

# Introduction aux méthodes de Monte Carlo par dynamique Hamiltonienne

Shmuel RAKOTONIRINA-RICQUEBOURG, Amaury DURAND

2 novembre 2017

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Méthodes de Monte Carlo	1
1.2	Monte Carlo Markov Chains (MCMC)	1
1.2.1	Loi invariante et réversibilité	1
1.2.2	Ergodicité	2
1.2.3	Algorithme de Metropolis-Hastings	2
<b>2</b>	<b>Hamiltonian Monte Carlo</b>	<b>2</b>
2.1	Principe général	2

## 1 Introduction

Ce rapport présente le travail effectué lors d'un projet du cours d'approfondissements en chaînes de Markov par Eric Moulines dans le cadre du master Mathématiques de l'aléatoires à l'université Paris-Sud. Le contenu présenté ci-dessous repose essentiellement sur [1] et [2].

Les méthodes de Monte Carlo par dynamique Hamiltonienne, plus communément appelées Hamiltonian Monte Carlo (HMC), font partie d'une grande famille de méthodes de simulation : les méthodes de Monte Carlo, et plus précisément dans la famille des méthodes de Monte Carlo par chaînes de Markov (ou Monte Carlo Markov Chains). Ces méthodes se placent dans le cadre suivant :

**Définition 1.1** (Carde général). Soit  $(\mathbb{X}, \mathcal{X})$  un espace mesurable. Soit  $\pi$  une loi de probabilité sur cet espace. On suppose que  $\pi$  n'est connue qu'à un facteur de proportionnalité près i.e on connaît  $\lambda\pi$  où  $\lambda \in \mathbb{R}$  une constante. Le but est de pouvoir approcher  $\pi f = \mathbb{E}[f(X)]$  avec  $X \sim \pi$  et  $f \in \mathbb{F}_+(\mathbb{X}, \mathcal{X}) \cup \mathbb{F}_b(\mathbb{X}, \mathcal{X})$ .

### 1.1 Méthodes de Monte Carlo

La méthode de Monte Carlo la plus simple (appelée Monte Carlo naïf) est la suivante : supposons que l'on sait simuler des variables aléatoires de loi  $\pi$  alors en tirant  $n$  échantillons  $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \pi$ , la loi des grands nombre nous indique qu'une bonne approximation de  $\pi f$  est  $\frac{1}{n} \sum_{i=1}^n f(X_i)$ . D'autres méthodes permettent d'atteindre le même but en ne sachant pas simuler de variables aléatoires de loi  $\pi$ . C'est le cas par exemple de l'échantillonnage d'importance qui consiste à simuler des variables i.i.d sous une loi différente de  $\pi$ .

Dans ces deux cas, l'approximation de  $\pi f$  repose sur une simulation de variables aléatoires i.i.d. Cette particularité permet alors de montrer des résultats de convergence notamment grâce à la loi des grands nombre. Les méthodes de Monte Carlo par chaînes de Markov ne reposent pas sur le caractère i.i.d des variables mais sur les propriétés des Chaînes de Markov.

### 1.2 Monte Carlo Markov Chains (MCMC)

Les méthode MCMC se basent sur les notions de loi invariante, de réversibilité et d'ergodicité.

#### 1.2.1 Loi invariante et réversibilité

On considère  $P$  un noyau de Markov sur  $\mathbb{X} \times \mathcal{X}$ .

**Définition 1.2.** Une loi de probabilité  $\pi$  sur  $(\mathbb{X}, \mathcal{X})$  est dite

- $P$ -invariante si  $\pi P = \pi$
- $P$ -réversible si  $\forall A, B \in \mathcal{X}, \pi \otimes P(A \times B) = \pi \otimes P(B \times A)$

**Proposition 1.1.** *Toute probabilité  $P$ -réversible est  $P$ -invariante.*

### 1.2.2 Ergodicité

**Théorème 1.1.** Soit  $(X_k)_{k \in \mathbb{N}}$  une chaîne de Markov de noyau  $P$  admettant une loi invariante  $\pi$ . Alors pour tout  $f \in \mathbb{F}_+(\mathbb{X}, \mathcal{X}) \cup \mathbb{F}_b(\mathbb{X}, \mathcal{X})$  et pour  $\pi$ -presque tout  $x \in \mathbb{X}$ ,

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}_x - p.s.} \pi f$$

Les algorithmes MCMC visent à construire, à partir de  $\pi$ , une chaîne de Markov vérifiant les hypothèses de ce théorème afin d'approcher  $\pi f$  par  $\frac{1}{n} \sum_{k=0}^{n-1} f(X_k)$ .

### 1.2.3 Algorithme de Metropolis-Hastings

L'algorithme HMC est une version améliorée de l'algorithme de Metropolis-Hastings. Nous allons donc d'abord définir ce dernier et nous l'utiliserons plus tard comme base de comparaison. On se place dans le cas où  $\mathbb{X} = \mathbb{R}^d$  et  $\mathcal{X} = \mathcal{B}(\mathbb{R}^d)$ .

On suppose que  $\pi$  a une densité  $h_\pi$  par rapport à une mesure  $\mu$ . On considère de plus un noyau markovien  $Q$  sur  $(\mathbb{X}, \mathcal{X})$  de densité  $q$  par rapport à  $\mu$ .  $Q$  est appelé noyau instrumental. La construction de la chaîne de Markov se fait en proposant un mouvement via  $Q$ , puis en acceptant ou en rejetant ce mouvement (algorithme 1.1).

---

#### Algorithme 1.1 : Random Walk Metropolis

---

$X_0 \leftarrow x \in \mathbb{X}$  arbitraire;

**répéter**

$Y_{k+1} \sim Q(X_k, \cdot)$  ;

*/\* Proposer un mouvement \*/*

$\alpha_{k+1} \leftarrow \alpha(X_k, Y_{k+1})$  où  $\alpha(x, y) = 1 \wedge \frac{h_\pi(y)q(y, x)}{h_\pi(x)q(x, y)}$ ;

$X_{k+1} \leftarrow \begin{cases} Y_{k+1} & \text{avec probabilité } \alpha_{k+1} \\ X_k & \text{with probability } 1 - \alpha_{k+1} \end{cases}$  ;

*/\* Accepter ou rejeter le mouvement \*/*

**jusqu'à une condition d'arrêt;**

---

Dans cet algorithme, on cherche à explorer l'espace  $\mathbb{X}$  en visitant moins souvent les régions où  $h_\pi$  est faible (peu chargées par  $\pi$ ) car ce sont les régions de forte probabilité qui donnent des informations sur la loi  $\pi$ . Ainsi, si le mouvement proposé vérifie  $h_\pi(Y_{k+1}) \geq h_\pi(X_k)$ , on se dirige vers une zone plus chargée par  $\pi$ , on accepte donc le mouvement. Dans le cas contraire, il est moins intéressant de bouger. On autorise quand même le mouvement avec une probabilité  $\frac{h_\pi(Y_{k+1})}{h_\pi(X_k)}$  d'autant plus faible que la position proposée est dans une région de probabilité faible.

*Remarque 1.1.* Il y a d'autres fonctions de rejet  $\alpha$  qui permettent à l'algorithme de fonctionner. De même, le choix du noyau  $Q$  est un degré de liberté de l'algorithme.

Le choix de  $Q$  est en fait une difficulté de l'algorithme, puisque le mouvement proposé doit permettre l'exploration de tout l'espace (afin de passer régulièrement par toutes les régions fortement chargées par  $\pi$ ). Nous verrons que l'algorithme HMC résout ce problème en proposant le mouvement selon une *dynamique hamiltonienne*.

**Définition 1.3.** On parle de *Random-walk Metropolis* quand  $Q$  est le noyau d'une marche aléatoire

$$Y_{k+1} \leftarrow X_k + U_{k+1}$$

avec  $(U_k)_{k \in \mathbb{N}}$  iid.

## 2 Hamiltonian Monte Carlo

On se place dans le cas où  $\mathbb{X} = \mathbb{R}^d$ ,  $\mathcal{X} = \mathcal{B}(\mathbb{R}^d)$  et  $\mu$  est la mesure de Lebesgue.

L'algorithme HMC fait partie des algorithmes MCMC. Il diffère de l'algorithme de Metropolis-Hastings en ce que le mouvement proposé ne sera pas tiré selon une chaîne de Markov de noyau  $Q$  mais selon une *dynamique hamiltonienne* associée à la loi à simuler.

### 2.1 Principe général

La dynamique hamiltonienne décrit le mouvement d'un objet qui glisse sans frottement le long d'une surface ou d'une courbe. L'objet est décrit par sa position  $x \in \mathbb{R}^d$  et sa vitesse  $v \in \mathbb{R}^d$  et on lui associe des énergies potentielle  $U(x)$  et cinétique  $K(v)$ . En physique, une énergie  $E$  et une température  $T$  est associée à une loi de probabilité par la loi de Boltzmann

$$\forall x \in \mathbb{R}^d, p(x) \propto \exp\left(-\frac{E(x)}{T}\right) \quad (1)$$

Ici on choisit donc l'énergie potentielle pour avoir

$$\forall x \in \mathbb{R}^d, h_\pi(x) \propto \exp\left(-\frac{U(x)}{T}\right) \quad (2)$$

En dimension 1, on décrit le glissement sans frottement d'un objet sur une rampe. Comme l'énergie potentielle est proportionnelle à la hauteur de la rampe, (2) donne que les creux de la rampe représentent les régions de forte probabilité. L'objet aura tendance à glisser vers les creux de la rampe, mais peut remonter une pente si sa vitesse  $v$  (et donc son énergie cinétique  $K(v)$ ) est assez grande.

L'algorithme HMC propose ses mouvements selon le principe suivant

- Considérer un objet placé en  $x = X_k$  de vitesse aléatoire  $v$  (de loi à préciser)
- Simulation la dynamique hamiltonienne (déterministe) de cet objet pendant une durée fixée, pour l'énergie potentielle  $U$  donnée par 2
- Considérer  $Y_{k+1}$  la position de cet objet à la fin de la simulation, et accepter ou rejeter le mouvement selon une fonction de rejet  $\alpha$  dépendant de  $U$

## Références

- [1] Neal Radford M. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 54 :113–162, 2010.
- [2] Douc Randal ; Moulines Eric ; Priouret Pierre ; Soulier Philippe. *Markov Chains*. 2017.