Méthodes de chaînes de Markov par méthodes de Monte Carlo

Loïk-johan ACAKPO-ADDRA Lilian TEL

Encadrant : Pierre-André ZITT

mercredi 2 juin 2021

Table des matières

1	Intr	coduction	3	
	1.1	Calcul déterministe	4	
	1.2	Observations	5	
	1.3	Conclusions		
2	Méthode de rejet			
	2.1	Idée et outils	8	
	2.2	Méthode		
	2.3	Observations	10	
	2.4	Conclusions		
3	Méthode de Monte Carlo par chaîne de Markov			
	3.1	Idées et outils	13	
	3.2	Métropolis-Hasting	13	
	3.3	Observations		
		Conclusions		

1 Introduction

Notre sujet d'étude est un sujet classique de la physique statistique. Le but de la physique statistique est de déduire le comportement et l'évolution d'un système donné à travers l'utilisation de méthodes probabilistes. Nous appelerons 'sites', les emplacements possibles des particules ; on suppose par ailleurs que chaque site ne peut accueillir qu'une seule particule.

Définition 1 (Configuration). Une configuration est un espace de particules statiques composé de N^2 sites. Dans notre système, nous modéliserons les sites par des couples (i,j) (que l'on notera s) appartenant à l'ensemble S_N défini par :

$$S_N = \{1, \cdots, N\} \times \{1, \cdots, N\}.$$

Une configuration est donc définie par une matrice

$$x = (x(s), s \in S_N).$$

avec

$$x(s) = \begin{cases} 1 & s'il \ y \ a \ une \ particule \ au \ site \ s \\ 0 & sinon \end{cases}$$

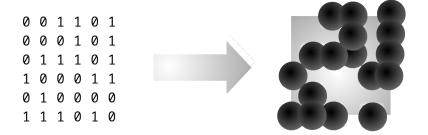


FIGURE 1 – Représentation d'une configuration

Définition 2 (Ensembles de configurations). Nous appelons E_N l'ensemble des configurations possibles, et on le note

$$E_N = \{0, 1\}^{S_N}.$$

Chacun des N^2 sites pouvant présenter deux états différents, il y a donc $|E_N| = 2^{N^2}$ sites possibles.

Définition 3 (Voisinage de s). Le voisinage de s, noté V(s) est la boule unité autour de s pour la norme 1. Si l'on pose $s = (i, j), 1 \le i, j \le N$

$$V(s) = \{(i-1, j), (i+1, j), (i, j-1), (i, j+1)\} \cap S_N$$

Définition 4 (Admissibilité). Une configuration est dite admissible si elle ne contient aucun couple de particules voisines. Autrement dit :

$$\forall s \in S_N, \forall s' \in V(s), \quad x(s)x(s') = 0.$$

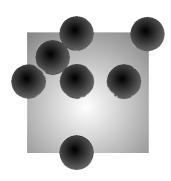


FIGURE 2 – Exemple d'une configuration admissible

1.1 Calcul déterministe

Posons A_N l'ensemble des configurations admissibles.

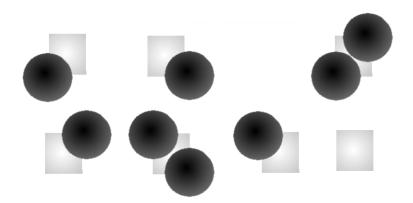


FIGURE 3 – Configurations admissibles pour N=2 La figure 3 représente $A_2,\,A_2$ est composé de sept éléments.

Définition 5 (Densité d'une configuration). La densité d'une configuration x est définie par la fonction : $d: E_N \longrightarrow \mathbb{R}$ $x \longmapsto \frac{1}{N^2} \sum_{s \in S_N} x(s)$

Notre objectif est de calculer la densité moyenne des configurations admissibles.

Définition 6. La densité moyenne des configurations admissibles, notée $D(A_N)$ est définie par

$$D(A_N) = \frac{1}{|A_N|} \sum_{x \in A_N} d(x)$$

Remarque 1. Si $X \sim \mathcal{U}(A_N)$ (la loi uniforme sur A_N), alors :

$$D(A_N) = \mathbb{E}[d(X)]$$

Quand N=2, la densité moyenne $D(A_2)=\frac{1}{7}\frac{1}{4}(0+4\times 1+2\times 2)=\frac{4}{7}$

1.2 Observations

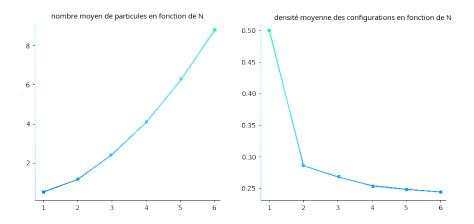


FIGURE 4 – Graphique

Dans la figure 4, pour N variant de 1 à 6, on constate que, lorsque N augmente, le nombre de particules augmente et la densité moyenne diminue.



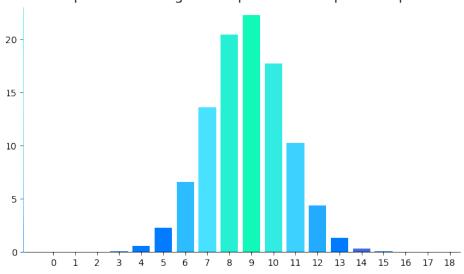


FIGURE 5 – Histogramme

La figure 5 est l'histogramme du nombre de particules sur les configurations admissibles. Ce graphique peut de prime abord faire penser que l'occurrence de certaines configurations est nulle aux valeurs extrêmes. Or chaque occurrence est différente de zéro, par exemple en ce qui concerne les configurations admissibles qui ne contiennent aucune particule, il y a la matrice nulle.

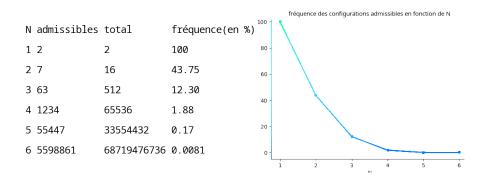


FIGURE 6 – Tableau et son graphique en pourcentage

Avec la figure 6, remarquons que lorsque N augmente la proportion des configurations admissibles diminue fortement et pour N=6 le pourcentage est proche de zéro.

1.3 Conclusions

Après nos premières observations, il semble que la densité moyenne converge sur les configurations admissibles vers un réel différent de zéro.

Remarquons que vérifier qu'une configuration est admissible est trivial mais donner l'ensemble des configurations admissibles et les énumérer systématiquement est difficile car A_N est beaucoup trop grand :

 $D\acute{e}monstration$. Montrer que $|A_N|$ est exponentiel.

On remarque que la configuration admissible avec le plus de particules est la configuration 'en échiquier', et que retirer des particules à une configuration admissible la laisse admissible. On pose p le nombre de particule de la configuration en échiquier, et A_N' l'ensemble des 'sous-configurations' de la configuration en échiquier. Pour chaque site on a le choix de mettre ou pas une particule, donc 2^p configurations différentes.

$$|A'_N| = 2^p \le |A_N| \le |E_N| = 2^{N^2}$$

D'après le théorème d'encadrement A_N est donc un espace de taille exponentielle proportionnellement à N.

Une question se pose alors : comment faire pour calculer la densité moyenne?

2 Méthode de rejet

2.1 Idée et outils

On va d'abord chercher à simuler des tirages de loi $\mathcal{U}(A_N)$ "uniforme sur A_N " et indépendants, ce qui permettra d'appliquer la loi des grands nombres, d'après la remarque 1. On pose M la taille de notre échantillon.

Théorème 1 (Loi des grands nombres (LGN)). Soit $(X_k)_{k\in\mathbb{N}}$ des variables aléatoires indépendantes, X de même loi que les X_k , $f: E_N \to \mathbb{R}$ telle que $\mathbb{E}[|f(X)|] < +\infty$. Alors:

$$\frac{1}{M} \sum_{k=1}^{M} f(X_k) \xrightarrow[M \to +\infty]{p.s} \mathbb{E}[f(X)]$$

Théorème 2 (Théorème central limite (TCL)). Soit $(X_k)_{k\in\mathbb{N}}$ des variables aléatoires indépendantes, X de même loi que les X_k , $f: E_N \to \mathbb{R}$ telle que $\mathbb{E}[f(X)^2] < +\infty$. Alors:

$$\frac{\sum_{k=1}^{M} f(X_k) - M\mathbb{E}[f(X)]}{\sqrt{M\operatorname{Var}(f(X))}} \xrightarrow[M \to +\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

Remarque 2.

$$\mathbb{E}[f(X)^2] < +\infty \Rightarrow \mathbb{E}[|f(X)|] < +\infty$$

par l'inégalité de Cauchy-Schwartz.

2.2 Méthode

Dans notre sujet la méthode de rejet engendrera indirectement $\mathcal{U}(A_N)$. Pour simuler une suite de variables aléatoires $(X_k)_{k\in\mathbb{N}}$ de distribution identique à celle $\mathcal{U}(A_N)$, il suffit dans un tirage de suite de tirage (Y_i) qui sont de loi $\mathcal{U}(E_N)$ par la méthode de Monte-Carlo, de sélectionner les Y_i qui sont admissibles et de refuser les autres.

La méthode de Monte-Carlo est une méthode algorithmique qui va simuler des variables indépendants identiquement distribuées. Pour simuler $\mathcal{U}(E_N)$ par méthode de Monte-Carlo, chaque site s suit la loi $\mathcal{U}(\{0,1\})$, et on obtient une configuration x qui suit la $\mathcal{U}(E_N)$:

$$x \sim \mathcal{U}(E_N) \Leftrightarrow \forall s \in S_N, x(s) \sim \mathcal{U}(\{0,1\})$$

Pour construire notre échantillon, on procède de manière itérative, on tire des variables aléatoires $Y_{1l} \sim \mathcal{U}(E_N)$, on pose $L = \inf\{l > 0 | Y_{1l} \in A\}$, on

écrit $X_1 = Y_{1L}$, on continue jusqu'à M la taille de notre échantillon. À la fin nos variables sont indépendants identiquement distribuées car elles suivent la même loi et ils sont indépendants par la méthode de Monte-Carlo.

De plus, pour toute fonction d sur E_N , $\mathbb{E}[|d(X)|] < +\infty$ car E_N est un espace fini

 $X_k \sim \mathcal{U}(A_N)$, les X_k sont indépendants et identiquement distribués, d fonction positive sur E_N , $\mathbb{E}[|d(X_1)|] < +\infty$

On peut appliquer la loi des grands nombres.

$$\mathbb{E}\left[\frac{1}{M}\sum_{k=1}^{M}d(X_{k})\right] = \frac{1}{M}\mathbb{E}\left[\sum_{k=1}^{M}d(X_{1})\right]$$

$$= \frac{1}{M}\mathbb{E}[Md(X_{1})]$$

$$= \mathbb{E}[d(X_{1})]$$

$$= \sum_{x \in A_{N}}d(x)P(X_{1} = x)$$

$$= \frac{1}{|A_{N}|}\sum_{x \in A_{N}}d(x)$$

$$= D(A_{N})$$

Algorithm 1:

Nous utiliserons cet algorithme pour nos expériences futures.

2.3 Observations

Comparaison de la densité moyenne entre la méthode exacte et la méthode de rejet au 1000è pas

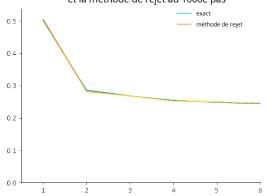


Figure 7 - Graphique comparatif

Sur la figure 7, on fait varier N de 1 jusqu'à 6 $M=10^3$, on obtient le graphique suivant - les deux courbes sont assez proches.

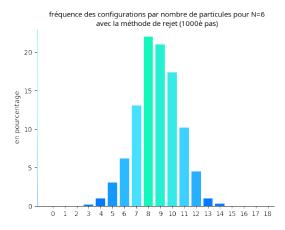


FIGURE 8 - Histogramme

Sur la figure 8, N=6 et $M=10^3$, on constate que l'histogramme ressemble très fortement à l'histogramme de la figure 5. Le théorème central limite nous indique que la méthode de Monté-Carlo converge en $O(\frac{1}{\sqrt{M}})$.

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

$$\Leftrightarrow \frac{\sum_{k=1}^{M} d(X_{k}) - M\mathbb{E}(d(X))}{\sqrt{M \operatorname{Var}(d(X))}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

$$\Leftrightarrow \sum_{k=1}^{M} d(X_{k}) - M\mathbb{E}(d(X)) \xrightarrow{\mathcal{L}} \sqrt{M \operatorname{Var}(d(X))} \mathcal{N}(0,1)$$

$$\Leftrightarrow \frac{1}{M} \sum_{k=1}^{M} d(X_{k}) - \mathbb{E}(d(X)) \xrightarrow{\mathcal{L}} \sqrt{\frac{\operatorname{Var}(d(X))}{M}} \mathcal{N}(0,1)$$

Dans le cas N=6, M=1000, pour le calcul déterministe on peut espérer une erreur de

$$\sqrt{\frac{\operatorname{Var}(d(X))}{M}} = \sqrt{\frac{\mathbb{E}(d(X)^2) - \mathbb{E}(d(X))^2}{M}} \approx 0.00155$$

soit une erreur d'environ 0.16%.

Sur un échantillon, la variance empirique est égale à

$$\sqrt{\frac{\frac{1}{M} \sum_{k=1}^{M} (d(X_k) - \frac{1}{M} \sum_{l=1}^{M} d(X_l))^2}{M}} \approx 0.00157$$

soit également une erreur d'environ 0.16%.

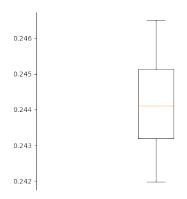


FIGURE 9 – Boîte à moustaches

La figure 9 correspond à la densité moyenne par la méthode de rejet ; l'expérience est répétée 10 fois avec M=1000, N=6.

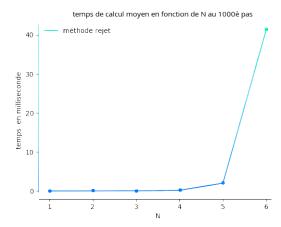


FIGURE 10 - Graphique

Sur la figure 10, on fait varier N de 1 jusqu'à 6 et M=1000, cette figure représente le temps moyen des calculs de méthode de rejet, pour chaque N l'expérience est répété cinq fois. On peut constater une grande différence de temps de calcul entre N=5 et N=6.

2.4 Conclusions

Cette méthode en application numérique devient rapidement inefficace pour de très grandes valeurs de N. Le temps de calcul devient trop grand, cela s'expliquant par la figure 6: la méthode de rejet est une loi géométrique, donc l'espérance du temps de calcul est 1/p soit $\frac{|E_N|}{|A_N|}$ et A_N est de taille exponentielle. En résumé avec cette méthode il est difficile de tirer une variable aléatoire $\mathcal{U}(A_N)$ pour ne pas dire impossible en pratique.

3 Méthode de Monte Carlo par chaîne de Markov

3.1 Idées et outils

On appelle méthode de Monte Carlo par chaîne de Markov (MCMC), une méthode produisant une chaîne de Markov (X_k) ergodique de loi cible. On définit la loi cible par

$$\pi(x) = \frac{\mathbf{1}_{\{x \in A_N\}}}{|A_N|}.$$

dans le but d'appliquer le théorème ergodique. Le théorème consiste à calculer une intégrale par rapport à π en utilisant la loi forte des grands nombres.

Définition 7 (Mesure réversible). Soit P une matrice de transition et π une mesure définis sur E_N . π est dite réversible si:

$$\forall x, y \in E_N \quad \pi(x)P(x, y) = \pi(y)P(y, x)$$

Définition 8 (Mesure invariante). Soit P une matrice de transition et π une mesure définis sur E_N . π est dite invariante si:

$$\pi P = \pi$$

Théorème 3 (Théorème ergodique). Soit (X_k) une chaîne de Markov issue d'une matrice de transition irréductible P, à valeurs dans E_N . Supposons qu'il existe une probabilité invariante, alors :

- (i) π est l'unique probabilité invariante, $\pi(x) > 0 \ \forall \ x \in E_N$.
- (ii) Tous les états sont récurrents, la chaîne est dite récurrente positive.
- (iii) Pour tous états $x \in E_N$ et toute fonction $d: E \to \mathbb{R}$ telle que $\sum_{x \in E_N} |d(x)| \pi(x) < +\infty$

$$\frac{1}{M} \sum_{k=0}^{M} d(X_k) \overset{M \to +\infty}{\longrightarrow} \sum_{x \in E_N} \pi(x) d(x) \quad P - p.s.$$

3.2 Métropolis-Hasting

La méthode de Métropolis-Hasting va simuler une chaîne de Markov de loi invariante. Suppose $\forall x, y \in E_N$, on se donne une matrice de transition Q symétrique définie par

$$Q(x,y) = \begin{cases} 0 & \text{si } x \text{ et } y \text{ sont différents en au moins deux sites} \\ 0 & \text{si } x = y \\ \frac{1}{N^2} & \text{si } x \text{ et } y \text{ sont différents en 1 site} \end{cases}$$

Remarque 3. Q est irréductible.

Description de l'algorithme Choisissons un $X_0 \in E_N$ tel que $\pi(X_0) > 0$ donc $X_0 \in A_N$. Construisons ensuite de manière récursive les X_k : supposons X_k calculé, on pose $X_k = x$, on tire un état y de E_N de loi Q(x, .) uniforme sur les transitions de x puis on simule une variable U_k uniforme sur [0, 1]. Pour le choix de X_{k+1}

Pour le choix de
$$X_{k+1}$$

$$X_{k+1} = \begin{cases} y & \text{si } U_k \leq \min(1, \frac{\pi(y)}{\pi(x)}) \text{ la proposition est acceptée} \\ x & \text{si } U_k > \min(1, \frac{\pi(y)}{\pi(x)}) \text{ la proposition est rejetée} \end{cases}$$
 On n'a pas de problème de division par zéro car on a supposé

On n'a pas de problème de division par zéro car on a supposé X_0 admissible. Dans notre cas, on remarque que si $y \in A_N$, $\min(1, \frac{\pi(x)}{\pi(y)}) = \min(1, 1) = 1$, donc la proposition est toujours acceptée. Sinon on a $y \notin A_N$, $\min(1, \frac{\pi(x)}{\pi(y)}) = \min(1, 0) = 0$ et la proposition est toujours rejetée. Définissons enfin P une matrice de transition pour x, y des configurations admissible défini par :

$$P(x,y) = \begin{cases} Q(x,y) & \text{si } x \neq y \\ 1 - \sum_{z \in E_N \mid z \neq x} Q(x,z) & \text{sinon} \end{cases}$$

Le processus (X_k) est une chaîne de Markov issue de la matrice P. E_N étant irréductible, il admet une unique mesure invariante, donc si l'on montre que π est réversible, on aura alors montré que π est la mesure invariante.

Démonstration. Supposons que $\forall x, y \in A_N$

$$\pi(x)P(x,y) = \pi(x)Q(x,y)$$

$$\pi(y)P(y,x) = \pi(y)Q(y,x)$$

$$= \pi(y)Q(x,y) \text{ car Q est symétique}$$

$$= \pi(x)Q(x,y) \text{ car } \pi(x) = \pi(y)$$

on remarque $\pi(x)P(x,y) = \pi(y)P(y,x)$ donc π est réversible

Proposition 1. Une mesure réversible implique une mesure invariante.

Démonstration. On suppose que π réversible, $\forall y \in E_N$

$$(\pi P)(y) = \sum_{x \in E_N} \pi(x) P(y, x)$$
$$= \sum_{x \in E_N} \pi(y) P(x, y)$$
$$= \pi(y) \sum_{x \in E_N} P(x, y)$$
$$= \pi(y)$$

Proposition 2. A_N est une classe fermée.

Démonstration. On suppose que $\forall x, y \in A_N$. Q est irréductible. On remarque que si l'on enlève une particule à x et à y, la nouvelle configuration reste dans A_N ; de plus enlever toutes les particules de x et de y, mène la matrice nulle. Notons 0 la matrice nulle. On a alors $x \rightsquigarrow 0$ et $y \rightsquigarrow 0$. De plus on a $0 \rightsquigarrow y$ car Q est symétrique et irréductible, on peut faire le chemin dans le sens inverse et on obtient $x \rightsquigarrow 0$ et $0 \rightsquigarrow y \Rightarrow x \rightsquigarrow y$ par transitivité. \square

Proposition 3. Tous les éléments de A_N sont récurrents.

Remarque 4. A_N est une classe fermée par conséquent la matrice P est irréductible.

Remarque 5. E_N est un espace fini et les éléments de A_N sont récurrents donc récurrents positifs.

On simule une loi uniforme sur les transitions de x; pour simuler cette loi on procède de la manière suivante : soit y une configuration identique à x, on choisit un site s qui suit la loi $\mathcal{U}(S_N)$, pour que y suit la loi uniforme sur les transitions de x, on modifie la configuration y au site s. De la manière suivante :

$$y(s) = 1 - x(s)$$

Un avantage de cette méthode est de simplifier la vérification de l'admissibilité de y: cela nécessite seulement de vérifier qu'il n'y ait pas d'autres aparticules au voisinage de s de y.

soit
$$s \in S_N, \forall s' \in V(s), \quad y(s)y(s') = 0.$$

On peut donner une estimation de $D(A_N)$ en utilisant le théorème ergodique. On a montré que P est irréductible et E_N est fini donc A_N également donc $\sum_{x \in A_N} \pi(x) d(x)$ converge avec d positive. Dans notre cas :

$$\frac{1}{M} \sum_{k=0}^{M} d(X_k) \overset{M \to +\infty}{\longrightarrow} \sum_{x \in A_N} \pi(x) d(x) = \sum_{x \in A_N} \frac{1}{|A_N|} d(x) = \frac{1}{|A_N|} \sum_{x \in A_N} d(x) = D(A_N) P - p.s.$$

Algorithm 2:

On utilise cet algorithme pour nos expériences futures.

3.3 Observations

Dans nos expériences, on choisit arbitrairement X_0 , quand X_0 correspond à la matrice nulle on dit que l'algorithme est MH-nulle. On parle de MH-échiquier si X_0 correspond à une configuration en échiquier.

Pour MH-aléa, les configurations de départ sont aléatoires avec un nombre $\lfloor \frac{N^2}{4} \rfloor$. X_0 est construit par l'algorithme de Métropolis-Hasting avec des conditions différentes. La configuration initiale est la matrice nulle dans cette algorithme et la condition d'arrêt est une configuration admissble avec $\lfloor \frac{N^2}{4} \rfloor$ éléments, MH-aléa est MH-nulle auquel on a supprimé les premiers éléments de l'échantillon et ajouté des éléments pour avoir la taille de l'échantillon conforme (préchauffage).

comparaison de la densité moyenne pour la méthode de Metropolis-Hastings au 1000è pas selon la configuration initiale

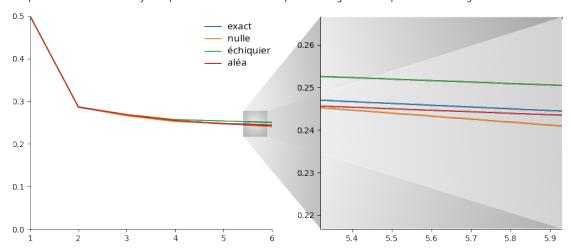


FIGURE 11 - Graphique

Sur la figure 11, on compare les estimations selon le X_0 , l'expérience pour chaque MH est répétée cinquante fois. On peut observer que la courbe échiquier se distingue à partir de N=4. Sur le zoom on peut mieux distinguer les courbes, on peut classer les courbes par rapport au X_0 . Dans l'ordre de la densité moyenne la plus élevé est celle de MH échiquier puis MH aléa et MH nulle. On peut ajouter que aléa est le plus proche de la courbe déterministe.

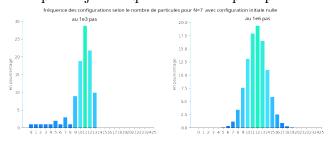


FIGURE 12 – Histogramme

Sur la figure 12 N = 7, X_0 correspond à une matrice nulle, dans cette expérience on a simulé deux échantillons le premier à gauche l'échantillon est de taille 10^3 , et à droite la taille de l'échantillon est 10^6 . On peut constater sur l'histogramme de gauche qu'il y a des occurrences qui ressortent beaucoup, ce phénomène n'est pas normal si l'on compare à la figure 5.

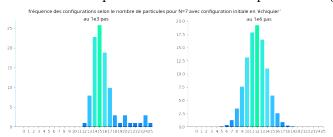


FIGURE 13 – Histogramme

Sur la figure 13 N=7, X_0 correspond à une configuration échiquier, dans cette expérience on a simulé deux échantillons : le premier à gauche l'échantillon est de taille 10^3 , et à droite la taille de l'échantillon est 10^6 . On peut observer le même phénomène que sur la figure 11 mais de manière symétrique.

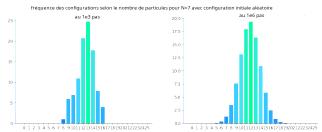


FIGURE 14 – Histogramme

Sur la figure 14 N = 7, X_0 correspond à une configuration aléa, dans cette expérience on a simulé deux échantillons le premier à gauche l'échantillon est de taille 10^3 , et à droite la taille de l'échantillon est 10^6 . On peut voir un phénomène différent des deux figures précédentes, l'histogramme est moins régulière que la figure 5, mais il y a une forte ressemblance.

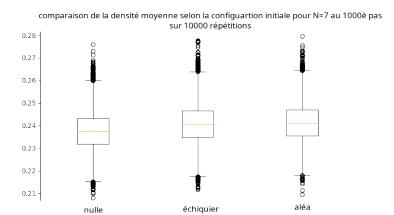


FIGURE 15 – Boîte à moustache

Sur la figure 15 N=7, on répète pour chaque boite l'expérience 10^4 , on observe que les médianes pour les trois boîtes sont assez proches.

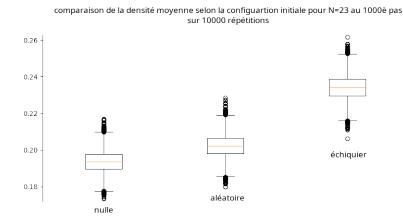


FIGURE 16 – Boîte à moustache

Sur la figure 16 N=23, on répète pour chaque boite l'expérience 10^4 , on observe que les médianes pour les trois boîtes sont assez éloignées.

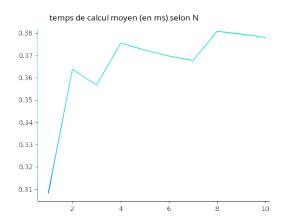


FIGURE 17 - Graphique

Figure 17, représente la moyenne du temps de calcul de l'algorithme MH nulle $M=10^5, N$ varie de 1 à 10 répétition 10^3 . Les temps de calcul sont relativement stables.

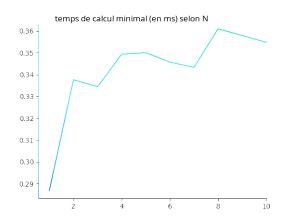


FIGURE 18 - Graphique

Figure 18, représente le temps minimum de calcul de l'algorithme MH nulle $M=10^5, N$ varie de 1 à 10 répétition 10^3 . Les temps de calcul sont stables.

3.4 Conclusions

Cette méthode en application numérique pour des valeurs très grandes donne, une bonne estimation de la densité de particules.

On remarque que MH nulle et MH échiquier gardent en mémoire d'où l'on part si l'effectif de l'échantillon n'est pas suffisant; on peut se permettre d'avoir un grand échantillon, puisque le temps de calcul est stable. Avec nos observations, il semblerait que MH aléa donne la meilleure estimation.

Charge de travail :

Loik : Programmation, démonstrations, rédaction...

Lilian : Rédaction, mise en page, relecture, programmation...

Un grand merci à Mr ZITT pour son implication dans notre TER #carry-

the team.