq 1$  une suite de variables aléatoires réelles indépendantes et équidistribuées.*

- 1- *La suite  $(\frac{S_n}{n})_n \geq 1$  converge presque sûrement si et seulement si les variables aléatoires réelles  $(X_n)_n \geq 1$  sont intégrables.*
- 2- *Si les variables aléatoires réelles sont intégrables, alors  $(S_n)_n \geq 1$  converge presque sûre et dans  $L^1(\Omega)$  vers  $\mathbb{E}(X)$ .*

### 1.3 Théorème limite central

La loi des grands nombre montre que la moyenne  $M_n = \frac{1}{n} \sum_{k=1}^n X_K$  d'une suite des variables aléatoires réelles  $X_K$  i.i.d. (ayant espérance et variance) tend vers un nombre : l'espérance commune  $\mu = \mathbb{E}(X_K)$  de ce variable aléatoire : bien entendu,  $(M_n)$  n'est que de variance petite et reste donc aléatoire : si l'on veut observer la variable de  $M_n$  (qui est proche de  $\mu$ ), il est nécessaire d'agrandir l'écart  $M_n - \mu$  en la multipliant par une grandeur  $\lambda(n)$  suffisamment grand avec  $n$ .

On a :

$$Z_n = \lambda(n)(M_n - \mu).$$

$$\begin{aligned} \text{Var}(\lambda(n)(M_n - \mu)) &= \text{Var}(Z_n) = 1 \\ &= \lambda^2(n) \text{Var}(M_n - \mu) \\ &= \lambda^2(n) \text{Var}\left(\frac{1}{n} \sum_{K=1}^n X_K\right) \\ &= \frac{\lambda^2(n)}{n^2} \sum_{K=1}^n \text{Var}(X_K) \quad \text{Par indépendance des } X_K \\ &= \frac{\lambda^2(n)}{n} \sigma^2. \end{aligned}$$

Il suffit donc de choisir  $\lambda(n) = \frac{\sqrt{n}}{\sigma}$  pour que  $Z_n$  soit de variance 1. Nous avons :

$$\begin{aligned} Z_n &= \lambda(n)(M_n - \mu) \\ &= \frac{\sqrt{n}}{\sigma} \frac{1}{n} \sum_{K=1}^n (X_K - \mu) \\ &= \frac{1}{\sigma\sqrt{n}} \left( \sum_{K=1}^n X_K - n\mu \right) \end{aligned}$$

qui est visiblement d'espérance nulle et la se produit une petite merveille : pour  $n$  grand, cette "remise à l'échelle" de  $M_n - \mu$  produit une variable aléatoire  $Z_n$  dont la loi est proche de la loi normale (centré, réduite, puisque tel est le cas pour la loi de tous les  $Z_n$ ) et ceci sans autre hypothèse : ceci explique pourquoi cette loi surgit si souvent en statistique cette merveille porte le nom de théorème limite central.

**Théorème 1.3.1** *Soit*

$$M_n = S_n/n = (X_1 + \cdots + X_n)/n,$$

*et*

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{M_n - \mu}{\sigma/\sqrt{n}},$$

*Alors la loi de  $Z_n$  converge vers la loi normale centrée réduite  $\mathcal{N}(0,1)$  lorsque  $n$  tend vers l'infini. Cela signifie que si  $F$  est la fonction de répartition de  $\mathcal{N}(0,1)$ , alors pour tout réel  $z$  :*

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n \leq z) = \Phi(z),$$

*ou, de façon équivalente :*

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{M_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = \Phi(z).$$

*On note*

$$Z_n \xrightarrow{L} \mathcal{N}(0,1)$$

**Démonstration :**

Pour une variable aléatoire  $Y$  d'espérance 0 et de variance 1, la fonction caractéristique de  $Y$  admet le développement limité :

$$\varphi_Y(t) = 1 - \frac{t^2}{2} + o(t^2), \quad t \rightarrow 0.$$

Si  $Y_i$  vaut  $\frac{X_i - \mu}{\sigma}$ , il est facile de voir que la moyenne centrée réduite des observations  $X_1, X_2, \dots, X_n$  est simplement :

$$Z_n = \frac{M_n - \mu}{\sigma/\sqrt{n}} = \sum_{i=1}^n \frac{Y_i}{\sqrt{n}}.$$

D'après les propriétés élémentaires des fonctions caractéristiques, la fonction caractéristique de  $Z_n$  est

$$\varphi_{Z_n}(t) = \left[ \varphi_Y\left(\frac{t}{\sqrt{n}}\right) \right]^n = \left[ 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) \right]^n \xrightarrow[n \rightarrow \infty]{} e^{-t^2/2}.$$

Mais cette limite est la fonction caractéristique de la loi normale centrée réduite  $\mathcal{N}(0,1)$ , d'où l'on déduit le théorème de la limite centrale grâce au théorème de continuité de Lévy, qui affirme que la convergence simple des fonctions caractéristiques implique la convergence en loi.

**Remarque 1.3.1** *Pour tout  $a, b \in \mathbb{R}$  avec  $a < b$  on a :*

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(a \leq \sqrt{n} \frac{M_n - \mu}{\sigma} \leq b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

# Chapitre 2

## Simulation de Variables aléatoires

Nous présentons ici quelques méthodes classique de simulation d'une variable aléatoire pour une approche plus approfondie.

Supposons que l'on dispose d'un générateur  $U$  de loi uniforme  $\mathcal{U}([0, 1])$  sur  $[0, 1]$  et que placé dans une boucle, ce générateur retourne des réalisation  $(U_n)$  i.i.d.  $U([0, 1])$ .

**Principe :** pour simuler des variables aléatoires de loi quelconque, l'idée est de ramener à la loi uniforme sur  $[0, 1]$ . La loi uniforme sur  $[0, 1]$  est donc à la base de toute simulation de variable aléatoire. Soit  $U$  une variable aléatoire de loi uniforme sur  $[0, 1]$ ,  $U \sim \mathcal{U}[0, 1]$ . Dans ce cas

$$F_U(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

$$f_u(x) = \mathbf{1}_{\{x \in [0,1]\}}$$

**Cadre général :** On suppose qu'on dispose d'un générateur de variables aléatoires de la loi uniforme sur  $[0, 1]$  indépendantes.

L'hypothèse d'indépendances des valeurs est parmi les conditions essentielles pour la validité de la plupart des algorithmes présentés dans la suite.

**Remarque 2.0.2** Une variable aléatoire de la loi uniforme sur  $[0, 1]$  est différente de 0 et 1 presque sûrement. En particulier si  $\Lambda(dx)$  note la mesure de lebesgue, on a :

$$\mathbb{P}(U \in \{0, 1\}) = \int_{\{0,1\}} \mathbf{1} d\Lambda = \Lambda(\{0, 1\}) = 0.$$

## 2.1 Simulation de la loi uniforme

### 2.1.1 Loi de probabilité discrète

Soit  $X$  une variable aléatoire à valeurs dans  $E$  avec  $E = \{x_i | i \in \mathbf{I}\}$  et  $\mathbf{I} = \mathbb{N}$  où  $\mathbf{I} = \{0, 1, \dots, n\}$  considérons le cas  $\mathbf{I}$  fini, le second se traite de la même manière. La loi de probabilité de  $X$  est donnée par :

$$\mu(x) = \sum_{i=0}^n p_i \delta_{x_i}(x)$$

où

$$p_i = \mathbb{P}(X = x_i) = \mu(x_i); \quad \forall i \in \mathbf{I} \quad \text{et} \quad \sum p_i = 1.$$

Notons par  $p_i$  la cumule des  $p_i$  pour  $0 \leq i \leq K$  i.e

$$p_K = \sum_{i=0}^K p_i.$$

On a alors  $P_0 = p_0$  et  $P_n = 1$ . Nous souhaitons simuler une variable aléatoire de même loi que  $X$ . Pour ce faire, on va construire une variable aléatoire  $Y$  telle que :

$$\mathbb{P}(X = x_i) = \mathbb{P}(Y = x_i) = p_i. \quad \forall i \in \mathbf{I}.$$

À l'aide d'une variable aléatoire  $U$  de loi uniforme sur  $[0, 1]$ .

**Algorithme :**

L'algorithme s'écrit de la manière suivante :

On tire  $U$  une variable de loi uniforme sur  $[0, 1]$  et on pose :

$$Y = K \quad \text{si} \quad P_{K-1} < U \leq P_K.$$

Ce qui s'exprime également sous la forme

$$Y = x_0 \mathbf{1}_{\{U \leq P_0\}} + \sum_{i=1}^n x_i \mathbf{1}_{\{P_{i-1} < U < P_i\}}.$$

**Corollaire 2.1.1** *Si  $U$  est une variable aléatoire de loi uniforme sur  $[0, 1]$  et si  $n \in \mathbb{N}^*$  alors la variable aléatoire  $[nU] + 1$  suit la loi uniforme discrète sur  $\{1, 2, \dots, n\}$ . La notation  $[X]$  désigne la partie entière du réel  $X$ .*

**Démonstration :**

On a pour  $K \in \{1, 2, \dots, n\}$

$$\begin{aligned} p([nU] + 1 = K) &= p([nU] = K - 1) \\ &= p(K - 1 \leq nU \leq K) \\ &= p\left(\frac{K-1}{n} \leq U \leq \frac{K}{n}\right) \\ &= \frac{1}{n} \end{aligned}$$

## 2.1.2 Simulation de la loi uniforme sur intervalle $[a, b]$

On souhaite simuler une variable aléatoire  $X$  de loi uniforme sur l'intervalle  $[a, b]$  pour  $a < b$ , avec :

$$f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x).$$

**Lemme 2.1.1** *Si  $U$  suit la loi uniforme sur  $[0, 1]$ , alors si  $a < b$  et  $a, b \in \mathbb{R}$ , la variable aléatoire  $X = a + (b-a)U$  suit la loi uniforme sur  $[a, b]$ .*

**Démonstration :**

Soit  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  une fonction continue bornée : en faisant le changement de variable  $X = a + (b - a)U$  on a :

$$\mathbb{E}(\varphi(a + (b - a)U)) = \int_0^1 \varphi(a + (b - a)u) du = \int_a^b \varphi(x) \frac{1}{b - a} dx.$$

Ce que signifie que  $X$  suit une variable aléatoire de densité :

$$f(x) = \frac{1}{b - a} \mathbf{1}_{[a,b]}(x).$$

Et reconnaître ainsi la densité de la loi uniforme sur  $[a, b]$ .

## 2.2 Simulation de la lois non uniformes

### 2.2.1 Simulation des variables aléatoires discrètes

Soit  $M$  une probabilité sur  $\mathbb{R}$  de la forme :

$$M = \sum_{i \in \mathcal{I}} p_i \delta_{x_i}.$$

1.  $(x_i)_{i \in \mathcal{I}}$  est une suite de réels deux à deux distincts ;
2.  $\delta_{x_i}$  est la mesure de Dirac en  $x_i$
3.  $(p_i)_{i \in \mathcal{I}}$  une famille de réels positifs ou nuls ( $\sum_{i \in \mathcal{I}} p_i = 1$ ) on prend

$$\mathbf{I} = \{0, 1, \dots, n\}.$$

Soit  $U$  une variable aléatoire réelle de loi uniforme sur  $[0, 1]$  et

$$X = x_0 \mathbf{1}_{U \leq p_0} + \sum_{i \geq 1} x_i \mathbf{1}_{\{p_0 + \dots + p_{i-1} \leq U \leq p_0 + \dots + p_i\}}$$

a pour loi de support  $X(\Omega) = \{x_i, i \in \mathbf{I}\}$  et pour tout  $i \in \mathbf{I}$ ,  $\mathbb{P}(X = x_i) = p_i$ .

**Preuve :**

On a :  $\mathbb{P}(X = x_0) = \mathbb{P}(U < p_0)$  et pour tout  $i \geq 1$

$$\mathbb{P}(X = x_i) = \mathbb{P}(p_0 + p_1 + \dots + p_{i-1} \leq U \leq p_0 + \dots + p_i) = p_i.$$

### 2.2.2 Loi de Bernoulli de paramètre $p$

Soit  $p \in [0, 1]$ , on veut simuler une variable aléatoire  $X \sim \mathcal{B}(p)$  de loi de probabilité

$$\mu(x) = p\delta_1(x) + (1 - p)\delta_0(x).$$

Pour cela on tire  $U$  une uniforme sur  $[0, 1]$  et on définit :

$$X = \begin{cases} 1 & \text{si } U \leq p \\ 0 & \text{sinon} \end{cases}$$

c'est-à-dire  $X = \mathbf{1}_{\{U \leq p\}}$ .

On propose deux fonctions pour simuler cette variable aléatoire, la première ne rend qu'une réalisation d'une variable aléatoire de la loi de Bernoulli de paramètre  $p$ , la seconde renvoie un  $K$ -échantillon.

**Exemple 2.2.1** (*Pile ou Face*) On veut simuler une variable aléatoire de loi  $X \sim \mathcal{B}(\frac{1}{2})$ ; on tire  $U$  une uniforme sur  $[0, 1]$  et

$$X = \begin{cases} \text{Pile} & \text{si } U \leq \frac{1}{2} \\ \text{Face} & \text{sinon} \end{cases}$$

formellement on peut écrire  $X = \mathbf{1}_{\{U \leq \frac{1}{2}\}}$ .

### 2.2.3 Loi Binomiale de paramètres $(n, p)$

Soient  $n \in \mathbb{N}$  et  $p \in [0, 1]$ , on veut simuler  $X \sim \mathcal{B}(n, p)$  donc de loi

$$\mu(x) = \sum_{K=0}^n \binom{n}{K} p^K (1-p)^{n-K} \delta_X(x).$$

On rappelle qu'une variable aléatoire binomiale peut être représentée comme la somme de  $n$  variables aléatoires indépendantes de bernoulli de paramètre  $p$

**Lemme 2.2.1** Soient  $(Y_i)_{1 \leq i \leq n}$  des variables aléatoires i.i.d. de la loi bernoulli de paramètre  $p$ . Alors

$$X = \sum_{i=1}^n Y_i$$

suit une loi binomiale de paramètre  $(n, p)$ . Il suffit donc de simuler  $n$  variables aléatoires indépendantes de loi  $\mathcal{B}(p)$  est d'en faire la somme

$$X = \sum_{i=1}^n \mathbf{1}_{\{U_i \leq p\}}.$$

**Remarque 2.2.1** Cette méthode s'appelle méthode de comparaison permet de simuler des variables aléatoires discrètes à valeur dans  $\mathbb{N}$ , elle consiste à découper l'intervalle  $[0, 1]$  en deux parties  $[0, p]$  et  $[p, 1]$ .

## 2.3 Simulation des variables aléatoires continue

**Définition 2.3.1** On appelle la loi uniforme sur  $[0, 1]$ , tout procédé permettant de produire une suite de variable aléatoire  $(U_i)_{i \geq 0}$  i.i.d. de loi uniforme sur  $[0, 1]$ . On dira qu'une loi de probabilité continue est simulable si on peut décrire un procédé explicite permettant de produire des échantillons observés de cette loi à partir d'un générateur de la loi uniforme sur  $[0, 1]$ .

**Proposition 2.3.1** Soit  $F$  une fonction de répartition continue et strictement croissante sur  $\mathbb{R}$  et soit  $F^{-1}$  sa fonction inverse définie sur  $]0, 1[$ . Alors

- Si  $X$  une variable aléatoire réelle de fonction de répartition  $F$ , la variable aléatoire  $Y = F(X)$  est de loi uniforme sur  $[0, 1]$
- Réciproquement, si  $U$  est une variable aléatoire uniforme sur  $[0, 1]$  la variable aléatoire  $Z = F^{-1}(U)$  admet une fonction de répartition égale à  $F$ .

**Démonstration :**

Soit  $t \in \mathbb{R}$ .

1. Si  $t \leq 0$  ( resp si  $t \geq 1$  ), on a  $\mathbb{P}(Y \leq t) = 0$  (resp  $\mathbb{P}(Y \leq t) = 1$ ) est si  $t \in ]0, 1[$  on a :

$$\mathbb{P}(Y \leq t) = \mathbb{P}(F(X) \leq t) = \mathbb{P}(X \leq F^{-1}(U) \leq F(t)) = F(F^{-1})(t) = t.$$

D'où le resultat de 1.

2.  $\mathbb{P}(Z < t) = \mathbb{P}(F^{-1}(U) \leq t) = \mathbb{P}(F(F^{-1}(U)) \leq F(t)) = \mathbb{P}(U \leq F(t)) = F(t).$

### 2.3.1 La Méthode d'inversion

Soit  $X$  une variable aléatoire de fonction de répartition  $F : \mathbb{R} \rightarrow [0, 1]$  définie, si  $x \in \mathbb{R}$  par  $F(x) = \mathbb{P}(X \leq x)$  :  $F$  est croissante (continue à droite, limitée à gauche) de limite 0 en  $-\infty$ , 1 en  $+\infty$  de plus  $F$  continue partout dès que  $X$  admet une densité  $g$ ; dans ce cas :

$$F(x) = \int_{-\infty}^x g(u) du.$$

Notons  $F^{-1} : [0, 1] \rightarrow \mathbb{R}$  la pseudo-inverse de  $F$  définie par :

$$F^{-1} = \inf \{x : F(x) \geq u\} \quad \text{si } u \in [0, 1].$$

**Proposition 2.3.2** *Si  $U$  uniforme sur  $[0, 1]$ , alors  $X = F^{-1}(U)$  a pour fonction de répartition  $F$ . En effet :*

$$\mathbf{P}(X \leq u) = \mathbf{P}(F^{-1}(U) \leq x) = \mathbf{P}(U \leq F(x)) = F(x).$$

Chaque fois que  $F$  est explicite ( $F^{-1}(U_n), n \geq 1$ ) est une suite i.i.d de loi  $X$ .

**Lemme 2.3.1** *Pour tout  $u \in [0, 1], x \in \mathbb{R}$ .*

$$F^{-1}(U) \leq x \Leftrightarrow u \leq F(x).$$

**Démonstration :**

Pour l'implication directe la croissance de  $F$  donne :

$$F(F^{-1}(u)) \leq F(x).$$

Puisque la fonction de répartition est continue à droite

$$F(F^{-1}(u)) \geq u.$$

Nous avons donc  $F(x) \geq u$ , quand à l'implication réciproque elle est triviale.

**Exemple 2.3.1 (Loi exponentielle)** *Soit  $X$  une variable aléatoire réelle qui suit une loi exponentielle de paramètre  $\lambda$ , la fonction réciproque vaut pour  $u \in [0, 1]$  :*

$$F(x) = 1 - e^{-\lambda x} \Leftrightarrow F^{-1}(u) = \frac{-\ln(1-u)}{\lambda}.$$

Et si  $U$  une variable aléatoire de loi uniforme sur  $[0, 1]$ ,  $X$  est de même loi que :

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

ou encore

$$X = \frac{-\ln u}{\lambda}.$$

### 2.3.2 Loi exponentielle de paramètre $\lambda$

Soit  $\lambda > 0$ , on souhaite simuler une variable aléatoire  $X$  de loi exponentielle de paramètre  $\lambda$ , avec :

$$f(x) = \lambda \exp(-\lambda x) \mathbf{1}_{\mathbb{R}_+}(x).$$

Sa fonction de répartition est

$$F(x) = \mathbb{P}(X < x) = 1 - \exp(-\lambda x).$$

Cette fonction est une bijection de  $]0; +\infty[$  dans  $]0, 1[$  d'inverse

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

Soit  $U$  variable aléatoire de loi uniforme sur  $[0, 1]$ . La variable aléatoire  $X$  a donc le même loi que  $\frac{-\ln(1-U)}{\lambda}$  et aussi par symétrie que :

$$\frac{-\ln U}{\lambda}.$$

## 2.4 Simulation par la méthode de rejet

**Définition 2.4.1** Soient  $(\Omega, \mathcal{F}, \mathbf{P})$  un espace probabilisé,  $B$  un borélien de  $\mathbb{R}^d$  tel que  $0 < \lambda_d(B) < +\infty$ , désignant la mesure de Lebesgue sur  $\mathbb{R}^d$ . Le vecteur aléatoire  $M : \Omega \rightarrow \mathbb{R}^d$  suit la loi uniforme sur  $B$  si la loi  $\mathbf{P}_M$  a une densité de la forme  $C \mathbf{1}_B$  par rapport à  $\lambda_B$ ,  $C$  étant une constante (notation  $M \sim U(B)$ ).

Il est clair que la seule valeur possible pour  $C$  est  $1/\lambda_d(B)$ . La caractérisation suivante est bien immédiate et bien plus utile en pratique que la définition

$$M \sim U(B) \Leftrightarrow \forall A \in \mathcal{B}(\mathbb{R}^d).$$

$$\mathbf{P}(M \in A) = \frac{\lambda_d(A \cap B)}{\lambda_d(B)}.$$

### 2.4.1 Algorithme de rejet

On veut simuler une variable aléatoire de loi de densité  $f$  (sur  $\mathbb{R}^d$ ), et on suppose qu'il existe une loi de densité  $g$  simulable facilement telle que :

$$\forall x \in \mathbb{R}^d, f(x) \leq K g(x),$$

où  $K$  est une constante réelle.

Soit  $U$  une variable aléatoire réelle de loi uniforme sur  $[0, 1]$  et  $Z$  une variable aléatoire indépendante de  $U$  de densité  $g$ . On pose  $V = KUg(Z)$ . Alors la loi de  $Z$  conditionnellement à l'évènement  $\{V < f(Z)\}$  a pour densité  $f$ .

**Démonstration :**

On fait la démonstration pour le cas  $\mathbb{R}$ . Notons que pour tout  $Z \in \mathbb{R}$ ,  $\frac{f(Z)}{Kg(Z)} \leq 1$ . On a tout d'abord :

$$\begin{aligned} \mathbb{P}(V < f(Z)) &= \mathbb{P}(kUg(Z) < f(Z)) \\ &= \int_{\mathbb{R}} g(Z) dZ \int_0^1 \mathbf{1}_{\{kUg(Z) < f(Z)\}} dU \\ &= \int_{\mathbb{R}} g(Z) \frac{f(Z)}{kg(Z)} dZ \\ &= \frac{1}{k} \end{aligned}$$

$$\mathbb{P}(\{Z \leq t\} \cap \{V < f(Z)\}) = \int_{-\infty}^t g(Z) dZ \int_0^1 \mathbf{I}_{\{kUg(Z) < f(Z)\}} dU.$$

Donc la densité de  $X$  est bien  $f$ .

$$\int_{-\infty}^t g(Z) \frac{f(Z)}{kg(Z)} = \frac{1}{k} \int_{-\infty}^t f(Z) dZ.$$

Donc :

$$P(Z \leq t/v < f(Z)) = \int_{-\infty}^t f(Z) dZ.$$

Alors la loi conditionnelle de  $Z$  sachant que  $\{V < f(Z)\}$  a bien pour densité  $f$ .

## 2.5 Simulation par la méthode de Box-Muller

Les composantes d'un vecteur aléatoire  $(X, Y)$  de  $\mathbb{R}^2$  sont indépendantes et de loi  $\mathcal{N}(0, 1)$  si  $X = r \cos(\theta)$  et  $Y = r \sin(\theta)$  où  $r$  et  $\theta$  sont deux variable aléatoire indépendantes avec  $r^2$  de loi exponentielle  $\varepsilon(1/2)$  et  $\theta$  de loi uniforme sur  $[0, 2\pi]$ . Ainsi si  $U$  et  $V$  sont indépendantes et de loi uniform sur  $[0, 1]$ , alors les variables aléatoires

$$X = \sqrt{-2 \ln U} \cos(2\pi V) \quad \text{et} \quad Y = \sqrt{-2 \ln U} \sin(2\pi V)$$

sont indépendantes et de loi gaussienne  $\mathcal{N}(0, 1)$ .

Si de plus  $m \in \mathbb{R}$  et  $\sigma \geq 0$ , alors  $m + \sigma X$  et  $m + \sigma Y$  sont deux variables aléatoires indépendantes de loi  $\mathcal{N}(m, \sigma^2)$ .

# Chapitre 3

## Calcul d'intégrale par La méthode de Monte-Carlo

Les méthodes de Monte-Carlo sont particulièrement utilisées pour calculer des intégrales en dimensions plus grandes de 1 (en particulier, pour calculer des surfaces, des volumes, ect...)

Dans ce chapitre on s'intéresse à la description de la méthode de Monte-Carlo.

### 3.1 Intégration Unidimensionnelle

Les méthodes de Monte-Carlo reposent sur une approximation probabiliste est non déterministe. En ce sens, on ne résout pas l'objet mathématique mais on cherche à l'approcher moyennant la loi forte des grands nombres, cet objet peut être un intégrale comme c'est le cas ici :

$$I(h) = \int_a^b h(x)dx.$$

Nous exprimons tout d'abord la loi fonction  $h(x)$  sous la forme :

$$h(x) = g(x)f(x)$$

où  $f(x)$  est une densité sur  $[a, b]$ . Ainsi par définition de l'espérance :

$$\begin{aligned} I(h) &= \int_a^b g(x)f(x)dx \\ &= \mathbf{E}[g(X)] \end{aligned}$$

où  $X$  est la variable aléatoire avec fonction de densité de probabilité  $f(x)$ , par la suite  $x_1, \dots, x_n$  : sont des observation simulées de la densité  $f(x)$ . Alors

$$I_n(h) \approx \frac{1}{n} \sum_{i=1}^n g(x_i).$$

Par souci de simplicité et parce qu'il est facile d'obtenir un échantillon aléatoire d'une loi uniforme, nous posons en général  $X \sim \mathcal{U}[a, b]$ , sont :

$$f(x) = \frac{1}{b-a} \quad , \quad a < x < b.$$

Dans ce cas ,  $g(x) = (b - a)h(x)$ , donc l'intégral se réécrit :

$$\begin{aligned} I(h) &= (b - a) \int_a^b h(x) \frac{1}{b-a} dx. \\ &= (b - a) \mathbf{E}[h(x)]. \end{aligned}$$

l'estimation de l'intégrale devient alors

$$I_n(h) = \frac{b - a}{n} \sum_{i=1}^n h(x_i).$$

où  $x_1, \dots, x_n$  est un échantillon aléatoire d'une loi  $\mathcal{U}[a, b]$ , par changement de variable posons

$$u = \frac{x - a}{b - a} \quad , \quad du = \frac{dx}{b - a},$$

équivalent à :

$$x = a + (b - a)u \quad , \quad dx = (b - a)du.$$

On remplace dans l'intégrale. On obtient alors :

$$\begin{aligned} I(h) &= (b - a) \int_0^1 h(a + (b - a)u) du. \\ &= (b - a) \mathbf{E}[h(a + (b - a)u)] \end{aligned}$$

où  $U \sim \mathcal{U}[0, 1]$ . Une estimation de l'intégrale est donc

$$I_n(f) = \frac{b - a}{n} \sum_{i=1}^n h(a + (b - a)u_i)$$

où  $u_1, \dots, u_n$  est un échantillon aléatoire d'une loi  $[0, 1]$ .

## 3.2 Intégration Multidimensionnelles

Le calcul de l'intégrale multidimensionnelles  $I(f) = \int_{\Omega} f(x) dx$ , avec  $\Omega \in \mathbb{R}^d$ ,  $d \in \mathbb{N}$ , par la méthode de Monte-Carlo. Se résoudre l'intégrale suivante :

$$I(f) = \int_{[0,1]^d} f(x) dx.$$

Dans ce cas, la méthode de Monte-Carlo consiste à écrire cette intégrale sous forme de la moyenne de  $f$  sur  $[0, 1]^d$ , et avec  $u$  une variable aléatoire suivant la loi uniforme sur  $[0, 1]^d$

$$I(f) = \mathbf{E}(f(u)).$$

Donc on doit

1. Mettre  $I(f)$  sous forme d'espérance : on pose  $X = f(U_1, \dots, U_d)$  ou  $U_1, \dots, U_d$  sont des réalisations de la loi uniforme sur l'intervalle  $[0, 1]$ . Alors :

$$\mathbf{E}(X) = \mathbf{E}(f(U_1, \dots, U_d)) = I(f).$$

2. Simulation de la variable aléatoire : on suppose que l'on dispose d'une suite  $(U_i)_{i \geq 1}$  de réalisation de la loi uniforme sur  $[0, 1]$ , on pose alors

$$X_1 = f(U_1, \dots, U_d) \quad , \quad X_2 = (U_{d+1}, \dots, U_{2d}), \dots .$$

Alors les  $(X_i)$  sont des réalisation de variable aléatoire  $X$  et

$$I(f) \simeq \frac{1}{n}(X_1 + \dots + X_n).$$

On approxime l'intégrale comme suit :

$$I(f) \simeq I_n(f) = \frac{1}{n} \sum_{i=1}^n f(x_i) = \frac{1}{n} \sum_{i=1}^n X_i.$$

les points  $x_i$  sont choisis dans l'intervalle  $\Omega$  donc quand le nombre des points  $n$  augmente l'approximation sera plus précise et on a :

$$I(f) = \int_{\Omega} f(x) dx = \lim_{n \rightarrow +\infty} I_n(f) = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n f(x_i).$$

# Méthode de Monte Carlo par chaîne de markov

Les méthodes de Monte-Carlo par chaîne de Markov, sont une classe de méthode d'échantillonnage à partir d'une distribution de probabilité qui ont une grande importance en mathématique. Pour cela on commence dans ce chapitre par rappeler les chaînes de markov, ainsi les propriétés importantes. Par la suite on verra les algorithmes.

## 4.1 Chaîne de Markov

### 4.1.1 Chaîne de Markov et processus stochastique

**Définition 4.1.1** Une Chaîne de Markov à temps discret est un processus stochastique à temps discret  $(X_0, X_1, \dots, X_n, \dots)$  qui vérifie, pour tout  $n$  et tous états  $(i_0, i_1, \dots, i_{n-1}, i, j \in E)$  la probabilité conditionnelle

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j | X_n = i)$$

ne dépend que de  $i$ , de  $j$  et de  $n$ .

On dit que la Chaîne de Markov est homogène si cette quantité ne dépend que de  $i, j$  et pas de  $n$ . En d'autres termes pour une chaîne de Markov homogène, l'état du système dans le futur (au temps  $n + 1$ ) ne dépend que de son état dans le présent (temps  $n$ ) et pas des états antérieurs dans le passé ( temps  $0 \leq K \leq n - 1$  ), on note :  $p_{ij} = P(X_{n+1} = j | X_n = i)$  et on appelle la probabilité de transition de  $i$  à  $j$ . La loi de  $X_0$  s'appelle la loi initial de la chaîne.

**Théorème 4.1.1** – Notons  $p_j^{(n)} = P(X_n = j)$  la loi de  $X_n$  est  $p^{(n)}$  le vecteur ligne associé. Si on note  $P$  la matrice  $r \times r$  dont les coefficients sont les  $p_{ij} = P(X_{n+1} = j | X_n = i)$ . On a pour tout  $k$  :

$$p^{(n+k)} = p^{(n)} p^{(k)}.$$

Ainsi, la loi de  $X_0$  est la probabilité de transitions déterminant la loi de  $X_n$ .

- Pour tout  $(i_1, \dots, i_n)$  :

$$\mathbf{P}(X_{s_n} = i_{s_n}, \dots, X_{s_1} = i_{s_1}) = \mathbf{P}(X_{s_n} = i_{s_n} | X_{s_{n-1}} = i_{s_{n-1}}) \dots \mathbf{P}(X_{s_2} = i_{s_2}) \mathbf{P}(X_{s_1} = i_{s_1}).$$

Par conséquent, la loi du vecteur aléatoire  $(X_{s_1}, \dots, X_{s_n})$  est déterminée par la loi de  $X_0$  et les probabilités de transition.

**Démonstration :**

Il suffit de remarquer que :

$$\begin{aligned} p_j^{n+1} &= \sum_{k=1}^r \mathbf{P}(X_{n+1} = j, X_n = k) \\ &= \sum_{k=1}^r \mathbf{P}(X_{n+1} = j | X_n = k) \mathbf{P}(X_n = k) \\ &= \sum_{k=1}^r \mathbf{P}_k^{(n)} \mathbf{P}_{kj}. \end{aligned}$$

Ce qui s'écrit matriciellement  $p^{(n+1)} = p^{(n)}p$ .

**Lemme 4.1.1** Si  $A$  est un évènement dans  $\mathfrak{F}(X_0, \dots, X_n)$ , alors :

$$\mathbf{P}(X_{n+k} = i_{n+k} | X_n = i_n, A) = \mathbf{P}(X_{n+k} = i_{n+k} | X_n = i_n)$$

**Démonstration :** On sait que cette formule est vraie quand  $k = 1$  et que  $A$  est de la forme  $C = \{X_{n-1} = i_{n-1}, \dots, X_0 = i_0\}$  (c'est la définition d'une chaîne de Markov). Tout évènement de  $\mathcal{F}(X_0, \dots, X_{n-1})$  est une union disjointe finie d'évènements  $C_l$  de cette forme  $C : A = \cup_l C_l$ .

$$\begin{aligned} \mathbf{P}(X_{n+1} = i_{n+1} | X_n = i_n, A) &= \frac{\mathbf{P}(X_{n+1}=i_{n+1}, X_n=i_n, A)}{\mathbf{P}(X_n=i_n, A)} \\ &= \sum_l \frac{\mathbf{P}(X_{n+1}=i_{n+1} | X_n=i_n, C_l)}{\mathbf{P}(X_n=i_n, A)} \\ &= \sum_l \frac{\mathbf{P}(X_{n+1}=i_{n+1} | X_n=i_n, C_l)}{\mathbf{P}(X_n=i_n, A)} \mathbf{P}(X_n = i_n, C_l) \\ &= \sum_l \frac{\mathbf{P}(X_{n+1}=i_{n+1} | X_n=i_n)}{\mathbf{P}(X_n=i_n, A)} \mathbf{P}(X_n = i_n, C_l) \\ &= \mathbf{P}(X_{n+1} = i_{n+1} | X_n = i_n) \sum_l \frac{\mathbf{P}(X_n=i_n, C_l)}{\mathbf{P}(X_n=i_n, A)} \\ &= \mathbf{P}(X_{n+1} = i_{n+1} | X_n = i_n). \end{aligned}$$

Le passage de la troisième à la quatrième ligne étant justifié par le fait évoqué plus haut que le lemme est vrai quand  $k = 1$  est pour des évènements de la forme  $C_l$ . Nous avons donc démontré que le lemme est vrai pour  $k = 1$ . Pour  $k$  quelconque on récurrence : supposons le lemme vrai et montrons qu'il est vrai pour  $k + 1$  comme précédemment :

$$\mathbf{P}(X_{n+(k+1)} = i_{n+(k+1)} | X_n = i_n, A) = \sum_{j \in E} \mathbf{P}(X_{n+k+1} = i_{n+k+1} | X_{n+1} = j, X_n = i_n, A) \mathbf{P}(X_{n+1} = j | X_n = i_n, A).$$

On applique l'hypothèse de récurrence avec  $n+1$  à la place de  $n$  et  $A' = \{X_n = i_n\} \cap A$  à la place de  $A$  ( $A' \in \mathcal{F}(X_0, \dots, X_{n-1}, X_n)$ ), ce que donne pour tout  $j \in E$ .

$$\begin{aligned} \mathbf{P}(X_{n+k+1} = i_{n+k+1} | X_{n+1} = j, X_n = i_n, A) &= \mathbf{P}(X_{n+k+1} = i_{n+k+1} | X_{n+1} = j, X_n = i_n, A') \\ &= \mathbf{P}(X_{n+k+1} = i_{n+k+1} | X_{n+1} = j, X_n = i_n). \end{aligned}$$

et on applique l'hypothèse pour  $k = 1, n$  et  $A$  pour obtenir :

$$\mathbf{P}(X_{n+1} = j | X_n = i_n, A) = \mathbf{P}(X_{n+1} = j | X_n = i_n).$$

Finalement :

$$\mathbf{P}(X_{(n+1)+k} = i_{(n+1)+k} | X_n = i_n, A) = \sum_{j \in E} \mathbf{P}(X_{(n+1)+k} = i_{(n+1)+k} | X_n = j) \mathbf{P}(X_{n+1} = j | X_n = i_n).$$

Et par conséquent :

$$\mathbf{P}(X_{n+(k+1)} = i_{n+(k+1)} | X_n = i_n, A) = \mathbf{P}(X_{n+(k+1)} = i_{n+(k+1)} | X_n = i_n).$$

## 4.2 Propriété de Markov

Nous démontrons dans cette section deux théorèmes qui seront d'un usage constant dans la suite.

**Théorème 4.2.1** *Soient  $n \geq 1$ ,  $A \in \mathcal{F}(X_0, \dots, X_{n-1})$  et  $B \in \mathcal{F}(X_{n+1}, X_{n+2}, \dots)$  pour tout  $i \in E$  on a :*

$$\mathbf{P}(B | X_n = i, A) = \mathbf{P}(B | X_n = i).$$

*Cette propriété signifie que le futur ne dépend de passé qu'à travers le présent.*

### 4.2.1 Matrice de transition et loi initiales

**Définition 4.2.1** *On dit qu'une matrice  $r \times r$  à coefficients réels  $P$  est stochastique si pour tous  $1 \leq i, j \leq r$  on a :*

- $0 \leq P_{ij} \leq 1$
- pour tous  $i$ ,  $\sum_{j=1}^r P_{ij} = 1$  (somme sur les lignes).

**Proposition 4.2.1** *Soit  $P$  une matrice stochastique pour toute loi de probabilité  $\mu$ , l'espace des états  $E$ , il existe un espace de probabilité,  $(\Theta, \mathcal{F}, \mathbf{P}_\mu)$  et une suite de variable aléatoire  $(X_n)_{n \geq 0}$  telle que  $(X_n)$  soit une chaîne de Markov admettant  $P$  comme matrice de transitions est  $\mu$  comme loi initiale.*

### 4.2.2 Récurrence

Les notions importantes sont celles d'état récurrent et d'état transitoire. Soit  $(X_n)_{n \geq 0}$ , une chaîne de Markov homogène de matrice de transition  $\mathbf{P}$ . On définit le temps de retour en  $i$  par :

$$T_i = \inf\{n \geq 1, X_n = i\}.$$

Avec la convention usuelle :  $\inf \emptyset = \infty$  (Ici :  $T_i = \infty$  si  $X_n \neq i$  pour tout  $n \geq 1$ ).

**Définition 4.2.2** *On dit que l'état  $i$  est récurrent si  $P_i(T_i < \infty) = 1$ . Un état  $i$  récurrent est dit récurrent positif si de plus  $E_i[T_i] < \infty$ , récurrent nul si  $E_i[T_i] = \infty$ . Un état qui n'est pas récurrent est dit transitoire pour un état  $i$  fixé, notons  $\{\tau_k\}_{k \geq 1}$  la suite des temps de retour successifs dans l'état  $i$ .*

*Formellement :*

$$\begin{aligned} \tau &= T_i \\ &\dots \\ \tau_{k+1} &= \inf\{n \geq \tau_k; X_n = i\} \\ &\dots \end{aligned}$$

**Théorème 4.2.2** Sachant que  $\tau_k < \infty$ , le processus  $\{X_{\tau_k+n}\}_{n>p}$  est une chaîne de Markov homogène de matrice de transition  $\mathbf{P}$  indépendant de  $\{X_{n\wedge\tau_k}\}_{n>0}$ .  
En particulier si l'état  $i$  est récurrent, les "cycles"  $\{X_n + \tau_k\}_{0 \leq n \leq \tau_{k+1}}$ ,  $k \geq 1$  sont i.i.d.

Notons  $f_{ji} = P_j(T_i < \infty)$  la probabilité de retourner à  $i$  en partant de  $j$ , et notons :

$$N_i = \sum_{n=1}^{\infty} 1_{\{X_n=i\}}.$$

Le nombre total de visite en  $i$  à partir de  $n = 1$ .

On a :  $P_j(N_i = 0) = 1 - f_{ji}$  et lorsque  $r \leq 1$  :

$$\begin{aligned} P_j(N_i = r) &= f_{ji} * (f_{ii})^{r-1} * (1 - f_{ii}) \\ &= (f_{ii})^r * (1 - f_{ii}) \end{aligned}$$

On déduit que :

1. Si  $f_{ii} = 1$ ,  $P_i(N_i = r) = 0$  pour tout  $r \geq 0$  et donc  $P_i(N_i = \infty) = 1$ , et bien entendu  $E_i[N_i] = \infty$ .
2. Si  $f_{ii} < 1$ , on a  $\sum_{r=0}^{\infty} P_i(N_i = r) = 1$ , et donc  $P_i(N_i = \infty) = 0$ , d'autre part un calcul élémentaire donne :  $E_i[N_i] = \frac{f_{ii}}{1-f_{ii}} < \infty$ .

**Théorème 4.2.3** Pour que l'état  $i$  soit récurrent, il faut et il suffit que :

$$\sum_{n=1}^{\infty} P_{ii}(n) = \infty.$$

**Démonstration :** Les remarques précédentes et l'énoncé du théorème montrent que  $f_{ii} = 1 \Leftrightarrow E_i[N_i] = \infty$ , d'autre part :

$$\begin{aligned} E_i[N_i] &= E_i \left[ \sum_{n=1}^{\infty} 1_{\{X_n = i\}} \right] \\ &= \sum_{n=1}^{\infty} E_i[1_{\{X_n = i\}}] \\ &= \sum_{n=1}^{\infty} P_i(X_n = i) \\ &= \sum_{n=1}^{\infty} P_{ii}(n). \end{aligned}$$

**Théorème 4.2.4** Soit une chaîne de Markov homogène de matrice de transition  $P$ . Deux états qui communiquent sont récurrents, et transitoires.

En particulier, si  $P$  est irréductible, les états sont soit tous récurrents, transitoires, la chaîne est dite alors récurrente et transitoire.

**Démonstration :** Si  $i$  et  $j$  communiquent il existe  $M$  et  $N$  tel que  $P_{ij}(M) > 0$  et  $P_{ji}(N) > 0$ . Posons  $\alpha = P_{ij}(M)P_{ji}(N) > 0$ . On a l'inégalité :

$$P_{ii}(M + n + N) \geq P_{ij}(M)P_{jj}(n)P_{ji}(N) = \alpha P_{jj}(n).$$

De même

$$P_{jj}(N + n + M) \geq \alpha P_{ii}(n).$$

Donc si  $i$  et  $j$  communiquent les séries  $\sum_{n=1}^{\infty} P_{ii}(n)$  et  $\sum_{n=1}^{\infty} P_{jj}(n)$  ont le même comportement.

Les états  $i$  et  $j$  sont donc, transitoires, ou récurrentes.

### 4.2.3 Irréductibilité

**Définition 4.2.3** On dit qu'un état  $i$  conduit à état  $j$  s'il existe  $n$  tel que  $(P^n)_{ij} > 0$ , c'est-à-dire s'il existe  $n$  tel que  $P(X_n = j | X_0 = i) > 0$ . On dit que  $i$  et  $j$  sont dans la même classe. Si  $i$  conduit à  $j$  et  $j$  conduit à  $i$ , c'est une relation d'équivalence et il est pertinent de considérer ses classes d'équivalence disjointes de  $E$ .

**Définition 4.2.4** Une classe d'état  $C \subset E$  est dite fermée si pour tout état tel que il existe un état  $i \in C$  qui conduit à  $j$ , on a  $j \in C$ .

Aurement dit, une classe  $C$  est fermée, si la chaîne de Markov quitte cette classe avec probabilité 0.

**Définition 4.2.5** Lorsque l'espace d'états se compose d'une seule classe (forcément fermée) c-à-d tous les états communiquent, la chaîne de Markov est dite irréductible.

### 4.2.4 Mesure Invariante

**Définition 4.2.6** Soit  $\mu$  une mesure positive sur  $E$ , on note aussi par  $\mu$  le vecteur ligne  $(\mu\{i\}, i \in E)$ .  $\mu$  est dit mesure invariante de la chaîne si  $\mu P = \mu$ . On prend garde au fait que  $\mu$  n'est pas nécessairement une probabilité. Observons que si  $\pi_0$  est une mesure invariante de la Chaîne,  $X_n$  est de loi  $\pi_0$  pour tout  $n \in \mathbb{N}$ .

**Notation :**

On note  $E_R$  l'ensemble des états récurrent,  $E_T$  l'ensemble des états transitoire.

$$N_n(j) = \sum_{i=1}^n 1_{\{X_i=j\}}, j \in E.$$

$$m_{ij} = E_i(T_j), i, j \in E.$$

**Théorème 4.2.5** Soit  $j \in E_R$ , on a :

$$\lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = \frac{1}{m_{jj}} 1_{\{T_i < \infty\}} p.s.$$

Pour  $i \in E$ , on a :

$$\lim_{n \rightarrow \infty} E_i\left(\frac{N_n(j)}{n}\right) = \frac{\rho_{ij}}{m_{jj}}.$$

**Théorème 4.2.6** Soit  $(X_n)_{n \geq 0}$  une chaîne de Markov homogène, irréductible récurrente positive de matrice de transition  $P$ . Alors  $(X_n)_{n \geq 0}$  admet une loi invariante  $\pi$ , donnée par :

$$\pi(\{i\}) = \frac{1}{m_{ii}}.$$

**Démonstration :**

Supposons que  $(X_n)_{n \geq 0}$  admet une loi invariante  $\pi$ , et montrons que

$$\pi(i) = \frac{1}{m_{ii}}.$$

$\pi$  loi invariante, elle vérifie :

$$\pi = \pi P.$$

C'est-à-dire pour tout  $i \in E$ ,

$$\pi(i) = \sum_{j \in E} \pi(j) P_{ji}.$$

Aussi

$$1 = \sum_{i \in E} \pi(i) = \sum_{i \in E} \sum_{j \in E} \pi(j) P_{ij}.$$

On a :

$$\begin{aligned} \sum_{j \in E} \pi(j) \left( \frac{1}{n} \sum_{m=1}^n P_{ji}^m \right) &= \frac{1}{n} \sum_{m=1}^n \sum_{j \in E} \pi(j) P_{ji}^m \\ &= \frac{1}{n} \sum_{m=1}^n \sum_{j \in E} \pi(j) P(X_m = i | X_0 = j) \\ &= \frac{1}{n} \sum_{m=1}^n \sum_{j \in E} P(X_m = i, X_0 = j) \\ &= \frac{1}{n} \sum_{m=1}^n \pi(i) \\ &= \pi(i). \end{aligned}$$

Or :

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P_{ji}^m &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P(X_n = i | X_0 = j) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P(X_n = i, X_0 = j) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n E_j(1_{\{X_m = i\}}) \\ &= \lim_{n \rightarrow \infty} E_j \left( \frac{N_n(j)}{n} \right) \\ &= \frac{\rho_{ji}}{m_{ii}}. \end{aligned}$$

$\rho_{ji} = 1$ , en effet :  
 $(X_n)_{n \geq 0}$  la Chaîne de Markov irréductible, alors

$$\forall i, j \in E, i \rightsquigarrow j \quad \text{et} \quad j \rightsquigarrow i.$$

Puisque la chaîne est récurrente, et  $\rightsquigarrow$  est symétrique sur l'ensemble des états récurrents, on a

$$\begin{aligned} \forall i, j \in E, i \rightsquigarrow j \text{ et } j \rightsquigarrow i. \\ i \rightsquigarrow j \Leftrightarrow \exists n \geq 1; p_{ij}^n > 0. \end{aligned}$$

On a

$$p_{ij}^n P_j(T_i = \infty) \leq P_i(T_i = \infty) = 0.$$

Ainsi

$$P_j(T_i = \infty) = 0.$$

Implique

$$P_j(T_i < \infty) = \rho_{ji} = 1.$$

Par passage à la limite quand  $n$  tend vers l'infini dans l'équation

$$\sum_{j \in E} \pi(j) \left( \frac{1}{n} \sum_{m=1}^n P_{ji}^m \right) = \pi(i).$$

On trouve

$$\sum_{j \in e} \pi(j) \frac{1}{m_{ii}} = \pi(i).$$

Implique

$$\pi(i) = \frac{1}{m_{ii}}.$$

Montrons maintenant l'existence de la loi invariante, soit  $E_1 \subset E$ , fini, on a :

$$\begin{aligned} \sum_{i \in E_1} \left( \frac{1}{n} \sum_{m=1}^n p_{ji}^m \right) &= \frac{1}{n} \sum_{m=1}^n \left( \sum_{i \in E_1} p_{ji}^m \right) \\ &\leq \frac{1}{n} \sum_{m=1}^n \left( \sum_{i \in E} p_{ji}^m \right) \\ &= 1. \end{aligned}$$

On prend la limite quand  $n$  tend vers l'infini, on obtient :

$$\sum_{i \in E} \frac{1}{m_{ii}} \leq 1.$$

Comme  $E_1$  est arbitraire; on a :

$$\sum_{i \in E} \frac{1}{m_{ii}} \leq 1.$$

D'autre part :

$$\begin{aligned} \sum_{i \in E_1} \left( \frac{1}{n} \sum_{m=1}^n p_{ki}^m \right) p_{ij} &= \frac{1}{n} \sum_{m=1}^n \sum_{i \in E_1} (p_{ki}^m p_{ij}) \\ &\leq \frac{1}{n} \sum_{m=1}^n \sum_{i \in E_1} (p_{ki}^m p_{ij}) \\ &= \frac{1}{n} \sum_{m=1}^n p_{ki}^{m+1}. \end{aligned}$$

Quand  $n$  tend vers l'infini; on a :

$$\sum_{i \in E_1} \frac{1}{m_{ii}} p_{ij} \leq \frac{1}{m_{jj}}.$$

Comme  $E_1$  est arbitraire :

$$\sum_{i \in E_1} \frac{1}{m_{ii}} p_{ij} \leq \frac{1}{m_{jj}}.$$

On somme sur  $j$  :

$$\sum_{i \in E} \frac{1}{m_{ii}} \leq \sum_{j \in E} \frac{1}{m_{jj}}.$$

Or :

$$\sum_{i \in E} \frac{1}{m_{ii}} = \sum_{j \in E} \frac{1}{m_{jj}}$$

Donc les inégalités précédentes sont égaux, on pose  $\pi(i) = \frac{1}{m_{ii}}$ ,

$$\sum_{i \in E} \pi(i) p_{ij} = \pi(j),$$

d'où l'existence de la loi invariante.

**Définition 4.2.7** Soit  $\pi$  une loi de probabilité sur  $E$ . La matrice de transition  $P$  est dite réversible par rapport à  $\pi$  si :

$$\pi(\{i\})p_{ij} = \pi(\{j\})p_{ji}.$$

Pour tout  $i, j \in E$ .

**Proposition 4.2.2** Si  $P$  est réversible par rapport à  $\pi$  alors  $\pi$  est une probabilité invariante.

**Démonstration :**

Pour montrer qu'une probabilité invariante il suffit de montrer que :

$$\begin{aligned} \forall i \in E, \pi(\{i\}) &= \sum_{j \in E} \pi(\{j\})p_{ji}. \\ \sum_{j \in E} \pi(\{j\})p_{ji} &= \sum_{j \in E} \pi(\{i\})p_{ij} \\ &= \sum_{j \in E} \pi(\{i\})p(X_1 = j | X_0 = i) \\ &= \pi(\{i\}). \end{aligned}$$

### 4.3 Simulation d'une Chaîne de Markov

Une autre façon de définir une Chaîne de Markov discrète est de la définir comme une suite de variables aléatoires  $(X_n)_{n \geq 0}$  défini par :

$$X_{n+1} = \mathcal{F}(X_n, U_n), X_0.$$

Où la suite de variable aléatoire  $(X_0, U_0, U_1, \dots, U_n, \dots)$  est indépendante, les variables aléatoires  $U_i$  étant à valeur dans un ensemble  $B$  et  $F : E \times B \rightarrow E$  est une application (mesurable).

**Lemme 4.3.1**  $U_n$  est indépendant de la tribu  $\mathcal{F}(X_n, \dots, X_0) \subset \mathcal{F}(X_0, U_0, U_1, \dots, U_{n-1})$ . Notons alors pour  $i, j \in E$

$$T_{i \rightarrow j} = \{b \in B; F(i, b) = j\}.$$

On a :

$$\begin{aligned} \mathbf{P}(X_{n+1} | X_n = i) &= \frac{\mathbf{P}(X_{n+1}, X_n = i)}{\mathbf{P}(X_n = i)} \\ &= \frac{\mathbf{P}(U_n \in T_{i \rightarrow j}, X_n = i)}{\mathbf{P}(X_n = i)} \\ &= \frac{\mathbf{P}(U_n \in T_{i \rightarrow j}) (\mathbf{P}(X_n = i))}{\mathbf{P}(X_n = i)} \\ &= \mathbf{P}(U_n \in T_{i \rightarrow j}). \end{aligned}$$

Où on a utilisé le fait que  $U_n$  et  $X_n$  soit indépendants. Calculons à présent :  $\mathbf{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \frac{\mathbf{P}(X_{n+1} = j, X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0)}{\mathbf{P}(X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0)}$ . Si on note  $A$  l'évènement  $A = (X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0)$ , on peut écrire :

$$\begin{aligned} \mathbf{P}(X_{n+1} | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) &= \frac{\mathbf{P}(X_{n+1} = j, X_n = i, A)}{\mathbf{P}(A)} \\ &= \frac{\mathbf{P}(U_n \in T_{i \rightarrow j}, A)}{\mathbf{P}(A)}. \end{aligned}$$

Et puisque  $U_n$  est indépendant de la tribu  $\mathcal{F}(X_n, \dots, X_0)$ .  
 $\mathbf{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbf{P}(U_n \in T_{i \rightarrow j})$ ,  
on a donc démontré que pour tout  $i, j \in E$ .  
 $\mathbf{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbf{P}(X_{n+1} = j | X_n = i)$ . Ce que signifie que  $(X_n)$  est une chaîne de Markov.

## 4.4 Loi stationnaire

**Définition 4.4.1** On dit qu'une loi  $\mu$  sur  $E$  est stationnaire (ou encore invariante) pour la matrice stochastique  $P$  si  $\mu P = \mu$ .

**Proposition 4.4.1** La loi  $\mu$  est stationnaire pour  $P$  si et seulement si la Chaîne de Markov de loi initial  $\mu$  et de matrice de transition  $P$  vérifie pour tous  $n, j$  :

$$\mathbf{P}(X_n = j) = \mu(j).$$

(cela signifie que la loi de  $X_n$  est également  $\mu$ ).

**Proposition 4.4.2** Soit la chaîne de Markov admettant une loi stationnaire  $\mu$ . Alors tout état  $i$  tel que  $\mu(\{i\}) > 0$  est récurrent. En outre la chaîne est irréductible elle est récurrente.

**Démonstration :** On a :  $\mu P^k = \mu$  pour tout entier  $k \geq 0$  est donc  $\mu(I + \dots + P^{n-1}) = n\mu$  par conséquent  $\sum_{j \in E} \mu_j V_j = \infty$  pour  $i$  tel que  $\mu(i) > 0$ .

**Théorème 4.4.1** Si l'espace des états est fini, toute chaîne de Markov sur  $E$  (de matrice stochastique  $P$ ) admet au moins une loi stationnaire.

**Démonstration :** Si  $\mu_0$  est une loi quelconque sur  $E$ , par exemple la loi  $\delta_i$ , notons  $\mu_n = \mu_0 P^n$  la loi de  $X_n =$  où  $(X_n)$  est la chaîne de Markov de matrice  $P$  est de loi initiale  $\mu_0$ . Notons encore  $\tilde{\mu}_n$  la loi  $\frac{(\mu_0 + \dots + \mu_{n-1})}{n}$ , si bien que :

$$\tilde{\mu} = \frac{1}{n}(\mu_0 + \mu_0 P + \dots + \mu_0 P^{n-1}).$$

Il existe sous-suite  $n_k$  telle que pour tout  $i \in E$ ,  $\mu_{n_k}(i)$  converge vers un réel  $\pi(i)$ . Il est clair que les  $\pi(i)$  sont dans  $[0,1]$  est vérifiant  $\sum_{i \in E} \pi(i) = 1$  :  $\pi$  est bien que une loi de probabilité. Elle est stationnaire. En effet :

$$\begin{aligned} \pi P &= \lim_{n_k \rightarrow \infty} \frac{1}{n_k}(\mu_0 + \mu_0 P + \dots + \mu_0 P^{n_k-1})P \\ &= \lim_{n_k \rightarrow \infty} \frac{1}{n_k}(\mu_0 P + \mu_0 P^2 + \dots + \mu_0 P^{n_k}) \\ &= \lim_{n_k \rightarrow \infty} \left( \frac{1}{n_k}(\mu_0 + \mu_0 P + \dots + \mu_0 P^{n_k-1}) + \frac{1}{n_k}(\mu_0 P^{n_k} - \mu_0) \right) \\ &= \pi. \end{aligned}$$

Quand l'espace des états est infini, l'existence d'une telle loi, de probabilité invariant n'est pas garantie, l'unicité n'est pas toujours vraie même si  $E$  est fini.

**Proposition 4.4.3** *Si la Chaîne de Markov est irréductible (et finie) alors tout mesure de probabilité stationnaire a pour support l'espace des états. En particulier, si  $i$  est un état récurrent il existe une mesure stationnaire  $\mu$  dont le support est la classe  $C(i)$  de  $i$ .*

**Démonstration :** Soit  $l$  un état où  $\mu(l) > 0$ , ( un tel état existe toujours puisque  $\sum_k \mu(k) = 1$ ). Si  $j$  est un état quelconque on sait qu'il existe  $n$  tel que  $(P^n)_{lj} > 0$ , or d'après l'invariance :

$$\mu_j = \sum_{k \in E} \mu_k (P^n)_{kj} \geq \mu_l (P^n)_{lj} > 0.$$

**Théorème 4.4.2** *Si une chaîne de Markov est finie alors les assertions suivant sont équivalentes :*

1. elle est irréductible.
2. elle admet une unique mesure de probabilité stationnaire.

**Démonstration :** Supposons que la chaîne irréductible et faisons l'hypothèse qu'il existe deux mesure de probabilité  $\mu_1, \mu_2$  différentes telle que  $\mu_1 = \mu_1 P$  et  $\mu_2 = \mu_2 P$ . Si on appelle  $x$  le vecteur ligne  $x = \mu_1 - \mu_2$  on a  $x = xP$  et  $\sum_{i \in E} x_i = 0$ ; on particulier il existe deux indices  $i, j$  tel que  $x_i$  et  $x_j$  sont de signes opposés puisque la matrice  $P$  est irréductible il existe un exposant  $m$  tel que  $(P^m)_{ij} > 0$ . Écrivons  $x = xP^m$  puis :

$$\begin{aligned} \sum_{j \in E} |x_j| &= \sum_{j \in E} \left| \sum_{k \in E} x_k (P^m)_{kj} \right| \\ &\leq \sum_{j \in E} \sum_{k \in E} |x_k| (P^m)_{kj} \\ &\leq \sum_{k \in E} |x_k| \sum_{j \in E} (P^m)_{kj} \\ &\leq \sum_{k \in E} |x_k|. \end{aligned}$$

**Théorème 4.4.3** *Pour une chaîne de Markov irréductible finie de loi de probabilité stationnaire  $\mu$ , on a  $\mathbf{E}(T^i | X_0 = i) = (\mu^i)^{-1}$*

**Démonstration :** Pour raccourcir les notation on va écrire ici  $\mathbf{E}(\cdot | X_0) = \mathbf{E}(\cdot)$  et  $\mathbf{P}(\cdot | X_0) = \mathbf{P}(\cdot)$ . On va prouver que le vecteur  $\gamma^i$  avec  $\gamma_j^i = \mathbf{E}_i \sum_{k=0}^{T^i-1} \mathbf{1}_j \circ X_k$  pour  $j \in E$  vérifie l'équation  $\gamma^i P = \gamma^i$ . Alors :

$$\begin{aligned} \sum_{j \in E} \gamma_j^i &= \sum_{j \in E} \mathbf{E}_i \sum_{k=0}^{T^i-1} \mathbf{1}_j \circ X_k \\ &= \mathbf{E}_i \sum_{k=0}^{T^i-1} \sum_{j \in E} \mathbf{1}_j \circ X_k \\ &= \mathbf{E}_i \sum_{k=0}^{T^i-1} 1 \\ &= \mathbf{E}_i T^i. \end{aligned}$$

Donc loi stationnaire  $\mu_j = \gamma_j^i | \mathbf{E}_i T^i$  comme  $\gamma_j^i = 1$ , le théorème sera prouvé. Pour prouver l'égalité  $\gamma^k = \gamma^k P$  notons d'abord que  $T^k < \infty$  p.s pour tout  $k$  nous pouvons donc écrire

$$\gamma_j^k = \mathbf{E}_k \sum_{n=0}^{T^k} \mathbf{1}_{\{X_n=j\}} + \mathbf{P}(X_0 = j) - \mathbf{P}_k(X_{T^k} = j) = \mathbf{E} \sum_{n=1}^{T^k} \mathbf{1}_{\{X_n=j\}}.$$

Car  $\mathbf{P}_k(X_0 = j) - \mathbf{P}_k(X_{T^k} = j) \text{ est } 1 - 1 = 0 \text{ si } j = k \text{ et } 0 - 0 = 0 \text{ si } j \neq k$ . (S la chaîne de Markov n'avait pas été récurrent on aurait eu  $\mathbf{P}(T^k < \infty) < 1$  est on n'aurait pas considéré  $X_{T^k}$ !) Alors :

$$\gamma_j^k = \mathbf{E}_k \sum_{n=1}^{\infty} 1_{\{X_n=j, n < T^k\}} = \sum_{n=1}^{\infty} \mathbf{P}_k(X_n = j, n \leq T^k).$$

Par la propriété de Markov

$$\begin{aligned} \mathbf{P}_k(X_{n-1} = i, X_n = j, T^k \geq n) &= \mathbf{P}_k(X_n = j | X_{n-1} = i, T^n \geq n) \mathbf{P}_k(X_{n-1} = i, T^k \geq n) \\ &= \mathbf{P}_k(X_n = j | X_{n-1} = i, X_1, \dots, X_{n-2} \neq k) \mathbf{P}_k(X_{n-1} = i, T^k \geq n) \\ &= \mathbf{P}_k(X_n = j | X_{n-1} = i) \mathbf{P}_k(X_{n-1} = i, T^k \geq n) \\ &= \mathbf{P}_{ij} \mathbf{P}_k(X_{n-1} = i, T^k \geq n). \end{aligned}$$

Alors :

$$\begin{aligned} \gamma_j^k &= \sum_{n=1}^{\infty} \mathbf{P}_k(X_{n-1} = i, T^k \geq n) \\ &= \sum_{n=1}^{\infty} \sum_{i \in I} \mathbf{P}_k(X_{n-1} = i, X_n = j, T^k \geq n) \\ &= \sum_{i \in I} \mathbf{P}_{ij} \sum_{n=1}^{\infty} \mathbf{P}_k(X_{n-1} = i, T^k \geq n) \\ &= \sum_{i \in I} \mathbf{P}_{ij} \mathbf{E}_k \sum_{m=0}^{\infty} 1_{\{X_m=i, m \leq T^k-1\}} \\ &= \sum_{i \in I} \mathbf{P}_{ij} \mathbf{E}_k \sum_{m=0}^{\infty} 1_{\{X_m=i\}} \\ &= \sum_{i \in I} \gamma_j^k \mathbf{P}_{ij}. \end{aligned}$$

## 4.5 Théorème Ergodique

Le théorème suivant est appelé aussi la loi des grands nombres pour les chaîne de Markov.

**Théorème 4.5.1** *Soit  $(X_n)_{n=0}^{\infty}$  une chaîne de Markov de matrice de transition  $P$  irréductible finie de loi stationnaire  $\mu$ . Alors :*

$$\frac{\sum_{k=0}^{n-1} 1_{\{X_k=i\}}}{n} \longrightarrow \mu_i, n \longrightarrow \infty \quad p.s.$$

De plus pour toute fonction  $f : E \longrightarrow \mathbf{R}$

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \longrightarrow \int_I f d\pi = \sum_{i \in E} \mu_i f(i), n \longrightarrow \infty, p.s.$$

**Démonstration :** On note  $\sum_{k=0}^{n-1} 1_{\{X_k=i\}} = N_i(n)$  le nombre de visite dans  $i$  avant l'instant  $n$ . Soit  $\sum_{k=0}^{\infty} 1_{\{X_k=i\}} = N_i$  est le nombre total de visite dans  $i$ .

Comme pour une chaîne de Markov irréductible finie  $\mathbf{P}(T^i < \infty) = 1$ , il suffit de considérer une chaîne de Markov de mesure initial  $\delta_i$ .

Soient  $Q_r^i = S(i)^{(r)} - S(i)^{(r-1)}$  les intervalles des temps entre le  $r$ ième et le  $(r-1)$ ième visite dans  $i$ , les variables aléatoires  $Q_1^i, Q_2^i, Q_3^i, \dots$  sont indépendentes de même loi et  $\mathbf{E}(Q_r^i | X_0 = i) = \mathbf{E}(T^i | X_0 = i)$  on a :

$$Q_1^i + Q_2^i + \dots + Q_{N_i(n)-1}^i \leq n - 1.$$

où  $Q_1^i + Q_2^i + \dots + Q_{N_i(n)-1}$  est l'instant de la dernière visite dans  $i$  avant l'instant  $n$ . On a aussi

$$Q_1^i + Q_2^i + \dots + Q_{N_i(n)} \geq n.$$

Où  $Q_1^i + Q_2^i + \dots + Q_{N_i(n)-1}$  est l'instant de la première visite dans  $i$  après l'instant  $n-1$ . On obtient

$$\frac{Q_1^i + Q_2^i + \dots + Q_{N_i(n)-1}}{N_i(n)} \leq \frac{n}{N_i(n)} \leq \frac{Q_1^i + Q_2^i + \dots + Q_{N_i(n)}}{N_i(n)}.$$

Par la loi des grand nombre :

$$\frac{Q_1^i + Q_2^i + \dots + Q_n^i}{n} \longrightarrow \mathbf{E}(T^i | X_0 = i), n \longrightarrow \infty .p.s.$$

Comme  $P$  est récurrente

$$\mathbf{P}(N_i(n) \longrightarrow \infty, n \longrightarrow \infty) = 1$$

doù

$$\frac{n}{N_i(n)} \longrightarrow \mathbf{E}(T^i | X_0 = i), n \longrightarrow \infty p.s.$$

## 4.6 Algorithme de Metropolis-Hastings

On se fixe une loi  $\pi$  suivant laquelle on aimerait simuler ou dont on voudrait calculer un intégral  $\int_E f(x)\pi d(x)$ . Nous appellerons  $\pi$  la loi cible.

On se donne un noyau de Markov  $Q$ . Nous appellerons  $Q$  le noyau de proposition. Nous allons construire une chaîne de Markov  $(X_n)_{n \geq 0}$ .

- Nous prenons  $X_0 = x_0$  tel que  $\pi(x_0) > 0$ .
- Si  $X_n = x$ , nous simulons  $Y_{n+1}$  et  $U_{n+1}$  indépendants (et indépendant des simulations passées) avec :

$$Y_{n+1} \sim Q(x, \cdot), U_{n+1} \sim \mathcal{U}([0, 1]).$$

La variable  $Y_{n+1}$  s'appelle une proposition, posons pour tout  $x, y$  :

$$\alpha(x, y) = \min \left( 1, \frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)} \right).$$

Alors :

$$X_{n+1} = \begin{cases} Y_{n+1} & \text{si } U_{n+1} \leq \alpha(X_n, Y_{n+1}) \\ X_n & \text{sinon} \end{cases}$$

Dans le cas où  $U_{n+1} \leq \alpha(X_n, Y_{n+1})$ , on dit qu'on accepte la proposition, et dans le cas contraire on dit qu'on refuse la proposition.

**Proposition 4.6.1** *La suite aléatoire  $(X_n)_{n \geq 0}$  construite ci-dessus est une chaîne de Markov est transition  $P$  avec :*

$$\begin{cases} P(x, y) = Q(x, y)\alpha(x, y) & \text{si } x \neq y \\ P(x, y) = 1 - \sum_{y \neq x} P(x, y). \end{cases}$$

La loi  $\pi$  est  $P$ -invariante.

**Démonstration :** Pour tout  $x$ ,  $\mathbb{P}(X_{n+1} = x | X_0, \dots, X_n) = \mathbb{P}(X_{n+1} = x | X_n)$  (d'après la construction ci-dessus).

Donc  $(X_n)_{n \geq 0}$  est bien une chaîne de Markov. Calculons pour tout  $x \neq y$  :

$$\begin{aligned} \mathbb{P}(X_{n+1}y | X_n = x) &= \mathbb{P}(Y_{n+1} = y, U_{n+1} \leq \alpha(X_n, Y_{n+1}) | X_n = x) \\ &= \mathbb{P}(Y_{n+1} = y, U_{n+1} \leq \alpha(x, Y_{n+1}) | X_n = x) \\ &= \mathbb{P}(Y_{n+1} = y, U_{n+1} \leq \alpha(x, y) | X_n = x) \end{aligned}$$

$$(Y_{n+1} \perp U_{n+1}) = Q(x, y)\alpha(x, y).$$

Nous en déduisons

$$\begin{aligned} \pi(x)P(x, y) &= \pi(x)Q(x, y)\alpha(x, y) \\ &= \pi(x)Q(x, y) \min\left(1, \frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)}\right) \\ &= \min(\pi(x), Q(x, y), \pi(x)Q(x, y)) \\ &= \min(\pi(y)Q(y, x), \pi(x)Q(x, y)) \\ &= \pi(y)Q(y, x) \end{aligned}$$

(on refait le calcul en inversant  $x$  et  $y$ ).

Donc  $\pi$  est symétrique par rapport à  $P$ . Nous avons donc  $\pi P = \pi$ .

**Lemme 4.6.1** Avec la notation définies ci-dessus si  $Q(x, y) \neq 0 \Rightarrow Q(y, x) \neq 0$ , est irréductible et vérifie :  $\forall x, y : Q(x, y) \neq 0 \Rightarrow Q(y, x) \neq 0$ , et  $\pi(x) > 0$ . Pour tout  $x, y$  alors  $P$  irréductible.

**Démonstration :**

Soient  $x, y$  dans  $E$ ,  $x \neq y$ . Il existe  $n$  dans  $\mathbb{N}$  et  $(x_1, \dots, x_n)$  dans  $E$  tels que :

$$Q(x, x_1)Q(x_1, x_2)\dots Q(x_n, y) > 0.$$

On peut toujours supposer  $x \neq x_1, \dots, x_n \neq y$ , on déduit des hypothèses que :

$$P(x, x_1)P(x_1, x_2), \dots, P(x_n, y) > 0.$$

**Application sous R :**

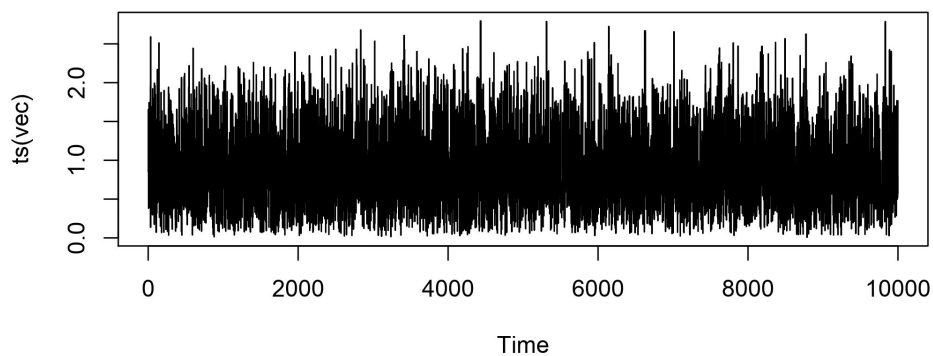
Voici un exemple de l'algorithme de Metropolis Hastings simulant la distribution Gamma  $Gam(2.3, 2.7)$  En utilisant 100000 échantillons de la distribution de proposition normale  $N(0, 85, 0, 30)$ . les graphiques ci-dessous montrent que la chaîne se mélange bien et qu'il y a indépendance parmi les échantillons générés. Aussi un histogramme des échantillons générés ressemblant à la courbe de densité gamma.

```
> gamm<-function (n, a, b)
+ {
+ mu <- a/b
+ sig <- sqrt(a/(b * b))
```

```

+ vec <- vector(numeric", n)"
+ x <- a/b
+ vec[1] <- x
+ for (i in 2:n) {
+ can <- rnorm(1, mu, sig)
+ aprob <- min(1, (dgamma(can, a, b)/dgamma(x, a, b))/(dnorm(can,
mu, sig)/dnorm(x,mu, sig)))
+ u <- runif(1)
+ if (u < aprob)
+ x <- can
+ vec[i] <- x
+ }
+ vec
+ }
> vec<- gamm(10000,2.3, 2.7)
> par(mfrow=c(2,1))
> plot(ts(vec))
> hist(vec,30)
> par(mfrow=c(1,1))

```



**Histogram of vec**

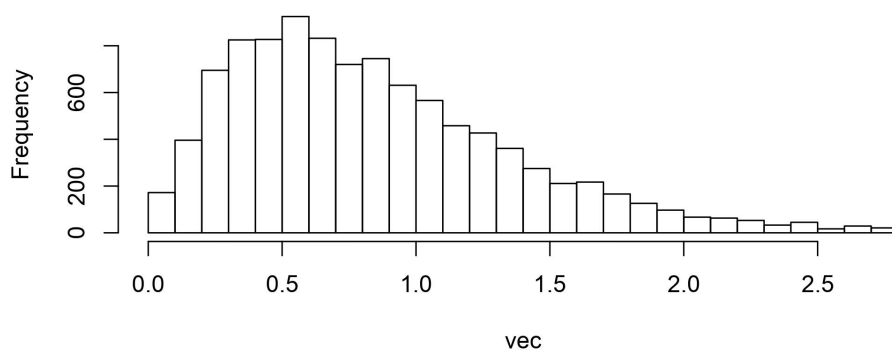


FIGURE 4.1 – La simulation de la loi Gamma par l’algorithme de Metropolis-Hasting

## 4.7 Algorithme de Metropolis

Si  $Q(x, y) = Q(y, x)$  pour tout  $x, y$  (on dit que le noyau  $Q$  est symétrique). Dans ce cas  $\alpha$  se simplifie en :

$$\alpha(x, y) = \min \left( 1, \frac{\pi(y)}{\pi(x)} \right).$$

Pour tout  $x, y$ .

**Proposition 4.7.1** *Si  $Q$  est un noyau de Markov irréductible symétrique et si  $\pi$  est une probabilité non constant telle que  $\pi(x) > 0$  pour tout  $x$ . Alors la chaîne de Markov de Metropolis de noyau de proposition  $Q$  et de loi cible  $\pi$  est irréductible, apériodique, de loi invariante  $\pi$ .*

**Démonstration :** Soit  $P$  le noyau de Metropolis de noyau de proposition  $Q$  et de loi cible  $\pi$ , pour tout  $x, y$   $\alpha(x, y) > 0$ . De plus, pour tout  $x, y$ , il existe  $p \in \mathbb{N}$  est une suite  $(x_1, \dots, x_n)$  de  $E$ .

$$P(x, x_1)P(x_1, x_2) \cdots P(x_n, y) = Q(x, x_1)Q(x_1, x_2) \cdots Q(x_n, y) > 0.$$

Supposons que  $P(x, x) = 0$  pour tout  $x$ .

$$\begin{aligned} P(x, x) &= 1 - \sum_{y: y \neq x} P(x, y) \\ &= \sum_y Q(x, y) - \sum_{y: y \neq x} Q(x, y)\alpha(x, y) \\ &= Q(x, x) + \sum_{y: y \neq x} Q(x, y)(1 - \alpha(x, y)) \\ &= Q(x, x) + \sum_{y: y \neq x} Q(x, y) \left( 1 - \frac{\pi(y)}{\pi(x)} \right) \end{aligned}$$

Donc :

$$P(x, x) - Q(x, x) = \sum_{y: y \neq x} Q(x, y) \left( 1 - \frac{\pi(y)}{\pi(x)} \right).$$

On fixe  $x \in E$ , les deux terms de l'équation ci-dessus sont de signes opposés, ils valent donc 0.

Nous avons donc pour tout  $y$  tel que  $y \neq x$  et  $Q(x, y) \neq 0$  (il en existe car  $Q$  est irréductible)

$$\left( 1 - \frac{\pi(y)}{\pi(x)} \right).$$

C'est-à-dire  $\pi(y) \geq \pi(x)$ .

### Application sous R :

Voici un exemple d'utilisation de l'algorithme Metropolis avec 10000 échantillons de la distribution de proposition uniforme  $U(-1, 1)$  pour générer la distribution normale  $N(0, 1)$ . Les graphiques ci-dessous montrent que la chaîne de Markov se mélange bien et qu'il existe une indépendance entre les échantillons générés. Aussi un histogramme des échantillons ressemblant à la courbe de densité normale.

```
> norm<-function (n, alpha)
```

```

+{
+  vec <- vector("numeric", n)
+  x<-0
+  vec[1] <- x
+  for (i in 2:n) {
rv <- runif(1,-alpha,alpha)
candidate <- x + rv
accept <- min(1, dnorm(candidate)/dnorm(x))
u <- runif(1)
if (u < accept)
  x <- candidate
vec[i] <- x
+}
+ vec +}
> normvec<-norm(10000,1)
> par(mfrow=c(2,1))
> plot(ts(normvec))
> hist(normvec,30)
> par(mfrow=c(1,1))
}

```

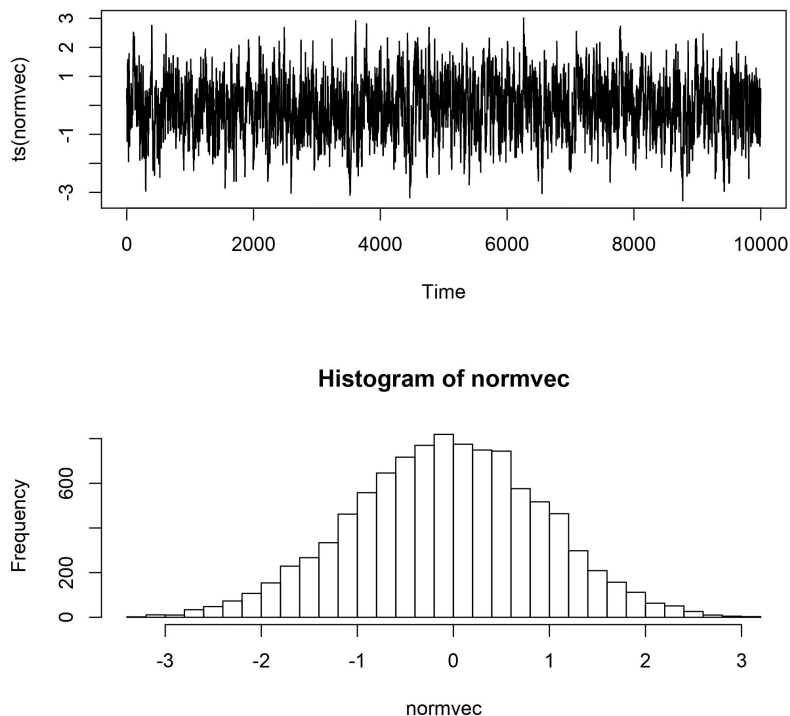


FIGURE 4.2 – Les graphes de la simulation de la loi Normale par la méthode de Metropolis

## 4.8 Algorithme de Gibbs

L'échantillonnage de Gibbs est un cas particulier de l'algorithme de Metropolis-Hastings. Cet algorithme peut être appliqué lorsque la distribution conjointe est inconnue ou difficile pour échantillonner directement, mais la distribution conditionnelle de chaque variable est connue et est facile à échantillonner. Il génère un nouvel échantillon de la distribution de chaque variable basée sur la distribution conditionnelle entre les valeurs actuelles de l'autre variable. Le but de l'algorithme de Gibbs est d'échantillonner à partir d'une distribution  $\pi$  pour approximer.

$$E_{\pi} f(x_1, x_2); \text{ ou } (x_1, x_2) \sim^{i.i.d} \pi.$$

En obtenant une séquence d'observations qui sont approximées à partir d'une distribution de probabilité multivariée spécifique. Cette séquence peut être utilisée pour approximer la distribution conjointe pour approximer la distribution marginale d'une des variables, ou d'un sous-ensemble de variables pour calculer une intégrale. Il faut que toutes les distributions conditionnelles de la distribution cible soient échantillonnées exactement. Basé sur des distributions conditionnelles complètes  $P(x_1|x_2)$  et  $P(x_2|x_1)$  avec la distribution cible  $\pi(x_1, x_2)$ .

### Stratégie générale :

- Alternner entre  $x_1$  et  $x_2$
- Échantillon de  $x_1 \sim p(x_1|x_2)$
- Échantillon de  $x_2 \sim p(x_2|x_1)$

L'utilisation de la procédure ne nécessite pas le calcul d'un taux d'acceptation car ce sera toujours 1.

**Normal Bivarié :** La distribution normale bivariée peut être générée à l'aide de l'échantillonneur Gibbs. Laisser  $Z = (x, y)$  être un échantillon normal bivarié (de taille 1), avec une moyenne  $\mu = (\mu_x, \mu_y)^T$  est inconnu. Les variances sont connues et égales à 1, et la corrélation  $\rho$  est connue. Prenons un a priori plat pour  $\mu$ , puis le postérieur est également normale bivariée, avec la même matrice de corrélation mais avec la moyenne  $Z$ .

# Chapitre 5

## Conclusion

Comme conclusion, nous retenons que la méthode de Monte-Carlo est une méthode d'approximation par introduction de procédés aléatoires. Cela permet d'estimer des valeurs numérique et de caractériser des systèmes complexes.

La méthode de Monte-Carlo est l'une des techniques de simulation offerant la possibilité de calculer des intégrales difficiles à calculer.

L'inconvénient est que cette méthode est lente, mais il existe des cas où c'est la seule technique accessible en effet elle ne dépend pas de la régularité de la fonction à intégrer.

D'autre part ne dépend pas de la dimension de l'espace en question ce qui évite les problèmes de déreglement de la convergence de l'estimateur lorsqu'il sagit de grand dimensions.

# Bibliographie

- [1] Abekhti. A. (2019). Calcul d'intégral par la méthode de Monte-Carlo [Mémoire de master]. Université d'Adrar.
- [2] Carlo. G. et Xavier G. (2008). Modélisation et statistique spatiales. Springer.
- [3] Christian P.R. et George C. (2010). Méthode de Monte-Carlo avec R. Springer.
- [4] Nicholas B.2018. Metropolis, Metropolis-Hastings and Gibbs Sampling Algorithms.Department of mathematical Sciences, Lakehead University, Ontario.
- [5] Pierre B. (2009). Initiation aux probabilité et aux chaîne de Markov, deuxième édition entierement révisée . Springer.
- [6] Robin S. (1997). Simulation de variable aléatoires, ING - PG.Biométrie.
- [7] Sylvian R. (2019). Méthode de Monte-carlo, Université Nice sophia antiplis.