

*République Algérienne Démocratique et Populaire*  
*Ministère de l'Enseignement Supérieur et de la Recherche Scientifique*  
UNIVERSITE MOULOUD MAMMARI DE TIZI-OUZOU  
FACULTE DE SCIENCES - DEPARTEMENT DE MATHEMATIQUES

*Mémoire de Master de Mathématiques*  
*Appliquées / Option : Processus*  
*Aléatoires et Statistique de la Décision*

ALGORITHME DE SIMULATION  
DE PROPP-WILSON

présenté par  
*DRIF Zina*,

*sous la Direction de Mr. BOUDIBA Mohand Arezki,*

le 01/10/2013, devant le jury :

HAMADOUCHE DJAMEL,	Pr, UMMTO,	Président ;
BOUDIBA MOHAND AREZKI,	MCA, UMMTO,	Rapporteur ;
FELLAG HOCINE,	Pr, UMMTO,	Examinateur

---

## *Remerciements*

J'adresse mes vifs remerciements à Monsieur Hamadouche Djamel, pour l'honneur qu'il me fait en présidant le jury.

Je remercie également Monsieur Fellag Hocine pour avoir accepté de lire ce mémoire et de faire partie du jury.

Je présente mes sincères remerciements à mon directeur de mémoire, Monsieur Boudiba Mohand Arezki, pour m'avoir proposer ce thème, pour ses orientations, et sa grande contribution à l'aboutissement de ce projet.

Enfin, il m'est agréable d'adresser mes remerciements à tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail.

# Table des matières

<b>Introduction</b>	<b>4</b>
<b>1 Quelques propriétés de chaînes de Markov</b>	<b>7</b>
1.1 Définitions et propriétés générales . . . . .	7
1.2 Distance en variation totale . . . . .	10
1.3 Application au cas fini ou dénombrable . . . . .	18
<b>2 Quelques méthodes de simulation d'une loi de probabilités</b>	<b>25</b>
2.1 Echantillon uniforme sur $[0, 1]$ . . . . .	25
2.2 Simulation d'une loi de fonction de répartition inversible . . . . .	26
2.3 Simulation de chaînes de Markov . . . . .	27
2.4 Simulation d'une loi de probabilité . . . . .	31
2.5 Application : Algorithme de Metropolis . . . . .	38
<b>3 Simulation exacte : L'algorithme de Propp-Wilson</b>	<b>43</b>
3.1 Présentation de l'algorithme de Propp et Wilson . . . . .	43
3.2 Validité du principe de l'algorithme . . . . .	45
<b>Conclusion</b>	<b>48</b>
<b>Bibliographie</b>	<b>49</b>

## *Introduction*

Nous nous intéressons aux méthodes de simulation d'une loi de probabilité, c'est à dire à l'échantillonnage d'une loi de probabilité donnée  $\pi$ . Les méthodes classiques de simulation sont en général basées sur une approximation de la loi  $\pi$ . De nombreux auteurs ont élaboré des méthodes (algorithmes) à cet effet ( [1], [2], [3], [10], [12], ... ). Nous faisons une synthèse de ces méthodes.

Nous nous intéressons en particulier aux méthodes actuelles de simulation exacte, plus particulièrement au premier algorithme dans ce sens qui est l'algorithme de Propp-Wilson (Cf. [6] et [7]), à notre connaissance. L'étude et le perfectionnement de cet algorithme, auxquels de nombreux auteurs se sont intéressés, sont basées sur l'étude des chaînes de Markov engendrées par l'itération de fonctions aléatoires.

Plus exactement, soit  $(E, \mathcal{E})$  un espace mesurable, et  $(Y_n)_n$  une suite de variables aléatoires, à valeurs dans  $E$ , indépendantes et identiquement distribuées. Pour  $y$  fixé dans  $E$ , soit  $x \mapsto f_y(x)$  une fonction de  $E$  dans  $E$ . Si  $X_0^x$  est une variable aléatoire indépendante des  $Y_n$ , on considère la chaîne de Markov  $(X_n^x)_n$  définie pour  $x \in E$  par

$$X_0^x = x;$$

et pour  $n > 0$

$$X_n^x = f_{Y_n}(X_{n-1}^x).$$

Soit  $(H_n^x)_n$  le processus associé à  $(X_n^x)_n$ , défini pour  $x \in E$ , par

$$H_0^x = x;$$

et pour  $n > 0$

$$H_n^x(x) = f_{Y_1} \circ \cdots \circ f_{Y_{n-1}}(x).$$

Dans le cas où  $E$  est fini, l'algorithme de Propp-Wilson découle de l'étude du processus  $(H_n^x)_n$  associé à  $(X_n^x)_n$ , appelé aussi "backward process."

## TABLE DES MATIÈRES

---

Si nous considérons la chaîne de Markov  $(H_n)_n$  d'espace des états les fonctions de  $E$  dans  $E$  définie par :

$$H_0 = h;$$

$$H_n = H_{n-1} \circ f_{Y_n}, \quad \forall n > 0;$$

l'algorithme de Propp-Wilson est fondé sur le fait que sous certaines conditions (Cf. [1], [6], [7], [8], [13], ...),  $\lim_n H_n$  est une fonction constante.

La recherche de ces conditions a fait l'objet de nombreux travaux, ces dernières années (Cf. [1], [6], [7], [8], ...).

Notre but est d'étudier et de synthétiser les résultats nécessaires pour la mise en œuvre de cet algorithme.

*TABLE DES MATIÈRES*

---

# Chapitre 1

## Quelques propriétés de chaînes de Markov

### 1.1 Définitions et propriétés générales

**Définition 1.** 1. Soit  $(E, \mathcal{E})$  un espace mesurable et  $(X_n)_{n \geq 0}$  une suite de variables aléatoires à valeurs dans  $E$ . On dit que  $(X_n)_{n \geq 0}$  est une chaîne de Markov d'espace des états  $E$  quelconque, si  $\forall B \in \mathcal{E}$ ,

$$P[X_{n+1} \in B | X_n, X_{n-1}, \dots, X_0] = P[X_{n+1} \in B | X_n].$$

2. Si  $E$  est un ensemble fini ou dénombrable, et  $(X_n)_{n \geq 0}$  une suite de variables aléatoire à valeurs dans  $E$ , alors on dit que  $(X_n)_{n \geq 0}$  est une chaîne de Markov (CM) d'espace des états  $E$  et de matrice de transition  $P = (p_{xy})_{x,y \in E}$  si :

$$\forall x_0, x_1, \dots, x_{n-1}, x, y \in E, \text{ et } \forall n \in \mathbb{N},$$

$$P[X_{n+1} = y | X_0 = x_0, X_1 = x_1, \dots, X_n = x] = P[X_{n+1} = y | X_n = x] = p_{xy}.$$

Si le nombre  $p_{xy}$  ne dépend pas de  $n$ , alors la chaîne de Markov est dite homogène.

N.B : Dans toute la suite, les chaînes de Markov considérées sont toutes homogènes et à espace des états au plus dénombrable.

## 1.1. DÉFINITIONS ET PROPRIÉTÉS GÉNÉRALES

---

Soit  $E = \{x, y, \dots\}$  un ensemble fini,  $(X_n)_n$  une chaîne de Markov sur  $E$  de matrice de transition  $P = (p_{xy})_{x,y \in E}$ .

**Définition 2.** Soit  $\pi = (\pi_x)_{x \in E}$  une loi de probabilité sur  $E$ .

1.  $\pi$  est une loi stationnaire (invariante) pour la chaîne de Markov  $(X_n)_n$ , si  $\pi = \pi P$ , c'est à dire, si pour tout  $y \in E$ , on a :

$$\pi_y = \sum_{x \in E} \pi_x p_{xy} \iff \pi_y = \sum_{x \in E} \pi_x p_{xy}^{(n)}.$$

2. La loi  $\pi$  est dite réversible ou que la matrice  $P$  est  $\pi$ -réversible, si

$$\forall x, y \in E, \quad \pi_x p_{xy} = \pi_y p_{yx}.$$

**Remarque 1.** Si  $\pi$  est réversible, alors elle est stationnaire.

En effet ;

$$\pi_x p_{xy} = \pi_y p_{yx}.$$

En sommant par rapport à  $x$ , on obtient :

$$\sum_{x \in E} \pi_x p_{xy} = \sum_{x \in E} \pi_y p_{yx} = \pi_y \sum_{x \in E} p_{yx} = \pi_y, \quad y \in E.$$

c'est à dire  $\forall y \in E, \quad \pi_y = \sum_x \pi_x p_{xy}$ .

**Proposition 1.** Soit  $\pi = (\pi_x)_{x \in E}$  la loi stationnaire d'une chaîne de Markov  $(X_n)_n$ . Si on désigne par  $T_x$  l'instant du premier retour en  $x$ , c'est à dire :

$$T_x = \begin{cases} \inf_n \{X_n = x\}, & \text{si } \cup_n \{X_n = x\} \neq \emptyset; \\ \infty, & \text{sinon.} \end{cases}$$

alors on a :

$$\pi_x = \frac{1}{\mathbb{E}_x[T_x]}.$$



## CHAPITRE 1. QUELQUES PROPRIÉTÉS DE CHAÎNES DE MARKOV

**Définition 3.** Une chaîne de Markov  $(X_n)_n$ , est dite irréductible si elle ne possède qu'une unique classe de communication, c'est à dire que tous ses éléments communiquent ( $x \leftrightarrow y$ ).

**Définition 4.** Soit  $(X_n)_n$  une chaîne de Markov irréductible. La période d'un état  $x \in E$  est le PGCD (plus grand commun diviseur) de  $\{n \in \mathbb{N}^*; p_{xx}^n > 0\}$ . Un état est dit apériodique si sa période est 1, sinon il est dit périodique.

**Note :** La périodicité est une propriété de classe. Si une classe de communication est de période 1, on dit qu'elle est apériodique. Une chaîne de Markov irréductible est dite apériodique si elle admet une classe apériodique.

**Exemple 1.** Marche aléatoire sur  $\mathbb{Z}$ .

Soit  $(Y_n)_n$  une suite de variables aléatoires indépendantes identiquement distribuées (i.i.d.) tel que :

$$P[Y_n = +1] = p, \quad \text{et} \quad P[Y_n = -1] = q, \quad p + q = 1,$$

et  $X_0$  une variable aléatoire indépendante de  $(Y_n)_n$ . On considère la suite  $(X_n)_n$  définie par :

$$X_0, \quad \text{et} \quad \forall n > 0, \quad X_n = \sum_{k=1}^n Y_k.$$

Alors  $(X_n)_n$  est une chaîne de Markov sur  $\mathbb{Z}$  de matrice de transition :

$$p_{xy} = \begin{cases} p, & \text{si } y=x+1; \\ q, & \text{si } y=x-1; \\ 0, & \text{sinon.} \end{cases}$$

De plus,  $(X_n)_n$  est irréductible et apériodique.

## 1.2 Distance en variation totale

A présent nous avons besoin de comparer deux mesures de probabilités. Si  $(E, \mathcal{E})$  est un espace mesurable et  $\mathcal{P}(E)$  l'ensemble des lois de probabilités sur cet espace, nous introduisons une distance sur  $\mathcal{P}(E)$  : La distance en variation totale, qui est adaptée pour la comparaison de lois discrètes, ce qui est notre cas, car  $E$  est un ensemble dénombrable comme espace d'états d'une chaîne de Markov. Il s'agit donc de munir  $\mathcal{P}(E)$  d'une topologie. Remarquons que  $\mathcal{P}(E)$  est déjà muni de la topologie de la convergence vague de mesures. Pour une comparaison entre ces deux topologies nous nous contentons de signaler les remarques faites dans [11].

Dans cette section, on considère que  $(E, \mathcal{E})$  est un espace mesurable et  $P, Q$  sont deux mesures de probabilité sur cet espace.

**Définition 5.** *La distance en variation totale entre  $P$  et  $Q$  est le nombre réel positif noté  $\|P - Q\|$ , défini par :*

$$\|P - Q\| = \sup_{A \in \mathcal{E}} \left| \int_E \varphi(\omega) d(P - Q) \right|.$$

avec  $\varphi$  une fonction  $\mathcal{E}$ -mesurable tel que :  $|\varphi(\omega)| \leq 1$ .

**Lemme 1.** (Cf.[11]).

*La distance en variation totale est donnée par :*

$$\|P - Q\| = 2 \sup_{A \in \mathcal{E}} |P(A) - Q(A)|.$$

**Démonstration.**

1. On montre d'abord que :

$$2 \sup_{A \in \mathcal{E}} |P(A) - Q(A)| \leq \|P - Q\|.$$

$\forall A \in \mathcal{E}$ , on a :

$$P(A) - Q(A) = 1 - P(\bar{A}) - 1 + Q(\bar{A}) = Q(\bar{A}) - P(\bar{A}),$$

CHAPITRE 1. QUELQUES PROPRIÉTÉS DE CHAÎNES DE MARKOV

donc :

$$\begin{aligned} 2|P(A) - Q(A)| &= |P(A) - Q(A)| + |Q(\bar{A}) - P(\bar{A})| \\ &= |P(A) - Q(A)| + |P(\bar{A}) - Q(\bar{A})| \\ &= \left| \int_A dP - \int_A dQ \right| + \left| \int_{\bar{A}} dP - \int_{\bar{A}} dQ \right| \\ &= \left| \int_A d(P - Q) \right| + \left| \int_{\bar{A}} d(P - Q) \right| \\ &= \left| \int_{A \cup \bar{A}} d(P - Q) \right| = \left| \int_E d(P - Q) \right|. \end{aligned}$$

Donc :  $2|P(A) - Q(A)| = \left| \int_E d(P - Q) \right|$ .

Or  $\forall \varphi$  une fonction  $\mathcal{E}$ -mesurable tel que  $|\varphi(\omega)| \leq 1$ , on a par définition :

$$\left| \int_E d(P - Q) \right| \leq \sup \left| \int_E \varphi(\omega) d(P - Q) \right|, \quad \forall A \subset \mathcal{E}$$

$$\implies 2|P(A) - Q(A)| \leq \|P - Q\|,$$

d'où :  $\forall A \subset \mathcal{E}, \quad 2 \sup_{A \subset \mathcal{E}} |P(A) - Q(A)| \leq \|P - Q\|$ .

## 1.2. DISTANCE EN VARIATION TOTALE

---

2. Montrons que :

$$\|P - Q\| \leq 2 \sup_{A \in \mathcal{E}} |P(A) - Q(A)|.$$

On va utiliser la décomposition de Hahn pour une mesure signée. Soit  $\mu = P - Q$  une mesure signée.  $\exists(A, B)$  une partition de  $E$  pour  $\mu$ , tel que  $A$  est l'ensemble positif de  $\mu$ , et  $B$  son ensemble négatif.

La variation totale de  $\mu$  est :  $\|\mu\| = |\mu|(E)$ , tel que :

$$|\mu|(E) = \mu^+(E) + \mu^-(E)$$

avec :

$$\mu^+(E) = \mu(A), \quad \mu^-(E) = -\mu(B).$$

Donc :

$$\begin{aligned} \|P - Q\| = \|\mu\| &= \mu^+(E) + \mu^-(E) \\ &= \mu(A) - \mu(B) \\ &= (P - Q)(A) - (P - Q)(B) \\ &= P(A) - Q(A) - P(B) + Q(B) \\ &= (P(A) - Q(A)) + (Q(B) - P(B)). \end{aligned}$$

Comme  $(A, B)$  est une partition de  $E$ , alors  $A \cup B = E$ , c'est à dire :

$B = \overline{A}$ , donc :

$$\|P - Q\| = (P(A) - Q(A)) + (Q(\overline{A}) - P(\overline{A})).$$

## CHAPITRE 1. QUELQUES PROPRIÉTÉS DE CHAÎNES DE MARKOV

Or on a déjà vu que :  $P(A) - Q(A) = Q(\bar{A}) - P(\bar{A})$ , donc :

$$\begin{aligned} \|P - Q\| = 2(P(A) - Q(A)) &\leq 2|P(A) - Q(A)| \\ &\leq 2 \sup_{A \in \mathcal{E}} |P(A) - Q(A)|. \end{aligned}$$

D'où :

$$\|P - Q\| \leq 2 \sup_{A \in \mathcal{E}} |P(A) - Q(A)|.$$

D'après (1) et (2), on conclue que :

$$\|P - Q\| = 2 \sup_{A \in \mathcal{E}} |P(A) - Q(A)|.$$

□

**Proposition 2.** (Cf. [4]).

Soient  $P$  et  $Q$  deux mesures de probabilité sur  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  de densités respectives (par rapport à la mesure de Lebesgue  $dx$ )  $p(x)$  et  $q(x)$ , c'est à dire :

$$p(x) = \frac{dP}{dx}, \quad \text{et} \quad q(x) = \frac{dQ}{dx}.$$

Alors :

$$\|P - Q\| = \int_{\mathbb{R}} |p(x) - q(x)| dx.$$

**Démonstration.** Il faut montrer que :  $\forall A \subset \mathbb{R}$ ,

$$2 \sup_A |P(A) - Q(A)| = \int_{\mathbb{R}} |p(x) - q(x)|.$$

## 1.2. DISTANCE EN VARIATION TOTALE

---

1. On montre en premier lieu que :

$$2 \sup_A |P(A) - Q(A)| \leq \int_{\mathbb{R}} |p(x) - q(x)|.$$

On a :

$$\begin{aligned} 2|P(A) - Q(A)| &= |P(A) - Q(A)| + |P(\bar{A}) - Q(\bar{A})| \\ &= \left| \int_A dP - \int_A dQ \right| + \left| \int_{\bar{A}} dP - \int_{\bar{A}} dQ \right| \\ &= \left| \int_A p(x)dx - \int_A q(x)dx \right| + \left| \int_{\bar{A}} p(x)dx - \int_{\bar{A}} q(x)dx \right| \\ &= \left| \int_A (p(x) - q(x))dx \right| + \left| \int_{\bar{A}} (p(x) - q(x))dx \right| \\ &\leq \int_A |p(x) - q(x)|dx + \int_{\bar{A}} |p(x) - q(x)|dx \\ &= \int_{A \cup \bar{A}} |p(x) - q(x)|dx = \int_{\mathbb{R}} |p(x) - q(x)|dx. \end{aligned}$$

Donc :

$$\forall A \subset \mathbb{R}, \quad 2|P(A) - Q(A)| \leq \int_{\mathbb{R}} |p(x) - q(x)|dx,$$

d'où :

$$2 \sup_A |P(A) - Q(A)| \leq \int_{\mathbb{R}} |p(x) - q(x)|dx.$$

## CHAPITRE 1. QUELQUES PROPRIÉTÉS DE CHAÎNES DE MARKOV

**Remarque 2.** Pour la suite, on a besoin de montrer que :

$$\int_{\mathbb{R}} |p(x) - q(x)| = 2 \int_{\{p(x) > q(x)\}} (p(x) - q(x)) dx = 2 \int_{\{p(x) < q(x)\}} (q(x) - p(x)) dx.$$

Comme :  $P$  et  $Q$  sont deux lois de probabilité, donc on a :

$$\int_{\mathbb{R}} dP = \int_{\mathbb{R}} dQ = 1,$$

c'est à dire :

$$\int_{\mathbb{R}} p(x) dx = \int_{\mathbb{R}} q(x) dx = 1,$$

donc :

$$\begin{aligned} 0 &= \int_{\mathbb{R}} p(x) dx - \int_{\mathbb{R}} q(x) dx \\ &= \int_{\mathbb{R}} (p(x) - q(x)) dx \\ &= \int_{\{p(x) > q(x)\}} (p(x) - q(x)) dx + \int_{\{p(x) < q(x)\}} (p(x) - q(x)) dx, \\ \\ \implies \int_{\{p(x) > q(x)\}} (p(x) - q(x)) dx &= - \int_{\{p(x) < q(x)\}} (p(x) - q(x)) dx \\ &= \int_{\{p(x) < q(x)\}} (q(x) - p(x)) dx. \end{aligned}$$

## 1.2. DISTANCE EN VARIATION TOTALE

---

$$\implies \int_{\{p(x)>q(x)\}} (p(x) - q(x))dx = \int_{\{p(x)<q(x)\}} (q(x) - p(x))dx.$$

Comme :  $\int_{\mathbb{R}} (p(x) - q(x))dx = 0$ , donc on a :

$$\begin{aligned} \int_{\mathbb{R}} |p(x) - q(x)|dx &= \int_{\mathbb{R}} (p(x) - q(x))dx \\ &= \int_{\{p(x)>q(x)\}} (p(x) - q(x))dx + \int_{\{p(x)<q(x)\}} (q(x) - p(x))dx \\ &= 2 \int_{\{p(x)>q(x)\}} (p(x) - q(x))dx \\ &= 2 \int_{\{p(x)<q(x)\}} (q(x) - p(x))dx. \end{aligned}$$

2. On montre que :

$$\int_{\mathbb{R}} |p(x) - q(x)|dx \leq 2 \sup_A |P(A) - Q(A)|.$$

On a :

$$\|P - Q\| = 2 \sup_A |P(A) - Q(A)| \implies \|P - Q\| \geq 2|P(A) - Q(A)|.$$

Posons  $B = \{x : p(x) > q(x)\}$ ; donc :  $\|P - Q\| \geq 2|P(B) - Q(B)|.$



CHAPITRE 1. QUELQUES PROPRIÉTÉS DE CHAÎNES DE MARKOV

$$\begin{aligned} \|P - Q\| &\geq 2|P(B) - Q(B)| \\ &= 2\left|\int_B dP - \int_B dQ\right| \\ &= 2\left|\int_B p(x)dx - \int_B q(x)dx\right| \\ &= 2\left|\int_{\{p(x)>q(x)\}} p(x)dx - \int_{\{p(x)>q(x)\}} q(x)dx\right| \\ &= 2\left|\int_{\{p(x)>q(x)\}} (p(x) - q(x))dx\right| \\ &= 2\int_{\{p(x)>q(x)\}} (p(x) - q(x))dx. \end{aligned}$$

Daprès la remarque 2, on aura :

$$\begin{aligned} 2 \sup_A |P(A) - Q(A)| &\geq 2 \int_{\{p(x)>q(x)\}} (p(x) - q(x))dx \\ &= \int_{\mathbb{R}} |p(x) - q(x)|dx. \end{aligned}$$

Donc :

$$2 \sup_A |P(A) - Q(A)| \geq \int_{\mathbb{R}} |p(x) - q(x)|dx.$$

### 1.3. APPLICATION AU CAS FINI OU DÉNOMBRABLE

---

De (1) et (2), on conclue que :

$$2 \sup_A |P(A) - Q(A)| = \int_{\mathbb{R}} |p(x) - q(x)| dx,$$

c'est à dire :

$$\|P - Q\| = \int_{\mathbb{R}} |p(x) - q(x)| dx.$$

□

**Corollaire 1.** *Dans le cas où  $P$  et  $Q$  sont deux mesures discrètes, on a :*

$$\|P - Q\| = \sum_{i=1}^{\infty} |p(x_i) - q(x_i)|.$$

## 1.3 Application au cas fini ou dénombrable

Si  $E$  est un ensemble fini ou dénombrable, alors nous avons (Cf. [5]) :

**Définition 6.** *Soient  $\mu$  et  $\nu$  deux mesures de probabilité sur  $E$ . La distance en variation totale entre  $\mu$  et  $\nu$  est le nombre réel positif noté  $d_{VT}(\mu, \nu)$ , défini par :*

$$d_{VT}(\mu, \nu) = \max_{A \subseteq E} |\mu(A) - \nu(A)|.$$

**Proposition 3.** *Etant donnée deux mesures de probabilité  $\mu$  et  $\nu$  sur  $E$ , on a l'identité suivante :*

$$d_{VT}(\mu, \nu) = \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)|.$$

CHAPITRE 1. QUELQUES PROPRIÉTÉS DE CHAÎNES DE MARKOV

**Démonstration.** Soit  $B = \{x : \mu(x) \geq \nu(x)\}$ . Si on considère que  $A \subset E$  est un ensemble quelconque, donc il peut s'écrire sous la forme suivante :

$$A = A_1 \cup B, \quad \text{tel que : } A_1 = \{x : \mu(x) < \nu(x)\}.$$

Comme :  $A_1 \cap B = \emptyset$ , alors :

$$\begin{aligned}(\mu - \nu)(A) &= (\mu - \nu)(A_1 \cup B) = (\mu - \nu)(A_1) + (\mu - \nu)(B), \\ \implies (\mu - \nu)(A) - (\mu - \nu)(A_1) &= (\mu - \nu)(B)\end{aligned}$$

or :

$$\begin{aligned}(\mu - \nu)(A_1) < 0 &\implies -(\mu - \nu)(A_1) > 0 \\ \implies (\mu - \nu)(A) &\leq (\mu - \nu)(B).\end{aligned}$$

De la même manière, on considère :  $\overline{B} = \{x : \mu(x) < \nu(x)\}$ , donc :

$$A = A_2 \cup \overline{B}, \quad \text{tel que : } A_2 = \{x : \mu(x) \geq \nu(x)\} = B.$$

Comme :  $A_2 \cap \overline{B} = \emptyset$ , alors :

$$\begin{aligned}(\nu - \mu)(A) &= (\nu - \mu)(A_2 \cup \overline{B}) = (\nu - \mu)(A_2) + (\nu - \mu)(\overline{B}) \\ &= (\nu - \mu)(B) + (\nu - \mu)(\overline{B})\end{aligned}$$

$$\implies (\nu - \mu)(A) - (\nu - \mu)(B) = (\nu - \mu)(\overline{B})$$

or :

### 1.3. APPLICATION AU CAS FINI OU DÉNOMBRABLE

---

$$(\nu - \mu)(B) > 0 \implies (\nu - \mu)(A) \geq (\nu - \mu)(\overline{B}).$$

Comme :  $(\mu - \nu)(B) = (\nu - \mu)(\overline{B})$ , donc :  $|(\mu - \nu)(A)| = (\mu - \nu)(B)$ .

$$\begin{aligned} |(\mu - \nu)(A)| &= (\mu - \nu)(B) \\ &= \frac{1}{2}(\mu(B) - \nu(B) + \nu(\overline{B}) - \mu(\overline{B})) \\ &= \frac{1}{2}[(\mu(B) - \nu(B)) - (\mu(\overline{B}) - \nu(\overline{B}))] \end{aligned}$$

Sur  $B$  :  $|(\mu - \nu)|(B) = \mu(B) - \nu(B)$ , sur  $\overline{B}$  :  $|(\mu - \nu)|(\overline{B}) = -[\mu(\overline{B}) - \nu(\overline{B})]$ ;

$$\begin{aligned} \frac{1}{2} [ (\mu(B) - \nu(B)) - (\mu(\overline{B}) - \nu(\overline{B})) ] &= \frac{1}{2} [ |\mu - \nu|(B) + |\mu - \nu|(\overline{B}) ] \\ &= \frac{1}{2} [ |\mu - \nu|(B \cup \overline{B}) ] \\ &= \frac{1}{2} [ |\mu - \nu|(E) ] \\ &= \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)|. \end{aligned}$$

CHAPITRE 1. QUELQUES PROPRIÉTÉS DE CHAÎNES DE MARKOV

Donc :

$$|(\mu - \nu)(A)| = \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)|.$$

Comme :  $|(\mu - \nu)(A)|$  est une constante, donc :

$$|(\mu - \nu)(A)| = \max_A |(\mu - \nu)(A)| = \|\mu - \nu\|.$$

d'où le résultat.

□

**Proposition 4.** *Avec ces notations, on a  $d_{VT}$  est une distance.*

**Démonstration.** (i)  $d_{VT}(\mu, \nu) = 0 \iff \mu = \nu$  :

$\implies$ ]

$$d_{VT}(\mu, \nu) = 0 \iff \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)| = 0$$

$$\implies \sum_{x \in E} |\mu(x) - \nu(x)| = 0$$

$$\implies |\mu(x) - \nu(x)| = 0 \tag{1.1}$$

$$\implies \mu = \nu, \quad \forall x \in E.$$

(1.1) est vraie, car on a les termes de la somme sont positifs ou nuls.

$\iff$ ]  $\mu = \nu$ , donc  $\forall x \in E$ , on a :

### 1.3. APPLICATION AU CAS FINI OU DÉNOMBRABLE

---

$$\mu(x) - \nu(x) = 0 \iff |\mu(x) - \nu(x)| = 0$$

$$\implies \sum_{x \in E} |\mu(x) - \nu(x)| = 0$$

$$\implies \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)| = 0 = d_{VT}(\mu, \nu).$$

(ii)  $d_{VT}(\mu, \nu) = d_{VT}(\nu, \mu)$  :

$$\begin{aligned} d_{VT}(\mu, \nu) &= \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)| \\ &= \frac{1}{2} \sum_{x \in E} |\nu(x) - \mu(x)| = d_{VT}(\nu, \mu). \end{aligned}$$

(iii)  $d_{VT}(\mu, \nu) \leq d_{VT}(\mu, \eta) + d_{VT}(\eta, \nu)$ ,  $\eta \in \mathcal{P}(E)$  :

On a :  $d_{VT}(\mu, \nu) = \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)|$ ,

$$\begin{aligned} |\mu(x) - \nu(x)| &= |\mu(x) - \eta(x) + \eta(x) - \nu(x)| \\ &\leq |\mu(x) - \eta(x)| + |\eta(x) - \nu(x)| \end{aligned}$$

CHAPITRE 1. QUELQUES PROPRIÉTÉS DE CHAÎNES DE MARKOV

$$\Rightarrow \sum_{x \in E} |\mu(x) - \nu(x)| \leq \sum_{x \in E} |\mu(x) - \eta(x)| + \sum_{x \in E} |\eta(x) - \nu(x)|$$

$$\Rightarrow \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)| \leq \frac{1}{2} \sum_{x \in E} |\mu(x) - \eta(x)| + \frac{1}{2} \sum_{x \in E} |\eta(x) - \nu(x)|$$

$$\Rightarrow d_{VT}(\mu, \nu) \leq d_{VT}(\mu, \eta) + d_{VT}(\eta, \nu).$$

**Conclusion :**

De (i), (ii) et (iii), on conclue que  $d_{VT}$  est une distance.

□

### *1.3. APPLICATION AU CAS FINI OU DÉNOMBRABLE*

---



# Chapitre 2

## Quelques méthodes de simulation d'une loi de probabilités

Nous présentons dans ce chapitre quelques méthodes de simulation de variables aléatoires dites méthodes de Monte-Carlo.

### 2.1 Echantillon uniforme sur $[0, 1]$

Le problème de la génération d'un échantillon aléatoire  $(U_1, U_2, \dots, U_n)$  où les  $U_i$  sont uniformément réparties dans  $[0, 1]$ , est en principe réglé par l'existence dans le système, d'un générateur de nombres aléatoires. De nombreux algorithmes existent, pour cela. Une méthode possible néanmoins, serait de s'appuyer sur le théorème de Borel, que nous rappelons cdi-après.

**Théorème 1** (Théorème de Borel). *Si  $x \in [0, 1] \cap \mathbb{Q}^c$ , et si  $x = 0.X_1X_2\dots X_n, \dots$  est le développement décimal de  $x$ , alors les  $X_i$  sont uniformément réparties sur  $\{0, 1, 2, \dots, 9\}$  c'est à dire :*

$$P[X_i = a] = \frac{1}{10},$$

avec  $a \in \{0, 1, \dots, 9\}$ .

Ce théorème contient un algorithme de génération des variables aléatoires  $U_i$  uniformément réparties sur  $[0, 1]$ .

## 2.2. SIMULATION D'UNE LOI DE FONCTION DE RÉPARTITION INVERSIBLE

---

Tout le problème est de disposer alors d'une méthode de développement décimal d'un irrationnel quelconque, par exemple :  $\sqrt{2}$ ,  $e$ ,  $\pi$ , ...

Si par exemple on veut calculer le développement décimal de  $\pi$ , alors on exploite la formule :

$$\frac{\pi}{4} = \sum_{n=0}^{\infty} (-1)^n \frac{1}{2n+1},$$

Pour des besoins de précision, on peut faire appel à cette méthode. Malgré la puissance de calcul accrue des machines actuelles, son caractère ardu (Le calcul des décimales d'un irrationnel est non trivial) et le caractère pratique des générateurs usuels de nombres aléatoires dans les machines et surtout leurs satisfactions aux tests courants, font que c'est un luxe inutile. On se satisfait donc des algorithmes usuels de génération de nombres aléatoires

Pour la suite, on suppose que nous disposons d'une machine muni d'un générateur de nombres aléatoires uniformément répartis sur  $[0, 1]$ .

## 2.2 Simulation d'une loi de fonction de répartition inversible

Notre problème est la simulation d'une variable aléatoire  $X$  suivant une loi de fonction de répartition  $F$ . Nous voulons donc construire un échantillon aléatoire  $(X_1, \dots, X_n)$  de  $X$ .

Notre point de départ est la proposition suivante :

**Proposition 5.** *Soit  $X$  une variable aléatoire de fonction de répartition  $F$ . Si  $F$  est inversible, alors la variable aléatoire  $F(X)$ , suit la loi uniforme  $\mathbb{U}[0, 1]$ .*

**Démonstration.** Si  $x \in [0, 1]$ , alors on a :

$$P[F(X) \leq x] = P[X \leq F^{-1}(x)] = F(F^{-1}(x)) = x.$$

□

**Conséquence :**

Si  $(U_0, \dots, U_n)$  une suite de variables aléatoires indépendantes suivant la loi uniforme sur  $[0, 1]$ , et si  $X_k = F^{-1}(U_k)$  sont indépendantes, alors les  $X_k$  sont indépendantes et  $(X_1, \dots, X_n)$  est un échantillon de taille  $n$  de  $X$  suivant la loi de fonction de répartition  $F$ .

En effet, d'après la proposition 1 et pour  $U_k \sim \mathbb{U}[0, 1]$ , on a :

$$P[X_k \leq x] = P[F^{-1}(U_k) \leq x] = P[U_k \leq F(x)] = F(x).$$

Ce qui signifie que chaque variable aléatoire  $X_k$  suit une loi dont la fonction de répartition est  $F$ . Les variables aléatoires  $X_k$  sont indépendantes puisque ils sont de la forme  $F^{-1}(U_k)$  où les  $U_k$  sont indépendantes.

**Résumé :** La proposition 1 contient un algorithme de simulation de la variable aléatoire  $X$ . Dans ce cas, pour appliquer cette méthode, on a l'algorithme suivant :

1. Générer  $U_k$  selon une loi uniforme sur  $[0, 1]$ ;
2. Effectuer la transformation  $X_k = F^{-1}(U_k)$ .

**Remarque 3.** *Cette méthode ne permet de simuler que des variables aléatoires dont on connaît leur fonction de répartition  $F$  et celle ci doit être inversible. Il existe cependant des cas où  $F^{-1}$  n'existe pas, donc la méthode est limitée dans ses applications.*

## 2.3 Simulation de chaînes de Markov

Supposons qu'on veut simuler une chaîne de Markov  $(X_n)_{n \geq 0}$  définie sur ensemble fini  $E = \{x, y, \dots\}$ , de matrice de transition  $P = (p_{xy})_{x,y \in E}$  et de loi initiale  $\mu^{(0)}$ .

### 2.3. SIMULATION DE CHAÎNES DE MARKOV

---

On suppose donnée une suite  $U_0, U_1, \dots, U_n$  de variables aléatoires indépendantes identiquement distribuées (i.i.d.) tel que :

$$\forall i, U_i \sim \mathbb{U}[0, 1].$$

Soit la fonction  $\psi$  (fonction d'initialisation) tel que :  $\psi : [0, 1] \longrightarrow E$ , définie par :

$$\psi(u) = \begin{cases} x, & \text{si } u \in [0, \mu^{(0)}(x) [ ; \\ y, & \text{si } u \in [ \mu^{(0)}(x), \mu^{(0)}(x) + \mu^{(0)}(y) [ ; \\ \vdots & \\ h, & \text{si } u \in [ \sum_{l=x}^{h-1} \mu^{(0)}(l), \sum_{l=x}^h \mu^{(0)}(l) [ ; \\ \vdots & \\ k, & \text{si } u \in [ \sum_{l=x}^{k-1} \mu^{(0)}(l), 1 ]. \end{cases}$$

On a alors  $X_0 = \psi(U_0)$  car nous avons d'après [1] :

**Proposition 6.** *La loi de  $\psi(U_0)$  est égale à  $\mu^{(0)}$ .*

**Démonstration.**

$$\begin{aligned} P[\psi(U_0) = h] &= \int_0^1 \mathbb{I}_{(\psi(u)=h)} d\mu \\ &= \mu(\psi(u) = h) \\ &= \sum_{l=x}^h \mu^{(0)}(l) - \sum_{l=x}^{h-1} \mu^{(0)}(l) \\ &= \mu^{(0)}(h) = P[X_0 = h]. \end{aligned}$$

□

De même, on considère la fonction  $\phi$  (fonction de mise à jour) tel que :

$\phi : [0, 1] \times E \longrightarrow [0, 1]$ , définie par :

$$\phi(x, u) = \begin{cases} x, & \text{si } u \in [0, p_{x1} [ ; \\ y, & \text{si } u \in [p_{x1}, p_{x1} + p_{x2} [ ; \\ \vdots & \\ h, & \text{si } u \in [ \sum_{l=x}^{h-1} p_{xl}, \sum_{l=x}^h p_{xl} [ ; \\ \vdots & \\ k, & \text{si } u \in [ \sum_{l=1}^{k-1} p_{xl}, 1 ]. \end{cases}$$

**Proposition 7.** Si  $X_{n+1} = \phi(X_n, U_{n+1})$ , alors on a

$$P[X_{n+1} = y / X_n = x] = p_{xy}.$$

**Démonstration.**

$$\begin{aligned} P[X_{n+1} = h / X_n = x] &= P[\phi(X_n, U_{n+1}) = h / X_n = x] \\ &= P[\phi(x, U_{n+1}) = h] \\ &= \int_0^1 \mathbb{I}_{(\phi(x,u)=h)} d\mu \\ &= \mu(\phi(x, u) = h) \\ &= \sum_{l=x}^h p_{xl} - \sum_{l=x}^{h-1} p_{xl} = p_{xh}. \end{aligned}$$

### 2.3. SIMULATION DE CHAÎNES DE MARKOV

---

Or la deuxième égalité est vraie car on a :  $U_{n+1}$  est indépendante de  $(U_0, \dots, U_n)$ ; et comme la suite  $(U_0, \dots, U_n)$  est indépendante de  $X_n$  (hypothèse), donc  $U_{n+1}$  est indépendante de  $X_n$ .

□

#### Application :

Cette méthode nous donne une définition d'une chaîne de Markov, et donne aussi un moyen de la simuler.

A l'aide d'une fonction  $\psi$ , nous avons  $X_0$ , et à l'aide de  $\phi$  nous avons  $X_n$  à partir de  $X_{n+1}$ , et on donne une suite de variables aléatoires  $(U_n)_n$  i.i.d. de loi uniforme sur  $[0, 1]$ . Pour simuler  $(X_n)_n$  :

- on simule  $X_0$  tel que :  $X_0 = \psi(U_0)$ ;
- on pose  $X_{n+1} = \phi(X_n, U_{n+1})$ ,  $\forall n \geq 0$ ;
- on affiche  $X_n$ .

L'algorithme donne alors une trajectoire  $(X_0, X_1, \dots, X_n)$  de  $(X_n)_n$ .

**Exemple 2.** Soit  $E = \{x, y, z\}$  un espace des états fini,  $(X_n)_{n \geq 0}$  une chaîne de Markov sur  $E$  de loi initiale  $\mu^{(0)} = (1/4, 1/2, 1/4)$  et de matrice de transition  $P$  telle :

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

$(U_i)_{0 \leq i \leq 5}$  une suite de variables aléatoires i.i.d. de loi uniforme  $\mathbb{U}[0, 1]$ , donc on a :

$$\psi(u) = \begin{cases} x, & \text{si } u \in [0, 1/4[ ; \\ y, & \text{si } u \in [1/4, 3/4[ ; \\ z, & \text{si } u \in [3/4, 1]. \end{cases}$$

$$\phi(x, u) = \begin{cases} x, & \text{si } u \in [0, 1/2[; \\ y, & \text{si } u = 1/2; \\ z, & \text{si } u \in [1/2, 1]. \end{cases}$$

$$\phi(y, u) = \begin{cases} x, & \text{si } u \in [0, 1/3[; \\ y, & \text{si } u \in [1/3, 2/3[; \\ z, & \text{si } u \in [2/3, 1]. \end{cases}$$

et  $\forall u \in [0, 1]$ , on a :  $\phi(z, u) = x$ .

D'où, on obtient l'échantillon suivant :

$$\begin{aligned} X_0 &= \psi(U_0); & X_1 &= \phi(X_0, U_1); \\ X_2 &= \phi(X_1, U_2); & X_3 &= \phi(X_2, U_3); \\ X_4 &= \phi(X_3, U_4); & X_5 &= \phi(X_4, U_5). \end{aligned}$$

## 2.4 Simulation d'une loi de probabilité

Nous avons la proposition suivante (Cf. [1])

**Proposition 8.** *Soit  $(X_n)_{n \geq 0}$  une chaîne de Markov irréductible et apériodique sur un ensemble fini  $E$ , de matrice de transition  $P = (p_{xy})_{x,y \in E}$ . Si on désigne par  $\pi = (\pi_x)_{x \in E}$  la loi stationnaire de  $(X_n)_{n \geq 0}$ , alors pour toute loi initiale  $\mu^{(0)}$ , on a :*

$$\mu^{(n)} \xrightarrow{VT} \pi, \quad n \rightarrow \infty$$

## 2.4. SIMULATION D'UNE LOI DE PROBABILITÉ

---

avec  $\mu^{(n)} = \mathcal{L}(X_n)$ .

**Démonstration.** Il faut montrer que :

$$d_{Vt}(\mu^{(n)}, \pi) \longrightarrow 0.$$

Soient  $(X_n)_{n \geq 0}$  et  $(Y_n)_{n \geq 0}$  deux chaînes de Markov indépendantes définies sur le même ensemble  $E$ .

$(X_n)_n$  est de loi initiale  $\mu^{(0)}$  quelconque,  $(Y_n)_n$  est de loi initiale  $\pi$  qui est stationnaire.

Soit  $T$  la variable aléatoire définie par :

$$T : \Omega \longrightarrow \mathbb{N} \cup \infty$$

$$T(\omega) = \begin{cases} \inf_n \{n > 0, X_n = Y_n\}, & \text{si } \bigcup_n \{X_n = Y_n\} \neq \emptyset; \\ \infty, & \text{sinon.} \end{cases}$$

$T$  est donc le premier instant où  $X_n = Y_n$ .

1. Montrons que :  $\lim_{n \rightarrow \infty} P[T > n] = 0$ .

$$P[T > n] = 1 - P[T \leq n].$$

$$\{T \leq n\} = \bigcup_{k=1}^n \{X_k = Y_k\} = \{X_1 = Y_1\} \cup \dots \cup \{X_n = Y_n\}.$$

$$P[X_n = Y_n] \leq P[T \leq n], \quad \text{car on a : } \{X_n = Y_n\} \subset \{T \leq n\}.$$



– Pour  $n = m$ ;  $P[T \leq m] \geq P[X_m = Y_m]$ .

$$\{X_m = Y_m\} = \{X_m = Y_m = x\} \cup \{X_m = Y_m = y\} \cup \dots \cup \{X_m = Y_m = z\}.$$

Or :  $P[X_m = Y_m = y] \leq P[X_m = Y_m]$ , car on a :

$$\{X_m = Y_m = y\} \subset \{X_m = Y_m\}.$$

$$\begin{aligned} P[T \leq m] &\geq P[X_m = Y_m] \\ &\geq P[X_m = Y_m = y] = P[X_m = y, Y_m = y] \\ &= P[X_m = y] P[Y_m = y] \text{ par indépendance.} \end{aligned}$$

D'après la formule des probabilités totales, on a :

$$\begin{aligned} P[X_m = y] &= \sum_{x \in E} P[X_m = y / X_0 = x] P[X_0 = x] = \sum_{x \in E} \mu^{(0)}(x) p_{xy}^{(m)}; \\ P[Y_m = y] &= \sum_{x \in E} P[Y_m = y / Y_0 = x] P[Y_0 = x] = \sum_{x \in E} \pi_x p_{xy}^{(m)}. \end{aligned}$$

donc :

$$P[T \leq m] \geq \left( \sum_{x \in E} \mu^{(0)}(x) p_{xy}^{(m)} \right) \left( \sum_{x \in E} \pi_x p_{xy}^{(m)} \right). \quad (2.1)$$

Comme  $(X_n)_n$  est irréductible, alors  $\exists m < \infty$  tel que :  $p_{xy}^{(m)} > 0$ .

On pose  $\alpha = \min_m(p_{xy}^{(m)}) \Rightarrow \alpha > 0$  et  $p_{xy}^{(m)} > \alpha$ , donc :

$$(1.2) \geq \left( \alpha \sum_{x \in E} \mu^{(0)}(x) \right) \left( \alpha \sum_{x \in E} \pi_x \right) = \alpha^2.$$

## 2.4. SIMULATION D'UNE LOI DE PROBABILITÉ

---

Car  $\mu, \pi \in \mathcal{P}(E) \Rightarrow \sum_{x \in E} \mu(x) = \sum_{x \in E} \pi_x = 1$ .

donc :

$$\begin{aligned} P[T \leq m] &\geq P[X_m = Y_m] \geq \alpha^2 \\ \implies [T > m] &\leq P[X_m \neq Y_m] \leq 1 - \alpha^2. \end{aligned} \quad (2.2)$$

$$\begin{aligned} - \text{ Pour } n = 2m; P[T > 2m] &= P[T > 2m, T > m] \\ &= P[T > 2m / T > m] P[T > m]. \end{aligned}$$

Comme  $P[T > m] \leq 1 - \alpha^2$ , on a :

$$\begin{aligned} P[T > 2m, T > m] &= P[T > 2m / T > m] P[T > m] \\ &\leq (1 - \alpha^2) P[T > 2m / T > m]. \end{aligned}$$

Or :  $P[T > 2m] \leq P[X_{2m} \neq Y_{2m}] \leq 1 - \alpha^2$ , d'après (2.2).

$$\begin{aligned} P[T > 2m / T > m] P[T > m] &\leq (1 - \alpha^2) P[T > 2m / T > m] \\ &\leq (1 - \alpha^2) P[X_{2m} \neq Y_{2m} / T > m] \\ &\leq (1 - \alpha^2) (1 - \alpha^2) = (1 - \alpha^2)^2. \end{aligned}$$

Donc :  $\forall l, n = lm, P[T > lm] \leq (1 - \alpha^2)^l$ .

CHAPITRE 2. QUELQUES MÉTHODES DE SIMULATION D'UNE LOI DE PROBABILITÉS

---

On a :  $\alpha = \min_m(p_{xy}^{(m)})$ ; de plus :  $p_{xy}^{(m)} \leq 1$  (matrice de transition),

donc :  $0 < \alpha \leq 1$ .

$$0 < \alpha \leq 1 \Rightarrow 0 < 1 - \alpha^2 \leq 1 \Rightarrow \lim_{l \rightarrow \infty} (1 - \alpha^2)^l = 0.$$

**Conclusion :**  $\forall n, P[T > n] \rightarrow 0$ .

2. On construit une chaîne de Markov  $(Z_n)_{n \geq 0}$  définie sur  $E$  tel que :

$$Z_n = \begin{cases} X_n, & \text{si } n \leq T; \\ Y_n, & \text{si } n > T. \end{cases}$$

3. Soient  $(Z_n)_n$  et  $(Y_n)_n$  deux chaînes de Markov indépendantes définies sur  $E$ .  $(Z_n)_n$  est de loi initiale  $\mu^{(0)}$ , et  $(Y_n)_n$  est de loi initiale la loi stationnaire  $\pi$ . Si  $\mu^{(n)} = \mathcal{L}(Z_n)$ , alors on a :

$$P[Z_n = x] = P[X_n = x, n \leq T] = \mu^{(n)}(x);$$

$$P[Y_n = x] = \pi_x.$$

$$\begin{aligned} \mu^{(n)}(x) - \pi_x &= P[X_n = x, n \leq T] - P[Y_n = x] \\ &= P[Z_n = x] - P[Y_n = x]. \end{aligned}$$

Or :

$$P[Z_n = x] - P[Y_n = x] = P[Z_n = x, Y_n \neq x]$$

car :

$$\{Y_n = x\} \subset \{Z_n = x\}.$$

## 2.4. SIMULATION D'UNE LOI DE PROBABILITÉ

---

De plus :

$$\begin{aligned} \{Z_n \neq Y_n\} &= \{Z_n = x, Y_n \neq x\} \cup \{Z_n \neq x, Y_n = x\} \cup \dots \cup \{Z_n \neq z, Y_n = z\} \\ \Rightarrow \{Z_n = x, Y_n \neq x\} &\subset \{Z_n \neq Y_n\} \Rightarrow P[Z_n = x, Y_n \neq x] \leq P[Z_n \neq Y_n]; \\ \text{donc :} \end{aligned}$$

$$\begin{aligned} \mu^{(n)}(x) - \pi(x) &= P[Z_n = x, Y_n \neq x] \\ &\leq P[Z_n \neq Y_n] \\ &= P[T > n]. \end{aligned}$$

Donc :

$$\mu^{(n)}(x) - \pi(x) \leq P[T > n].$$

Comme :  $P[T > n] \rightarrow 0$ , (d'après conclusion 1) on a :

$$\begin{aligned} \mu^{(n)}(x) - \pi(x) \rightarrow 0 &\iff \lim_{n \rightarrow \infty} |\mu^{(n)}(x) - \pi_x| = 0. \\ \lim_{n \rightarrow \infty} |\mu^{(n)}(x) - \pi_x| = 0 &\iff \lim_{n \rightarrow \infty} \left( \frac{1}{2} \sum_{x \in E} |\mu^{(n)}(x) - \pi_x| \right) = 0 \\ &\iff d_{VT}(\mu^{(n)}, \pi) \rightarrow 0. \end{aligned}$$

□

Nous avons le théorème suivant (Cf.[1]).

**Théorème 2.** *Soit  $(X_n)_n$  une chaîne de Markov irréductible et apériodique sur un ensemble fini  $E$ , alors elle possède une unique loi stationnaire  $\pi$ .*

**Démonstration.** On démontre par l'absurde. Soit  $(X_n)_n$  une chaîne de Markov irréductible et apériodique de lois stationnaires  $\pi$  et  $\pi'$ , tel que  $\pi \neq \pi'$ .

On suppose que la loi initiale de  $(X_n)_n$  est  $\mu^{(0)}$ , telque  $\mu^{(0)} = \pi'$ . Si on désigne par  $\mu^{(n)}$  la loi de  $(X_n)_n$  à l'instant  $n$ , alors on a  $\mu^{(n)} = \pi'$ , car  $\pi'$  est la loi stationnaire de  $(X_n)_n$ .

D'après la proposition précédente, on a :

$$\mu^{(n)} \xrightarrow{VT} \pi, \quad n \rightarrow \infty$$

donc :

$$\lim_n d_{VT}(\mu^{(n)}, \pi) = 0.$$

Comme  $\mu^{(n)} = \pi'$  (hypothèse), donc :

$$\lim_n d_{VT}(\pi', \pi) = 0.$$

Or  $d_{VT}(\pi, \pi')$  ne dépend pas de  $n$ , donc :

$$\lim_n d_{VT}(\pi', \pi) = 0 \iff d_{VT}(\pi', \pi) \longrightarrow 0.$$

donc  $\pi = \pi'$  (contradiction).

□

D'après la proposition 8, on a alors :

$$\mu^{(n)}(j) \rightarrow \pi(j), \quad n \rightarrow \infty$$

donc on peut conclure que :  $\forall \epsilon > 0$ ,  $n$  grand,

$$|\mu^{(n)}(j) - \pi(j)| < \epsilon$$

c'est à dire, on peut prendre comme valeur de  $\pi(j)$ ,  $\mu^{(n)}(j)$ .

**Problème : Comment construire la chaîne  $(X_n)_n$  et donc construire des variables aléatoires de loi  $\pi$ ?**

## 2.5 Application : Algorithme de Metropolis

Soit  $\pi$  une loi de probabilité sur un ensemble fini  $E$ . On cherche à construire un échantillon de valeurs tirées selon la loi  $\pi$ . L'idée est de construire une chaîne de Markov qui admet  $\pi$  comme loi stationnaire (Cf.[2]).

On se donne une matrice de transition  $Q = (q_{xy})_{x,y \in E}$  sur  $E$ , appelée matrice de selection tel que :

$$q_{xy} > 0 \Rightarrow q_{yx} > 0, \quad \forall x, y \in E.$$

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

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

**Note :** On peut prendre :  $h(u) = \min(u, 1)$ .

Pour  $x \neq y$ , on pose :

$$\alpha(x, y) = \begin{cases} h\left(\frac{\pi_y q_{yx}}{\pi_x q_{xy}}\right) & \text{si } q_{xy} \neq 0; \\ 0 & \text{sinon.} \end{cases}$$

CHAPITRE 2. QUELQUES MÉTHODES DE SIMULATION D'UNE LOI DE PROBABILITÉS

---

Donc si  $h(u) = \min(u, 1)$ , alors on a :

$$\alpha(x, y) = \begin{cases} \min\left(\frac{\pi_y q_{yx}}{\pi_x q_{xy}}, 1\right) & \text{si } q_{xy} \neq 0; \\ 0 & \text{sinon.} \end{cases}$$

L'algorithme de Metropolis correspondant est le suivant :

**Etape 0** : Initialiser  $X_0$

**Etape  $n + 1$**  :

(Selection) Choisir  $y$  avec la loi  $Q(X_n, y)$

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

Si  $U < \alpha(X_n, y)$  accepter la selection :

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

Sinon, refuser la selection :

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

**Etude de la chaîne construite :**

$(X_n)_n$  est une chaîne de Markov de matrice de transition  $P$  définie par :

$$p_{xy} = \begin{cases} q_{xy} \alpha(x, y) & \text{si } x \neq y; \\ 0 & \text{sinon.} \end{cases}$$

Alors on a la proposition suivante (Cf.[2]) :

**Proposition 9.** *La matrice de transition  $P$  définie ci-dessus, est réversible par rapport à  $\pi$ .*

*Si  $Q$  est irréductible, alors  $P$  est irréductible. Si de plus  $h(u) < 1$ , alors elle est apériodique.*

**Exemple 3.** *Utilisons l' algorithme de Metropolis à marche aléatoire équiprobable pour simuler une loi de Poisson de paramètre  $\lambda$ .*

*On cherche à simuler la loi de Poisson,*

$$\pi_x = \exp(-\lambda) \frac{\lambda^x}{x!}, \quad \text{pour } x \in \mathbb{N}.$$

*Dans le cas de la marche aléatoire,  $Q(x, y) = q(y - x)$  ; donc :*

$$\alpha = \min\left(1, \frac{\pi_y q(x - y)}{\pi_x q(y - x)}\right).$$

*Si  $q$  est symétrique,  $q(t) = q(-t)$ , on a :*

$$\alpha(x, y) = \min\left(1, \frac{\pi_y}{\pi_x}\right) = \min\left(1, \lambda^{y-x} \frac{x!}{y!}\right).$$



## CHAPITRE 2. QUELQUES MÉTHODES DE SIMULATION D'UNE LOI DE PROBABILITÉS

---

On considère ici le cas de la marche aléatoire, dont la matrice de transition est alors donnée par :

$$q_{xy} = \frac{1}{2}\delta_{x-1}(y) + \frac{1}{2}\delta_{x+1}(y).$$

L'algorithme s'écrit alors sous la forme suivante :

**Etape 0** : Initialiser  $X_0$ ;

**Etape  $n + 1$**  :

space\*2.3cm On pose  $y = x_n - 1$  ou  $y = x_n + 1$  avec probabilité  $\frac{1}{2}$  ;

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

Si  $U < \min(1, \lambda^{y-X_n} \frac{X_n!}{y!})$  accepter la selection :

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

Sinon, refuser la selection :

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

## 2.5. APPLICATION : ALGORITHME DE METROPOLIS

---

# Chapitre 3

## Simulation exacte : L'algorithme de Propp-Wilson

Les méthodes classiques de simulation présentées dans le chapitre précédent, sont en général basées sur des méthodes d'approximation d'une loi de probabilité. En 1996, Propp et Wilson (P-W) ont développé une nouvelle méthode (algorithme) (Cf. [6] et [7]), dite de simulation *exacte ou parfaite*. Cette méthode rompt avec la conception classique (approximation) car elle donne la possibilité de construire un échantillon distribué exactement suivant la loi cible. Elle est fondée sur les propriétés des chaînes de Markov itératives.

### 3.1 Présentation de l'algorithme de Propp et Wilson

Soit  $E$  un ensemble fini et  $\pi$  une loi de probabilités sur  $E$ . L'algorithme de P-W est basé sur la construction d'une chaîne de Markov  $(X_n^x)_n$  itérative de loi stationnaire  $\pi$  (Cf. [8]), et du processus  $(H_n^x)_n$  associé appelé "*backward process*". Plus précisément :

- Soit  $(Y_n)_n$  une suite de variables aléatoires i.i.d. à valeurs dans  $E$ , et de loi  $\mu$ .
- Pour  $x, y \in E$ , soit la fonction  $x \mapsto f_Y(x)$  définie de  $E$  dans  $E$ .

### 3.1. PRÉSENTATION DE L'ALGORITHME DE PROPP ET WILSON

---

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

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

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

**Définition 7.** On appelle "backward process" associé à  $(X_n^x)_n$ , le processus  $(H_n^x)_n$  définie par :

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

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

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

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

alors  $T$  est fini, p.s.

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

$$\begin{aligned} Z_1 &= \lim_n H_n^{x_1}, \\ Z_2 &= \lim_n H_n^{x_2}, \\ &\vdots \\ Z_n &= \lim_n H_n^{x_n}, \end{aligned}$$

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

### 3.2 Validité du principe de l'algorithme

Diaconis et Freedman ont donné les conditions (Cf. [8]), pour que  $\lim_n H_n^x$  existe, et que cette limite soit une constante.

Soit  $(E, d)$  un espace métrique,  $(f_y)_{y \in E}$  une famille de fonctions Lipschitzienne de rapport  $K_y$ , c'est à dire :

$$\forall x, x' \in E, \quad d(f_y(x), f_y(x')) \leq K_y d(x, x').$$

Selon Diaconis et Freedman (Cf. [8]), si on a les conditions suivantes :

1. Soit la fonction  $f \mapsto K_f$  tel que :

$$\forall u > 0, \exists \alpha, \beta > 0, \text{ telque : } \mu\{K_f > u\} < \frac{\alpha}{u^\beta}.$$

2. Soit la fonction  $f \mapsto A(f) = d(f(x_0), x_0)$ ,  $x_0 \in E$ , tel que :

$$\forall u > 0, \exists \alpha, \beta > 0, \text{ telque : } \mu\{d(f(x_0), x_0) > u\} < \frac{\alpha}{u^\beta}.$$

3. Les fonctions  $f_y$  sont contractantes en moyenne, c'est à dire :

$$\int_E \log(K_f) \mu(dy) < 0.$$

### 3.2. VALIDITÉ DU PRINCIPE DE L'ALGORITHME

---

alors on a :

1. La CM  $(X_n^x)_n$  admet une unique distribution stationnaire.
2. Le processus  $(H_n^x)_n$  converge p.s et sa limite est une constante.

#### Remarque :

Si l'algorithme de P-W s'arrête, alors le résultat est la variable aléatoire

$Z = H_T^x(x)$ , tel que :

$$T = \inf\{n \in \mathbb{N} : H_n^x = cte\}.$$

C'est le temps de coalescence.

Dans la proposition suivante, Benaïme (Cf.[2]), montre que  $T$  est fini p.s.

**Proposition 10.** *Si  $(X_n^x)_n$  est une chaîne de Markov irréductible et apériodique, et si la variable aléatoire  $Z$  est distribuée selon la loi stationnaire  $\pi$ , alors la variable aléatoire  $T$  est fini p.s.*

Soit  $\mathcal{F} = E^E = \{f : E \rightarrow E\}$ ,  $h \in \mathcal{F}$ , et soit  $(H_n)_n$  le processus défini sur  $\mathcal{F}$  par :

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

Alors  $(H_n)_n$  est une chaîne de Markov sur  $\mathcal{F}$  (Cf.[3]).

La proposition de Benaïme peut se formuler alors de la façon suivante :

**Proposition 11.** *Si  $E$  est fini, alors les seules classes récurrentes pour  $(H_n)_n$  sont les constantes.*

**Démonstration.** Remarquons que si  $h : E \rightarrow E$  est constante, avec :

$$h = h_a, \text{ avec } a \in E; \text{ et } h_a(x) = a, \forall x \in E,$$

alors si :  $H_n = h_a$ , on a :

$$H_{n+1} = h_a \circ H_n = h_a;$$

d'où  $\{h_a\}$  est un ensemble fermé, donc  $h_a$  est un point absorbant. Donc :

$$\text{Si } H_n = h_a, \text{ alors } H_{n+k} = h_a, \forall k \geq 1.$$

Comme la chaîne de Markov  $(H_n)_n$  est finie, donc elle est absorbée dans une classe récurrente au bout d'un temps fini. Ce qui est l'implication dans un sens. Dans l'autre nous n'arrivons pas à conclure.

□

## *Conclusion*

Dans ce mémoire nous avons introduit et résumé les principales méthodes de simulation de lois de Probabilités. Notre but est une introduction à l'algorithme de simulation exacte de Propp-Wilson. L'étude des conditions de validité de cet algorithme est une motivation à l'étude des chaînes de Markov itératives. Les perspectives qui s'ouvrent sont nombreuses et de deux types :

1. Trouver des systèmes itératifs explicites qui vérifient les conditions de validité de l'algorithme en liaison avec un temps moyen d'arrêt acceptable ( temps moyen que met le backward process pour être constant ) ;
2. Les conditions de Diaconis et Freedman sont assez lourdes, il s'agit de trouver des systèmes itératifs avec des conditions moins contraignantes ;

Ces deux directions de travail nous semblent retenir le plus d'attention dans les travaux actuels (Cf. [13],...).



# Bibliographie

- [1] Häggström, O. *Finite Markov chains and algorithmic applications*. Cambridge University Press 2002.
- [2] Michel Benaïm et Nicole El Karoui. *Promenade aléatoire, Chaînes de Markov et simulations ; martingales et stratégies*. Editions de l'école Polytechnique - Novembre 2007.
- [3] Bernard. Ycart. *Algorithmes markoviens*. Département de l'ingénieur de mathématique. Université de Chile. Decembre 1997.
- [4] LACHAUD Béatrice. *Détection de la convergence de processus de Markov*. Thèse de doctorat. Option probabilité. Université RENE DESCARTES - PARIS 5. Septembre 2005.
- [5] David A. Levin, Yuval Peres, Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. 2005. <http://www.uoregon.edu/~dlevin>.
- [6] J.G. Propp, D.B. Wilson. *Exact sampling with coupled Markov chains and application to statistical mechanics*, Random structures and algorithms, 9, pp. 223-252, 1996.
- [7] J.G. Propp, D.B. Wilson. *How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph*. Journal of Algorithms 27, 170-217, 1998.
- [8] Persi Diaconis et David Freedman. *Iterated Random Functions*. SIAM Review, 41, 1, pp. 45- 76, 1999.
- [9] B. BESSAD, F. LADJIMI, M.A. BOUDIBA. *Itération de fonctions aléatoires et application à la simulation*. Communication en workshop, EP-QOS, Bejaia. Mai 2013.
- [10] J.N. Corcoran, R.L. Tweedie *Perfect sampling from independent Metropolis-Hastings chains*. Journal of Statistical Planning and Inference 104, 297-314, 2002.

*BIBLIOGRAPHIE*

---

- [11] A.N. Shiryaev. *Probability*. Second Edition. Springer. 1989.
- [12] P. Diaconis et L. Saloff-Coste. *What Do We Know about the Metropolis Algorithm ?* Journal of Computer and System Sciences 57, 20-36 (1998).
- [13] KRISHNA B. ATHREYA et ORJAN STENFLO. *Perfect sampling for Doblin chains*. (Prepublication).