

# MACS 205 Projet 4 : Metropolis Hastings adaptatif

Rafael Brutti, Priam Cardouat, Caroline Xia

26 avril 2021

## Table des matières

<b>1</b>	<b>Metropolis Hastings classique</b>	<b>3</b>
1.1	Description . . . . .	3
1.2	Intérêt . . . . .	3
1.3	Quelques exemples . . . . .	5
1.4	Problèmes . . . . .	8
<b>2</b>	<b>Metropolis Hastings adaptatif (Haario et al.)</b>	<b>9</b>
2.1	Description . . . . .	9
2.2	Avantages de la méthode . . . . .	10
2.3	Quelques exemples . . . . .	10
<b>3</b>	<b>Métropolis-Hastings Mixture</b>	<b>12</b>
3.1	Principe . . . . .	12
3.2	Résultats . . . . .	13
<b>4</b>	<b>Metropolis Hastings avec l'estimateur de densité à noyau</b>	<b>15</b>
4.1	Description . . . . .	15
4.2	Quelques exemples . . . . .	15
<b>5</b>	<b>Metropolis Hastings avec l'estimateur de densité à noyau adaptatif</b>	<b>17</b>
5.1	Principe . . . . .	17
5.2	Illustration . . . . .	17

<b>6</b>	<b>Tests et comparaison</b>	<b>18</b>
6.1	Test avec une double gaussienne . . . . .	18
6.2	Comparaison . . . . .	19
<b>7</b>	<b>Conclusion</b>	<b>22</b>
<b>8</b>	<b>Pour l'avenir...</b>	<b>22</b>
<b>9</b>	<b>Références bibliographiques</b>	<b>23</b>

# 1 Metropolis Hastings classique

## 1.1 Description

Nous allons rapidement rappeler le principe de la méthode de Hasting Metropolis. Il s'agit d'une méthode de Monte Carlo par chaînes de Markov qui permet d'obtenir un échantillon à partir d'une densité  $f$  lorsque l'échantillonnage direct est compliqué.

On a une densité cible  $f : R^d \rightarrow R_+$  et un noyau de Markov  $q : R^d \rightarrow R^d$ .  $X_0$  est tiré selon une distribution initiale  $\nu$ . On tire les  $X_i$  de la manière suivante :

Etape 1 : Tirage de  $Y : Y \sim q(\cdot, X_{i-1})$

Etape 2 : 2 possibilités : Acceptation de  $Y$  et  $X_i = Y$  ou rejet et  $X_i = X_{i-1}$

L'acceptation pour le point  $X_i$  suit une loi de Bernoulli de paramètre :

$$p_i = \min\left(1, \frac{f(Y)q(X_{i-1}, Y)}{f(X_{i-1})q(Y, X_{i-1})}\right)$$

et  $p_i = 1$  si  $f(X_{i-1})q(Y, X_{i-1}) = 0$

## 1.2 Intérêt

La méthode de Metropolis Hastings est une méthode de Monte Carlo par Chaîne de Markov (MCMC) : il s'agit de générer une chaîne de Markov de loi stationnaire  $f$ .

Il faut trouver des probabilités de transition :

$$P(x, dy) = P(X_{j+1} \in dy \mid X_j = x)$$

telle que :

$$\forall y \in \mathcal{X} \quad \int_{x \in \mathcal{X}} f(dx) \cdot P(x, dy) dx = f(dy)$$

En effet, cela découle de certaines propriétés des chaînes de Markov.

**Définition :** Une chaîne de Markov avec probabilité de transition  $P(x, dy)$  sur un espace  $\mathcal{X}$  est dite réversible par rapport à une distribution  $\pi(\cdot)$  si :

$$\pi(dx) \cdot P(x, dy) = \pi(dy) \cdot P(y, dx) \quad \forall x, y \in \mathcal{X}$$

**Propriété :** Une mesure réversible est invariante.

Preuve :

$$\begin{aligned}\int_{x \in \mathcal{X}} \pi(dx) \cdot P(x, dy) dx &= \int_{x \in \mathcal{X}} \pi(dy) \cdot P(y, dx) dx \\ &= \pi(dy) \int_{x \in \mathcal{X}} P(y, dx) dx = \pi(dy)\end{aligned}$$

Générer une chaîne de Markov de loi stationnaire  $f$  est donc possible si notre chaîne de Markov est réversible par rapport à  $f$  grâce à cette propriété. Ainsi, la loi de Bernoulli de paramètre  $p_i = \min(1, \frac{f(Y)q(X_{i-1}, Y)}{f(X_{i-1})q(Y, X_{i-1})})$  rentre en jeu.

### **Hypothèse sur le noyau**

On suppose que le noyau de transition s'écrit de la façon suivante :

$$P(x, dy) = p(x, y)dy + r(x)\delta_x(dy)$$

Le deuxième terme provient du fait que la probabilité de rester au point  $X_i$  n'est pas nécessairement nulle.

Avec les notations de Metropolis-Hastings, on obtient alors :

$$P_i(x, dy) = q(y, x) \cdot p_i(x, y)dy + \delta_x(dy) \cdot (1 - \int_{x \in \mathcal{X}} q(y, x) \cdot p_i(x, y)dy)$$

Dans le cas où  $x=y$ , on a que la loi est réversible.

Dans le cas maintenant où  $x$  est différent de  $y$ , on a :

$$\begin{aligned}f(dx) \cdot P_i(x, dy) &= f(x)dx \cdot q(y, x) \cdot p_i(x, y)dy \\ &= f(x) \cdot q(y, x) \cdot \min(1, \frac{f(y)q(x, y)}{f(x)q(y, x)})dx dy \\ &= \min(f(x) \cdot q(y, x), f(y)q(x, y))dx dy \\ &= f(dy) \cdot P_i(y, dx)\end{aligned}$$

Ainsi, dans les deux cas,  $f$  est réversible donc invariante par rapport à la probabilité de transition  $P$ . Une telle loi  $f$  était recherchée, toutefois, cela

ne suffit pas. En effet, une loi stationnaire  $f$  n'est pas forcément unique et la loi de la chaîne de Markov associée ne converge pas forcément vers  $f$ . Un théorème sur les chaînes de Markov stipule que si la chaîne est irréductible et apériodique avec loi stationnaire  $f$  alors elle converge asymptotiquement vers  $f$ . Ces conditions seront vérifiées dans nos cas pratiques.

Finalement, une caractéristique de l'approche Metropolis-Hastings est le fait que la constante de normalisation de la densité cible ne doit pas nécessairement être connue puisque cette densité n'apparaît que sous forme de ratio dans l'expression de  $p_i$ . Dans le contexte bayésien, cette constante de normalisation est difficilement calculable pour une fonction de densité à postériori donc la méthode de Metropolis-Hastings est très attrayante.

### 1.3 Quelques exemples

Nous avons donc cherché à illustrer et représenter le comportement de cette méthode sur quelques exemples. Pour des raisons liées au temps de calcul, nous nous sommes limités à des densités en dimension 2, même si les codes utilisés fonctionnent pour une dimension  $d$  quelconque.

Dans un premier temps, nous avons considéré une densité cible qui était une gaussienne de variance  $(1, 1)$  et de moyenne  $(4, 4)$ .

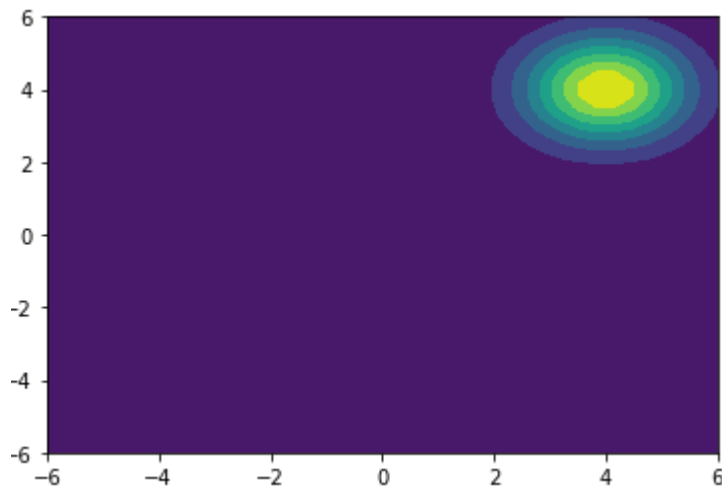


FIGURE 1 – Densité cible

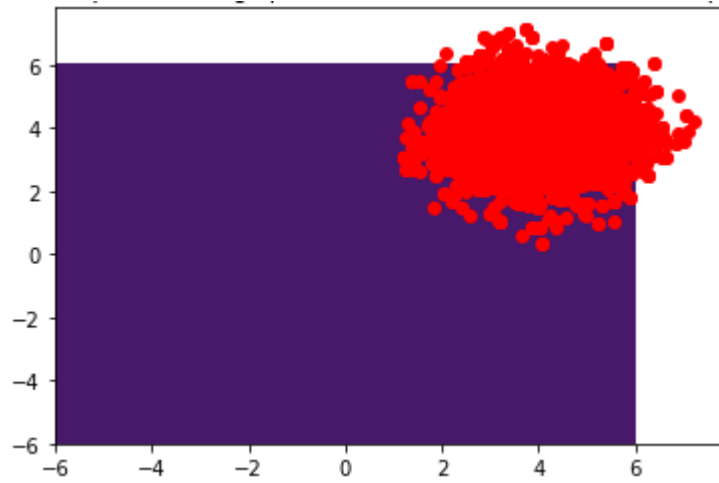


FIGURE 2 – Metropolis Hastings pour 5000 échantillons et bon départ en  $(4,4)$

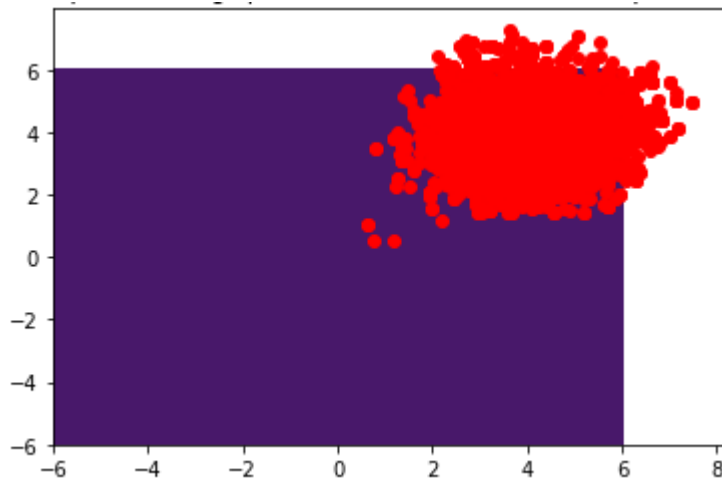


FIGURE 3 – Metropolis Hastings pour 5000 échantillons et départ à froid en  $(0,0)$

La distribution initiale est une loi normale de covariance la matrice identité et de moyenne  $(4,4)$  dans le premier cas et  $(0,0)$  dans le second. On parlera de "cold start" dans le deuxième cas, compte tenu du fait que le premier échantillon tiré est "loin" de la distribution cible. Nous pouvons dès

lors voir sur la figure 3 que cette méthode est sensible au problème du cold start.

Par ailleurs, nous avons cherché à illustrer son comportement dans le cas où la densité cible est une double gaussienne.

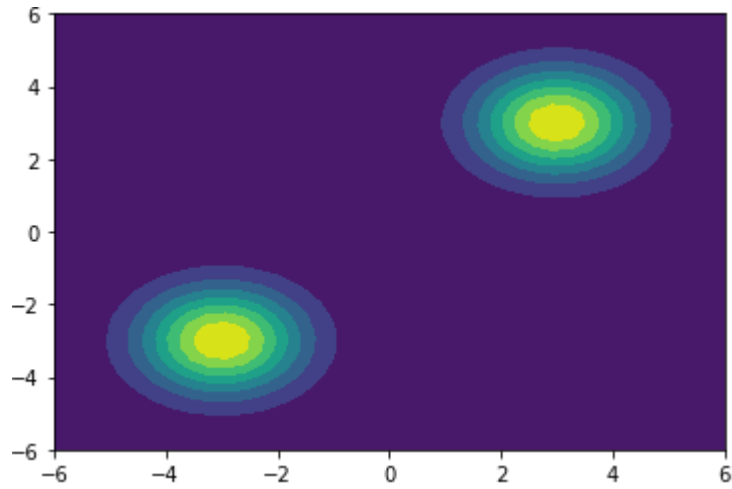


FIGURE 4 – Densité cible

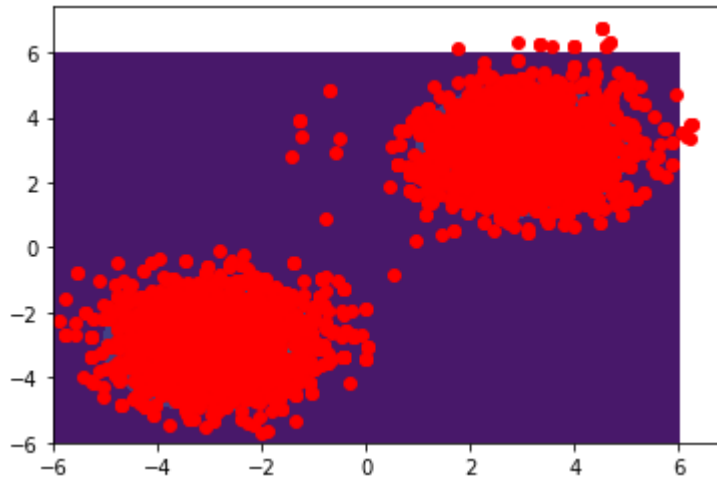


FIGURE 5 – Metropolis Hastings pour 5000 échantillons, distribution initiale suivante une loi normale centrée réduite

Tant que les deux gaussiennes ne sont pas trop éloignées, la méthode fonc-

tionne encore relativement correctement (figure 5). En revanche dès qu'elles commencent à s'éloigner (figure 6), la méthode ne va généralement que vers une seule gaussienne et s'y "enferme" pour ne plus en ressortir, ce problème étant d'autant plus important dans le cadre d'un cold start.

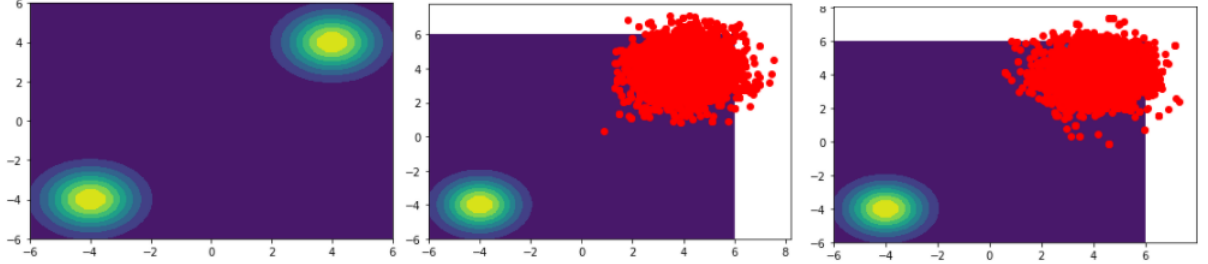


FIGURE 6 – Metropolis Hastings pour 5000 échantillons et une densité cible une double gaussienne de moyennes (4,4) et (-4,-4), distribution initiale selon une loi normale centrée réduite

## 1.4 Problèmes

A la vue des illustrations du comportement de la méthode précédente, nous pouvons relever quelques problèmes de cette méthode :

1. Metropolis Hastings classique ne visite pas tout l'espace (exemple de la double gaussienne sur la fig 6) : une fois que l'on tire des échantillons dans une gaussienne, nous allons plus en sortir, notamment dû au fait que la "variance" est fixe et donc la "distance" jusqu'où on peut tirer est limitée.

2. Une possibilité pour le noyau est de prendre par exemple  $X_i \sim \mathcal{N}(X_{i-1}, 1)$ . L'algorithme ne s'adapte alors pas à la forme de la densité cible car la variance est constante (cf Haario et al.)

3. La méthode est sensible au choix de la distribution initiale (problème du cold start).

Afin de résoudre au moins le premier problème, Haario et al. ont proposé une version adaptative de la méthode précédente que nous allons maintenant décrire.



## 2 Metropolis Hastings adaptatif (Haario et al.)

### 2.1 Description

L'idée de l'algorithme de Metropolis adaptatif est de mettre à jour la distribution pour choisir le point suivant en utilisant toutes les connaissances ultérieures sur la densité cible. C'est-à-dire que la variance est calculée à chaque étape selon les tirages précédents. Ainsi, la méthode est moins "rigide" et nous nous sommes plus limités par une variance fixe, mais une qui dépend des tirages précédents. Par ailleurs, on ne travaille plus avec une chaîne de Markov. Sinon, la définition de l'algorithme est identique au processus habituel de Metropolis.

On a une densité cible  $f : R^d \rightarrow R_+$  et à l'étape  $i$ , on a déjà tiré  $X_0, X_1, \dots, X_{i-1}$  :

Etape 1 : Tirage de  $Y : Y \sim q_i(\cdot | X_0, X_1, \dots, X_{i-1})$

Etape 2 : 2 possibilités : Acceptation de  $Y$  et  $X_i = Y$  ou rejet et  $X_i = X_{i-1}$

L'acceptation pour le point  $X_i$  suit une loi de Bernoulli de paramètre :

$$p_i = \min(1, \frac{f(Y)}{f(X_{i-1})})$$

La densité  $q_i(\cdot | X_0, X_1, \dots, X_{i-1})$  suit une distribution normale d'espérance  $X_{i-1}$  et de covariance  $C_i$  qui dépend de  $X_0, X_1, \dots, X_{i-1}$ . Dans l'algorithme étudié :

$$C_i = \begin{cases} C_0 & \text{si } t \leq t_0 \\ s_d(\text{Cov}(X_0, X_1, \dots, X_{i-1}) + \epsilon I_d) & \text{sinon.} \end{cases}$$

Après calcul :

$$C_{i+1} = \frac{i-1}{i} C_i + \frac{s_d}{i} (i \bar{X}_{i-1} \bar{X}_{i-1}^\top - (i+1) \bar{X}_i \bar{X}_i^\top + X_i X_i^\top + \epsilon I_d)$$

Avec :  $\bar{X}_k = \frac{1}{k+1} \sum_{i=0}^k X_i$

Ce qui permet d'obtenir les covariances sans trop de calculs car les moyennes  $\bar{X}_k$  se calculent récursivement.

Concernant le choix des différents paramètres :

- Le choix de  $t_0$  est libre, cependant plus  $t_0$  est grand plus tard l'effet adaptatif aura lieu. Cela dépend de la justesse du choix du  $C_0$  initial par rapport à notre densité cible
- On prend  $s_d = \frac{2.4^2}{d}$  qui est une valeur qui optimise la méthode pour des densités cibles gaussiennes
- Le choix de  $\epsilon > 0$  petit permet d'éviter que la matrice  $C$  soit singulière (non inversible) et permet l'ergodicité de la chaîne.

## 2.2 Avantages de la méthode

L'avantage de cette méthode est que le tirage des points comme défini précédemment permet d'avoir l'ergodicité de la chaîne. L'élément clé est le fait que la matrice  $C_t$  comme défini précédemment est inversible grâce au  $\epsilon > 0$ .

Théorème (preuve dans le papier de Haario et al.) :

Soit  $\pi$  la densité cible défini sur  $S \subset R^d$  borné et mesurable, et supposons que  $\pi$  est bornée sur  $S$ . Soit  $\epsilon > 0$  et  $\mu_0$  une distribution initiale sur  $S$ . Soit  $(X_i)$  un tirage fait selon l'algorithme de Metropolis adaptatif comme défini dans la partie 2.1. Alors le tirage simule bien la densité cible et on a pour toute fonction bornée et mesurable  $f : S \rightarrow R$ , on a :

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} (f(X_0) + \dots + f(X_n)) = \int_S f(x) \pi(dx) \quad p.s.$$

## 2.3 Quelques exemples

De même que précédemment, nous avons illustré le comportement de la méthode pour une gaussienne simple, selon un bon départ ou non. La matrice de covariance initiale est dans les deux cas une matrice identité multipliée par un scalaire  $h$ . En choisissant  $h$ , nous choisissons notre covariance initiale. Enfin, nous avons pris  $\epsilon = 0.5$ . Nous avons ici retenu une matrice identité comme covariance initiale et une période initiale de 100 tirages. Le résultat est donnée par la figure 7.

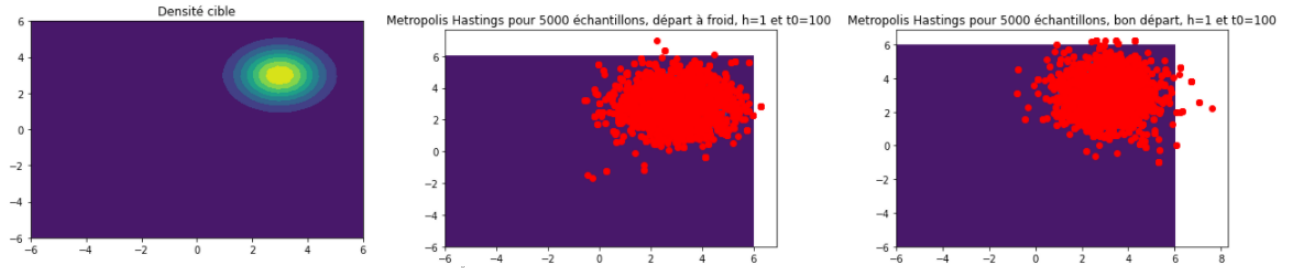


FIGURE 7 – Metropolis Hastings adaptatif pour 5000 échantillons, période initiale de 100 itérations

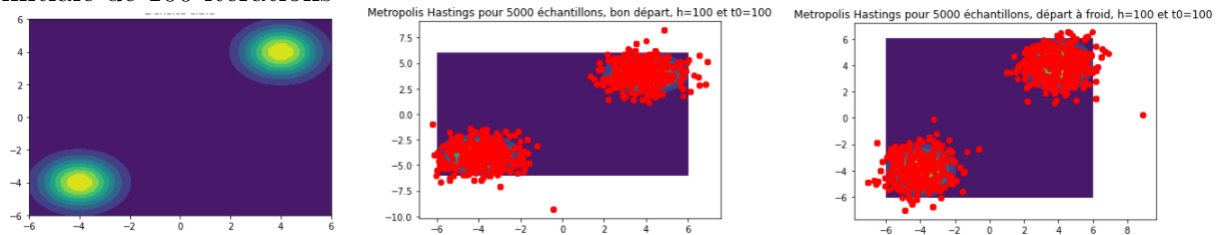


FIGURE 8 – Metropolis Hastings adaptatif pour 5000 échantillons, période initiale de 100 itérations

En ce qui concerne la double gaussienne, nous avons choisi de retenir une double gaussienne avec une moyenne  $(4, 4)$  et une moyenne  $(-4, -4)$ . La covariance initiale est une matrice identité multipliée par 100 et la période initiale est de 100 tirages. la valeur de  $\epsilon$  est la même que précédemment. Nous pouvons déjà voir sur la figure 8 que la méthode parvient à parcourir tout l'espace nécessaire, et ce en dépit de la qualité du départ, ce qui est déjà un grand pas en avant en comparaison à la méthode de Metropolis Hastings classique. Par ailleurs, s'il venait à arriver que les moyennes soient plus "écartées", augmenter la période initiale permet quand même d'obtenir un bon résultat.

## 3 Métropolis-Hastings Mixture

### 3.1 Principe

Nous allons présenter désormais une nouvelle méthode de Métropolis-Hastings. Tout d'abord, donnons quelques définitions des appellations des noyaux classiques utilisés lors de la méthode de Métropolis-Hastings.

#### Définitions :

Random Walk :  $q(x, y) = q(y - x)$

- Cette méthode tire un point dit "proche" ou "dépendant" du point précédemment tiré.
- Son avantage : elle permet de viser une partie de la densité cible très rapidement.
- Son inconvénient : une fois une trajectoire suivie, elle ne visite pas tout l'espace. On a une convergence partielle.

Independent :  $q(x, y) = q(y)$

- Cette méthode tire un point qui ne dépend plus du précédent.
- Son avantage : elle peut visiter tout l'espace de densité.
- Son inconvénient : son caractère aléatoire et indépendant des tirages précédent fait qu'elle ne converge pas forcément vers la densité cible.

L'idée derrière la Métropolis-Hastings Mixture est de créer une densité instrumentale conditionnelle  $q$  qui mélange plusieurs types de densité instrumentale. Par exemple, mélanger la Random Walk et la Independent.

Le but de ce mélange est de prendre les avantages de chaque noyau afin d'obtenir des résultats plus intéressants que si un seul noyau était choisi. Dans le cas de la mixture entre la MH Random Walk (MHRW) et la MH Independent (MHI), on espère que la MHRW permettra la convergence vers une partie de la densité cible et que la MHI grâce à son caractère aléatoire permettra de visiter tout l'espace de densité sans rester bloqué dans une trajectoire particulière.

Nous allons expliquer comment cette méthode fonctionne. Nous voulons écrire notre noyau  $q$  sous la forme :

$$\forall x, y \in \mathcal{X} \quad q(y, x) = \alpha \cdot q_{rw}(y, x) + (1 - \alpha) \cdot q_{ind}(y, x) \text{ avec } \alpha \in [0, 1]$$

Or, il est facile de montrer que :

Si une variable aléatoire  $B$  suit une loi de Bernouilli  $\mathcal{B}(\alpha)$  et si  $X$  et  $Y$  sont deux variables aléatoires avec pour densité respective  $f_X$  et  $f_Y$  alors la variable aléatoire  $Z = B.X + (1 - B).Y$  a pour densité  $f_Z = \alpha.f_X + (1 - \alpha).f_Y$

L'idée de la Métropolis-Hastings Mixture est donc de choisir à chaque itération un des deux noyaux selon une loi de Bernoulli.

## 3.2 Résultats

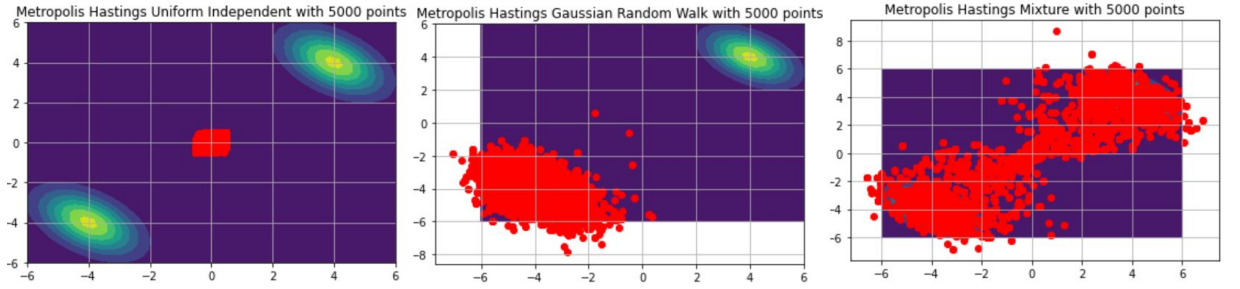


FIGURE 9 – Metropolis Hastings Mixture pour 5000 échantillons avec (1) Uniform Independent, (2) Gaussian Random Walk et (3)  $\alpha = 0.5$

Commentaires : Nous observons que le choix des noyaux et du paramètre  $\alpha$  influent sur les résultats graphiques. La Métropolis Hastings Independent tire un point indépendant des points précédemment tirés (ici proche du point  $(0,0)$ ), ce qui permet à la Métropolis Hastings Random Walk de partir dans de nouvelles directions et ainsi viser la densité cible entièrement. Nous voyons que prendre une Gaussian RW converge plus rapidement vers une des parties de la densité cible que la Uniform RW car cette dernière peut tirer des points à une distance maximale  $\frac{1}{2}$  du point précédemment tiré alors que la Gaussian RW a plus de portée. Toutefois, la Uniforme RW mélangée avec la Gaussian Independent (fig. 10) donne un meilleur résultat que la Gaussian RW et la Uniform Independent (fig. 9) car elle retourne moins souvent vers le point  $(0,0)$ .

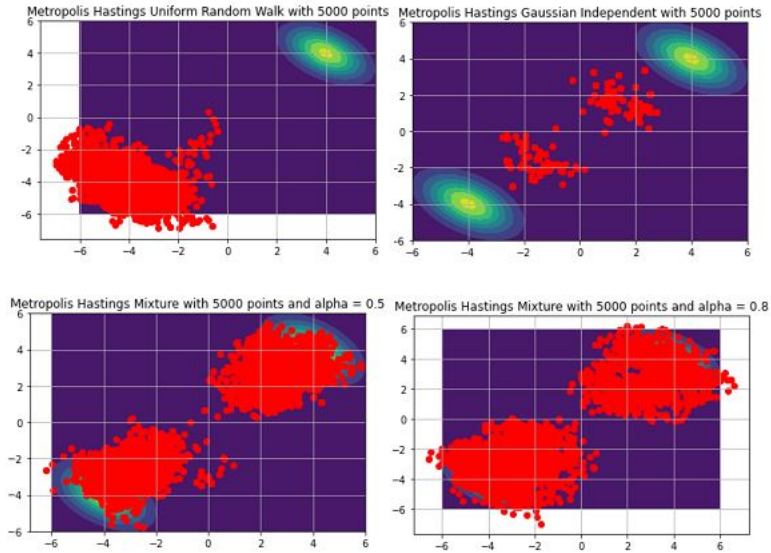


FIGURE 10 – Metropolis Hastings Mixture pour 5000 échantillons avec (1) Uniform Random Walk, (2) Gaussian Independent et alpha = (3) 0.5 (4) 0.8

De plus, prendre  $\alpha = 0.8$  (fig. 10) permet de favoriser la Random Walk plutôt que la Independent, ce qui permet une meilleure convergence car nous retournons moins souvent vers le point (0,0) et on laisse le temps à la Métropolis-Hastings de tirer des points proches des points précédemment tirés.

## 4 Metropolis Hastings avec l'estimateur de densité à noyau

### 4.1 Description

Par rapport aux travaux de Haario et al. nous avons cherché à étudier un nouveau type de méthode de Métropolis Hastings avec une politique différente. Celle-ci est maintenant donnée par :

$$q_n(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i)$$

avec  $K_h(u) = h^d K(\frac{u}{h})$  où  $K : R^d \rightarrow R_+$  une densité. L'algorithme reste fondamentalement le même par ailleurs.

Pour mieux comprendre comment fonctionne cette densité, explicitons la pour  $n=2$ . Prenons :  $q = \frac{1}{2}q_1 + \frac{1}{2}q_2$  On cherche à tirer  $X$  selon la densité  $q$ , comment fait-on ?

Soit  $B \sim B(\frac{1}{2})$  :

$$\begin{cases} X \text{ est tiré selon } q_1 & \text{si } B = 1 \\ X \text{ est tiré selon } q_2 & \text{sinon.} \end{cases}$$

Pour la mixture de  $n$  noyaux :

Etape 1 : Tirage de  $M$  un vecteur de  $R^n$  variable suivant une loi multinomiale de paramètre  $n = 1$  et  $\forall j \in 1, \dots, n, p_j = \frac{1}{n}$ , on récupère  $t$  l'index où  $M_t = 1$ . On tire :  $Y \sim q(\cdot | X_t)$

Etape 2 : 2 possibilités : Acceptation de  $Y$  et  $X_i = Y$  ou rejet et  $X_i = X_{i-1}$

L'acceptation pour le point  $X_i$  suit une loi de Bernoulli de paramètre :

$$p_i = \min(1, \frac{f(Y)}{f(X_{i-1})})$$

### 4.2 Quelques exemples

Tout d'abord pour le noyau, nous avons choisi de prendre le noyau d'une loi normale. Nous avons pris une fenêtre  $h = 10$ . Nous pouvons déjà constater

que cette méthode marche pour une gaussienne simple, que nous ayons un bon ou mauvais départ.

Par ailleurs, en ce qui concerne la double gaussienne, nous pouvons voir figure 12 que la méthode arrive à balayer l'espace de la densité cible. Ainsi par rapport à la méthode de Metropolis Hastings classique, elle permet de résoudre ce problème de densité "non continue", comme pour la méthode de Haario et al., et cela peu importe la qualité du départ une fois encore.

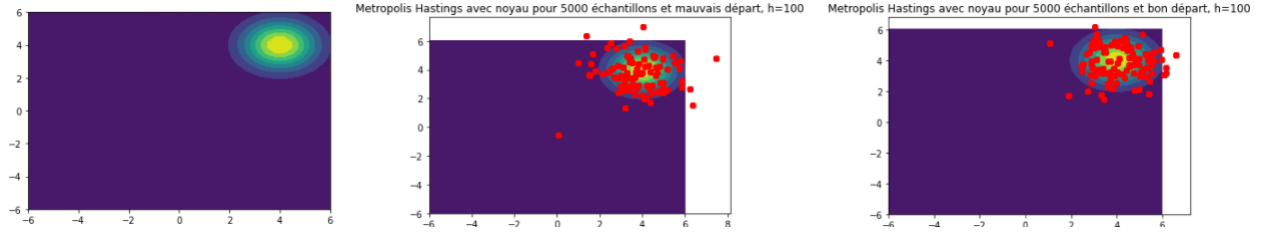


FIGURE 11 – Metropolis Hastings avec estimateur par noyau pour 5000 échantillons

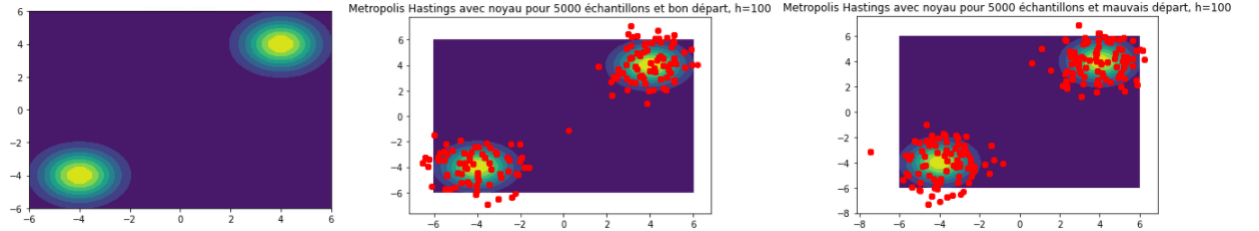


FIGURE 12 – Metropolis Hastings avec estimateur par noyau pour 5000 échantillons



## 5 Metropolis Hastings avec l'estimateur de densité à noyau adaptatif

### 5.1 Principe

Nous avons eu l'idée de rendre cette dernière méthode encore plus adaptative à la façon de Haario et al. et nous avons donc transposé celle-ci à la méthode précédente avec l'estimateur à noyau. En effet, nous considérons de même une période initiale  $t_0$ , où nous recourrons à une covariance initiale choisie arbitrairement et qui peut dépendre de notre connaissance a priori, puis utilisons ensuite une covariance des vecteurs tirés précédemment, tout comme dans la méthode de Haario. Le tirage de la nouvelle variable aléatoire se fait alors selon une loi normale de moyenne un tirage d'indice  $j$ , avec  $j$  tiré selon une loi multinomiale parmi les tirages réalisés jusqu'alors, et de covariance la covariance alors utilisée dans notre cas.

### 5.2 Illustration

Nous avons repris les mêmes paramètres que pour la méthode de Haario et al. avec un noyau gaussien. Comme précédemment, la méthode est efficace pour une gaussienne simple (figure 13), mais elle l'est également pour une double gaussienne (figure 14) et ici aussi en dépit du problème de la qualité du départ. Il nous reste donc à comparer les différentes méthodes.

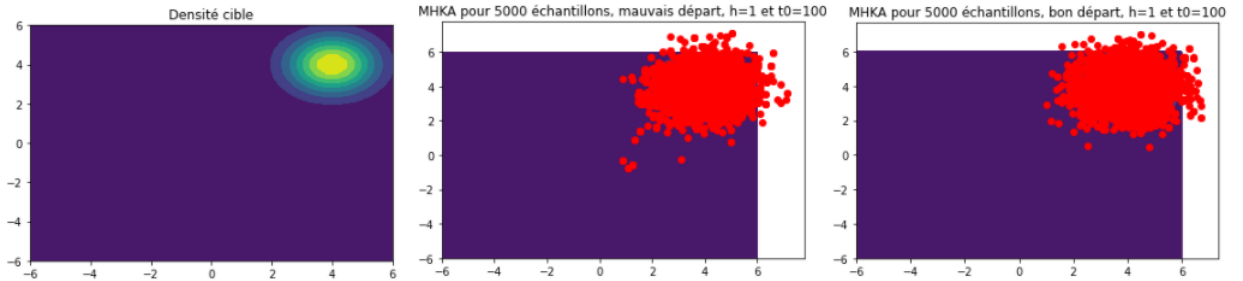


FIGURE 13 – Metropolis Hastings avec noyau adaptatif pour 5000 échantillons, période initiale de 100 itération,  $\epsilon = 0.5$ ,  $h = 100$

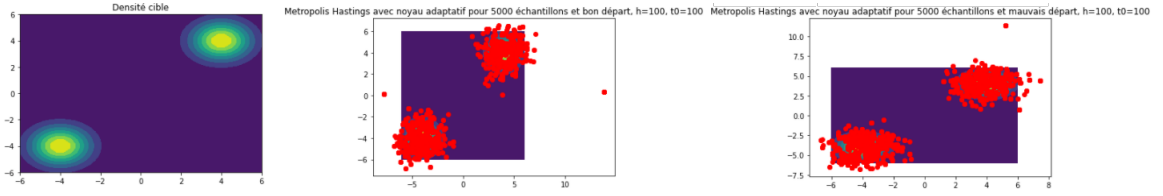


FIGURE 14 – Metropolis Hastings avec noyau adaptatif pour 5000 échantillons, période initiale de 100 itération,  $\epsilon = 0.5$ ,  $h = 100$

## 6 Tests et comparaison

Nous avons cherché à comparer les différentes méthodes dans différents cas et à comparer leurs performances. Pour des raisons de puissance de calcul de nos ordinateurs à disposition, nous nous sommes limités à des densités en dimension 2 (même si les algorithmes implémentés fonctionnent pour des dimensions supérieures) et pour les valeurs de  $n$  suivantes :  $n = \{1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000\}$ . Pour réaliser les tests, nous avons choisi une densité cible gaussienne. L'erreur moyenne quadratique est donnée par :

$$MSE = \|\mu - \hat{\mu}_n\|_2$$

avec  $\mu$  la moyenne de la densité cible et  $\hat{\mu}_n$  la moyenne estimée à partir des particules issues de la méthode de Métropolis Hastings.

### 6.1 Test avec une double gaussienne

La où le test nous a semblé le plus intéressant fût pour une densité cible qui était une double gaussienne. En effet, l'algorithme de Métropolis Hastings a du mal à explorer une densité "discontinue" en raison de la variance qui est fixée. Ce problème est quelque peu corrigé par les algorithmes de Métropolis Hastings adaptatifs comme celui proposé par Haario et al., mais toutefois il peut rencontrer rapidement des problèmes : lors de certaines exécutions, seule une gaussienne était explorée (comme sur la figure 15). Enfin l'algorithme adaptatif que nous avons proposé s'adapte mieux pour de tels densités.

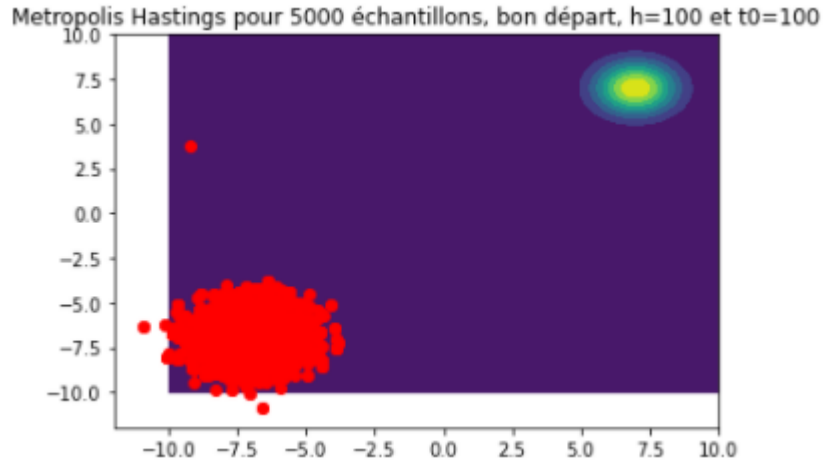


FIGURE 15 – Metropolis Hastings adaptatif pour 5000 échantillons et une densité cible formée de deux gaussienne de moyenne  $(6,6)$  et  $(-6,-6)$

## 6.2 Comparaison

Pour comparer les résultats, nous avons déterminé la moyenne de l'erreur quadratique moyenne sur 20 exécutions sur chaque nombre de points. Les résultats sont représentés sur la figure 16

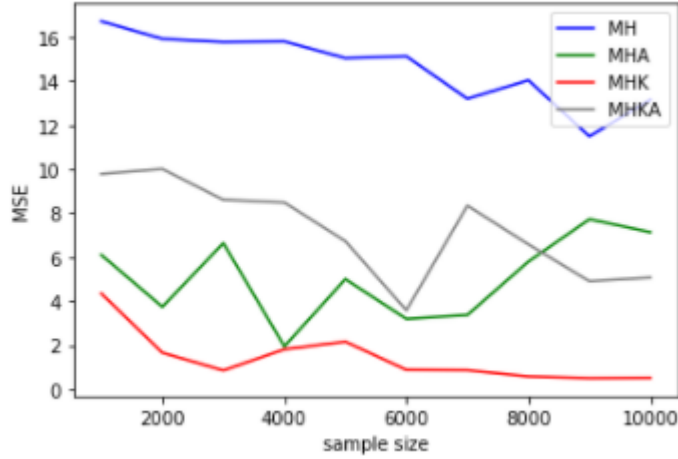


FIGURE 16 – Comparaison des erreurs quadratiques moyennes moyennées sur 20 exécutions pour chaque méthode

Il apparaît donc que les méthodes classique et adaptatives (comme celle proposée par Haario et al.) suffisent pour des densités cibles "simples", c'est-à-dire avec une variance relativement faible. En revanche, si la densité cible est plus "exotique", la méthode avec noyaux que nous proposons s'adapte mieux à la situation et cela rapidement. En revanche, le temps de calcul est plus important. Enfin, à durée initiale égale, la méthode de Haario et notre méthode à noyau adaptative (en gris sur la figure) semblent se valoir. En revanche, il semble qu'il y ait une durée initiale optimale, qui empiriquement se situerait aux alentours de  $\frac{n}{5}$  avec  $n$  le nombre d'échantillons tirés, et dans ce cas, la méthode à noyau adaptative converge et l'erreur est plus petite en comparaison avec celle de Haario (cf figures 17 et 18).

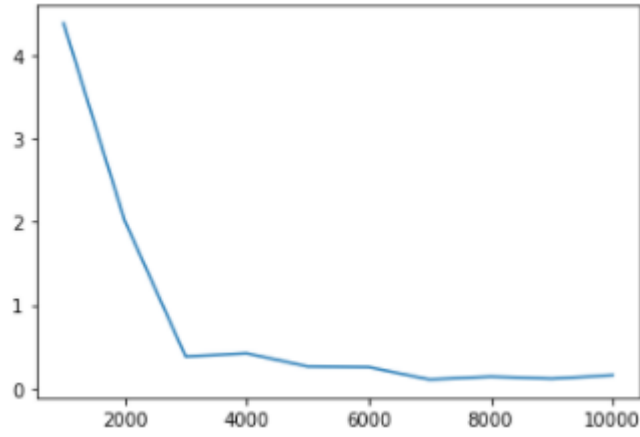


FIGURE 17 – Erreurs quadratiques moyennes moyennées sur 20 exécutions pour la méthode de Haar et un temps d'arrêt de  $\frac{n}{5}$

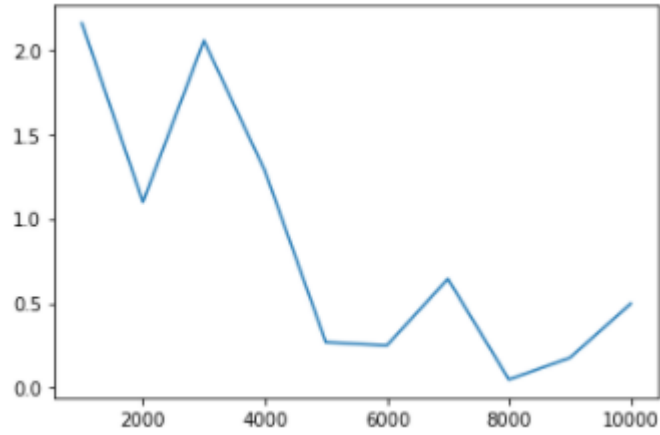


FIGURE 18 – Erreurs quadratiques moyennes moyennées sur 20 exécutions pour la méthode à noyaux adaptative et un temps d'arrêt de  $\frac{n}{5}$

## 7 Conclusion

Nous remarquons que la nature de la méthode de Metropolis Hastings ainsi que le choix de la fonction densité utile influent énormément sur les résultats.

La méthode de Metropolis Hastings avec l'estimateur de densité à noyau est intéressante en comparaison à la méthode classique et même par rapport à la méthode adaptative proposée par Haario et al puisqu'elle est bien plus précise, et cela rapidement. Par ailleurs, par rapport à la méthode classique, elle permet notamment de surmonter le problème du "cold start" : ainsi nous avons pas à nous préoccuper du choix de la distribution initiale et du tirage initial. Par conséquent, cette méthode est bien plus avantageuse.

Toutefois, cette méthode est longue en temps de calcul : en effet, pour le "peu" de tirages que nous avons effectués, le temps de calcul est long par rapport à la méthode classique. Toutefois, les avantages ci-dessus l'emportent et permettent de justifier le choix de cette méthode par rapport à la méthode classique.

## 8 Pour l'avenir...

Nous nous sommes ici limités à l'utilisation de  $n$  noyaux gaussiens à l'étape  $n$ , or il pourrait être intéressant d'utiliser des mixtures plus variées afin d'exploiter les points forts de chaque noyau.

Par ailleurs, il pourrait être intéressant de voir s'il est possible de déterminer une période initiale optimale, notamment en ce qui concerne la méthode avec noyau adaptative. Empiriquement, il semblerait qu'une période initiale égale au nombre total d'échantillon divisé par 5 semble mener à des résultats convaincants

Finalement, il serait intéressant d'utiliser d'autres moyens de comparaisons que la moyenne d'erreur quadratique car cette méthode ne pénalise pas les densités utiles qui se comportent en moyenne comme la densité cible. Par exemple, une loi uniforme centrée ne sera pas forcément pénalisée lorsque la densité cible est composée de deux gaussiennes de moyenne opposée. Une autre piste d'investigation pourrait être de trouver une fenêtre  $h$  optimale, notamment en cherchant à minimiser à chaque étape une distance de Kullback Leibler entre notre politique et la densité cible.

## 9 Références bibliographiques

Biau, G., B. Cadre, M. Sangnier, and U. Tanielian (2018). Some theoretical properties of gans.arXiv preprint arXiv :1803.07819 .

Delyon, B. and F. Portier (2018). Asymptotic optimality of adaptive importance sampling. Thirty-second Conference on Neural Information Processing Systems .

Mireuta, M. (2011). Mémoire présenté à la Faculté des études supérieures en vue de l'obtention du grade de Maître des sciences (M.Sc.) en mathématiques. Université de Montréal. [https://dms.umontreal.ca/~bedard/Mireuta\\_Matei\\_2011\\_memoire.pdf](https://dms.umontreal.ca/~bedard/Mireuta_Matei_2011_memoire.pdf)

Delyon, B., Portier, F. (2021). Safe adaptive importance sampling : A mixture approach. The Annals of Statistics, 49(2), 1. <https://doi.org/10.1214/20-aos1983>

Chib, S., Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. The American Statistician, 49(4), 327-335. <https://doi.org/10.1080/00031305.1995.10476177>

Haario, H., Saksman, E., Tamminen, J. (2001a). An Adaptive Metropolis Algorithm. Bernoulli, 7(2), 223. <https://doi.org/10.2307/3318737>