

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche  
Scientifique  
Université Abou Bakr Belkaid- Tlemcen



Faculté des Sciences  
Département des Mathématiques

**MEMOIRE DE MASTER  
EN MATHÉMATIQUES**  
Option : Probabilités-Statistiques

**Méthodes de Monte-Carlo**

Présenté par : **MEDOUAKH Fatima Zahra**

Mémoire soutenu devant le jury composé de :

<b>M. T. Mourid</b>	<b>Professeur</b>	<b>Universsité de Tlemcen</b>	<b>Président</b>
<b>Mme. M. Dali youcef</b>	<b>M.C.A</b>	<b>Universsité de Tlemcen</b>	<b>Encadreure</b>
<b>M. A. Allam</b>	<b>M.C.A</b>	<b>Universsité de Tlemcen</b>	<b>Examineur</b>
<b>M. A. Labbas</b>	<b>M.C.A</b>	<b>Universsité de Tlemcen</b>	<b>Examineur</b>

Année universitaire 2016-2017

## Remerciements

*Tout d'abord, je remercie ALLAH qui m'a donné la volonté, la patience, et surtout la santé durant toutes mes années d'étude.*

*Mes vifs remerciements sont adressés à mon encadreuse Malika DALI YOUCEF pour sa patience, ses conseils, ses orientations, et sa disponibilité durant la préparation de ce mémoire de fin d'études de Master.*

*Je voudrais remercier tous les professeurs qui ont contribué à ma formation et à la réussite de mes études universitaires ainsi que les membres de mon Jury de soutenance :*  
*monsieur Tahar MOURID qui m'a fait le grand honneur d'accepter, d'être le président de mon jury, monsieur Ahmed LABBAS et monsieur Abdelaziz ALLAM qui ont pris la peine de me lire et d'évaluer mon travail.*

*Enfin ma pensée et ma gratitude vont à l'ensemble des responsables et de l'administration du Département de Mathématiques, sans oublier tout le personnel employé au Département .*

*Je tiens aussi à remercier tout particulièrement mes parents d'abord et ma famille ensuite pour leur encouragements continus.*

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>iii</b>
<b>2</b>	<b>Simulation de variables aléatoires.</b>	<b>1</b>
2.1	Simulation de la loi uniforme . . . . .	1
2.2	Méthode d'inversion . . . . .	1
2.3	Méthode de rejet . . . . .	5
2.4	Simulation des lois normales . . . . .	10
<b>3</b>	<b>Calcul d'une intégrale par la méthode de Monte-Carlo.</b>	<b>13</b>
<b>4</b>	<b>Méthode de Monte-Carlo par Chaînes de Markov.</b>	<b>20</b>
4.1	La propriété de Markov . . . . .	20
4.2	Calcul des lois marginales . . . . .	21
4.3	Récurrence et transience . . . . .	24
4.4	Mesure invariante . . . . .	26
4.5	Période d'une chaîne de Markov . . . . .	30
4.6	Méthode de Monte-Carlo . . . . .	31
<b>5</b>	<b>Conclusion</b>	<b>33</b>
<b>6</b>	<b>Annexe</b>	<b>34</b>

# 1 Introduction

La méthode de Monte-Carlo a vu son essor à partir de la fin de la seconde guerre mondiale, essentiellement dans le cadre du projet américain "Manhattan" concernant le développement de l'arme nucléaire. Cette époque correspond également à la construction des premiers "ordinateurs". Ce projet étant été classé "secret défense", il est difficile de savoir exactement qui parmi ses pionniers : Von Neumann, Ulam, Metropolis a proposé le nom de "Monte-Carlo". Quoi qu'il en soit, ce terme fait référence aux jeux de hasard : la capitale de la principauté de Monaco.

Le terme méthode de Monte-Carlo désigne une famille de méthodes algorithmiques visant à calculer une valeur numérique approchée en utilisant des procédés aléatoires, c'est-à-dire des techniques probabilistes. D'un point de vue purement mathématique, une méthode de Monte-Carlo peut servir au calcul d'intégrales (simples ou multiples) et à la résolution d'équations aux dérivées partielles, de systèmes linéaires et de problèmes d'optimisation. D'une manière plus globale, elle permet d'avancer toutes les possibilités liées à une observation, en quantifiant le risque associé (ou l'incertitude statistique). Pratiquement, ces techniques sont couramment utilisées dans divers domaines (physique, ingénierie, finance...).

Dans ce mémoire, nous avons présenté la méthode de Monte-Carlo du point de vue de son application pour le calcul d'intégrale. En particulier, pour appliquer la méthode en question, il faut savoir simuler les variables aléatoires, c'est la raison pour la quelle nous avons introduit dans la première section la simulation de variables aléatoires dont la méthode d'inversion, méthode de rejet, ainsi que la méthode de Box-Muller.

Le problème traité est donc le calcul de  $\int h(x)d\pi(x)$ , où  $\pi$  est une loi de probabilité et  $h$  est une fonction donnée. Il existe des cas où on ne peut pas simuler  $\pi$  par les méthodes précédentes de simulation, la méthode de Monte Carlo par chaîne de Markov pour résoudre de tels problèmes "plus fins" et c'est le thème de la troisième section.

## 2 Simulation de variables aléatoires.

Pour appliquer une méthode de Monte-Carlo, il faut savoir simuler suivant une loi donnée, dans cette section nous allons voir quelques méthodes de simulation de variables aléatoires.

### 2.1 Simulation de la loi uniforme

La distribution de base qu'il faut savoir simuler numériquement est la loi uniforme sur  $[0, 1]$ . Tous les langages de programmation disposent d'un générateur de nombres au hasard dans  $[0, 1]$ . Par exemple, en Matlab le générateur est **rand**.

Maintenant, pour simuler  $X$  une variable aléatoire qui suit la loi uniforme sur  $[a, b]$ , il suffit de simuler  $Y = a + (b - a)U$ , où  $U$  suit la loi uniforme sur  $[0, 1]$ .

Il existe plusieurs méthodes de simulation : Nous rappelons dans le paragraphe suivant quelques une de ces méthodes de simulation.

### 2.2 Méthode d'inversion

#### principe

La méthode d'inversion est la plus simple des méthodes générales de simulation. Elle consiste à composer l'inverse de la fonction de répartition  $F$  de la distribution à simuler avec un générateur de la loi uniforme sur  $[0, 1]$ .  $F$  n'étant pas toujours inversible au sens classique, on définit son inverse de la façon suivante :

$$\forall u \in [0, 1], F^{-1}(u) = \inf\{x; F(x) \geq u\}.$$

De point de vue théorique, le principe de la méthode est justifié dans la proposition suivante :

**Proposition 2.1.** *Soient  $F$  une fonction de répartition sur  $\mathbb{R}$  et  $U$  une variable aléatoire de loi uniforme sur  $[0, 1]$ . Alors la variable aléatoire*

$$X = F^{-1}(U),$$

*a pour fonction de répartition  $F$ .*

Pour la preuve nous avons besoin du lemme suivant :

**Lemme 2.1.** *Pour tout  $u \in [0, 1], x \in \mathbb{R}$*

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

*Démonstration du lemme.* 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  $F(x) \geq u$ , quant à l'implication réciproque elle est triviale.  $\square$

*Démonstration de la proposition.*

$$\forall x \in \mathbb{R}, P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x).$$

$\square$

Cette proposition permet d'obtenir les formules analytiques utiles pour simuler un grand nombre de lois, nous introduisons quelques exemples illustratifs :

## Exemples

1. La loi exponentielle : soit  $X$  une variable aléatoire réelle qui suit une loi exponentielle de paramètre  $\lambda$ . La fonction réciproque de sa fonction de répartition vaut pour  $u \in [0, 1]$ ,

$$F^{-1}(u) = \frac{-\ln(1-u)}{\lambda},$$

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

$$\frac{-\ln(1-U)}{\lambda},$$

ou encore

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

Le code R suivant donne 100 observations de la loi exponentielle de paramètre 1

```

N = 100
U = runif(N)
X = -log(1 - U)

```

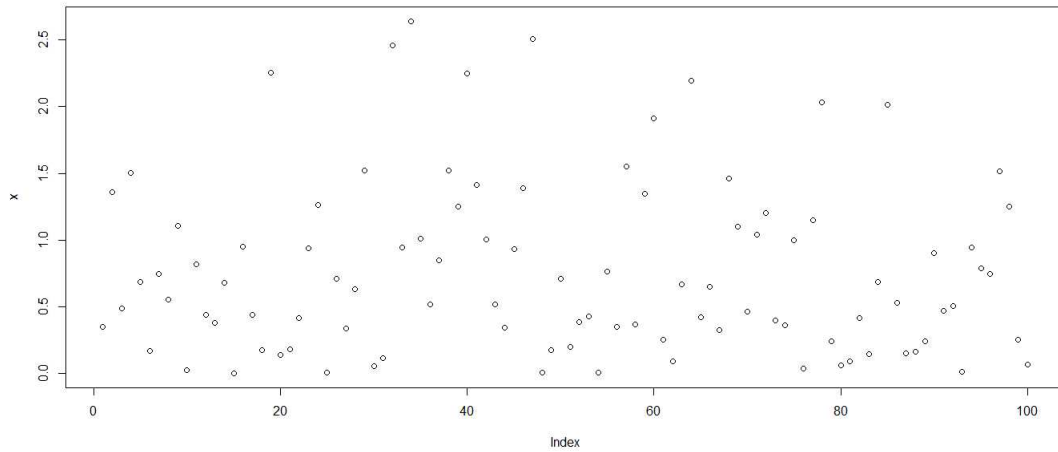


FIGURE 1 – 100 observations de la loi exponentielle de paramètre 1.

2. La Loi de Cauchy : soit  $X$  une variable aléatoire réelle de loi de Cauchy de paramètre  $a$ , alors la variable

$$F^{-1}(U) = a \tan(\pi(U - \frac{1}{2})),$$

suit une loi de Cauchy de paramètre  $a$ .

Nous avons proposé le programme suivant qui permet de simuler 100 observations de la loi de Cauchy de paramètre 1 :

```

N = 100
U = runif(100)
X = numeric(N)
for(i in 1 : N)
  {X[i] = tan(pi * (U[i] - 1/2))}
plot(X)

```

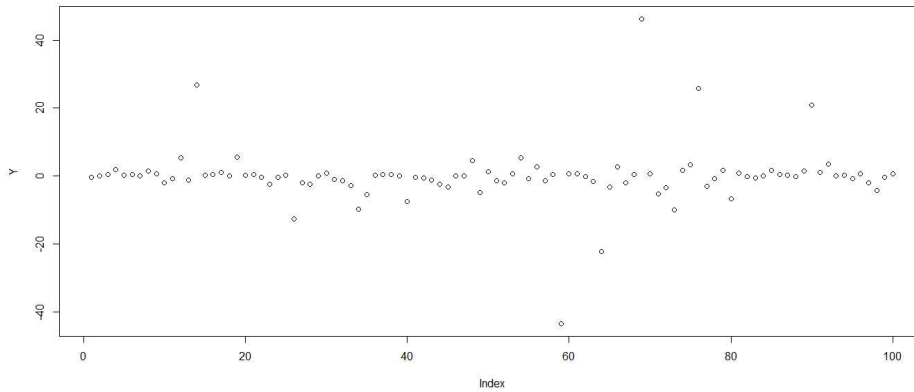


FIGURE 2 – 100 observations de la loi de Cauchy de paramètre 1.

On peut aussi simuler les variables discrètes par la méthode d'inversion.

### Lois discrètes

Soit  $X$  une variable aléatoire réelle discrète de loi  $p_i = P(X = x_i)$ , nous supposons que les éventualités sont rangé par ordre croissant

$$x_i < x_{i+1}, i \in \{1 \dots n\}, n \in \mathbb{N},$$

considérons La fonction de répartition de  $X$  est définie par :

$$F(x) = \begin{cases} 0 & \text{si } x < x_1 \\ p_1 + \dots p_i = F_i & \text{si } x_i \leq x < x_{i+1} \end{cases}$$

Nous pouvons définir l'algorithme de simulation par la méthode d'inversion comme suit :

```

i ← 1
choix ← Random
Tant que (choix > Fi) faire
i ← i + 1
fin Tant que
X ← xi.

```

— Exemple : simulation de la loi de Poisson



Si  $X$  suit une loi de Poisson de paramètre  $\lambda$ , il n'y a pas d'expression simple pour la fonction de répartition et l'ensemble des valeurs possibles est infini. Il faut donc calculer les valeurs  $F_i$  au fur et à mesure.

L'algorithme relatif à une telle simulation est donc :

```
X ← 0
P ← exp -λ
F ← P
choix ← Random
Tant que (choix > F) faire
X ← X + 1
P ← Pλ/X
F ← F + P
fin Tant que
X ← X - 1.
```

Remarquons que la méthode d'inversion n'est exacte qu'à condition de connaître l'expression explicite de  $F^{-1}$ . C'est rarement la plus efficace pour les variables à densité. Une seconde méthode de simulation est introduite dans le paragraphe suivant.

### 2.3 Méthode de rejet

La méthode de rejet est utilisée pour engendrer indirectement une variable aléatoire  $X$ , de densité de probabilité  $f$ , lorsqu'on ne peut pas utiliser la simulation par inversion. Le plus simple est de l'introduire en premier pour une loi uniforme.

#### Méthode de rejet pour la loi uniforme

On suppose que l'on sait simuler une loi uniforme sur  $D$ , et que l'on veut simuler une loi uniforme sur  $C \subset D$ . Alors l'algorithme de simulation par la méthode de rejet est le suivant :

```
répéter
tirer un point  $X$  au hasard dans  $D$ 
jusqu'à ( $X \in C$ ).
```

Pour le choix du support  $D$ , on a intérêt à le choisir le plus proche possible de  $C$ , mais suffisamment simple pour ne pas ralentir la simulation. Citons comme exemple la simulation de la loi uniforme sur le disque unité :  $C =$  le disque unité.

On choisit  $D =$  le carré  $[-1, 1]^2$ .

L'algorithme est :

répéter

$X \leftarrow 2 \text{ Random} - 1$

$Y \leftarrow 2 \text{ Random} - 1$

$S \leftarrow X^2 + Y^2$

jusqu'à ( $S \leq 1$ ).

La simulation des lois uniformes revient souvent dans les applications. En ce sens que le résultat suivant montre que la simulation d'une loi de densité quelconque peut toujours se ramener à la simulation d'une loi uniforme.

**Proposition 2.2.** *Soit  $f$  une densité de probabilité par rapport à la mesure de Lebesgue sur  $\mathbb{R}^d$ , continue par morceaux, et*

$$D = \{(x, u) \in \mathbb{R}^d \times \mathbb{R}^+; 0 < u < f(x)\}.$$

*Soit  $X$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^d$  et  $U$  une variable aléatoire réelle. le couple  $(X, U)$  suit alors la loi uniforme sur  $D$  si et seulement si :*

1.  $X$  a pour densité  $f$ .
2. La loi conditionnelle de  $U$  sachant " $X = x$ " est la loi uniforme sur  $[0, f(x)]$ .

*Démonstration.* Pour l'implication direct ,  $(X, U)$  suit une loi uniforme sur  $D$ , alors sa fonction de densité est :

$$f_{(X,U)}(x, u) = \frac{1}{\text{vol}(D)} 1_D(x, u),$$

$\text{vol}(D) = 1$  car  $f$  est une densité.

Pour montrer le point 1 de la proposition, on intègre  $f_{(X,U)}$  par rapport à  $u$

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f_{(X,U)}(x, u) du = \int_{\mathbb{R}} 1_D(x, u) du \\ &= \int_0^{f(x)} du = f(x), \end{aligned}$$

pour le point 2 :

$$\begin{aligned} f_U^{X=x}(u) &= \frac{f_{(X,U)}(x,u)}{f_X(x)} \\ &= \frac{1}{f(x)} 1_D(x,u) \\ &= \frac{1}{f(x)} 1_{[0,f(x)]}(u). \end{aligned}$$

Et pour l'implication inverse, remarquons que la loi du couple  $(X, U)$  est le produit de la loi de  $X$  avec la loi conditionnelle de  $U$  sachant " $X = x$ "

$$\begin{aligned} f_{(X,U)}(x,u) &= f_X(x) f_U^{X=x}(u) \\ &= f(x) \frac{1_{[0,f(x)]}(u)}{f(x)} \\ &= 1_{[0,f(x)]}(u). \end{aligned}$$

□

En d'autres termes, une variable aléatoire de densité  $f$  est l'abscisse d'un point au hasard sous le graphe de  $f$ . Ce résultat établi pour des loi à densités par rapport à la mesure de Lebesgue, s'étend naturellement à des mesures quelconques.

Dans le paragraphe suivant, la méthode de rejet est appliquée pour la simulation de loi à densité.

### Méthode de rejet pour les lois à densité

Nous allons voir une proposition qui permet à partir d'une densité  $g$  facile à simuler, simuler une loi de densité  $f$ .

**Proposition 2.3.** *Soit  $\mu$  une mesure positive sur  $\mathbb{R}^d$ ,  $f$  et  $g$  deux densités de probabilité sur  $(\mathbb{R}^d, \mu)$  telles qu'il existe une constante  $c$  vérifiant :*

$$\forall x \in \mathbb{R}^d, cg(x) \geq f(x).$$

*Soit  $X$  une variable aléatoire de densité  $g$  (par rapport à  $\mu$ ) et  $U$  une variable aléatoire de loi uniforme sur  $[0, 1]$ , indépendante de  $X$ . Alors la loi conditionnelle de  $X$  sachant l'événement  $\{cUg(X) < f(X)\}$ , a pour densité  $f$  par rapport à  $\mu$ .*

*Démonstration.* On cherche la loi conditionnelle de  $X$  sachant l'évènement

$$A = \{cUg(X) < f(X)\},$$

pour cela calculons d'abord la probabilité de  $A$ .

Remarquons que  $c \geq 1$  car

$$c \int_{\mathbb{R}^d} g(x) d\mu(x) \geq \int_{\mathbb{R}^d} f(x) d\mu(x) = 1,$$

donc  $c \neq 0$  et  $\frac{f(x)}{cg(x)} \leq 1$ .

D'autre part, le support de  $f$  est inclus dans le support de  $g$ , alors

$$\begin{aligned} P(A) &= P((X, U) \in \{(x, u), cug(x) < f(x)\}) \\ &= \int_{\{(x, u), cug(x) < f(x)\}} g(x) 1_{[0,1]}(u) du d\mu(x) \\ &= \int_{\{x, g(x) > 0\}} g(x) \left( \int_0^{\frac{f(x)}{cg(x)}} du \right) d\mu(x) \\ &= \int_{\{x, g(x) > 0\}} g(x) \frac{f(x)}{cg(x)} d\mu(x) = \frac{1}{c}. \end{aligned}$$

On veut montrer que la densité conditionnelle de  $X$  sachant  $A$  est  $f$ , c'est-à-dire que pour tout borélien  $B$  de  $\mathbb{R}^d$  :

$$P(X \in B/A) = \int_B f(x) d\mu(x)$$

$$\begin{aligned} P(X \in B/A) &= c P(\{X \in B\} \cap A) \\ &= c \int_{\{x, g(x) > 0\} \cap B} g(x) \left( \int_0^{\frac{f(x)}{cg(x)}} du \right) d\mu(x) \end{aligned}$$

par le calcul précédent, nous avons

$$\begin{aligned} P(X \in B/A) &= \int_{\{x, g(x) > 0\} \cap B} f(x) d\mu(x) = \int_{\mathbb{R}^d \cap B} f(x) d\mu(x) \\ &= \int_B f(x) d\mu(x). \end{aligned}$$

□

L'algorithme de simulation de la méthode de rejet est alors :  
répéter  
simuler  $X$  de densité  $g$   
 $U \leftarrow \text{Random}$   
jusqu'à  $(cUg(X) < f(X))$ .

On remarque que l'algorithme comporte une boucle dont la condition porte sur des variables aléatoires. le nombre d'itérations, noté  $N$  est donc lui-même aléatoire. On peut montrer que  $N$  suit la loi géométrique de paramètre  $\frac{1}{c}$ , car :

$$P(N = k) = (1 - p)^{k-1}p,$$

d'après la preuve de la proposition précédente nous avons :

$$p = P(cUg(X) < f(X)) = \frac{1}{c}.$$

L'espérance de  $N$  c'est-à-dire le nombre moyen d'itération à effectuer avant d'obtenir une réalisation de la densité  $f$  vaut  $c$ .

$$E(N) = c.$$

On a donc tout intérêt à choisir  $c$  le plus petit possible. Comme exemple de la simulation par la méthode de rejet, citons la simulation de la densité gaussienne centrée réduite à partir de la densité de Laplace.

Tout d'abord pour la simulation de la densité de Laplace, nous pouvons utiliser la méthode d'inversion :

La densité de Laplace est :

$$g(x) = \frac{e^{-|x|}}{2}, x \in \mathbb{R},$$

et sa fonction de répartition :

$$G(x) = \begin{cases} \frac{e^x}{2} & \text{si } x < 0 \\ 1 - \frac{e^{-x}}{2} & \text{si } x \geq 0 \end{cases}$$

son inverse est défini comme suit :

$$G^{-1}(u) = \begin{cases} \ln(2u) & \text{si } 0 < u < \frac{1}{2} \\ -\ln(2(1-u)) & \text{si } u \geq \frac{1}{2} \end{cases}$$

L'algorithme de simulation de  $g$  est le suivant :

$U \leftarrow \text{Random}$

si  $U < \frac{1}{2}$   
 $X \leftarrow \ln(2U)$   
 sinon  
 $X \leftarrow -\ln(2(1-U))$ .

Soit  $f$  la densité gaussienne, nous avons :

$$f(x) \leq \sqrt{\frac{2e}{\pi}} g(x), \forall x \in \mathbb{R}.$$

En effet :

$$e^{-\frac{1}{2}x^2+|x|} = e^{\frac{1}{2}} e^{-\frac{1}{2}(|x|-1)^2} \leq e^{\frac{1}{2}},$$

$$\implies \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \leq \sqrt{\frac{2e}{\pi}} \frac{e^{-|x|}}{2},$$

alors il existe  $c = \sqrt{\frac{2e}{\pi}}$  tel que :

$$f(x) \leq c g(x),$$

nous pouvons alors appliquer l'algorithme de simulation par rejet.

## 2.4 Simulation des lois normales

Pour la simulation de lois gaussiennes, nous rappelons l'algorithme de Box-Muller consiste à engendrer des couples de variables indépendantes en simulant leurs coordonnées polaires. Le principe de la base est établi dans la proposition suivante :

**Proposition 2.4.** *Soit  $(X, Y)$  un couple de variables aléatoires indépendantes, de même loi  $\mathcal{N}(0, 1)$ , et  $(R, \Theta)$  le couple de coordonnées polaires correspondant :*

$$X = R \cos \Theta; \quad Y = R \sin \Theta.$$

*Les variables aléatoires  $R$  et  $\Theta$  sont indépendantes, le module  $R$  a pour densité :*

$$f_R(r) = r \exp(-r^2/2) 1_{\mathbb{R}^+}(r).$$

*et l'argument  $\Theta$  suit la loi uniforme sur  $[0, 2\pi]$ .*

*Démonstration.* Posons :

$$\phi : \mathbb{R}^2 \longrightarrow \mathbb{R}^+ \times [0, 2\pi]$$

$$(x, y) \longrightarrow (r, \theta)$$

$r$  et  $\theta$  les coordonnées polaires associées à  $(x, y)$ , alors  $\phi$  est bijective.

$$\forall (r, \theta) \in \mathbb{R}^+ \times [0, 2\pi], \phi^{-1}(r, \theta) = (r \cos \theta, r \sin \theta),$$

$$Jac(\phi^{-1}) = r.$$

Pour le calcul de la densité de  $(R, \Theta)$ , considérons une fonction  $h : \mathbb{R}^2 \longrightarrow \mathbb{R}$  mesurable bornée ou positive

$$\begin{aligned} E(h(R, \Theta)) &= E(h(\phi(X, Y))) \\ &= \int_{\mathbb{R}^2} h(\phi(x, y)) f_{(X, Y)}(x, y) dx dy \\ &= \int_{\mathbb{R}^2} h(\phi(x, y)) f_X(x) f_Y(y) dx dy \text{ de l'indépendance de } X \text{ et } Y \\ &= \int_{\mathbb{R}^2} h(\phi(x, y)) \frac{1}{2\pi} \exp(-\frac{1}{2}(x^2 + y^2)) dx dy. \end{aligned}$$

Il en découle, par changement de variable

$$E(h(R, \Theta)) = \int_0^\infty \int_0^{2\pi} h(r, \theta) \frac{1}{2\pi} r \exp(-\frac{1}{2}r^2) d\theta dr,$$

donc

$$f_{(R, \Theta)}(r, \theta) = \frac{1}{2\pi} r \exp(-\frac{1}{2}r^2) 1_{\mathbb{R}^+}(r) 1_{[0, 2\pi]}.$$

Alors on déduit que  $R$  et  $\Theta$  sont indépendantes, De plus  $f_R(r) = r \exp(-r^2/2) 1_{\mathbb{R}^+}(r)$ , et  $\Theta$  suit la loi uniforme sur  $[0, 2\pi]$ .  $\square$

L'algorithme de simulation peut alors être formulé :

$$R \longleftarrow \sqrt{(-2 \ln(Random))}$$

$$\Theta \longleftarrow 2\pi Random$$

$$X \longleftarrow R \cos \Theta$$

$$Y \longleftarrow R \sin \Theta.$$

La formule de  $R$  dans l'algorithme précédent est justifié par :  $R$  a pour densité  $f_R$  d'après la proposition 2.4, alors  $R^2$  suit une loi exponentielle de paramètre  $\frac{1}{2}$ , en effet :

Soit  $h : \mathbb{R} \rightarrow \mathbb{R}$  mesurable bornée ou positive, posons  $Z = R^2$

$$\begin{aligned} E(h(Z)) &= E(h(R^2)) \\ &= \int_{\mathbb{R}} h(r^2) f_R(r) dr \\ &= \int_{\mathbb{R}} rh(r^2) e^{-\frac{r^2}{2}} 1_{\mathbb{R}^+}(r) dr, \end{aligned}$$

et le changement de variable  $z = r^2$ , on obtient :

$$E(h(Z)) = \int_{\mathbb{R}} h(z) \frac{e^{-\frac{z}{2}}}{2} 1_{\mathbb{R}^+}(z) dz.$$

Le carré de  $R$  est simulé par la méthode d'inversion par  $-2 \ln(\text{Random})$ .  
Le code R suivant permet de générer par exemple 100 observations d'une variable aléatoire gaussienne centrée réduite en utilisant la méthode de Box-Muller.

Nous proposons l'algorithme suivant :

```
N = 50
X <- numeric(N)
Y <- numeric(N)
for(i in 1 : N)
{U <- runif(1)
V <- runif(1)
X[i] = sqrt(-2 * log(V)) * cos(2 * pi * U)
Y[i] = sqrt(-2 * log(V)) * sin(2 * pi * U)}
plot(X,Y)
```



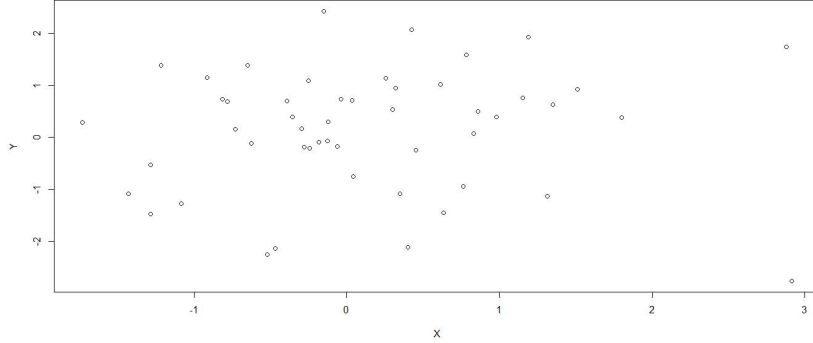


FIGURE 3 – 50 observations du couple de variables indépendantes  $\mathcal{N}(0, 1)$ .

### 3 Calcul d'une intégrale par la méthode de Monte-Carlo.

Les méthodes de Monte-Carlo sont particulièrement utilisées pour approximer les intégrales (en particulier, pour calculer des surfaces, des volumes, etc).

Supposons que l'on veuille estimer une quantité :

$$I = \int_{\mathbb{R}^d} g(x_1, x_2, \dots, x_d) f(x_1, x_2, \dots, x_d) dx_1 dx_2 \dots dx_d,$$

où  $f$  est une densité.

L'idée de la méthode est de présenter  $I$  sous forme d'une espérance

$$I = E(g(X)),$$

où  $X$  variable dans  $\mathbb{R}^d$  ayant pour densité  $f$ .

La justification théorique de la méthode est la loi forte des grands nombres qui permet de ne faire appel qu'à une réalisation d'un échantillon, c'est à dire à la suite  $(X_n(\omega))_n$  pour un seul  $\omega$ .

La vitesse de convergence est bien sûr un problème crucial pour maîtriser l'erreur commise en approxinant la valeur souhaitée  $E(X)$  par

$$\bar{X}_n(\omega) = \frac{S_n(\omega)}{n},$$

que l'on peut simuler. le théorème de la limite centrale précise la vitesse de convergence.

**Théorème 3.1.** : *limite centrale*

Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires réelles, indépendantes et de même loi avec  $E(X_1^2) < \infty$ , alors

$$\frac{S_n - nE(X_1)}{\sqrt{n}\sigma_{X_1}}$$

converge en loi vers une variable  $\mathcal{N}(0, 1)$ .

*Démonstration.* Supposons au départ que  $E(X_1) = 0$  et  $Var(X_1) = 1$ ,

$$\frac{S_n - nE(X_1)}{\sqrt{n}\sigma_{X_1}} = \frac{S_n}{\sqrt{n}},$$

a comme fonction caractéristique :

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = E(e^{it\frac{S_n}{\sqrt{n}}}) = (E(e^{it\frac{X_1}{\sqrt{n}}}))^n = (\varphi_{X_1}(\frac{t}{\sqrt{n}}))^n.$$

$X_1$  à carré intégrable, nous pouvons donc dériver deux fois sa fonction caractéristique

$$\varphi'_{X_1}(0) = 0, \quad \varphi''_{X_1}(0) = -1$$

pour tout  $u$  au voisinage de 0, un développement limité de la fonction caractéristique donne

$$\varphi_{X_1}(u) = 1 - \frac{u^2}{2} + o(u^2).$$

Donc

$$\begin{aligned} \lim_{n \rightarrow \infty} \varphi_{\frac{S_n}{\sqrt{n}}}(t) &= \lim_{n \rightarrow \infty} (\varphi_{X_1}(\frac{t}{\sqrt{n}}))^n \\ &= \lim_{n \rightarrow \infty} (1 - \frac{u^2}{2n} + o(\frac{u^2}{n}))^n \\ &= e^{-\frac{t^2}{2}}, \end{aligned}$$

qui est la fonction caractéristique d'une loi  $\mathcal{N}(0, 1)$ . On en déduit alors la convergence en loi de  $\frac{S_n}{\sqrt{n}}$  vers la loi gaussienne centrée réduite.

Si  $E(X_1) \neq 0$  ou  $Var(X_1) \neq 1$ , posons pour tout  $i \in \mathbb{N}$ ,

$$Y_i = \frac{X_i - E(X_i)}{\sigma_{X_i}},$$

les  $Y_i$  sont des variables centrées réduites, on applique alors la première partie de la démonstration.  $\square$

Par le théorème de la limite centrale nous avons :  
 $\forall h : \mathbb{R} \rightarrow \mathbb{R}$  mesurable bornée avec  $P(N \in D) = 0$ , où  $N$  est une variable aléatoire réelle de loi  $\mathcal{N}(0, 1)$ , et  $D$  l'ensemble des point de discontinuité de  $h$ ,

$$\lim_{n \rightarrow \infty} E(h(\frac{\sqrt{n}}{\sigma}(\frac{1}{n} \sum_1^n g(X_k) - E(g(X_1)))))) = \int_{-\infty}^{+\infty} h(x) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx,$$

$$\sigma = \sigma_{g(X_1)}.$$

Alors pour tous réels  $a$  et  $b$  avec  $a < b$ , nous avons aussi

$$\lim_{n \rightarrow \infty} P(a \frac{\sigma}{\sqrt{n}} \leq \frac{1}{n} \sum_1^n g(X_k) - E(g(X_1)) \leq b \frac{\sigma}{\sqrt{n}}) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

La vitesse de convergence est en  $\frac{1}{\sqrt{n}}$ , ce qui n'est pas très rapide, mais c'est parfois la seule méthode accessible, car elle ne dépend pas de la régularité de la fonction intégrée de plus la vitesse ne change pas si nous sommes en grande dimension. L'erreur de la méthode de Monte-Carlo est souvent présentée soit en donnant leur écart-type, c'est à dire  $\frac{\sigma}{\sqrt{n}}$ , soit en donnant l'intervalle de confiance à 95%. De la table de la fonction de répartition de la loi gaussienne centrée réduite, nous avons :

$$P(|\frac{1}{n} \sum_1^n g(X_k) - E(g(X_1))| \leq 1.96 \frac{\sigma}{\sqrt{n}}) \simeq 0.95,$$

c'est à dire que nous avons un intervalle de confiance à 95%

$$[\frac{1}{n} \sum_1^n g(X_k) - 1.96 \frac{\sigma}{\sqrt{n}}, \frac{1}{n} \sum_1^n g(X_k) + 1.96 \frac{\sigma}{\sqrt{n}}].$$

Il faut se souvenir que la variance  $\sigma^2$  n'est pas plus connue que la solution cherchée  $E(g(X_1))$ . Nous pouvons alors faire une estimation de cette variance.

### Estimation de la variance

Soient  $X_1, X_2, \dots, X_n$  un échantillon de variables aléatoires réelles, indépendantes et de même loi de variance  $\sigma^2$  inconnue,

$$\sigma_n^2 = \frac{1}{n-1} \sum_1^n (g(X_i^2) - (\frac{1}{n} \sum_1^n g(X_k))^2),$$

et c'est un estimateur sans biais de  $\sigma^2$ .

L'inégalité de Chebychev donne une estimation de la probabilité que la moyenne empirique soit « loin » de l'espérance, mais c'est une estimation très grossière, et peut être nettement améliorée par l'inégalité de Hoeffding.

**Théorème 3.2. : Hoeffding**

Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires indépendantes. Supposons que pour tout  $i$ , il existe  $a_i, b_i \in \mathbb{R}$  tels que  $a_i \leq X_i \leq b_i$ . Alors pour tout  $c > 0$ ,

$$P(|\sum_1^n X_i - E(\sum_1^n X_i)| \geq c) \leq 2 \exp\left(\frac{-2c^2}{\sum_1^n (b_i - a_i)^2}\right).$$

*Démonstration.* Nous avons pour tout  $h > 0$  et  $x \in \mathbb{R}$ ,

$$1_{[0; \infty[}(x) \leq e^{hx},$$

remplaçons  $x$  par  $\sum_1^n (X_i - E(X_i)) - c$

$$1_{\sum_1^n (X_i - E(X_i)) - c \geq 0} \leq e^{h(\sum_1^n (X_i - E(X_i)) - c)},$$

prenons l'espérance

$$P(\sum_1^n (X_i - E(X_i)) \geq c) \leq e^{-hc} E(e^{h(\sum_1^n (X_i - E(X_i)))}),$$

par indépendance des  $X_i$  il en découle

$$P(\sum_1^n (X_i - E(X_i)) \geq c) \leq e^{-hc} \prod_1^n E(e^{h(X_i - E(X_i))}).$$

Posons alors

$$Y_i = X_i - E(X_i),$$

et montrons que

$$E(e^{hY_i}) \leq e^{\frac{h^2}{8}(b_i - a_i)^2}.$$

Pour tout  $i \geq 1$ ,

$$a_i - E(X_i) \leq Y_i \leq b_i - E(X_i),$$

que l'on peut réécrire sous la forme

$$Y_i = \alpha_i(a_i - E(X_i)) + (1 - \alpha_i)(b_i - E(X_i)),$$

avec

$$\alpha_i = \frac{Y_i - b_i + E(X_i)}{a_i - b_i} = \frac{b_i - X_i}{b_i - a_i}.$$

D'autre part, la fonction  $x \mapsto e^{hx}$  est convexe, alors

$$e^{hY_i} \leq \alpha_i e^{h(a_i - E(X_i))} + (1 - \alpha_i) e^{h(b_i - E(X_i))},$$

ce qui implique

$$\begin{aligned} E(e^{hY_i}) &\leq E(\alpha_i) e^{h(a_i - E(X_i))} + E(1 - \alpha_i) e^{h(b_i - E(X_i))}, \\ &= \frac{b_i - E(X_i)}{b_i - a_i} e^{h(a_i - E(X_i))} - \frac{a_i - E(X_i)}{b_i - a_i} e^{h(b_i - E(X_i))} \\ &= e^{-hE(X_i)} \left[ \frac{b_i - E(X_i)}{b_i - a_i} e^{ha_i} - \frac{a_i - E(X_i)}{b_i - a_i} e^{hb_i} \right]. \end{aligned}$$

Notons  $h_i = h(b_i - a_i)$ ,

$$\begin{aligned} E(e^{hY_i}) &\leq e^{-\frac{h_i}{b_i - a_i} E(X_i)} \left[ \frac{b_i - E(X_i)}{b_i - a_i} e^{\frac{h_i}{b_i - a_i} a_i} - \frac{a_i - E(X_i)}{b_i - a_i} e^{\frac{h_i}{b_i - a_i} b_i} \right] \\ &= e^{\frac{h_i}{b_i - a_i} (a_i - E(X_i))} \left[ \frac{b_i - E(X_i)}{b_i - a_i} - \frac{a_i - E(X_i)}{b_i - a_i} e^{h_i} \right] \\ &= e^{L_i(h_i)}, \end{aligned}$$

avec 
$$L_i(h_i) = \frac{h_i}{b_i - a_i} (a_i - E(X_i)) + \ln \left( \frac{b_i - E(X_i)}{b_i - a_i} - \frac{a_i - E(X_i)}{b_i - a_i} e^{h_i} \right).$$

Posons

$$p_i = \frac{E(X_i) - a_i}{b_i - a_i},$$

$$L_i(h_i) = -p_i h_i + \ln((1 - p_i) + p_i e^{h_i}).$$

Le développement de Taylor Lagrange d'ordre 1 de  $L_i$  au voisinage de 0 donne

$$\begin{aligned} L_i(u) &= -p_i u + \ln((1 - p_i) + p_i e^u), \\ L_i'(u) &= -p_i + \frac{p_i e^u}{1 - p_i + p_i e^u}, \\ L_i''(u) &= \frac{p_i(1 - p_i) e^u}{((1 - p_i) + p_i e^u)^2}. \end{aligned}$$

donc

$$\begin{aligned} L_i(h_i) &= L_i(0) + L_i'(0)h_i + \frac{L_i''(\xi_i)}{2}h_i^2, \quad \xi_i \in ]0, h_i[ \\ &= \frac{L_i''(\xi_i)}{2}h_i^2. \end{aligned}$$

Pour montrer

$$E(e^{hY_i}) \leq e^{\frac{h^2}{8}(b_i - a_i)^2},$$

il suffit de vérifier que

$$L_i''(\xi_i) \leq \frac{1}{4}.$$

La fonction  $L_i''$  possède un maximum au point  $\ln(\frac{1-p_i}{p_i})$ , en effet :

$$L_i^{(3)}(u) = \frac{p_i(1-p_i)e^u(1-p_i-p_ie^u)}{(1-p_i+p_ie^u)^3},$$

$L_i^{(3)}(u) = 0$  donne  $u = \ln(\frac{1-p_i}{p_i})$  qui le point critique unique de  $L_i''$ ,

$$\begin{aligned} L_i^{(4)}(u) &= \frac{p_i(1-p_i)e^u(1-p_i-p_ie^u) - p_i^2(1-p_i)e^{2u} - 3p_i^2(1-p_i)e^{2u}}{(1-p_i+p_ie^u)^6} \\ &= \frac{p_i(1-p_i)(1-p_i-5p_ie^u)}{(1-p_i+p_ie^u)^3}, \end{aligned}$$

$$L_i^{(4)}(\ln(\frac{1-p_i}{p_i})) = -2(1-p_i) < 0.$$

Ainsi

$$\forall u \in \mathbb{R}, \quad L_i''(u) \leq L_i''(\ln(\frac{1-p_i}{p_i})) = \frac{(1-p_i)^2}{4(1-p_i)^2} = \frac{1}{4}.$$

Nous avons donc

$$E(e^{hY_i}) \leq e^{L_i(h_i)} = e^{\frac{L_i''(\xi_i)}{2}h_i^2} \leq e^{\frac{1}{8}h_i^2} = e^{\frac{h^2}{8}(b_i - a_i)^2}.$$

Il en découle alors

$$\begin{aligned} P\left(\sum_1^n (X_i - E(X_i)) \geq c\right) &\leq e^{-hc} \prod_1^n e^{\frac{1}{8}h^2(b_i - a_i)^2} \\ &= e^{-hc + \frac{h^2}{8} \sum_1^n (b_i - a_i)^2}. \end{aligned}$$

Posons

$$g(h) = -hc + \frac{h^2}{8} \sum_1^n (b_i - a_i)^2,$$

$$g'(h) = -c + \frac{h}{4} \sum_1^n (b_i - a_i)^2,$$

$g'(h) = 0$  implique

$$h_0 = \frac{4c}{\sum_1^n (b_i - a_i)^2}$$

,

$$g''(h) = \frac{1}{4} \sum_1^n (b_i - a_i)^2 \geq 0,$$

donc  $h_0$  est un minimum de  $g$ .

Finalement,

$$\begin{aligned} P\left(\sum_1^n (X_i - E(X_i)) \geq c\right) &\leq e^{-h_0 c + \frac{h_0^2}{8} \sum_1^n (b_i - a_i)^2} \\ &= \exp\left(\frac{-2c^2}{\sum_1^n (b_i - a_i)^2}\right). \end{aligned}$$

Pour conclure le résultat du théorème, montrons que

$$P\left(\sum_1^n (X_i - E(X_i)) \leq -c\right) \leq \exp\left(\frac{-2c^2}{\sum_1^n (b_i - a_i)^2}\right),$$

pour cela, il suffit de poser  $Z_i = -X_i$ , puis appliquer le résultat précédent aux variables  $Z_i$ ,  $i \in \mathbb{N}$ .  $\square$

## 4 Méthode de Monte-Carlo par Chaînes de Markov.

L'objectif est toujours de calculer  $I = \int h(x)d\pi(x)$ , où  $\pi$  est une loi de probabilité. Dans la section précédente, on a estimé  $I$  par

$$\frac{1}{n} \sum_1^n h(X_k),$$

où  $X_1, \dots, X_n$  est un échantillon de variables aléatoires indépendantes de même loi  $\pi$ . le problème est dans la simulation d'un échantillon i.i.d de la loi  $\pi$ , car elle n'est pas toujours explicitable.

La méthode de Monte-Carlo par chaîne de Markov donne une estimation de  $I$  égale

$$\frac{1}{n} \sum_1^n h(X_k),$$

où  $X_1, \dots, X_n$  des états successifs d'une chaîne de Markov. Avant d'aller à la justification de la méthode, on passe par quelques rappels sur les chaînes de Markov.

Dans tout ce chapitre,  $E$  est un ensemble dénombrable,  $\mathcal{P}(E)$  est l'ensemble de ses parties.

### 4.1 La propriété de Markov

**Définition 4.1.** *On dit qu'une suite de variables aléatoires  $(X_n)_{n \in \mathbb{N}}$ , à valeurs dans  $(E, \mathcal{P}(E))$  et définie sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , est une chaîne de Markov si, pour tout  $(n+1)$  uplet  $(i_0, \dots, i_n)$  de points de  $E$  tel que*

$$P(\cap_{j=0}^{n-1} \{X_j = i_j\}) > 0,$$

$$P\{X_n = i_n / \cap_{j=0}^{n-1} \{X_j = i_j\}\} = P\{X_n = i_n / X_{n-1} = i_{n-1}\} \quad (1)$$

*Autrement dit, la loi de  $X_n$  conditionnellement à  $(X_0, \dots, X_{n-1})$  et la loi de  $X_n$  conditionnellement à  $X_{n-1}$  sont identiques.*

*On appelle  $E$  l'espace des états. La loi de  $X_0$  est appelé la loi ou la mesure initiale, et l'égalité (1) s'appelle propriété de Markov.*



**Définition 4.2.** On dit qu'une chaîne de Markov  $(X_n)_{n \in \mathbb{N}}$  est homogène si, pour tout couple  $(i, j)$  de points de  $E$ ,  $P(X_{n+1} = j / X_n = i)$  est indépendant de  $n$ ,  $n$  décrivant l'ensemble des entiers pour lesquels  $P(X_n = i) > 0$ .

**Définition 4.3.** Une matrice  $M = (M_{ij})_{i, j \in E}$  est une matrice stochastique si elle vérifie

- (i)  $M_{ij} \geq 0$  pour tout  $i, j \in E$
- (ii)  $\sum_{j \in E} M_{ij} = 1$  pour tout  $i \in E$

**Proposition 4.1.** Le produit de deux matrices stochastiques est encore une matrice stochastique.

*Démonstration.* Soient  $A = (a_{ij})_{i, j \in E}$  et  $B = (b_{ij})_{i, j \in E}$  deux matrices stochastiques, montrons que  $AB = (c_{ij})_{i, j \in E}$  est une matrice stochastique, pour tout  $i, j \in E$ ,

$$c_{ij} = \sum_{k \in E} a_{ik} b_{kj},$$

on a

- i-  $c_{ij} \geq 0$ ,
  - ii-  $\sum_{j \in E} c_{ij} = \sum_{j \in E} \sum_{k \in E} a_{ik} b_{kj} = \sum_{k \in E} (a_{ik} \sum_{j \in E} b_{kj}) = 1$ ,
- donc  $AB$  est une matrice stochastique. □

Soit  $(X_n)_{n \in \mathbb{N}}$  une chaîne de Markov homogène. On note alors  $p_{i,j}$  la valeur commune des  $P(X_{n+1} = j / X_n = i)$  et  $P = (p_{ij})_{i, j \in E}$ . La matrice  $P$  est appelée matrice de transition de la chaîne, et c'est une matrice stochastique.

Dans la suite,  $X = (X_n)_{n \geq 0}$  est une chaîne de Markov homogène, définie sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ , à valeurs dans  $(E, \mathcal{P}(E))$ , de matrice de transition  $P = (p_{ij})_{i, j \in E}$  et de loi initiale  $\pi_0$ .

## 4.2 Calcul des lois marginales

**Théorème 4.1.** :

- 1- Pour tout  $n \geq 1$  et tous  $i_0, \dots, i_n \in E$ ,

$$P(X_0 = i_0, \dots, X_n = i_n) = \pi_0(i_0) p_{i_0 i_1} \dots p_{i_{n-1} i_n}.$$

- 2- Pour tous  $A_0, \dots, A_{n-1}, B_1, \dots, B_m \in \mathcal{P}(E)$ ,  $i \in E$

$$\begin{aligned} P(X_{n+1} \in B_1, \dots, X_{n+m} \in B_m / X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = i) \\ = P(X_{n+1} \in B_1, \dots, X_{n+m} \in B_m / X_n = i) \\ = P(X_1 \in B_1, \dots, X_m \in B_m / X_0 = i). \end{aligned}$$

*Démonstration* . Le point 1 du théorème :

$$\begin{aligned}
& P(X_0 = i_0, \dots, X_n = i_n) \\
&= P(X_n = i_n / X_0 = i_0, \dots, X_{n-1} = i_{n-1}) P(X_0 = i_0, \dots, X_{n-1} = i_{n-1}) \\
&= P(X_n = i_n / X_{n-1} = i_{n-1}) P(X_0 = i_0, \dots, X_{n-1} = i_{n-1}) \text{ par la propriété de Markov} \\
&= p_{i_{n-1}i_n} P(X_0 = i_0, \dots, X_{n-1} = i_{n-1}) \\
&= p_{i_{n-1}i_n} P(X_{n-1} = i_{n-1} / X_0 = i_0, \dots, X_{n-2} = i_{n-2}) P(X_0 = i_0, \dots, X_{n-2} = i_{n-2}) \\
&= p_{i_{n-1}i_n} p_{i_{n-2}i_{n-1}} P(X_0 = i_0, \dots, X_{n-2} = i_{n-2}),
\end{aligned}$$

après plusieurs itérations, nous avons

$$\begin{aligned}
P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) &= p_{i_{n-1}i_n} \dots p_{i_0i_1} P(X_0 = i_0) \\
&= p_{i_{n-1}i_n} \dots p_{i_0i_1} \pi_0(i_0).
\end{aligned}$$

Le point 2 :

$$\begin{aligned}
& P(X_{n+1} \in B_1, \dots, X_{n+m} \in B_m / X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = i) \\
&= \frac{\sum_{i_0 \in A_0} \dots \sum_{i_{n-1} \in A_{n-1}} \sum_{j_1 \in B_1} \dots \sum_{j_m \in B_m} p_{i_0i_1} \dots p_{i_{n-1}i} p_{ij_1} \dots p_{j_{m-1}j_m}}{\sum_{i_0 \in A_0} \dots \sum_{i_{n-1} \in A_{n-1}} p_{i_0i_1} \dots p_{i_{n-1}i}} \\
&= \sum_{j_1 \in B_1} \dots \sum_{j_m \in B_m} p_{ij_1} \dots p_{j_{m-1}j_m} \\
&= P(X_1 \in B_1, \dots, X_m \in B_m / X_0 = i).
\end{aligned}$$

La matrice  $P^n = \underbrace{P \times \dots \times P}_{n \text{ fois}}$  est une matrice stochastique, d'après la proposition 4.1. On note  $p_{ij}^n$  ses éléments. □

**Corollaire 4.1.** *Pour tous entiers naturels  $n, m$  et tous états  $i, j \in E$ ,*

*i* –  $P(X_n = j) = \sum_{k \in E} \pi_0(k) p_{kj}^n,$

*ii* –  $P(X_{n+m} = j / X_m = i) = p_{ij}^n,$

*iii* –  $P(X_{n+m} = j / X_0 = i) = \sum_{k \in E} P(X_m = j / X_0 = k) P(X_n = k / X_0 = i).$

*L'égalité (iii) est appelée équation de Chapman-Kolmogorov.*

*Démonstration.* i- Elle se fait par récurrence, pour  $n = 1$

$$\begin{aligned}
 P(X_1 = j) &= P(X_1 = j, \Omega) \\
 &= \sum_{k \in E} P(X_1 = j, X_0 = k) \\
 &= \sum_{k \in E} P(X_1 = j / X_0 = k) P(X_0 = k) \\
 &= \sum_{k \in E} \pi_0(k) p_{kj}.
 \end{aligned}$$

On suppose qu'elle est vraie pour  $n$  fixé, et on montre qu'elle est vraie pour  $n + 1$ ,

$$\begin{aligned}
 P(X_{n+1} = j) &= P(X_{n+1} = j, \Omega) \\
 &= \sum_{i \in E} P(X_{n+1} = j, X_n = i) \\
 &= \sum_{i \in E} P(X_{n+1} = j / X_n = i) P(X_n = i) \\
 &= \sum_{i \in E} p_{ij} \sum_{k \in E} \pi_0(k) p_{ki}^n \\
 &= \sum_{k \in E} \pi_0(k) p_{kj}^{n+1}.
 \end{aligned}$$

ii- Elle se fait aussi par récurrence, pour  $n = 1$

$$P(X_{m+1} = j / X_m = i) = p_{ij}.$$

Supposons la vraie au rang  $n$ , et donc

$$\begin{aligned}
 P(X_{m+n+1} = j / X_m = i) &= \sum_{k \in E} P(X_{m+n+1} = j, X_{n+m} = k / X_m = i) \\
 &= \sum_{k \in E} P(X_{m+n+1} = j / X_{n+m} = k, X_m = i) P(X_{n+m} = k / X_m = i) \\
 &= \sum_{k \in E} p_{kj} p_{ik}^n = p_{ij}^{n+1}.
 \end{aligned}$$

$$\begin{aligned}
iii - P(X_{n+m} = j/X_0 = i) &= \sum_{k \in E} P(X_{n+m} = j, X_n = k/X_0 = i) \\
&= \sum_{k \in E} P(X_{n+m} = j/X_n = k)P(X_n = k/X_0 = i) \\
&= \sum_{k \in E} P(X_m = j/X_0 = k)P(X_n = k/X_0 = i).
\end{aligned}$$

□

**Remarque 4.1.** *La donnée de la matrice de transition et de la loi initiale suffit à caractériser la loi d'une chaîne de Markov homogène jusqu'à tout instant fixé.*

### 4.3 Récurrence et transience

**Notation 4.1.** :

Soit  $A \in \mathcal{P}(E)$ ,

$$\begin{aligned}
T_A &= \inf\{n > 0, X_n \in A\}. \\
T_A^{(k)} &= \inf\{n > T_A^{(k-1)}, X_n \in A\}.
\end{aligned}$$

Soient  $i, j \in E$

$$\rho_{ij} = P(T_j < \infty/X_0 = i) = P_i(T_j < \infty).$$

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

**Définition 4.4.** *Soient  $i$  et  $j$  deux éléments de  $\mathbb{E}$ . On dit que  $i$  conduit à  $j$ , noté  $i \rightsquigarrow j$ , si  $\rho_{ij} > 0$ ; on dit que  $i$  et  $j$  communiquent, noté  $i \longleftrightarrow j$ , si  $i$  conduit à  $j$  et  $j$  conduit à  $i$ .*

**Définition 4.5.** *Un point  $i \in E$  est dit récurrent pour la chaîne de Markov si  $\rho_{ii} = 1$ . Il est dit transient dans le cas contraire.*

**Définition 4.6.** *Un ensemble  $C \subset E$  est dit clos si :*

$$\rho_{ij} = 0 \quad \forall i \in C, j \notin C.$$

**Définition 4.7.** *Un ensemble  $C \subset E$  est dit irréductible si :  $i \rightsquigarrow j$  ou  $j \rightsquigarrow i$ ;  $\forall i, j \in C$ .*

**Remarque 4.2. :**

1-  $\forall i, j \in E$

$$\rho_{ij} > 0 \iff \exists n \geq 1, p_{ij}^n > 0.$$

En effet :

$$\begin{aligned} \rho_{ij} > 0 &\iff P_i(T_j < \infty) > 0 \\ &\iff P_i(\cup_{n \geq 1} \{X_n = j\}) > 0 \\ &\iff \exists n \geq 1, P_i(X_n = j) > 0 \\ &\iff \exists n \geq 1, p_{ij}^n > 0. \quad \text{par le point 2 du corollaire 4.1} \end{aligned}$$

2- Si  $E$  est irréductible, la chaîne est dite irréductible.

3- Si tous les états sont récurrents la chaîne est dite récurrente. Elle est dite transiente si les états sont transients.

4- La relation " $\rightsquigarrow$ " est une relation d'équivalence sur l'ensemble des états récurrents, en effet :

Soient  $i, j, k$  des états récurrents. Commençons par " $\rightsquigarrow$ " réflexive, on a

$$\rho_{ii} = 1,$$

implique

$$i \rightsquigarrow i.$$

Pour " $\rightsquigarrow$ " symétrique, montrons que si  $i \rightsquigarrow j$  alors  $j \rightsquigarrow i$

$$i \rightsquigarrow j \iff \exists n \geq 1, p_{ij}^n > 0,$$

on a

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

car l'ensemble des trajectoires qui passent de l'états  $i$  à l'états  $j$  en  $n$  étapes, puis qui n'atteignent jamais  $i$  est inclus dans l'ensemble des trajectoires qui partant de  $i$  n'atteignent jamais  $i$ , ainsi

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

implique

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

alors  $j \rightsquigarrow i$ .

Il reste à montrer " $\rightsquigarrow$ " est transitive, on a :

$i \rightsquigarrow j$  et  $j \rightsquigarrow k$ , montrons que  $i \rightsquigarrow k$

$$i \rightsquigarrow j \iff \exists n \geq 1, p_{ij}^n > 0,$$

$$j \rightsquigarrow k \iff \exists m \geq 1, p_{jk}^m > 0,$$

$$p_{ik}^{n+m} \geq p_{ij}^n p_{jk}^m > 0,$$

donc  $i \rightarrow k$ .

**Définition 4.8.** Soit  $i$  un état récurrent, on note  $E_i(T_i) = E(T_i/X_0 = i)$ ,

*i*–  $i$  est dit récurrent positif si  $E_i(T_i) < \infty$ .

*ii*–  $i$  est dit récurrent nul si  $E_i(T_i) = +\infty$ .

#### 4.4 Mesure invariante

**Définition 4.9.** 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 prendra 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 4.2.** On note  $E_R$  l'ensemble des états récurrent,  $E_T$  l'ensemble des états transitoires.

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

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

**Théorème 4.2.** Soit  $j \in E_R$ , on a

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

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 4.3.** 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 d'abord que  $(X_n)_{n \geq 0}$  admet une loi invariante  $\pi$ , et montrons que

$$\pi(i) = \frac{1}{m_{ii}}, \quad \forall i \in E$$

$\pi$  loi invariante, elle vérifie :

$$\pi = \pi P,$$

c'est-à-dire pour tout  $i \in E$ ,

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

ainsi

$$1 = \sum_{i \in E} \pi(i) = \sum_{i \in E} \sum_{j \in E} \pi(j) p_{ji},$$

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

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n p_{ji}^m &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P(X_m = i / X_0 = j) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P_j(X_m = i) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n E_j(1_{X_m=i}) \\ &= \lim_{n \rightarrow \infty} E_j \left( \frac{N_n(i)}{n} \right) \\ &= \frac{\rho_{ji}}{m_{ii}}, \text{ d'après le théorème 4.2.} \end{aligned}$$

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

$$\forall i, j \in E, i \rightsquigarrow j \text{ ou } j \rightsquigarrow i,$$

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

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

$$i \rightsquigarrow j \iff \exists n \geq 1, p_{ij}^n > 0,$$

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

$$\sum_{i \in E_1} \frac{1}{m_{ii}} \leq 1,$$

comme  $E_1$  est arbitraire, on a

$$\sum_{i \in E} \frac{1}{m_{ii}} \leq 1. \quad (2)$$

D'autre part,

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

quand  $n \rightarrow \infty$ , on a

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

comme  $E_1$  est arbitraire

$$\sum_{i \in E} \frac{1}{m_{ii}} p_{ij} \leq \frac{1}{m_{jj}}, \quad (3)$$

en sommant 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 des égalités, 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.  $\square$

**Définition 4.10.** 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.** Si  $P$  est réversible par rapport à  $\pi$  alors  $\pi$  est une probabilité invariante.

*Démonstration.* Pour montrer que  $\pi$  est une probabilité invariante, il suffit de montrer que :

$$\forall i \in E, \quad \pi(\{i\}) = \sum_{j \in E} \pi(\{j\})p_{ji}.$$

$$\begin{aligned} \sum_{j \in E} \pi(\{j\})p_{ji} &= \sum_{j \in E} \pi(\{i\})p_{ij} \quad \text{car } \pi \text{ est réversible} \\ &= \sum_{j \in E} \pi(\{i\})P(X_1 = j / X_0 = i) \\ &= \pi(\{i\}). \end{aligned}$$

□

## 4.5 Période d'une chaîne de Markov

**Définition 4.11.** Soit  $i \in E$  tel que  $\rho_{ii} > 0$ , on pose

$$d_i = \text{pgcd}\{n \geq 1 / p_{ii}^n > 0\}$$

$d_i$  est appelé période de  $i$ .

**Théorème 4.4.** Soient  $i, j \in E$

Si  $i \leftrightarrow j$  alors  $d_i = d_j$ .

*Démonstration.* Soient  $i, j \in E$  tel que  $i \leftrightarrow j$ , on a  $\rho_{ij} > 0$  et  $\rho_{ji} > 0$ , alors

$$\exists k \geq 1, p_{ij}^k > 0,$$

et

$$\exists l \geq 1, p_{ji}^l > 0,$$

ainsi

$$p_{ii}^{k+l} \geq p_{ij}^k p_{ji}^l > 0,$$

alors  $d_i$  divise  $k + l$ .

Soit  $n \geq 1$  tel que  $p_{jj}^n > 0$

$$p_{ii}^{k+l+n} \geq p_{ij}^k p_{jj}^n p_{ji}^l > 0,$$

donc  $d_i$  divise  $k + l + n$ .

On a  $d_i$  divise  $k + l$  et divise  $k + l + n$ , donc divise  $n$ , ce qui implique que

$$d_i \leq d_j,$$

de même on montre que

$$d_j \leq d_i,$$

d'où

$$d_i = d_j.$$

□

**Conséquence 4.1.** *Si  $(X_n)_{n \geq 0}$  est une chaîne de Markov irréductible récurrente, alors*

$$\forall i, j \in E \quad d_i = d_j = d.$$

*$d$  est appelé la période de la chaîne, si  $d = 1$  la chaîne est dite apériodique.*

## 4.6 Méthode de Monte-Carlo

Soit  $E$  un espace d'état fini et  $\pi$  la loi invariante de matrice de transition  $P$  irréductible et apériodique. Supposons que l'on cherche à calculer  $\int h(x) d\pi(x)$ , où  $h$  est une fonction donnée. Une méthode naturelle dite de Monte Carlo par chaîne de Markov consiste à générer les états successifs  $X_0 \dots X_n$  d'une chaîne de Markov de matrice de transition  $P$  à partir d'un état initial  $X_0$  de loi quelconque. Dans le théorème suivant, on établit que

$$S_n(h) = \frac{1}{n} \sum_1^n h(X_k)$$

est une bonne approximation de  $\int h(x) d\pi(x)$ , et la loi de  $X_n$  est proche de  $\pi$ .

**Théorème 4.5.** *Soit  $(X_n)_{n \geq 0}$  une chaîne de Markov sur  $E$  irréductible récurrente positive,  $\pi$  la loi invariante de la chaîne alors pour toute fonction  $f$  définie sur  $E$  telle que  $f \geq 0$  ou  $\int f(x) d\pi(x) < \infty$ , on a :*

$$\frac{1}{n} \sum_1^n f(X_k) \longrightarrow \int f(x) d\pi(x) \text{ p.s}$$

## Algorithme de Metropolis

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  $E$ , appelée matrice de sélection, telle que pour tout couple  $(i, j) \in E$

$$Q_{ij} > 0 \implies Q_{ji} > 0$$

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

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

Par exemple

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

ou bien

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

Pour  $i \neq j$ , posons

$$R_{ij} = \begin{cases} h\left(\frac{\pi(j)Q(j,i)}{\pi(i)Q(i,j)}\right) & \text{si } Q(i, j) \neq 0 \\ 0 & \text{sinon} \end{cases}$$

On construit alors une matrice de transition  $P$  définie par

$$p_{ij} = Q_{ij}R_{ij}$$

pour  $i \neq j$ , et

$$p_{ii} = 1 - \sum_{i \neq j} p_{ij}$$

**Proposition 4.3.** *On suppose que  $\pi$  charge tous les points de  $E$ . La matrice  $P$  définie précédemment est réversible par rapport à  $\pi$ . Elle est irréductible si  $Q$  est irréductible. Si de plus  $h(u) < 1$ , elle est apériodique.*

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

Etape 0. Initialiser  $X_0$ .

Etape n+1.

(sélection) choisir  $y$  avec la loi  $Q(X_n, y)$ .

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

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

$X_{n+1} = y$ .

Sinon refuser la sélection

$X_{n+1} = X_n$ .

## 5 Conclusion

Comme conclusion à cette modeste étude, nous retenons que la méthode de Monte-Carlo est une méthode d'approximation par introduction de procédés aléatoires. Cela permet d'estimer des valeurs numériques et de caractériser des systèmes complexes. 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 s'agit de grandes dimensions.

## 6 Annexe

**Théorème 6.1.** *Loi forte des grands nombres*

Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires indépendantes, identiquement distribuées intégrables. Alors  $\frac{S_n}{n}$  converge presque sûrement vers  $E[X_1]$ , avec

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

*Démonstration.* La démonstration consiste à prouver dans un premier temps le résultat sous l'hypothèse plus forte  $E(|X_1|^4) < \infty$  et  $E(X_1) = 0$ . Dans ce cas, dont on peut se contenter en première lecture,  $P(|S_n|/n \geq \epsilon)$  peut être majoré en utilisant l'inégalité de Markov. La borne ainsi obtenue est le terme général d'une série convergente, ce qui permet de conclure grâce au lemme de Borel-Cantelli. Sous l'hypothèse plus faible du théorème, on approxime toute variable de  $L^1$  par des variables de  $L^4$ , puis on se ramène au cas traité. Commençons donc par montrer le résultat lorsque  $E(|X_1|^4) < \infty$  et  $E(X_1) = 0$ . Dans ce cas, l'inégalité de Markov montre que pour tout  $n \geq 1$  et tout  $\delta > 0$ ,

$$P(|S_n| \geq \delta n) \leq \frac{1}{\delta^4 n^4} E(S_n^4).$$

Observons que

$$\begin{aligned} S_n^4 &= \left( \sum_{i=1}^n X_i \right)^4 \\ &= \left( \sum_{i=1}^n X_i^2 + \sum_{i \neq j} X_i X_j \right)^2 \\ &= \left( \sum_{i=1}^n X_i^2 \right)^2 + 2 \sum_{i=1}^n X_i^2 \sum_{j \neq k} X_j X_k + \left( \sum_{i \neq j} X_i X_j \right)^2 \\ &= \sum_{i=1}^n X_i^4 + \sum_{i \neq j} X_i^2 X_j^2 + 2 \left( \sum_{1 \leq i, j, k \text{ distincts} \leq n} X_i X_j X_k^2 + 2 \sum_{1 \leq i \neq j \leq n} X_i^3 X_j \right) \\ &\quad + 2 \sum_{i \neq j} X_i^2 X_j^2 + 4 \sum_{1 \leq i, j, k \text{ distincts} \leq n} X_i X_j X_k^2 + \sum_{1 \leq i, j, k, l \text{ distincts} \leq n} X_i X_j X_k X_l \\ &= \sum_{1 \leq i \leq n} X_i^4 + 4 \sum_{1 \leq i \neq j \leq n} X_i^3 X_j + 3 \sum_{1 \leq i \neq j \leq n} X_i^2 X_j^2 + 6 \sum_{1 \leq i, j, k \text{ distincts} \leq n} X_i X_j X_k^2 \\ &\quad + \sum_{1 \leq i, j, k, l \text{ distincts} \leq n} X_i X_j X_k X_l. \end{aligned}$$

Donc, par linéarité de l'espérance, indépendance et centrage des  $X_i$ ,

$$\begin{aligned} E(S_n^4) &= \sum_{1 \leq i \leq n} E(X_i^4) + 4 \sum_{1 \leq i \neq j \leq n} E(X_i^3)E(X_j) + 3 \sum_{1 \leq i \neq j \leq n} E(X_i^2)E(X_j^2) \\ &\quad + 6 \sum_{1 \leq i, j, k \text{ distincts} \leq n} E(X_i)E(X_j)E(X_k^2) + \sum_{1 \leq i, j, k, l \text{ distincts} \leq n} E(X_i)E(X_j)E(X_k)E(X_l) \\ &= nE(X_1^4) + 3n(n-1)(E(X_1^2))^2. \end{aligned}$$

La série

$$\sum_{n \geq 1} \frac{[nE(X_1^4) + 3n(n-1)(E(X_1^2))^2]}{n^4 \delta^4} = \frac{E(X_1^4)}{\delta^4} \sum_{n \geq 1} \frac{1}{n^3} + 3 \frac{(E(X_1^2))^2}{\delta^4} \sum_{n \geq 1} \frac{n-1}{n^3}$$

est convergente, ce qui donne la convergence de la série  $\sum_{n \geq 1} P(|S_n| > n\delta)$ . D'après la proposition suivante nous avons la convergence de  $\frac{S_n}{n}$  vers 0 presque sûrement.

**Proposition 6.1.** Soient  $(X_n)_{n \in \mathbb{N}}$ ,  $X$  des variables aléatoires réelles.

Si  $\forall \epsilon > 0$

$$\sum_{n \geq 0} P(|X_n - X| > \epsilon) < \infty$$

alors  $X_n \rightarrow X$  p.s.

*Démonstration.* Posons :

$$A_n = \{|X_n - X| > \epsilon\}.$$

D'après Borel-Cantelli

$$P(\limsup_{n \rightarrow \infty} A_n) = 0,$$

i.e

$$P(\forall n \geq 0, \exists k \geq n, |X_k - X| < \epsilon) = 1,$$

ce qui implique que  $X_n \rightarrow X$  p.s. □

Supposons maintenant que les  $X_i$  sont intégrables et centrées, sans autre hypothèses. Soit  $\epsilon > 0$  fixé. Il existe, pour tout  $i \geq 1$ , des variables  $Y_i$  étagées, centrées, indépendantes et de même loi, telles que  $E(|X_i - Y_i|) \leq \epsilon$ .

Si  $T_n = \sum_{1 \leq i \leq n} Y_i$ , nous avons

$$\begin{aligned} \frac{1}{n}|S_n| &= \frac{1}{n} \left| \sum_1^n (X_i - Y_i) + \sum_1^n Y_i \right| \\ &\leq \frac{1}{n} \sum_{1 \leq i \leq n} |X_i - Y_i| + \frac{1}{n} |T_n|. \end{aligned}$$

Par le point précédent  $\frac{T_n}{n}$  converge p.s. vers 0, car les  $Y_i$  sont indépendantes, de même loi, centrées et étagées (étagées donc dans  $L^4$ ). Donc il suffit de montrer que :

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{1 \leq i \leq n} |X_i - Y_i|,$$

peut être rendu arbitrairement petit en prenant  $\epsilon$  arbitrairement petit. Posons  $Z_i = |X_i - Y_i|$ ,  $i \geq 1$ , soient  $k \in \mathbb{N}$  et  $\delta > 0$ ,

$$\begin{aligned} & P\left(\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{1 \leq i \leq n} Z_i \geq 2E(Z_1) + \delta\right) \\ &= P\left(\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{1 \leq i \leq n} Z_i \geq 2E(Z_1) + \delta, (\cup_{1 \leq i \leq 2^{k+1}} \{Z_i \leq 2^k\}) \cup (\cup_{1 \leq i \leq 2^{k+1}} \{Z_i > 2^k\})\right) \\ &\leq P\left(\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{1 \leq i \leq n} Z_i \geq 2E(Z_1) + \delta, \cup_{1 \leq i \leq 2^{k+1}} \{Z_i \leq 2^k\}\right) \\ &+ P\left(\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{1 \leq i \leq n} Z_i \geq 2E(Z_1) + \delta, \cup_{1 \leq i \leq 2^{k+1}} \{Z_i > 2^k\}\right) \\ &\leq P\left(\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{1 \leq i \leq n} Z_i \geq 2E(Z_1) + \delta, \cup_{1 \leq i \leq 2^{k+1}} \{Z_i \leq 2^k\}\right) + P(\cup_{1 \leq i \leq 2^{k+1}} \{Z_i > 2^k\}) \\ &= P\left(\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{1 \leq i \leq n} Z_i 1_{[0, 2^k]}(Z_i) \geq 2E(Z_1) + \delta\right) + P(\exists i \in \{1, 2, \dots, 2^{k+1}\} : Z_i > 2^k). \end{aligned}$$

Nous avons

$$\begin{aligned} \max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{1 \leq i \leq n} Z_i 1_{[0, 2^k]}(Z_i) &\leq \max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \max_{2^k < n \leq 2^{k+1}} \sum_{1 \leq i \leq n} Z_i 1_{[0, 2^k]}(Z_i) \\ &= \frac{1}{2^k} \sum_{1 \leq i \leq 2^{k+1}} Z_i 1_{[0, 2^k]}(Z_i), \end{aligned}$$

Ce qui implique

$$\left\{ \max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{1 \leq i \leq n} Z_i 1_{[0, 2^k]}(Z_i) \geq 2E(Z_1) + \delta \right\} \subset \left\{ \frac{1}{2^k} \sum_{1 \leq i \leq 2^{k+1}} Z_i 1_{[0, 2^k]}(Z_i) \geq 2E(Z_1) + \delta \right\}.$$

De plus on a

$$\left\{ \sum_1^{2^{k+1}} Z_i 1_{[0, 2^k]}(Z_i) \geq 2^{k+1} E(Z_1) + 2^k \delta \right\} \subset \left\{ \sum_1^{2^{k+1}} Z_i 1_{[0, 2^k]}(Z_i) \geq 2^{k+1} E(Z_1 1_{[0, 2^k]}(Z_1)) + 2^k \delta \right\}$$



Donc

$$\begin{aligned}
& P\left(\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_1^{2^{k+1}} Z_i > 2E(Z_1) + \delta\right) \\
& \leq 2^{k+1}P(Z_1 > 2^k) + P\left(\sum_{1 \leq i \leq 2^{k+1}} Z_i 1_{[0,2^k]}(Z_i) \geq 2^{k+1}E(Z_1) + 2^k\delta\right) \\
& \leq 2^{k+1}P(Z_1 > 2^k) + P\left(\sum_{1 \leq i \leq 2^{k+1}} [Z_i 1_{[0,2^k]}(Z_i) - E(Z_i 1_{[0,2^k]}(Z_i))] \geq 2^k\delta\right).
\end{aligned}$$

Par l'inégalité de Tchebitchev

$$\begin{aligned}
& P\left(\sum_{1 \leq i \leq 2^{k+1}} [Z_i 1_{[0,2^k]}(Z_i) - E(Z_i 1_{[0,2^k]}(Z_i))] \geq \delta 2^k\right) \\
& \leq \frac{E\left(\sum_{1 \leq i \leq 2^{k+1}} [Z_i 1_{[0,2^k]}(Z_i) - E(Z_i 1_{[0,2^k]}(Z_i))]\right)^2}{2^{2k}\delta^2} \\
& = \frac{2}{2^k\delta^2} [E(Z_1 1_{[0,2^k]}(Z_1))^2 - E^2(Z_1 1_{[0,2^k]}(Z_1))] \\
& \leq \frac{2}{2^k\delta^2} E(Z_1^2 1_{[0,2^k]}(Z_1)).
\end{aligned}$$

Alors

$$P\left(\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{1 \leq i \leq n} Z_i \geq 2E(Z_1) + \delta\right) \leq 2^{k+1}P(Z_1 > 2^k) + \frac{2}{2^k\delta^2} E(Z_1^2 1_{[0,2^k]}(Z_1)).$$

Pour  $t < 2^{k+1}$ ,

$$P(Z_1 > t) \geq P(Z_1 > 2^{k+1}),$$

implique

$$\int_{2^k}^{2^{k+1}} P(Z_1 > t) dt \geq 2^k P(Z_1 > 2^{k+1}),$$

ce qui donne

$$\begin{aligned}
\sum_{k \geq -1} 2^k P(Z_1 > 2^{k+1}) &= \sum_{k \geq 0} 2^{k-1} P(Z_1 > 2^k) \\
&\leq E(Z_1),
\end{aligned}$$

multiplions les deux cotés par 4

$$\sum_{k \geq 0} 2^{k+1} P(Z_1 > 2^k) \leq 4E(Z_1).$$

De plus

$$\begin{aligned} \sum_{k \geq 0} 2^{-k} E(Z_1^2 1_{[0, 2^k]}(Z_1)) &= E(Z_1^2 \sum_{k \geq 0} 2^{-k} 1_{[0, 2^k]}(Z_1)) \\ &\leq 4E(Z_1). \end{aligned}$$

En effet :

pour chaque  $\omega$  soit  $0 \leq Z_1(\omega) \leq 1$  ou bien  $\exists p \geq 0 / 2^p < Z_1(\omega) \leq 2^{p+1}$ ,  
pour  $2^p < Z_1(\omega) \leq 2^{p+1}$

$$\begin{aligned} Z_1^2(\omega) \sum_{k \geq 0} 2^{-k} 1_{[0, 2^k]}(Z_1(\omega)) &= Z_1^2(\omega) \sum_{k \geq p+1} 2^{-k} 1_{[0, 2^k]}(Z_1(\omega)) \\ &\leq 2^{2p+2} \sum_{k \geq p+1} 2^{-k} = 2^{p+2} \\ &\leq 4Z_1(\omega), \end{aligned}$$

de même pour  $0 \leq Z_1(\omega) \leq 1$ , donc  $\forall \omega$

$$Z_1^2(\omega) \sum_{k \geq 0} 2^{-k} 1_{[0, 2^k]}(Z_1(\omega)) \leq 4Z_1(\omega),$$

Ce qui implique le résultat cherché.

Finalement,

$$\begin{aligned} \sum_{k \geq 0} P\left(\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{1 \leq i \leq n} Z_i \geq 2E(Z_1) + \delta\right) &\leq 4(1 + 2\delta^{-2})E(Z_1) \\ &< \infty, \end{aligned}$$

d'après le lemme de Borel-Cantelli

$$P(\exists k \geq 0, \forall p \geq k / \max_{2^p < n \leq 2^{p+1}} \frac{1}{n} \sum_{1 \leq i \leq n} Z_i \geq 2E(Z_1) + \delta) = 0,$$

alors pour  $k$  assez grand, nous avons :

$$\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{1 \leq i \leq n} Z_i < 2E(Z_1) + \delta \text{ p.s.},$$

implique pour  $n$  assez grand,

$$\frac{1}{n} \sum_{1 \leq i \leq n} Z_i < 2E(Z_1) + \delta \text{ p.s.},$$

puisque  $\delta$  est arbitraire,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{1 \leq i \leq n} Z_i \leq 2E(Z_1).$$

Nous pouvons maintenant finir la démonstration

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} |S_n| &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{1 \leq i \leq n} |X_i - Y_i| + \limsup_{n \rightarrow \infty} \frac{1}{n} |T_n| \\ &\leq 2E(|X_1 - Y_1|) \\ &\leq 2\epsilon. \end{aligned}$$

Puisque  $\epsilon$  est arbitraire, ceci conclut la démonstration. □

## Références

- [1] M. Benaïm, N. El Karoui. Promenade aléatoire. Chaines de Markov et simulation, martingales et stratégies. Ecole Polytechnique. Novembre 2006.
- [2] P. Barbe, M. Ledoux. Probabilité. EDP Sciences. 2007.
- [3] C. Edouard Pfister. Théorie des probabilités, cours d'introduction avec application à la statistique mathématique. Presses polytechniques et universitaires romandes. 2012
- [4] L. Elie, B. Lapeyre. Introduction aux Méthodes de Monte-Carlo. Lien : [cermics.enpc.fr/bl/PS/SIMULATION-X/poly-monte-carlo-x.pdf](http://cermics.enpc.fr/bl/PS/SIMULATION-X/poly-monte-carlo-x.pdf). Septembre 2001.
- [5] J. François Delmas, B. Jourdain. Modèles aléatoires. Application aux sciences de l'ingénieur et du vivant. Springer. 2006.
- [6] M. Ledra. La méthode Monté Carlo et ses application. Lien : <https://www.researchgate.net/publication/310828308>. Mai 2016
- [7] T. Mourid. Cours de M2 probabilités statistiques. Module "Processus stochastique". 2016/2017.
- [8] B. Ycart. Modèles et Algorithmes Markoviens. Springer. 2000.