

# Méthodes de Monte-Carlo.



Michel ROGER

Service de Physique de l'Etat Condensé  
CEA Saclay

13 octobre 2008





# Préface

---

Les chapitres 1 à 8 de ce cours concernent les bases essentielles de la méthode de Monte-Carlo. Elles sont enseignées dans le cadre commun de trois Masters de l'INSTN :

- “Physique Médicale”
- “Rayonnement & Energie”
- “Physique & Environnement”

Cet enseignement est précédé de six séances d'une heure et demie destinées à des rappels fondamentaux de théorie des probabilités (Annexes A, B) et de Physique Statistique (Annexes C, D).

Les chapitres 9 à 12 ne sont enseignés qu'aux étudiants des masters “Rayonnement & Energie” et “Physique & Environnement”. Ils abordent des aspects plus spécifiques de la méthode de Monte-Carlo, en particulier les problèmes d'optimisation, la simulation des polymères et le traitement de problèmes quantiques.





# Table des matières

---

<b>1</b>	<b>Introduction</b>	<b>11</b>
<b>2</b>	<b>Calcul d'intégrales</b>	<b>13</b>
2.1	Travail dirigé introductif : l'aiguille de Buffon . . . . .	13
2.1.1	Expérience . . . . .	13
2.1.2	Montrer que $M/N$ "tend vers" $2L/(\pi D)$ . . . . .	14
2.2	Calcul d'intégrales multidimensionnelles. . . . .	17
<b>3</b>	<b>Techniques d'échantillonnage</b>	<b>19</b>
3.1	Génération de nombres "pseudo-aléatoires" . . . . .	19
3.1.1	Introduction . . . . .	19
3.1.2	La méthode des congruences linéaires . . . . .	20
3.1.3	Le "test spectral" . . . . .	23
3.1.4	Suites de Fibonacci "retardées" . . . . .	26
3.1.5	Puis-je utiliser la fonction "RAND(), random(), ..." de mon ordinateur ?	27
3.2	Echantillonnage d'une loi de probabilité non uniforme . . . . .	28
3.2.1	Méthode de transformation . . . . .	28
3.2.2	Méthode de réjection de Von Neumann . . . . .	32
<b>4</b>	<b>Echantillonnage suivant l'importance.</b>	<b>35</b>
4.1	Calcul stochastique d'une intégrale . . . . .	35
4.2	Réduction de la variance . . . . .	36
<b>5</b>	<b>Simulation de systèmes statistiques</b>	<b>41</b>

5.1	Introduction . . . . .	41
5.2	Chaînes de Markov . . . . .	42
5.2.1	Définition . . . . .	42
5.3	Convergence vers une loi de distribution invariante imposée . . . . .	43
5.3.1	Echantillonnage de $\mathbf{f}(\mathbf{x}) = \mathbf{e}^{-\beta \mathbf{E}(\mathbf{x})} / \mathbf{Z}$ . . . . .	44
5.4	Algorithme de Metropolis . . . . .	45
<b>6</b>	<b>Le modèle d'Ising</b>	<b>47</b>
6.1	L'Hamiltonien d'Ising . . . . .	47
6.2	Modélisation par la méthode de Monte-Carlo . . . . .	47
6.3	Analyse des résultats . . . . .	48
6.4	Evolution en fonction de la température . . . . .	52
<b>7</b>	<b>Utilisation de biais. Algorithme de Wolff</b>	<b>55</b>
7.1	Introduction de biais . . . . .	55
7.2	Exemple : algorithme de Wolff pour le modèle d'Ising . . . . .	56
<b>8</b>	<b>Applications à la physique médicale</b>	<b>59</b>
8.1	Introduction . . . . .	59
8.2	Modélisation d'un rayonnement . . . . .	59
8.3	Modélisation de la source . . . . .	60
8.4	Modélisation d'une trajectoire . . . . .	60
8.5	Interaction photon-matière . . . . .	62
8.5.1	Diffusion élastique de Rayleigh . . . . .	62
8.5.2	Diffusion inélastique de Compton . . . . .	63
8.5.3	Absorption photoélectrique . . . . .	64
8.5.4	Choix d'un des événements . . . . .	64
8.5.5	Pour en savoir plus... . . . .	64
8.6	Propagation dans un milieu inhomogène . . . . .	64
<b>9</b>	<b>Méthode de Monte Carlo et Dynamique Moléculaire</b>	<b>67</b>
9.1	Introduction . . . . .	67
9.2	Equation stochastique . . . . .	67
9.3	Comparaison avec la Dynamique Moléculaire . . . . .	69

<b>10 Optimisation</b>	<b>73</b>
10.1 Méthode du “Recuit Simulé” (Simulated Annealing)	73
10.2 “Recuit Parallèle” (Parallel Tempering)	73
<b>11 Polymères</b>	<b>77</b>
11.1 Introduction	77
11.2 Modélisation de polymères	77
11.2.1 Modèle du “collier de perles”	77
11.2.2 Modèle du “Chapelet de ressorts” (String beads)	78
11.3 Mouvements dans un algorithme de Metropolis	79
11.3.1 Reptation	79
11.3.2 Mouvements de pivot	79
11.4 “Reconstruction biaisée”	79
<b>12 Monte-Carlo quantique</b>	<b>83</b>
12.1 Monte-Carlo Variationnel	83
12.2 “Diffusion Monte-Carlo”	84
12.3 “Path Integral Monte-Carlo”	88
12.3.1 Valeur moyenne d’une grandeur physique	88
12.3.2 La Matrice Densité	89
12.3.3 Echantillonnage de $\rho(R, R'; \beta)$	90
<b>A Rappels de théorie des probabilités</b>	<b>95</b>
A.1 Expérience, ensemble des résultats possibles.	95
A.2 Notion d’événement aléatoire.	95
A.3 Notion de probabilité.	96
A.3.1 Définition heuristique	96
A.3.2 Définition mathématique	97
A.4 Probabilités conditionnelles. Evénements indépendants	97
A.5 Variable aléatoire.	98
A.6 Loi de probabilité d’une variable aléatoire	99
A.7 Fonction de répartition d’une variable aléatoire réelle	99
A.8 Densité de probabilité.	100

A.9	Espérance, moments d'ordre $n$ , variance . . . . .	101
A.9.1	Valeur moyenne ou "Espérance" d'une variable aléatoire réelle . . .	101
A.9.2	Moment d'ordre $n$ . Variance d'une variable aléatoire réelle . . . . .	102
A.10	Fonction Caractéristique . . . . .	102
A.10.1	Définition . . . . .	102
A.10.2	Exemple : . . . . .	103
A.11	Inégalité de Bienaymé-Tchebichef . . . . .	104
A.12	Couples de variables aléatoires . . . . .	105
A.12.1	Covariance . . . . .	105
A.12.2	Indépendance . . . . .	105
A.12.3	Somme de 2 variables aléatoires indépendantes . . . . .	105
A.13	Généralisation à $N$ variables aléatoires . . . . .	107
A.14	Somme de variables aléatoires indépendantes . . . . .	107
A.14.1	Lois "faible" des grands nombres. . . . .	107
A.14.2	Théorème de la limite centrale . . . . .	109
A.15	Divers modes de convergence . . . . .	111
<b>B</b>	<b>Chaînes de Markov</b>	<b>113</b>
B.1	Définition . . . . .	113
B.2	Chaînes dans un ensemble discret . . . . .	114
B.2.1	Propriétés . . . . .	114
B.2.2	Rayon spectral de la matrice $\mathbf{p}$ . . . . .	115
B.2.3	Convergence vers une loi de distribution stationnaire . . . . .	116
B.2.4	Chaîne de Markov ergodique . . . . .	116
B.2.5	Généralisation à un ensemble d'états continus . . . . .	117
<b>C</b>	<b>Rappels de Physique statistique</b>	<b>119</b>
C.1	Entropie statistique . . . . .	119
C.1.1	Définition de Shannon . . . . .	119
C.1.2	Théorème de Khinchin (1957). . . . .	120
C.2	Systèmes macroscopiques à l'équilibre . . . . .	120
C.2.1	Ensemble micro-Canonique . . . . .	120
C.2.2	Ensemble Canonique . . . . .	121



---

C.2.3	Ensemble Grand Canonique . . . . .	123
C.3	Evolution vers l'équilibre . . . . .	124
C.3.1	Probabilité de transition . . . . .	124
C.3.2	Equation maîtresse . . . . .	124
C.3.3	Système isolé . . . . .	125
C.3.4	Système $\mathcal{S}$ en contact avec un thermostat $\mathcal{T}$ . . . . .	125
C.3.5	Propriétés à l'équilibre. Relation du bilan détaillé. . . . .	126
<b>D</b>	<b>Valeurs moyennes d'observables</b>	<b>129</b>
D.1	Définition . . . . .	129





# Introduction

---

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

En l’occurrence, la méthode de Monte-Carlo consiste à simuler sur ordinateur un phénomène stochastique par essence : la diffusion de neutrons dans des matériaux fissiles.

*Un ordinateur effectue une suite d’opérations parfaitement déterministes, comment alors simuler le hasard sur un ordinateur ? On génère des suites de nombres “pseudo aléatoires” qui ont les mêmes propriétés statistiques que des nombres aléatoires, mais qui ne sont pas aléatoires. Les algorithmes les plus souvent utilisés seront étudiés au Chapitre 3.*

Les applications en physique médicale que nous verrons au Chapitre 8, sont du même ordre. Par exemple, à partir d’une source de rayon X ou  $\gamma$  dont on connaît les caractéristiques (distribution des photons en énergie, en direction) et des processus physiques bien maîtrisés d’interaction des photons avec la matière (tissus, milieu osseu, ...), on va par exemple simuler un ensemble de photons arrivant sur une tumeur cancéreuse, leur diffusion sur les cellules, et on va en déduire un certains nombres de grandeurs statistiques, telles que le nombre moyen de photons absorbés à une profondeur  $x$  etc...

D’autre part, nous verrons des applications de la méthode de Monte Carlo à la physique statistique (Chapitre 5). En mécanique statistique classique, la loi de Boltzman nous dit que pour un système à grand nombre de degrés de libertés, en contact avec un thermostat qui le maintient à la température  $T$ , la probabilité pour que le système soit dans un état de configuration  $X$  ( $X$  représente l’ensemble de ses degrés de liberté : positions de particules etc ...) est proportionnelle à  $\exp[-E(X)/kT]$  où  $E(X)$  est l’énergie correspondant à la configuration  $X$ . La méthode de Monte-Carlo va nous permettre d’engendrer un grand nombre de configurations, suivant une loi de probabilité proportionnelle à  $\exp[-E(X)/kT]$ , et de faire des moyennes statistiques sur les grandeurs macroscopiques caractéristiques du

système : énergie moyenne, chaleur spécifique, etc...

Mais nous allons commencer par un autre domaine très important de la méthode de Monte-Carlo : l'utilisation de méthodes stochastiques pour une réponse approchée à un problème mathématique exact : par exemple l'estimation d'une intégrale multidimensionnelle (qui, à priori, n'a rien d'aléatoire !).

Dans ce domaine, il est d'usage de citer comme premier exemple historique de la méthode (qui ne s'appellait pas encore "de Monte-Carlo" !) le jeu mathématique proposé par le Comte de Buffon (plus célèbre pour ses travaux naturalistes) en 1777. Ce jeu consiste à jeter  $N$  fois au hasard une aiguille sur un parquet formé de lattes parallèles et identiques et à compter le nombre  $M$  de fois où l'aiguille intercepte la ligne de jointure entre deux lattes. Si la longueur  $L$  de l'aiguille est inférieure à la largeur  $D$  d'une latte, alors le rapport  $M/N$  "tend" vers  $2L/(\pi D)$ . Et Laplace a fait remarquer un peu plus tard (1820) que ce jeu constituait une méthode expérimentale de détermination du nombre  $\pi$ .



# Calcul d'intégrales

---

## 2.1 Travail dirigé introductif : l'aiguille de Buffon

A travers cet exemple simple, nous allons revoir l'utilisation des notions et principaux résultats de la théorie des probabilités : loi des grands nombres, théorème de la limite centrale, convergences, estimation d'erreurs etc... fondamentales pour ce cours ! Un rappel de toutes ces notions est effectué en annexe A.

### 2.1.1 Expérience

Chacun dispose d'une feuille de papier sur laquelle sont tracées des lignes équidistantes de  $D$  et d'une aiguille de longueur  $L$ . Lancer l'aiguille  $N = 100$  fois et compter le nombre  $M$  de fois où celle-ci intercepte une ligne. Disposer dans un tableau les valeurs  $M_i/N$  obtenues par chacun des  $n$  étudiants. Calculer une estimation de la valeur moyenne globale :

$$m = \frac{1}{n} \sum_{i=1}^{i=n} \frac{M_i}{N}$$

ainsi qu'une estimation de la variance  $\sigma_N^2$  pour  $N=100$  lancers

$$\sigma_N^2 = \frac{1}{n-1} \sum_{i=1}^{i=n} \left[ \frac{M_i}{N} - m \right]^2$$

En déduire une valeur approchée du nombre  $\pi$  et en estimer la "précision" obtenue sur les  $n$  répétitions des  $N$  lancers. .

*En langage probabiliste :*

- Lancer l'aiguille est une expérience aléatoire, elle est caractérisée par l'ensemble des résultats possibles  $\Omega$  constitué par toutes les positions possibles  $\omega$  de l'aiguille sur la feuille.
- Le fait de couper une ligne définit un événement aléatoire, représenté par le sous ensemble  $A$  de résultats de  $\Omega$  qui le réalisent.

- Si on répète l'expérience un très grand nombre de fois, on observera que le rapport  $M/N$  tend vers une limite, un nombre compris entre 0 et 1, que l'on définira comme la probabilité de l'événement  $A$ .

### 2.1.2 Montrer que $M/N$ “tend vers” $2L/(\pi D)$

Pour chaque résultat  $\omega$  de l'expérience on mesure l'angle  $0 \leq \Theta(\omega) < \pi$  entre la direction de l'aiguille et celle des droites parallèles, et la distance  $0 \leq Y(\omega) < D/2$  du centre de l'aiguille à la droite la plus proche.  $\Theta$  et  $Y$  sont deux fonctions de  $\Omega$  dans  $R$ , dont la valeur dépend du résultat  $\omega$  d'une expérience aléatoire. Ce sont donc deux variables aléatoires à valeurs dans  $R$ .

On peut considérer que la loi de probabilité de  $Y$  est uniforme sur le segment  $[0, D/2]$  (toute valeur de  $Y$  entre 0 et  $D/2$  est équiprobable), c'est à dire :

$$dP_Y = \text{Prob}(\{y \leq Y(\omega) < y + dy\}) = \frac{2dy}{D}$$

et que la loi de distribution de  $\Theta$  est elle-aussi uniforme sur  $[0, \pi]$  :

$$dP_\Theta = \text{Prob}(\{\theta \leq \Theta(\omega) < \theta + d\theta\}) = \frac{d\theta}{\pi}$$

L'événement “couper une ligne” se produit si et seulement si

$$Y < \frac{L}{2} \sin \Theta$$

Le décompte que nous opérons peut lui-même être représenté par une variable aléatoire  $Z(\omega)$  fonction des deux précédentes et à valeurs dans l'ensemble  $\{0,1\}$  :

$$Z(\omega) = 1 \quad \text{si} \quad Y(\omega) < \frac{L}{2} \sin \Theta(\omega) \quad \text{et} \quad 0 \quad \text{sinon}$$

L'espérance  $E(Z)$  de la variable aléatoire  $Z(Y, \Theta)$  est l'intégrale de  $Z(Y, \Theta)$  par rapport aux “mesures de probabilités”  $dP_Y, dP_\theta$  dans le pavé  $\mathcal{D}$  à deux dimensions représenté sur la figure 2.1 c'est à dire :

$$I = E(Z) = \int \int_{\mathcal{D}} Z(X, \Theta) dP_Y dP_\Theta$$

soit :

$$I = E(Z) = \int \int_{\mathcal{D}} 1_{[y < \frac{L}{2} \sin \theta]} \frac{d\theta}{\pi} \frac{2dy}{D}$$

où

$$1_{[y < \frac{L}{2} \sin \theta]}$$

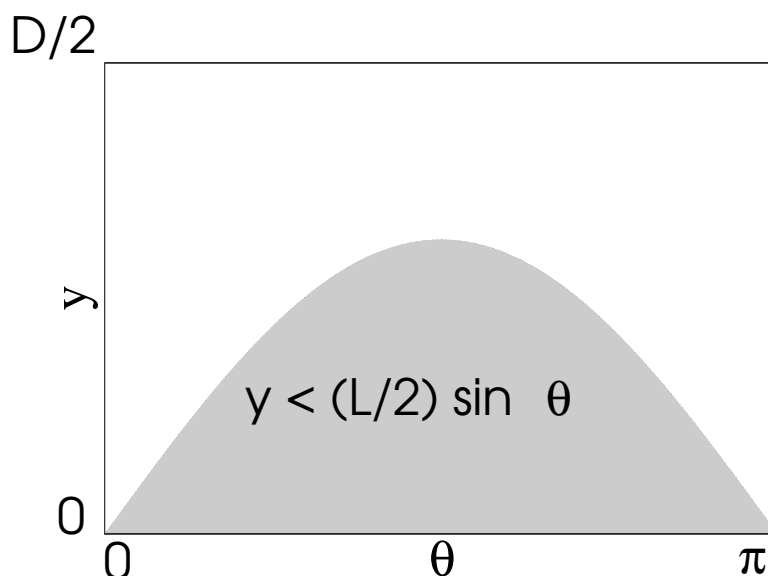


FIG. 2.1 – L'angle  $\theta$  est réparti uniformément sur le segment  $[0, \pi]$ , la distance  $y$  à la ligne la plus proche est répartie uniformément sur le segment  $[0, D/2]$ . La proportion de points qui tombent sous la courbe mesure l'aire de la surface correspondante.

est la “Fonction Indicatrice” du sous ensemble de points du pavé  $\mathcal{D} = [0, \pi][0, D]$  tels que  $y < \frac{L}{2} \sin \theta$ .

On en déduit :

$$I = E(Z) = \int_0^\pi \frac{d\theta}{\pi D} L \sin \theta = \frac{2L}{\pi D}$$

Si on répète  $N$  fois l'expérience et qu'on effectue la somme  $S_N$  des  $N$  résultats (0 ou 1) obtenus pour  $Z$ , alors  $S_N$  définit une nouvelle variable aléatoire, somme de  $N$  variables aléatoires indépendantes  $Z_1, Z_2, \dots, Z_N$  ayant la même loi de probabilité que  $Z$  :

$$S_N = Z_1 + Z_2 + \dots + Z_N$$

Alors la “loi forte des grands nombres” nous assure que  $S_N/N$  converge “presque sûrement” vers l'espérance  $E(Z)$ . C'est dans ce sens que le rapport  $M/N$  “tend” vers  $\frac{2L}{\pi D}$ . Cette loi ne nous dit rien sur la vitesse de convergence.

A titre anecdotique : Mario Lazzarini annonce en 1901 avoir calculé  $\pi = 3.1415929$  en lançant 3408 fois une aiguille de 2,5 cm sur un feuille de papier comportant des lignes parallèles équidistantes de 3 cm. VRAI ou FAUX ?

Pour répondre à cette question, nous allons déterminer la “vitesse de convergence”, c'est à dire estimer “l'erreur” :

$$\epsilon_N = \frac{S_N}{N} - E[Z]$$

lorsque  $N \rightarrow \infty$ . Cette “erreur”  $\epsilon_N$  définit une nouvelle variable aléatoire

Or le “Théorème de la Limite Centrale” qui nous dit que la variable aléatoire :

$$\frac{\sqrt{N}}{\sigma} \epsilon_N$$

“converge en loi” vers la loi normale (loi Gaussienne réduite), de densité de probabilité :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$\sigma$  est “l’écart type” correspondant à la variable aléatoire  $Z$ ,

$$\sigma = \sqrt{E[\{Z - E[Z]\}^2]} = \sqrt{E[Z^2] - (E[Z])^2}$$

$E[Z^2]$  représente le moment d’ordre deux de la variable aléatoire  $Z$ , et  $E[(Z - E[Z])^2] = \sigma^2$  sa “variance” .

Calculons exactement le moment d’ordre deux de la variable aléatoire  $Z(\Theta, Y)$ .

$$E[Z^2] = \int \int_{\mathcal{D}} Z(X, \Theta)^2 dP_Y dP_{\Theta}$$

soit :

$$E[Z^2] = \int \int_{\mathcal{D}} \left\{ 1_{[y < \frac{L}{2} \sin \theta]} \right\}^2 \frac{d\theta}{\pi} \frac{2dy}{D}$$

qui peut s’écrire

$$E[Z^2] = \int_0^\pi \frac{d\theta}{\pi D} L \sin \theta = \frac{2L}{\pi D}$$

On a donc :

$$\sigma = \sqrt{\frac{2L}{\pi D} \left[ 1 - \frac{2L}{\pi D} \right]}$$

En appliquant le théorème de la limite centrale, nous déduisons, par exemple, que la probabilité pour que  $\frac{\sqrt{N}}{\sigma} |\epsilon_N|$  soit inférieur à 2 est égale à la probabilité pour qu’une variable aléatoire gaussienne réduite  $X$  de loi

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

soit de module inférieur à 2, soit

$$Prob(|X| < 2) = \int_{-2}^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0.95$$

On définit ainsi un “intervalle de confiance à 95%” pour  $I$  :

$$Prob \left[ \frac{S_N}{N} - \frac{2\sigma}{\sqrt{N}} < I < \frac{S_N}{N} + \frac{2\sigma}{\sqrt{N}} \right] = 0.95$$



Revenons donc à l'expérience de Lanzarini. Le rapport  $L/D$  est  $2.5/3$ , on en déduit

$$\sigma = 0.499$$

avec  $N = 3408$ , il obtient avec une probabilité de 95% une erreur inférieure à  $2 \cdot 0.499 / \sqrt{3408} = 0.017$  sur le nombre  $2L/(\pi D) = 0.5305$  soit une erreur relative inférieure à 3%. Il ne peut donc estimer  $\pi$  qu'avec un seul chiffre significatif après la virgule... c'est un imposteur !

*Pour l'expérience réalisée en cours, comparer votre écart type expérimentalement estimé à la valeur théorique exacte. Avec le nombre total  $n.M$  de lancers effectués déduire "l'incertitude" sur votre mesure de  $\pi$ .*

## 2.2 Calcul d'intégrales multidimensionnelles.

Les méthodes usuelles de quadrature (méthode de Simpson, méthodes de Gauss...) sont efficaces pour le calcul d'intégrales à une dimension et se généralisent facilement à plusieurs dimensions lorsque le domaine d'intégration est simple (hypercube, par exemple).

Nous allons voir, sur un exemple : la méthode de Simpson, qu'il est par contre illusoire de généraliser ces méthodes à un grand nombre de dimensions.

La méthode de Simpson à une dimension revient à couper l'intervalle d'intégration (que nous supposons  $[0,1]$ ) en  $N$  segments infinitésimaux de longueur  $h = 1/N$ . Nous considérons un petit segment particulier, et par une translation nous recentrons l'origine au milieu de cet intervalle. Dans ce nouveau système d'axes, nous approximations la fonction  $\Phi(x)$  par son développement de Taylor d'ordre deux :

$$\Phi(x) = \Phi(0) + \frac{\partial \Phi}{\partial x} x + \frac{\partial^2 \Phi}{\partial x^2} \frac{x^2}{2} + \mathcal{O}(x^3)$$

et nous intégrons cette expression entre  $-h/2$  et  $h/2$ . les trois premiers termes sont pris en compte exactement. Le premier terme d'erreur d'ordre  $x^3$  donne un terme en  $[x^4/4]_{-h/2}^{h/2}$  qui s'annule par symétrie. L'erreur provient donc de l'intégration du terme suivant en  $x^4$  qui donne une erreur d'ordre  $h^5$ . Le domaine d'intégration est divisé en  $N$  intervalles égaux, de longueur  $h = 1/N$  l'erreur totale ( $N$  fois la précédente) est d'ordre  $Nh^5 \approx 1/N^4$ .

A deux dimensions, on divise le pavé d'intégration  $[0,1] \times [0,1]$  en  $N$  pavés élémentaires de côté  $h = (1/N)^{1/2}$ . Dans un carré élémentaire, on recentre les axes sur le centre de ce carré infinitésimal  $[-h/2, h/2] \times [-h/2, h/2]$  et on approxime la fonction  $\Phi(x, y)$  par un polynôme d'ordre 2 en  $x, y$ . L'erreur provient des termes d'ordre 3 et 4 négligés. Comme précédemment le terme d'ordre 3 donne une contribution nulle par symétrie. L'erreur provenant des termes d'ordre 4 s'écrit :

$$\frac{1}{4!} \sum_{\alpha=0}^4 \frac{\partial^4 \Phi(x, y)}{\partial y^{4-\alpha} \partial x^\alpha} \int_{-h/2}^{h/2} dy y^{4-\alpha} \int_{-h/2}^{h/2} dx x^\alpha \approx \mathcal{O}(h^6)$$

Et à  $d$  dimensions, l'erreur dans une cellule élémentaire est d'ordre  $h^{4+d}$ . Si le nombre de cellules (c'est à dire le nombre de points où on doit calculer la fonction) est  $N$ , on a  $h = \frac{1}{N^{1/d}}$  et, par rapport à  $N$ , l'erreur totale est

$$Nh^{4+d} \approx \mathcal{O}\left(\frac{1}{N^{4/d}}\right)$$

Lorsque  $d$  est grand, la convergence, en fonction du nombre  $N$  de points où la fonction est évaluée (ce qui détermine le temps CPU) devient très lente. Pour cette règle de Simpson, la convergence devient plus lente que  $1/\sqrt{N}$  lorsque  $d$  est supérieur à 8. Nous voyons apparaître l'intérêt d'une méthode stochastique, où, comme nous allons le montrer, l'erreur est en  $1/\sqrt{N}$  quelle que soit la dimension  $d$ .

Considérons une intégrale à  $d$  dimensions :

$$I = \int dx \Phi(x) \quad (2.1)$$

où  $x = (\alpha_1, \alpha_2, \dots, \alpha_d)$  est un vecteur à  $d$  dimensions et

$$\int dx = \int \int \dots \int d\alpha_1 d\alpha_2 \dots d\alpha_d$$

Nous pouvons écrire  $\Phi(x)$  sous la forme :

$$\Phi(x) = g(x)f(x) \quad (2.2)$$

où la fonction  $f(x)$  est une *densité de probabilité*, c'est à dire :

$$f(x) > 0 \quad \forall x \quad \text{et} \quad \int dx f(x) = 1$$

Il y a évidemment une infinité de possibilités (que nous exploiterons) pour une telle décomposition. Nous pouvons donc réécrire :

$$I = E[g(X)] = \int g(x) dP \quad \text{avec} \quad dP = f(x) dx \quad (2.3)$$

Considérons un ensemble de  $N$  points  $\{x_1, x_2, \dots, x_N\}$  répartis suivant la *densité de probabilité*  $f(x)$ . Alors, d'après la "*loi des grands nombres*", la quantité :

$$\frac{S_N}{N} = \frac{1}{N} \sum_{i=1}^N g(x_i) \quad (2.4)$$

tend vers  $I$  lorsque  $N \rightarrow \infty$ , et le théorème de la limite centrale nous assure que l'erreur  $(S_N/N - I)$  est d'ordre  $1/\sqrt{N}$



# Techniques d'échantillonnage

---

## 3.1 Génération de nombres “pseudo-aléatoires”

*“Anyone who consider arithmetical methods of producing random digits is, of course, in a state of sin”.*

John von Neumann (1951)

### 3.1.1 Introduction

Il convient de faire la distinction entre “nombres aléatoires” et “nombres pseudo-aléatoires”.

- “nombres aléatoires” :

Ils peuvent être engendrés par traitement d’une source d’entropie extérieure. Par exemple, un élément radioactif constitue une excellente source d’entropie. L’intervalle de temps entre deux désintégrations est aléatoire. Autre exemple : un site internet : ( [www.random.org](http://www.random.org) ) fournit gratuitement des fichiers préenregistrés (10 Moctets) ou personnalisés de nombres aléatoires. Ces nombres aléatoires sont obtenus en réglant un récepteur radio sur une fréquence où personne ne diffuse.

- “nombres pseudo-aléatoires” :

Ce sont des nombres obtenus par un algorithme mathématique et qui ont l’apparence de nombres aléatoires mais dont la suite est parfaitement reproductible.

A priori, le jugement moral de John von Neumann pourrait nous donner mauvaise conscience. Sachant toutefois qu’il a lui-même “beaucoup péché”, nous le tempérerons par cette autre citation de Donald Knuth dont l’un des ouvrages[1] reste la référence en la matière :

*“Random number generators should not be chosen at random”*

Nous illustrerons ce conseil sur les deux algorithmes mathématiques les plus utilisés actuellement :

- La méthode des congruences linéaires

– Les suites de Fibonacci “retardées”

### 3.1.2 La méthode des congruences linéaires

#### a) Définition

Cet algorithme proposé par Lehmer[2] génère une suite de nombres entiers compris entre 0 et  $(m - 1)$  par la relation :

$$X_{n+1} = (aX_n + c) \pmod{m} \quad (3.1)$$

à partir d'un nombre initial  $X_0$  appelé “semence”.

On peut toujours se ramener à des nombres réels  $U_n$  appartenant à  $[0, 1[$  en considérant la suite  $U_n = X_n/m$

Il est évident que cette suite est périodique, de période inférieure ou égale à  $m$ . Deux exemples avec  $m = 16$  et  $a = 5$  sont illustrés sur la Figure 3.1 : le premier avec  $c = 1$ , le second avec  $c = 0$ . Les nombres successivement obtenus sont représentés sur un cercle. Quel que soit le point de départ du cycle (semence  $X_0$ ), on y revient après avoir parcouru le cercle dans le sens trigonométrique.

– Exemple avec  $c=1$  :

On obtient la période maximale  $m = 16$  et on a engendré une répartition uniforme dans le sous ensemble des entiers appartenant à  $[0,15]$ , ce qui semble satisfaisant. Par contre,

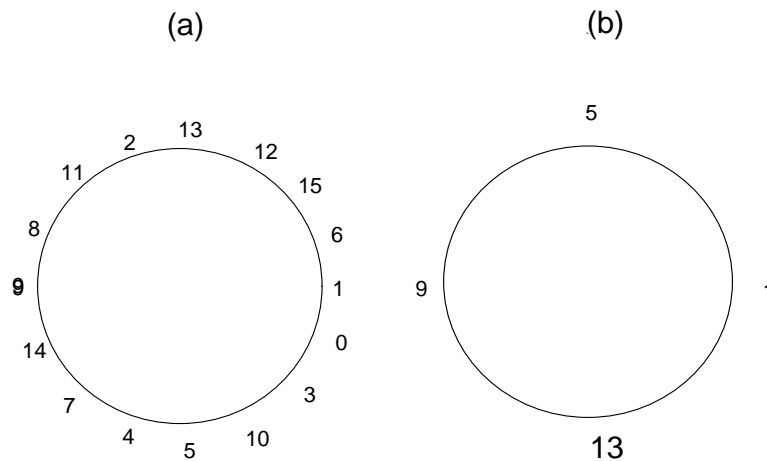


FIG. 3.1 – (a) Congruence linéaire  $X_{n+1} = 5X_n + 1 \pmod{16}$ . (b) Congruence linéaire  $X_{n+1} = 5X_n \pmod{16}$ .

on observe une alternance parfaite de nombres pairs et impairs. Dans une représentation binaire des nombres obtenus, le dernier digit alterne entre 0 et 1 et n’a aucun caractère aléatoire.

Nous retiendrons de cet exemple que pour engendrer une suite aléatoire de nombres 0 ou 1, il est fortement déconseillé d’utiliser un générateur de nombres entiers pour ne conserver que le dernier digit. De même, dans le but de gagner du temps CPU, pour obtenir, par exemple, deux nombres aléatoires de 16 digits on peut imaginer d’engendrer un nombre de 32 digits puis former 2 nombres avec respectivement les 16 digits les plus bas et les 16 les plus hauts. Ceci est fortement déconseillé : il faut engendrer deux nombres de 32 digits dont on ne conservera que les 16 digits les plus hauts (généralement les plus “aléatoires”)

- Exemple avec  $c=0$  :

Par rapport à l’exemple précédent, on a juste changé la constante additive  $c$ . La suite obtenue a une très courte période 4, et elle est très fortement corrélée puisque la différence entre deux nombres successifs est 4.

Après cette illustration triviale de l’avertissement de Knuth, cité plus haut, nous allons donner quelques règles simples pour le choix des paramètres  $m$ ,  $a$  et  $c$ . La justification de ces règles dépasse le cadre de cette introduction, nous renvoyons le lecteur au livre de Knuth[1] pour la démonstration des théorèmes énoncés.

#### *b) Choix du module $m$*

- Le module  $m$  constituant une borne supérieure à la période, on a intérêt à le prendre le plus grand possible, par exemple on peut choisir le plus grand entier représenté sur la machine. Pour une machine 32 bits, en langage Fortran, on peut prendre  $m = 2^{31}$  (le 32<sup>ième</sup>) digit étant réservé au signe. En langage C, on pourra prendre  $2^{32}$  en travaillant avec des entiers positifs (“unsigned int”).
- Avantages et inconvénients à prendre pour  $m$  une puissance de 2 :  
Si  $m = 2^\alpha$ , l’opération “modulo  $m$ ” sur ordinateur devient très rapide, elle correspond à retenir les  $\alpha$  derniers digits du nombre considéré, on la réalise avec un masque et une opération binaire “AND”. On évite une coûteuse division. De même, il n’y a pas à se soucier de dépassement éventuel de l’entier maximal représenté lors de la multiplication de  $a$  par  $X_n$ , dans ce cas la machine conserve les 32 derniers digits du résultat de la multiplication et jette le reste.

Par contre, si  $m = 2^l$  est une puissance de 2, on montre[1] que les  $l$  derniers digits de  $X_n$  ont une période inférieure ou égale à  $2^l$ . (Dans l’exemple de la Figure 1a, les deux derniers digits parcourent le cycle 01, 10, 11, 00 )

- Knuth montre, qu’on peut encore programmer de manière efficace l’opération “modulo  $m$ ” lorsque  $m = 2^\alpha - 1$  ou  $m = 2^\alpha + 1$ . Par contre, dans ce cas, les digits les plus bas sont “aussi aléatoires” que les plus hauts. Ce choix est donc préférable.

#### *c) Choix du facteur multiplicatif $a$*

Le module  $m$  étant fixé, il convient de choisir  $a$  de manière à obtenir la plus grande période

(inférieure ou égale à  $m$ ). Ceci n'est pas suffisant pour avoir un bon générateur de nombres quasi-aléatoires. Il faudra encore passer avec succès un certain nombre de tests statistiques dont les principaux ont été recommandés par Knuth.

Nous énonçons deux théorèmes dont les démonstrations figurent dans le livre de Knuth[1].

### **Théorème A**

La période de la congruence linéaire définie par  $\{m, a, c, X_0\}$  est  $m$ , si et seulement si les trois conditions suivantes sont simultanément vérifiées :

- i)  $c$  est différent de zéro et est premier par rapport à  $m$
- ii)  $(a - 1)$  est multiple de  $p$  pour tout nombre premier qui divise  $m$
- iii)  $(a - 1)$  est multiple de 4, si  $m$  est multiple de 4

Le cas où  $c = 0$  est plus complexe

### **Théorème B**

Lorsque  $c = 0$ , la période est toujours strictement inférieure à  $m$ .

Si

$$m = 2^\alpha p_1^{\beta_1} \cdots p_r^{\beta_r} \quad (3.2)$$

où les  $p$  sont des nombres premiers, distincts, impairs, la période maximale possible est le plus petit multiple commun de :

$$\lambda(2^\alpha), \lambda(p_1^{\beta_1}), \cdots, \lambda(p_r^{\beta_r}) \quad (3.3)$$

où

$$\lambda(p^\beta) = p^{\beta-1}(p-1) \text{ si } p \text{ est impair}$$

$$\lambda(1) = \lambda(2) = 1; \lambda(4) = 2$$

$$\lambda(2^\alpha) = 2^{\alpha-2} \quad (\alpha > 2)$$

On obtient la période maximale si les trois conditions suivantes sont simultanément vérifiées :

- i)  $a^n \not\equiv 1 \pmod{p_j^{\beta_j}}$  pour  $0 < n < \lambda(p_j^{\beta_j})$
- ii)  $a \equiv 1 \pmod{2}$  si  $\alpha = 1$   
 $a \equiv 3 \pmod{4}$  si  $\alpha = 2$   
 $a \equiv 3 \text{ ou } 5 \pmod{8}$  si  $\alpha = 1$
- iii)  $X_0$  et  $m$  sont premiers entre eux

### 3.1.3 Le “test spectral”

Parmi les nombreux tests statistiques qui permettent de s’assurer du caractère aléatoire des nombres générés par un algorithme numérique [1], nous avons choisi le test spectral, non seulement parce qu’il peut être illustré graphiquement, mais parce qu’il apparaît comme l’un des tests les plus forts. Un mauvais algorithme peut passer avec succès un certain nombre de tests statistiques, excepté celui-ci.

Le test spectral examine les corrélations entre  $k$  nombres consécutifs  $\{X_{n+1}, X_{n+2}, \dots, X_{n+k}\}$ , fournis par un algorithme.

Même si un algorithme de congruence linéaire fournit des nombres individuellement “quasi-aléatoires”, les  $k$  – *uples* de nombres consécutifs sont loin d’être exempts de corrélations !

Si  $\{X_{n+1}, X_{n+2}, \dots, X_{n+k}\}$  représentent les coordonnées d’un point  $P_n$  dans un espace à  $k$  dimensions, alors les points  $P_n$  se répartissent sur des hyperplans à  $n - 1$  dimensions et on montre [3] qu’il y a **au plus**  $m^{1/k}$  tels plans (mais il peut y en avoir beaucoup moins !...).

Reprenons le même exemple de la Figure 3.1a, et reportons sur un plan les coordonnées des doublets (1,6), (6,15), (15,2) etc... . Les points obtenus se répartissent sur 4 lignes (Fig. 3.2). Puisque  $m = 16 = 2^4$ , on en attend *au plus*  $2^{4/2} = 4$  lignes.

Si on considère, non pas la suite des entiers  $X_n$ , mais les réels  $U_n = X_n/m$  appartenant à  $[0, 1[$ , pour les  $k$  – *uples* consécutifs  $P_n = (U_{n+1}, U_{n+2}, \dots, U_{n+k})$  représentés par des points d’un hypercube à  $k$  dimensions de côté  $[0, 1]$ , la distance maximale entre hyperplans

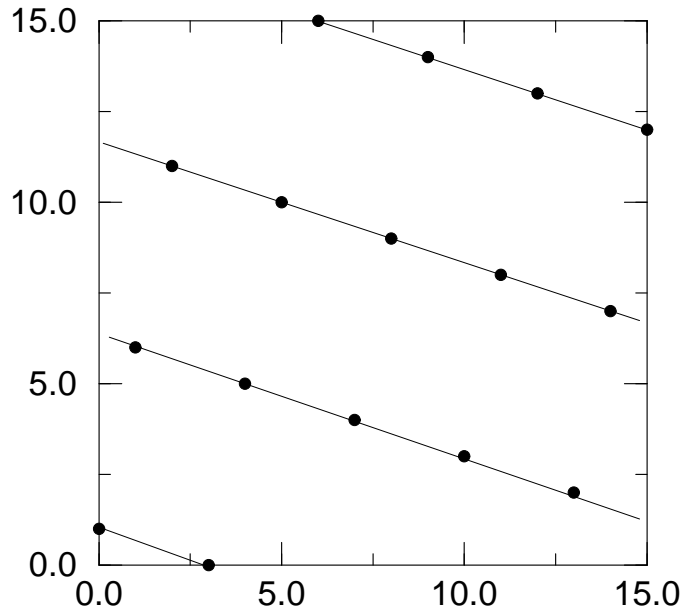


FIG. 3.2 – Corrélations entre paires pour la congruence linéaire :  $X_{n+1} = 5X_n + 1 \pmod{16}$ .

notée  $1/\nu_k$  par Knuth constitue une mesure des corrélations.

Nous donnons maintenant un exemple non trivial de générateur de nombres quasi-aléatoires, qui a été utilisé pendant plusieurs décennies sur des machines 32 bits (en particulier IBM) et qui subsistait encore comme l'un des sous programmes intrinsèques de génération de nombre aléatoires dans la version 1998 de "DIGITAL FORTRAN90" ! Son nom est "RANDU()". Il utilise un algorithme de congruence linéaire avec  $c = 0$ ,  $a = 2^{16} + 3$  et  $m = 2^{32}$ . Il est rapide puisque l'opération  $[\text{mod } m]$  s'effectue par une opération binaire AND et la multiplication par  $a$  peut être implémentée de manière efficace puisque  $(a - 3)$  est une puissance de 2. Par contre, les corrélations entre  $k$ -uplets  $(X_n, X_{n+1}, \dots, X_{n+k})$  sont catastrophiques. La Figure 3.3 illustre les corrélations entre triplets. L'utilisation de ce générateur de nombres quasi-aléatoires pour le calcul Monte-Carlo de l'aimantation dans le modèle d'Ising ferromagnétique à deux dimensions donne une valeur 10% plus forte que la valeur exacte !

Même pour des générateurs de nombres quasi-aléatoires relativement satisfaisant, des corrélations existent, mais à une échelle plus petite. L'une de bibliothèques mathématiques les plus utilisées : IMSL, propose un algorithme de congruence linéaire avec  $c = 0$ ,  $a = 7^5$  et  $m = 2^{31} - 1$ . Ici,  $a$  est suffisamment faible pour que la multiplication par  $a$  puisse être implémentée de manière efficace. Les corrélations entre triplets (Fig. 3.4) n'apparaissent qu'à une échelle 100 fois plus faible.

Knuth donne un tableau des  $\nu_k$ ,  $k = 2$  à 6 pour une trentaine d'algorithmes dont ceux

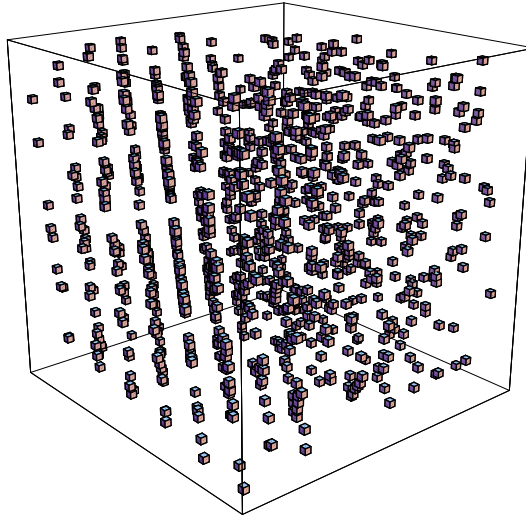


FIG. 3.3 – Corrélations entre triplets représentées dans un cube  $[0,1] \times [0,1] \times [0,1]$  pour la congruence linéaire  $X_{n+1} = (2^{16} + 3)X_n \pmod{2^{32}}$



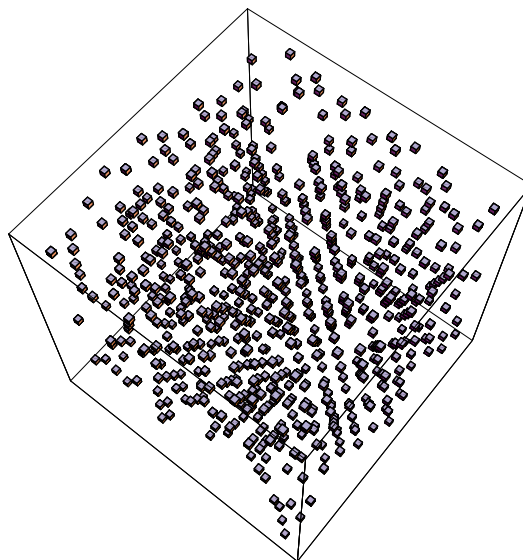


FIG. 3.4 – Corrélations entre triplets représentées dans un cube  $[0,0.01] \times [0,0.01] \times [0,0.01]$  pour la congruence linéaire  $X_{n+1} = 7^5 X_n \pmod{2^{31} - 1}$ .

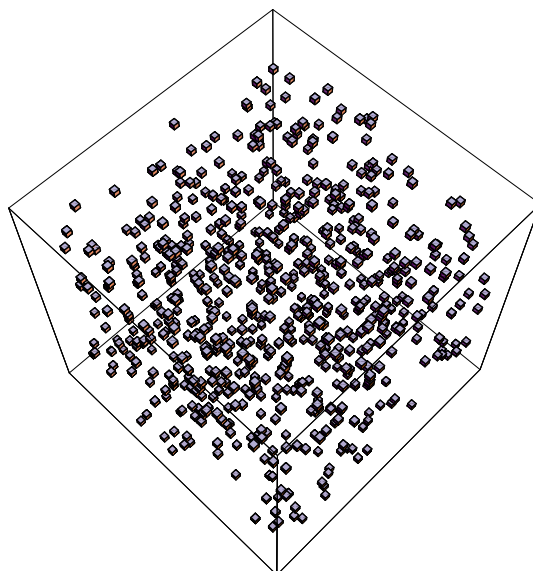


FIG. 3.5 – Corrélations entre triplets représentées dans un cube  $[0,0.01] \times [0,0.01] \times [0,0.01]$  pour la congruence linéaire  $X_{n+1} = 7^5 X_n \pmod{2^{31} - 1}$  après “shuffling”.

décrits ci-dessus.

Pour briser ces corrélations, une technique appelée “shuffling” (traduction : “battre les cartes”) peut être utilisée. L'algorithme correspondant est simple [1, 3] et consomme peu de temps CPU. Il consiste à garder constamment en mémoire un tableau des derniers  $p$  nombres tirés et les permuer aléatoirement en fonction de nouveaux nombres tirés au hasard par un autre algorithme.

La bibliothèque IMSL propose l'algorithme précédemment cité avec shuffling superposé. Le coût supplémentaire en temps CPU est de l'ordre de 50%, mais les corrélations (Fig. 3.5) disparaissent.

### 3.1.4 Suites de Fibonacci “retardées”

Dans les congruences considérées précédemment,  $X_n$  ne dépend que du nombre précédent  $X_{n-1}$ . On peut accroître la période en choisissant une relation où  $X_n$  dépend de plus d'une valeur précédente. La relation la plus simple, à cet effet, est :

$$X_{n+1} = (X_n + X_{n-1}) \pmod{m} \quad (3.4)$$

*En choisissant  $X_0 = X_1 = 1$ , la récurrence  $X_{n+1} = (X_{n-1} + X_{n-1})$  définit la “suite de Fibonacci” et on sait que lorsque  $n \rightarrow \infty$  le rapport  $X_{n+1}/X_n$  de deux termes successifs tend vers le nombre d'or. Elle ne peut conduire à un bon générateur de nombres aléatoires !*

Par contre on peut généraliser :

$$X_{n+1} = (X_{n-l} + X_{n-k}) \pmod{m} \quad (3.5)$$

Les nombres  $k$  et  $l$  sont appelés “lags” (i.e “retards” ou “décalages”) On montre que si  $m = 2^\alpha$  est une puissance de deux, et si  $l$  et  $k$  sont tels que le trinôme :

$$x^k + x^l + 1$$

est un polynôme premier (i.e. n'a aucun autre polynôme diviseur que lui-même ou une constante) dans le corps des entiers modulo 2, alors la période est  $2^{\alpha-1}(2^k - 1)$  ( $k > l$ ) et le nombre de cycles différents est  $2^{(k-1) \times (\alpha-1)}$

Les couples de “lags” (24,55), (38,89), (37,100), (30,127), (83,258) satisfont cette propriété[1, 5].

Pour  $k$  et  $l$  suffisamment grands, tous les test statistiques s'avèrent en général excellents, bien supérieurs aux résultats des congruences linéaires. Cette méthode présente de nombreux avantages :

- Calcul rapide, puisqu'il n'y a pas de multiplication et si  $m = 2^\alpha$ , l'opération “modulo  $m$ ” s'effectue par un cache et une opération binaire AND.
- La période peut être très longue.

- Le nombre de cycles indépendants  $(2^{(k-1) \times (\alpha-1)})$  est extrêmement grand, ce qui en fait actuellement une méthode de choix pour implantation sur ordinateurs massivement parallèles.

G.J Mitchell and D.P. Moore ont, les premiers proposé ce type d’algorithme [4, 1] avec  $m$  pair,  $p = 24$ ,  $q = 55$  et  $\{X_0, X_1, X_2, \dots, X_{54}\}$  entiers arbitraires, non tous pairs. Pour  $m = 2^\alpha$  sa période est  $2^{\alpha-1}(2^{55} - 1)$ . Il figure aussi parmi les générateurs de nombres aléatoires proposé dans “Numérical Recipes” [3].

Au lieu de la relation additive précédente, on peut préférer une relation multiplicative :

$$X_{n+1} = (X_{n-l} \times X_{n-k}) \pmod{m} \quad (3.6)$$

Les performances statistiques sont encore meilleures, mais au prix d’un accroissement du temps CPU.

Enfin, on peut encore utiliser des algorithmes hybrides en associant cette méthode et celle des congruences linéaires :

$$X_{n+1} = (a_1 X_{n-1} + a_2 X_{n-2} + \dots + a_k X_{n-k}) \pmod{m} \quad (3.7)$$

Avec  $m = 2$ , l’équation précédente peut fournir un excellent générateur de bits (0 ou 1) quasi-aléatoires. Dans ce cas, les coefficients  $a_i$  sont 0 ou 1 et le polynome

$$1 + a_1 x + a_2 x^2 + \dots + a_k x^k$$

doit être un polynome premier dans le corps des entiers modulo 2 (cf. les travaux de Tausworthe [6])

### 3.1.5 Puis-je utiliser la fonction “RAND(), random(), ...” de mon ordinateur ?

La réponse dans les années 80 était dans la plupart des cas négative et à cette époque, bon nombre de calculs stochastiques publiés étaient erronés par suite de l’usage de générateurs de nombres peu aléatoires !

Les choses on maintenant évolué. Sur les stations 32 bits, la plupart des constructeurs fournissent un choix de générateurs de nombre quasi-aléatoires dont certains sont satisfaisants (mais il faut encore se méfier -cf. fonction RANDU(), qui, malgré ses performances pitoyables n’a pas encore complètement disparu-).

Quelquefois on a accès à une notice détaillée sur le ou les générateurs de nombres aléatoires proposés avec description de l’algorithme utilisé et résultats de test statistiques (ceux préconisés par Knuth).

Dans le cas de générateurs “boite noire”, il est prudent de se livrer soi-même à quelques tests (ex : test spectral) avant de faire un choix.

L'avantage à utiliser un générateur de nombre aléatoires fourni par le constructeur est que l'algorithme est programmé en langage machine en tenant compte de son architecture spécifique. Le même algorithme, programmé en langage évolué (Fortran, C...) serait plus lent.

Par contre, si on souhaite qu'un programme soit portable et donne exactement les mêmes résultats, quelle que soit la machine, on peut programmer soi-même l'algorithme.

Les bibliothèques IMSL et NAG contiennent actuellement de bon sous-programmes de génération de nombres aléatoires.

Tous ces programmes fonctionnent de la manière suivante :

- Si on ne précise pas “la semence”, c'est à dire le ou les nombres initiaux pour démarrer l'algorithme, l'ordinateur le fait lui-même. Soit il utilise toujours les mêmes nombres (dans ce cas, si on relance le programme on obtient le même résultat), soit il les calcule par un algorithme, en prenant pour point de départ un nombre formé à partir de la date de son horloge (dans ce cas on a un résultat différent à chaque exécution).
- Il y a toujours possibilité de choisir les paramètres initiaux et les passer à l'ordinateur par un sous programme prévu à cet effet.

Nous terminons par quelques comparaisons de temps de calculs obtenus pour divers générateurs de nombre aléatoires sur un processeur Pentium IV cadencé à 3 GHz

Temps CPU pour le tirage de  $10^9$  nombres pseudo-aléatoires.

- (i) Bibliothèque mathématique IMSL, congruence linéaire avec  $c = 0$ ,  $m = 2^{31} - 1$ ,  $a = 7^5$  : 50 s CPU.
- (ii) Bibliothèque mathématique IMSL, méthode précédente + “shuffling” : 75 s CPU.
- (iii) Bibliothèque mathématique IMSL, congruence linéaire avec  $c = 0$ ,  $m = 2^{31} - 1$ ,  $a = 397204094$  : 100 s CPU.
- (iv) Fibonacci :  $X_n = (X_{n-24} + X_{n-50}) \pmod{10^9}$ , programme Fortran de “Numerical Recipes” [3] : 20 s CPU.

## 3.2 Echantillonnage d'une loi de probabilité non uniforme

### 3.2.1 Méthode de transformation

#### Changement de variable

Soit  $X$  une variable aléatoire réelle possédant la densité de probabilité  $f(x)$ . Soit  $\Psi(x)$  une fonction réelle, monotone croissante, continue et dérivable.

A  $X$  nous associons la variable aléatoire  $Y$  définie par :

$$Y = \Psi(X)$$

Nous allons montrer que la variable aléatoire  $Y$  possède une densité de probabilité  $g(y)$  que nous allons expliciter.

D'après la définition de la densité de probabilité :

$$P(\{x < X < x + dx\}) = f(x)dx \quad \text{pour } dx \rightarrow 0$$

mais puisque  $\Psi(x)$  est une fonction monotone croissante, elle réalise une bijection et le premier membre de cette égalité peut aussi s'écrire :

$$P(\{\Psi(x) < Y < \Psi(x + dx)\}) = P(\{\Psi(x) < Y < \Psi(x) + \frac{d\Psi(x)}{dx}dx\})$$

En notant  $y = \Psi(x)$  et  $dy = \frac{d\Psi(x)}{dx}dx$

$$P(\{y < Y < y + dy\}) = f(x)dx$$

Puisque  $\Psi$  est monotone, elle admet une fonction réciproque  $\Psi^{-1}$ , donc

$$x = \Psi^{-1}(y)$$

On a

$$dx = \frac{dy}{\frac{d\Psi(x)}{dx}} = dy \frac{d\Psi^{-1}(y)}{dy}$$

d'où le résultat final

$$P(\{y < Y < y + dy\}) = f(\Psi^{-1}(y)) \frac{d\Psi^{-1}(y)}{dy} dy$$

Ce qui prouve que  $Y$  admet une densité de probabilité  $g$

$$g(y) = f(\Psi^{-1}(y)) \frac{d\Psi^{-1}(y)}{dy}$$

Dans le cas d'une fonction monotone décroissante, on obtient :

$$g(y) = -f(\Psi^{-1}(y)) \frac{d\Psi^{-1}(y)}{dy} \quad (3.8)$$

Ce type de transformation peut se généraliser pour des variables aléatoires à valeurs dans  $R^d$ . Elle fait intervenir un Jacobien.

**Cas particulier où  $\psi(x) = F(x)$ ,  $F(x)$  représentant la fonction de répartition d'une variable aléatoire  $X$  de densité de probabilité  $f(x)$**

$$F(x) = \int_{-\infty}^x f(u) du$$

Par construction, la fonction de répartition  $F$  est positive, monotone, croissante, de dérivée :

$$\frac{dF(x)}{dx} = f(x) \quad (3.9)$$

Elle varie entre  $F(-\infty) = 0$  à  $F(\infty) = 1$

La fonction  $F(x)$  admet donc pour  $0 \leq y \leq 1$  une fonction inverse  $F^{-1}(y)$  ayant pour dérivée

$$\frac{dF^{-1}(y)}{dy} = \frac{1}{f(x)} = \frac{1}{f(F^{-1}(y))}$$

Et la densité de probabilité  $g(y)$  de la variable aléatoire  $Y = F(X)$  s'écrit donc :

$$g(y) = \frac{f(F^{-1}(y))}{f(F^{-1}(y))} = 1 \quad 0 \leq y \leq 1$$

Elle correspond à la loi uniforme sur le segment  $[0,1]$ .

### Echantillonnage d'une densité de probabilité quelconque $f$

- Etant donnée une densité de probabilité  $f(x)$ , on détermine la fonction de répartition  $F(x) = \int_{-\infty}^x f(u) du$ .
- On tire des nombres  $\eta$  au hasard suivant la loi uniforme sur le segment  $[0,1]$
- D'après ce qui précède les nombres  $x = F^{-1}(\eta)$  sont répartis suivant la densité de probabilité  $f$ .

### Quelques exemples

#### *loi exponentielle*

La loi exponentielle de densité de probabilité :

$$f(x) = \lambda \exp(-\lambda x) \quad 0 \leq x < \infty$$

intervient souvent en physique, elle régit le temps qui sépare deux désintégrations successives d'un élément radioactif. Sa fonction de répartition est :

$$F(x) = \lambda \int_0^x \exp(-\lambda u) du = -[\exp(-\lambda u)]_0^x = 1 - \exp(-\lambda x)$$

La fonction réciproque  $F^{-1}$  est donnée par l'inversion de la relation :

$$y = 1 - \exp(-\lambda x)$$

en

$$\exp(-\lambda x) = 1 - y$$

soit

$$x = -\text{Ln}(1 - y)/\lambda$$

Donc on tire un nombre  $\eta$  uniformément dans l'intervalle  $[0, 1]$  et les nombres  $x = -\text{Ln}(1 - \eta)/\lambda$  sont répartis suivant une loi exponentielle.

*loi gaussienne*

La loi gaussienne a pour densité :

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

sa fonction de répartition :

$$F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x \exp\left(-\frac{u^2}{2\sigma^2}\right)$$

s'exprime à l'aide de la fonction "erreur", qui elle même n'admet pas de fonction inverse analytique. On va donc devoir utiliser une astuce.

Il s'avère qu'il est plus simple de tirer un couple  $(X_1, X_2)$  de deux variables aléatoires gaussiennes indépendantes.

Puisque  $X_1$  et  $X_2$  sont deux variables aléatoires indépendantes, la densité de probabilité du couple  $(X_1, X_2)$  est le produit  $f(x_1)f(x_2)$  des densités de probabilité de chacune des deux variables aléatoires. C'est à dire que :

$$P(\{x_1 < X_1 < x_1 + dx_1\} \cap \{x_2 < X_2 < x_2 + dx_2\}) = \exp\left(-\frac{x_1^2 + x_2^2}{2\sigma^2}\right) \frac{dx_1 dx_2}{2\pi\sigma^2}$$

En coordonnées polaires :

$$x_1 = \rho \cos \theta$$

$$x_2 = \rho \sin \theta$$

$$dx_1 dx_2 = \rho d\rho d\theta$$

le second membre s'écrit :

$$\exp\left(-\frac{\rho^2}{2\sigma^2}\right) \frac{\rho}{\sigma^2} d\rho \frac{d\theta}{2\pi}$$

c'est à dire que pour échantillonner deux variables aléatoires gaussiennes indépendantes, on peut :

- Tirer un angle  $\theta$  uniformément sur le segment  $[0, 2\pi]$
- Tirer un rayon  $\rho$  suivant la loi de probabilité

$$h(\rho) = \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \frac{\rho}{\sigma^2}$$

les deux nombres  $x_1 = \rho \cos \theta$  et  $x_2 = \rho \sin \theta$  seront répartis suivant des distributions gaussiennes indépendantes.

Echantillonner  $\rho$  suivant la loi de probabilité  $h(\rho)$  est facile. La fonction de répartition correspondante est :

$$F(\rho) = \int_0^\rho \exp\left(-\frac{u^2}{2\sigma^2}\right) \frac{u}{\sigma^2} du = -\left[\exp\left(-\frac{u^2}{2\sigma^2}\right)\right]_0^\rho = 1 - \exp\left(-\frac{\rho^2}{2\sigma^2}\right)$$

Son inverse est obtenue en écrivant :

$$F(\rho) = r = 1 - \exp\left(-\frac{\rho^2}{2\sigma^2}\right)$$

d'où

$$\rho = \sqrt{-2\sigma^2 \text{Ln}(1-r)}$$

Donc on tire  $r$  uniformément sur le segment  $[0, 1]$  et le nombre  $\rho$  obtenu par la transformation précédente apparaît avec la densité de probabilité  $h(\rho)$ .

Pour les cas où la fonction de répartition n'admet pas d'inverse qui puisse s'exprimer de manière analytique, il nous reste une méthode qui coûte plus cher en tant de calcul, mais qui s'applique de manière tout à fait générale.

### 3.2.2 Méthode de réjection de Von Neumann

Cette méthode élémentaire est illustrée sur la Figure 3.6.

Soit à échantillonner la densité de probabilité proportionnelle à la fonction  $f(x)$  représentée par le trait plein sur le segment  $[xmin, xmax]$  et égale à 0 en dehors. Soit  $C$  une constante supérieure ou égale à  $Fmax$ , valeur maximale de  $f(x)$ . La méthode consiste à engendrer  $N$  couples de réels, uniformément dans le pavé  $[xmin, xmax] \times [0, C]$  et à ne retenir que ceux qui “tombent” sous la courbe  $f(x)$ .

Soit  $M$  le nombre total de couples retenus et  $\nu_M(x)dx$  le nombre de ceux d'entre eux dont l'abscisse est comprise entre  $x$  et  $x + dx$  ( $dx$  fixé, arbitrairement petit). Alors lorsque  $N$  tend vers l'infini le rapport

$$\frac{\nu_M(x)dx}{M} \quad \text{tend vers :} \quad \tilde{f}(x)dx = \frac{f(x)dx}{\int_{xmin}^{xmax} f(x)dx}$$

Concrètement :

- La valeur une valeur de  $x$  est “tirée” suivant une loi uniforme sur le segment  $[xmin, xmax]$



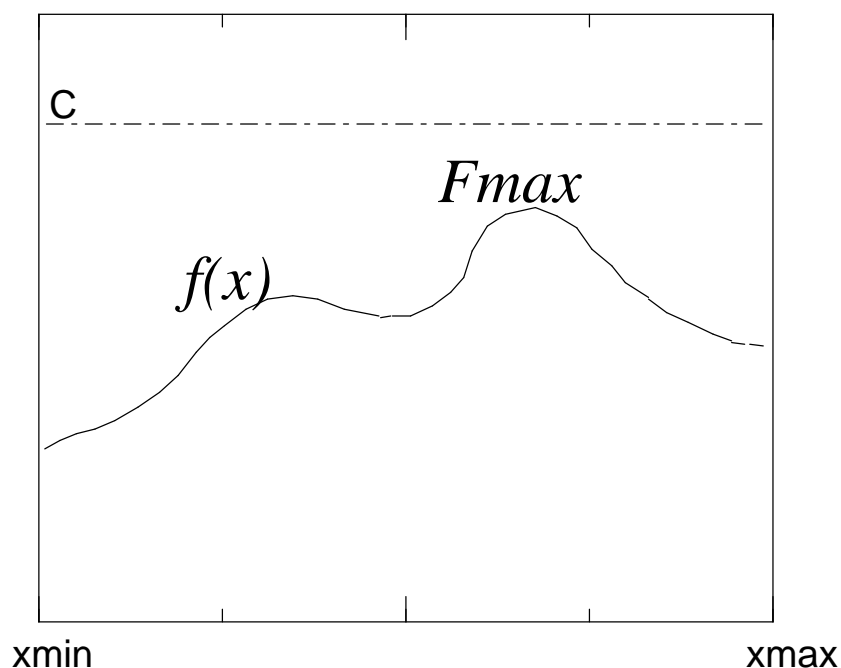


FIG. 3.6 – Méthode de réjection de Von Neuman

- Un nombre aléatoire  $\eta$  distribué uniformément dans le segment  $[0,1]$  est généré.
- La valeur  $x$  est acceptée si  $f(x)/C > \eta$ ; sinon, elle est rejetée.

La relation ci-dessus montre que la méthode peut permettre plus généralement d'engendrer une distribution de probabilité  $\tilde{f}(x)$ , proportionnelle à une fonction positive quelconque  $f(x)$ , même lorsque l'intégrale de cette dernière n'est pas normalisée à un.

La méthode de réjection est évidemment plus coûteuse en temps de calcul que la méthode de transformation, puisqu'une partie des tirages est simplement rejetée !

Lorsque  $Fmax$  est connu, on minimisera le nombre de rejections en prenant  $C = Fmax$ .





# Echantillonnage suivant l'importance.

## 4.1 Calcul stochastique d'une intégrale

Considérons une intégrale à  $d$  dimensions :

$$I = \int dx \Phi(x) \quad (4.1)$$

où  $x = (\alpha_1, \alpha_2, \dots, \alpha_d)$  est un vecteur à  $d$  dimensions et

$$\int dx = \int \int \dots \int d\alpha_1 d\alpha_2 \dots d\alpha_d$$

$\Phi(x)$  est une fonction de  $R^d$  dans  $R$ .

Choisissons arbitrairement une fonction  $f(x)$  ayant les propriétés d'une densité de probabilité, c'est à dire positive et telle que

$$\int dx f(x) = 1$$

En définissant :

$$g(x) = \frac{\Phi(x)}{f(x)}$$

L'intégrale s'écrit :

$$I = \int g(x) dP \quad \text{avec} \quad dP = f(x) dx \quad (4.2)$$

Elle représente la moyenne stochastique (ou "Espérance") de la variable aléatoire  $Y = g(X)$  transformée de la variable aléatoire  $X$  par la fonction  $g$

Si nous "tirons"  $N$  vecteurs  $x_i$  de  $R^d$  suivant la loi de probabilité  $P$ , nous pouvons associer au résultat de cette expérience la variable aléatoire :

$$\frac{S_N}{N} = \frac{Y_1 + Y_2 + \dots + Y_N}{N} = \frac{g(X_1) + g(X_2) + \dots + g(X_N)}{N}$$

D'après la loi forte des grands nombres (Annexe A), la "moyenne expérimentale" :

$$\frac{y_1 + y_2 + \dots + y_N}{N} = \frac{g(x_1) + g(x_2) + \dots + g(x_N)}{N} \quad \text{avec} \quad y_i = g(x_i)$$

converge "presque sûrement" vers  $I$  lorsque  $N \rightarrow \infty$ .

Si  $\sigma_g^2$  représente la variance de la variable aléatoire  $g(X)$  :

$$\sigma_g^2 = \int [g(x) - I]^2 f(x) dx = \int \left[ \frac{\Phi(x)}{f(x)} - I \right]^2 f(x) dx \quad (4.3)$$

alors la variance de la variable aléatoire  $S_N/N$  est  $\sigma_g^2/N$  (cf. Annexe A), et l'écart type  $\sigma_g/\sqrt{N}$  qui caractérise l'erreur décroît en  $1/\sqrt{N}$ .

## 4.2 Réduction de la variance

Nous allons maintenant mettre à profit le caractère arbitraire dans le choix de la densité de probabilité  $f$  pour réduire la variance, donc l'erreur.

Il est toutefois illusoire d'essayer d'annuler la variance. Ceci ne peut se faire que si  $\Phi(x)/f(x) = I$  dans la relation précédente, mais pour cela il faut connaître  $I$ , c'est à dire avoir résolu le problème !

Nous allons d'abord illustrer l'optimisation du choix de  $f$  sur un exemple à une dimension et nous reprendrons celui proposé par Hammersley et Handscomb dans leur ouvrage introductif aux méthodes de Monte-Carlo[7].

Considérons la fonction dans  $R$  :

$$\Phi(x) = 2 \frac{e^x - 1}{e - 1}$$

variant de 0 à 2 lorsque  $x$  varie de 0 à 1 et représentée sur la figure 4.1.

Nous nous proposons de calculer par la méthode de Monte-Carlo l'intégrale :

$$I = \int_0^1 \Phi(x) dx$$

Le premier choix qui vient à l'esprit pour la densité de probabilité  $f_u(x)$  est la loi uniforme sur le segment  $[0, 1]$  :

$$f_u(x) = 1 \quad \text{si} \quad x \in [0, 1] \quad \text{et} \quad f(x) = 0 \quad \text{sinon.}$$

La fonction choisie est intégrable analytiquement, on a  $I=0.836...$  et l'écart-type :

$$\sigma_g^u = \sqrt{\int_0^1 (\Phi(x) - I)^2 dx} = 0.5726$$

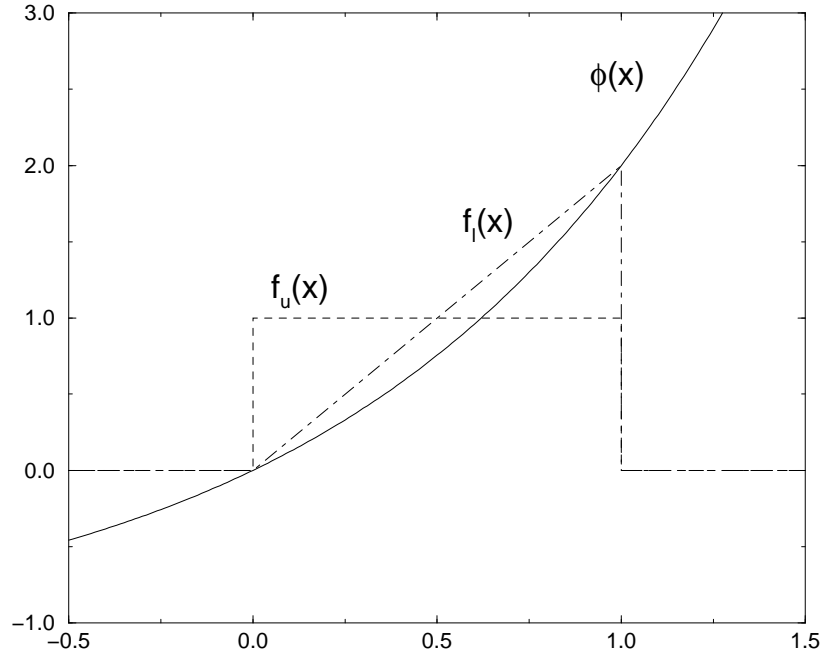


FIG. 4.1 – Trait plein : fonction  $\Phi(x)$  à intégrer entre 0 et 1. Tirets : densité de probabilité  $f_u$  uniforme. Traits mixtes : une densité de probabilité  $f_l$  simple réalisant un échantillonnage suivant l'importance.

On en déduit, par exemple, que pour obtenir  $I$  à 1% près, dans un “intervalle de confiance” à 95%, on doit sommer sur un nombre de points  $N$  défini par  $2\sigma/(\sqrt{N}I) = 0.01$  soit  $N = 18800$  points

Ce premier choix n'est pas très astucieux, car en “tirant” des points avec une loi uniforme, on a la même densité de points vers  $x = 0$  où la fonction est faible que vers  $x = 1$  où elle est maximale. On a plutôt intérêt à choisir une loi de probabilité qui donne une densité de points “tirés” plus importante là où la fonction est maximale, d'où l'idée “d'échantillonnage suivant l'importance”.

Prenons la densité de probabilité linéaire (cf. Fig. 4.1) :

$$f_l(x) = 2x \quad \text{si } x \in [0, 1] \quad \text{et} \quad f_l(x) = 0 \quad \text{sinon.}$$

qui est beaucoup plus proche de la fonction  $\Phi(x)$  que le choix précédent. Alors l'écart type est :

$$\sigma_g^l = \sqrt{\int_0^1 \left( \frac{\Phi(x)}{2x} - I \right)^2 2x dx} = 0.1048$$

soit 5.49 fois plus faible. On obtiendra donc la même précision avec  $5.49^2 = 30$  fois moins de points soit seulement  $N = 630$ . Cet exemple à une dimension reste toutefois académique, dans la mesure où une méthode de Gauss à 5 points donne une précision nettement supérieure !

L'échantillonnage suivant l'importance devient par contre spectaculaire pour des intégrales à grand nombre de dimensions. Pour rester dans le cadre de calculs que l'on peut encore faire analytiquement, généralisons l'exemple précédent à  $d$  dimensions en prenant sur  $R^d$  (cf. Negele et Orland[8]) :

$$\Psi(\alpha_1, \alpha_2, \dots, \alpha_d) = \Phi(\alpha_1)\Phi(\alpha_2) \cdots \Phi(\alpha_d)$$

que nous intégrons sur un hypercube  $[0, 1] \times [0, 1] \times \cdots \times [0, 1]$  Et prenons pour densité de probabilité le produit :

$$f(\alpha_1)f(\alpha_2) \cdots f(\alpha_d)$$

Du fait de la séparabilité de l'intégrale, on obtient aisément la valeur de l'intégrale à  $d$  dimensions :  $I_d = I^d$  ainsi que la variance :

$$\sigma_d^2 = \prod_{i=1}^d \int \left[ \frac{\Phi(\alpha_i)}{f(\alpha_i)} \right]^2 f(\alpha_i) d\alpha_i - \left[ \prod_{i=1}^d \int f(\alpha_i) d\alpha_i \right]^2$$

soit :

$$\sigma_d^2 = (\sigma_g^2 + I^2)^d - I^{2d}$$

et en divisant chaque membre par  $I^{2d}$

$$\frac{\sigma_d^2}{I^{2d}} = \left( \frac{\sigma_g^2}{I^2} + 1 \right)^d - 1$$

avec  $\sigma_g^u = 0.5726$  pour la loi uniforme on obtient, pour  $d$  grand :

$$\frac{\sigma_d^u}{I^d} \approx 1.4691^{d/2}$$

avec  $\sigma_g^l = 0.1048$  pour un échantillonnage suivant l'importance (loi linéaire) on a, pour  $d$  grand

$$\frac{\sigma_d^l}{I^d} \approx 1.0157^{d/2}$$

Pour  $d = 100$ , ce qui reste modeste pour un calcul Monte-Carlo, on améliore l'écart-type, donc l'erreur statistique, d'un facteur

$$(1.469/1.016)^{50} \approx 10^8$$

par l'échantillonnage suivant l'importance.

Pour fixer les idées, pour obtenir l'intégrale à  $d=100$  dimensions, dans un intervalle de confiance à 95%, avec une précision relative de 1%, il nous faut, avec la loi uniforme un nombre  $N$  tel que :

$$\frac{2 * 1.469^{50}}{\sqrt{N}} = 0.01$$

soit  $N = 2 * 10^{21}$  points. Et pour la même précision, il ne nous faudra avec la loi linéaire que  $N = 2 * 10^{21} / 10^{16} = 2 * 10^5$  points

Pour fixer les idées, en supposant qu'avec un Pentium 4, cadencé à 3 GHz, le "tirage" d'un point  $x$  et le calcul de  $\Psi(x)$  nécessite 300 cycles. L'échantillonnage avec  $N = 2 * 10^{21}$  points nécessiterait  $2 * 10^{14}$  s, soit six millions d'années.

Avec un échantillonnage suivant l'importance élémentaire, on obtient le même résultat en vingt millisecondes !...







# Simulation de systèmes statistiques

## 5.1 Introduction

Dans le cadre de l'ensemble "canonique", considérons un système  $\mathcal{S}$  en contact avec un réservoir d'énergie (ou thermostat) qui impose sa température  $T$ . Supposons d'abord que le système est décrit par un ensemble d'états discrets. La probabilité pour le système  $\mathcal{S}$  de se trouver dans un état  $l$  d'énergie  $E_l$  est :

$$P_l = \frac{\exp[-E_l/(k_B T)]}{Z}$$

où

$$Z = \sum_l \exp[-E_l/(k_B T)]$$

est la fonction de partition du système.

La valeur moyenne de l'énergie est donc :

$$\langle E \rangle = \frac{\sum_l E_l \exp[-E_l/(k_B T)]}{Z}$$

et plus généralement, la valeur moyenne de toute grandeur  $\mathcal{O}_l$  associée à l'état  $l$  est :

$$\langle \mathcal{O} \rangle = \frac{\sum_l \mathcal{O}_l \exp[-E_l/(k_B T)]}{Z}$$

Si nous passons à la limite continue : l'état du système  $\mathcal{S}$  est caractérisé par un ensemble de variables que nous noterons collectivement  $\mathbf{x}$ ,  $\mathbf{x} \in \mathbb{R}^d$  représentant un vecteur à  $d$  dimensions (par exemple l'ensemble des coordonnées  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  des particules qui le constituent). L'énergie correspondant à l'état  $\mathbf{x}$  est  $E(\mathbf{x})$ . La probabilité de trouver le système dans un état dont les variables  $\mathbf{x}$  sont comprises entre  $\mathbf{x}$  et  $\mathbf{x} + d\mathbf{x}$  est :

$$\frac{\exp[-E(\mathbf{x})/(k_B T)]}{Z} d\mathbf{x}$$

La fonction de partition  $Z$  s'exprimant alors comme une intégrale multiple :

$$Z = \int \exp[-E(\mathbf{x})/(k_B T)] d\mathbf{x}$$

La valeur moyenne de toute grandeur  $\mathcal{O}(\mathbf{x})$  associée à l'état  $\mathbf{x}$  (par exemple, l'énergie  $\mathcal{O}(\mathbf{x}) = E(\mathbf{x})$ ) est définie par :

$$\langle \mathcal{O} \rangle = \frac{\int \mathcal{O}(\mathbf{x}) \exp[-E(\mathbf{x})/(k_B T)] d\mathbf{x}}{Z} \quad (5.1)$$

L'intérêt de la méthode de Monte-Carlo, avec "échantillonnage suivant l'importance" pour le calcul de l'intégrale multidimensionnelle 5.1 est immédiat :

- Si on est capable d'engendrer un ensemble de configurations  $\mathbf{x}$  correspondant à une variable aléatoire  $\mathbf{X}$ , à valeurs dans  $\mathbb{R}^d$  ayant pour densité de probabilité :

$$\frac{\exp[-E(\mathbf{x})/(k_B T)]}{Z} \quad (5.2)$$

- Alors la valeur moyenne  $\langle \mathcal{O} \rangle$  de la grandeur physique  $\mathcal{O}(\mathbf{x})$  apparaît comme l'espérance de la variable aléatoire  $\mathcal{O}(\mathbf{X})$ .

Le problème crucial est que l'on ignore dans l'expression de la densité de probabilité 5.2 la valeur de la constante de normalisation  $Z$ , c'est à dire la fonction de partition ! On pourrait éventuellement s'en sortir en utilisant la méthode de réjection de Von Neumann, mais dans ce problème à très grand nombre de dimensions, cette méthode s'avère tout à fait inefficace. Une méthode élégante et efficace a été proposée par Métropolis. Elle construit un ensemble de configurations suivant la densité de probabilité 5.2, à partir des états successifs d'une chaîne de Markov ayant atteint sa distribution d'équilibre.

## 5.2 Chaînes de Markov

### 5.2.1 Définition

Une suite infinie *ordonnée*  $(X_1, X_2, \dots, X_t, \dots)$  de variables aléatoires constitue une chaîne de Markov si la loi de probabilité conditionnelle de  $X_{t+1}$  lorsqu'on se donne les valeurs de  $X_1, X_2, \dots, X_t$ , se réduit à la loi de probabilité conditionnelle de  $X_{t+1}$  lorsqu'on se donne seulement la valeur de  $X_t$ .

[ $t$  désigne ici un entier naturel, faisant référence à un temps discrétisé]

Remarque importante : pour une suite ainsi définie, la connaissance du passé influe sur l'avenir ; mais, d'un ensemble d'événements passés, seule subsiste l'influence du plus récent. Si l'on se donne  $X_t$ , la connaissance de  $X_{t'}$  pour  $t' > t$  n'est en rien précisée par la donnée des valeurs de  $X_{t-1}, X_{t-2}, \dots, X_1$ .

Nous nous restreindrons aux Chaînes de Markov "homogènes" où la loi de probabilité conditionnelle de  $X_{t+1}$  lorsqu'on se donne la valeur de  $X_t$  ne dépend pas de  $t$ . Alors la chaîne de Markov est entièrement déterminée par cette loi de probabilité conditionnelle appelée "*loi de transition*" et par la loi de probabilité initiale de la variable aléatoire  $X_1$ .

Nous supposons aussi que la loi de probabilité conditionnelle de  $X_{t+1}$  lorsqu'on se donne la valeur de  $X_t$  possède la densité de probabilité  $p(y/x)$ ,

$$P(\{y < X_{t+1} < y + dy\} / \{X_t = x\}) = p(y/x)dy \quad (5.3)$$

Et nous noterons :

$$p(y/x) = p(x \rightarrow y)$$

Comme toute densité de probabilité, la fonction  $p(y/x) = p(x \rightarrow y)$  satisfait quel que soit  $x$  à la condition de normalisation :

$$\int p(y/x)dy = \int p(x \rightarrow y)dy = 1 \quad (5.4)$$

Au cours des rappels de théorie des probabilités (Annexe B), nous avons montré, dans le cas de chaînes de Markov à valeurs dans un ensemble discret, que lorsque la chaîne de Markov possède la propriété *d'ergodicité*, c'est à dire que pour tout couple d'états discrets  $(a_i, a_j)$ , il existe une probabilité non nulle de passer de l'un à l'autre après  $n$  pas, alors cette chaîne converge vers une loi de distribution invariante *unique*.

Nous admettons que ce résultat reste vrai dans le cas continu : si pour tout couple d'état  $(x, y)$  il existe une probabilité non nulle de passer de l'un à l'autre en un nombre fini d'étapes  $n$ , alors il y a convergence vers une loi de distribution invariante *unique*.

Il nous reste alors à imposer une condition à la loi de transition  $p(y/x) = p(x \rightarrow y)$  pour que cette loi de distribution invariante limite soit :

$$f(\mathbf{y}) = \frac{\exp[-E(\mathbf{y})/(k_B T)]}{Z}$$

### 5.3 Convergence vers une loi de distribution invariante imposée

Théorème

- Si “la loi de transition”  $p(x \rightarrow y)$  est *ergodique*, c'est à dire que pour tout couple d'états  $(x, y)$  il existe une probabilité non nulle de passer de l'un à l'autre en  $n$  étapes [–donc tous les états sont “visités”–]
- si  $p(x \rightarrow y)$  satisfait à la condition “*de microréversibilité*”

$$[f(x)p(x \rightarrow y) - f(y)p(y \rightarrow x)] = 0 \quad (5.5)$$

*Cette condition de microréversibilité est encore appelée principe du bilan détaillé par référence à l'équation maîtresse et au principe du bilan détaillé qui en découle en thermodynamique (cf. Annexe C)*

Alors la loi de distribution de la chaîne de Markov correspondante converge vers une distribution invariante qui est proportionnelle à  $f(x)$ .

Preuve :

Puisque la chaîne de Markov est ergodique, nous savons que sa loi de distribution converge vers une distribution invariante et que cette distribution invariante est unique. Il nous suffit donc de montrer que la distribution  $f(x)$  est invariante par rapport à la loi de transition  $p(x \rightarrow y)$ .

Si “à l’instant  $t$ ”,  $X_t$  est distribuée suivant la loi  $\pi^t(x) = f(x)$ , alors, en vertu de l’équation de Chapman-Kolmogorov B.3, “à l’instant  $(t + 1)$ ”,  $X_{t+1}$  est distribuée suivant la loi de probabilité :

$$\pi^{t+1}(x) = \int du f(u) p(u \rightarrow x)$$

Si  $p(u \rightarrow x)$  satisfait la relation du bilan détaillé 5.5, la relation précédente peut encore s’écrire :

$$\pi^{t+1}(x) = \int du f(x) p(x \rightarrow u) = f(x) \int du p(x \rightarrow u)$$

et le second membre est simplement :

$$\pi^{t+1}(x) = f(x)$$

puisque  $\int du p(x \rightarrow u) = 1 \quad \forall x$

### 5.3.1 Echantillonnage de $f(\mathbf{x}) = e^{-\beta E(\mathbf{x})}/Z$

Etant donnée une chaîne de Markov de loi de transition  $p(y \rightarrow x)$ , nous allons “échantillonner” une suite de points  $\{x_1, x_2, \dots, x_t, \dots\}$  ( $x_i \in R^d$ ) suivant cette chaîne, i.e. :  $x_t$  ayant été obtenu, le point suivant  $x_{t+1}$  est “tiré” suivant la loi de probabilité  $p(x_t \rightarrow x_{t+1})$ .

Si  $p(x_t \rightarrow x_{t+1})$  satisfait la relation 5.5 du bilan détaillé, avec  $f(x) = e^{-\beta E(x)}/Z$ , c’est à dire :

$$\left[ \frac{e^{-\beta E(x_t)}}{Z} p(x_t \rightarrow x_{t+1}) - \frac{e^{-\beta E(x_{t+1})}}{Z} p(x_{t+1} \rightarrow x_t) \right] = 0$$

alors après un certain “temps d’équilibre”, c’est à dire pour  $n > N$  ( $N$  suffisamment grand) les  $x_t$  sont distribués suivant la densité de probabilité  $e^{-\beta E}/Z$ .

*Remarque importante : on peut multiplier chaque membre de l’équation ci-dessus par  $Z$  :*

$$\left[ e^{-\beta E(x_t)} p(x_t \rightarrow x_{t+1}) - e^{-\beta E(x_{t+1})} p(x_{t+1} \rightarrow x_t) \right] = 0$$

*La connaissance de la constante de normalisation (ici fonction de partition  $Z$ ) devient absolument inutile pour l’échantillonnage suivant la loi de probabilité  $e^{-\beta E}/Z$ .*

## 5.4 Algorithme de Metropolis

L'algorithme le plus utilisée pour engendrer une chaîne de Markov qui satisfasse le *principe du bilan détaillé* a été proposé par Metropolis et al.[9].

Partant d'une configuration  $x_t = x$ , d'énergie  $E(x)$ , on définit une configuration d'essai  $y$  en déplaçant une ou plusieurs particules et on calcule sa nouvelle énergie  $E(y)$ . Cette configuration d'essai est acceptée pour nouvelle configuration avec la probabilité  $p(x \rightarrow y) = \min\{1, e^{-\beta E(y)} / e^{-\beta E(x)}\}$ , sinon, elle est rejetée et on prend pour nouvelle configuration  $x_{t+1} = x_t = x$  c'est à dire :

- $x_{t+1} = y$  avec la probabilité  $p = \min\{1, e^{-\beta E(y)} / e^{-\beta E(x)}\}$
- $x_{t+1} = x_t = x$  avec la probabilité  $(1 - p)$

En pratique :

- Si  $E(y) \leq E(x)$ , alors  $x_{t+1} = y$ , la configuration d'essai est acceptée avec certitude ( $p = 1$ )
- Si  $E(y) > E(x)$ , alors on "tire" un nombre  $\eta$  suivant la loi uniforme dans le segment  $[0,1]$  et on accepte pour nouvelle configuration la configuration d'essai  $y$  si  $e^{-\beta E(y)} / e^{-\beta E(x)} > \eta$ , sinon la nouvelle configuration est identique à l'ancienne :  $x_{t+1} = x_t = x$ .

Cette technique simple, similaire à la méthode de réjection de Von Neuman décrite précédemment, impose le fait que la configuration d'essai est acceptée avec la probabilité  $p(x \rightarrow y) = \min\{1, e^{-\beta E(y)} / e^{-\beta E(x)}\}$ .

Il nous reste à vérifier que cette probabilité  $p(x \rightarrow y)$  satisfait effectivement l'équation du bilan détaillé.

Supposons  $E(y) > E(x)$ , alors :

$$p(y \rightarrow x) = 1 \quad \text{et} \quad p(x \rightarrow y) = e^{-\beta E(y)} / e^{-\beta E(x)}$$

la relation

$$p(y \rightarrow x)e^{-\beta E(y)} = p(x \rightarrow y)e^{-\beta E(x)}$$

est bien vérifiée.

La vérification pour le cas  $E(y) \leq E(x)$  est analogue.





# Le modèle d'Ising

Nous allons “expérimenter” l’algorithme de Metropolis sur un modèle simple, dont les propriétés thermodynamiques sont déjà d’une extrême richesse.

## 6.1 L’Hamiltonien d’Ising

On considère un ensemble de spins  $s_i$ , porteurs d’un moment magnétique, qui ne peut prendre que deux valeurs :

- $+1$  si le moment magnétique est dirigé “vers le haut”
- $-1$  si le moment magnétique est dirigé “vers le bas”

Deux spins premiers voisins subissent une interaction  $J$  qui tend à les aligner. A température nulle, les spins seront donc tous alignés pour minimiser l’énergie : ceci constitue le modèle le plus simple pour un corps ferromagnétique. En l’absence de champ magnétique, l’état fondamental du système est doublement dégénéré car les spins peuvent être alignés tous vers le haut ou tous vers le bas. Si on applique une induction magnétique  $B$  dirigée vers le haut, par exemple, on lève la dégénérescence : tous les spins s’orientent vers le haut. L’Hamiltonien s’écrit :

$$H = -J \sum_{\langle \alpha, \beta \rangle}^{1er\ vois.} s_\alpha s_\beta - B \sum_{\alpha} s_\alpha \quad (6.1)$$

La première somme est effectuée sur toutes les paires distinctes de premiers voisins.

Nous nous limiterons, en Travaux Pratiques au cas où l’induction magnétique est nulle  $B = 0$ .

## 6.2 Modélisation par la méthode de Monte-Carlo

On considère un réseau carré de taille  $L \times L$  comprenant un nombre  $N = L^2$  de sites. Les sites sont repérés par leurs coordonnées  $\alpha = (i, j)$ ,  $i$  et  $j$  variant de  $0$  à  $L - 1$ . On applique des conditions aux limites périodiques :

$$s_{(i+pL, j+qL)} = s_{(i, j)}$$

On travaille dans l'ensemble canonique à une température donnée  $T$ .

On utilise l'algorithme de Metropolis en envisageant des mouvements élémentaires locaux (retournement d'un spin).

1. On choisit une configuration initiale, par exemple :
  - Tous les spins alignés
  - Spins aléatoirement distribués vers le haut ou vers le bas
2. On choisit un spin, au hasard (suivant une loi de distribution uniforme)
3. On envisage comme tentative de mouvement le retournement de ce spin
4. On calcule la différence d'énergie d'interaction des spins  $\Delta$  entre la nouvelle configuration avec un spin retourné et la configuration de départ.
5. Si  $\Delta$  est négatif, ou nul c'est à dire, si retourner un spin diminue l'énergie, on la laisse invariante, on accepte la nouvelle configuration
6. Si  $\Delta$  est positif :
  - On tire un nombre  $\eta$ , au hasard, suivant une loi uniforme sur le segment  $[0, 1]$
  - Si  $\eta < \exp(-\Delta/kT)$  on accepte la configuration avec spin retourné comme nouvelle configuration. Sinon on la rejette et la configuration à l'étape suivante est identique à la configuration précédente.
7. On calcule un certain nombre de variables :
  - énergie d'interaction  $E = -J \sum_{\langle \alpha, \beta \rangle} s_\alpha s_\beta$ ,
  - Aimantation par spin  $m = \sum_\alpha s_\alpha / N$
  - etc...
8. On réitère

### 6.3 Analyse des résultats

Fixons la température à  $kT/J = 3$  et partons d'une configuration initiale entièrement désordonnée. On obtient, en fonction du nombre de pas Monte-Carlo l'énergie  $E$  représentée sur la figure 6.1. On peut estimer que l'équilibre est atteint lorsque l'énergie ne dérive plus, c'est à dire environ  $20 \cdot N$  itérations (i.e après avoir tenté en moyenne 20 fois de retourner chaque spin)

L'image instantanée du réseau au bout de  $100 \cdot N$  itérations est représentée sur la figure 6.2. Les spins ne sont pas répartis au hasard, on voit se développer des domaines  $s = +1$  et des domaines  $s = -1$  dont la taille va croître si on diminue la température.

L'évolution de l'aimantation par spin est représentée sur la figure 6.3.

A partir du moment où l'équilibre est atteint, on va pouvoir calculer les valeurs moyennes des grandeurs thermodynamiques intéressantes :

- L'énergie moyenne  $\langle E \rangle$
- L'aimantation moyenne par site  $\langle m \rangle$



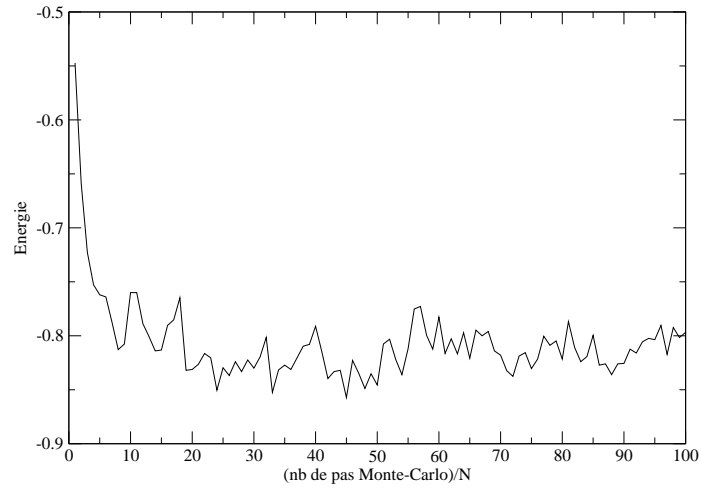


FIG. 6.1 – Evolution de l'énergie "instantanée" en fonction du nombre de pas Monte-Carlo pour  $kT/J = 3$ . Une estimation grossière du nombre de pas nécessaires pour atteindre l'équilibre est donné par le nombre d'itérations au bout duquel l'énergie ne dérive plus, c'est à dire environ  $20 \cdot N$  pas (i.e on a tenté de retourner chaque spin en moyenne 20 fois)

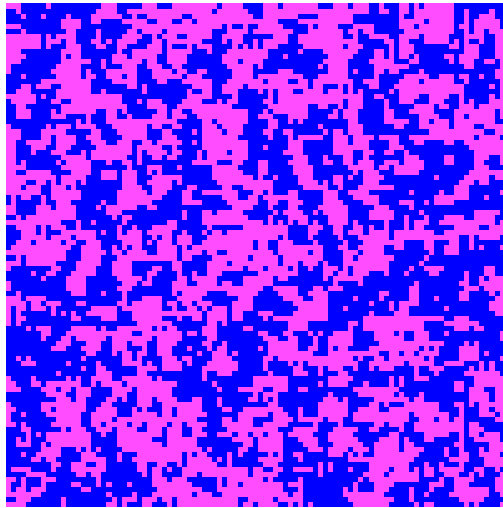


FIG. 6.2 – Image instantanée du réseau au bout de  $100 \cdot N$  itérations.

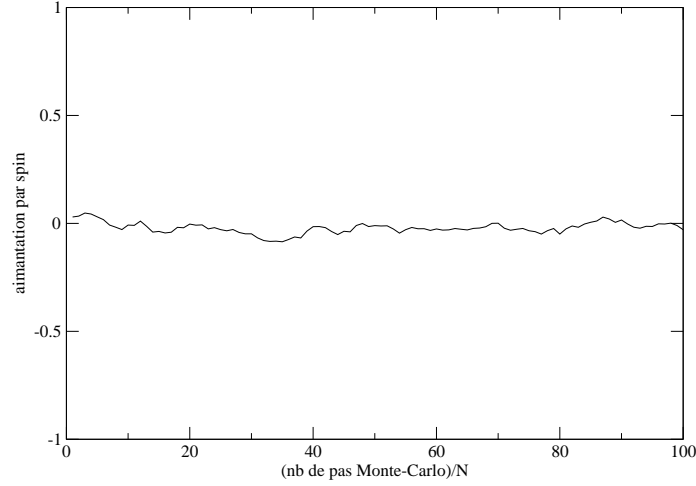


FIG. 6.3 – Evolution de l'aimantation  $m$  pour  $k_B T/J = 3$ .

– La chaleur spécifique à volume constant :

$$C_v = \frac{\partial \bar{E}}{\partial T} = \frac{\partial}{\partial T} \left[ \sum_i E_i \exp(-E_i/kT) / Z \right]$$

$$C_v = (\langle E^2 \rangle - \langle E \rangle^2) / kT^2$$

Il nous faut maintenant chiffrer l'erreur sur le calcul de ces valeurs moyennes en fonction du nombre d'itérations.

Considérons, par exemple l'aimantation moyenne par site :

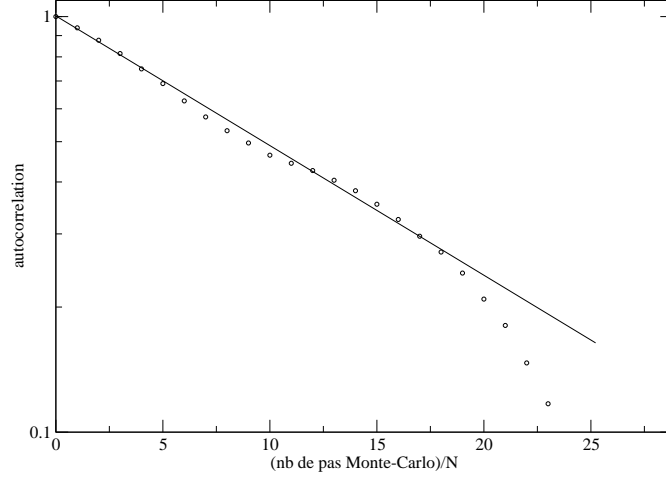
$$\langle m \rangle = \frac{\sum_{t=t_{eq}}^{t_{max}} m(t)}{t_{max} - t_{eq}}$$

(Chaque itération est repérée par la variable discrète  $t$ ,  $t_{eq}$  correspond au nombre d'itérations au bout desquelles on estime avoir atteint l'équilibre et  $t_{max}$  correspond au nombre total d'itérations).

On peut estimer un intervalle de confiance pour cette valeur moyenne en calculant l'écart quadratique moyen :

$$\sigma = \sqrt{\frac{\sum_{t=t_{eq}}^{t_{max}} (m(t) - \langle m \rangle)^2}{t_{max} - t_{eq}}} = \sqrt{\langle m^2 \rangle - \langle m \rangle^2}$$

Mais le théorème de la limite centrale suppose que les  $(t_{max} - t_{eq})$  mesures de  $m$  sont indépendantes, ce qui n'est pas vrai. Ayant fait une mesure, il nous faut estimer le "temps"

FIG. 6.4 – Autocorrélation de l'aimantation  $\chi(t)/\chi(0)$ 

$\tau$  (i.e. nombre d'itérations Monte-Carlo) qu'il faut attendre pour obtenir une nouvelle mesure indépendante de la première.

Ceci peut être chiffré en calculant une “fonction d'autocorrélation”. Pour l'aimantation par site, cette fonction est :

$$\chi(t) = \int dt' [m(t') - \langle m \rangle] [m(t' + t) - \langle m \rangle]$$

soit

$$\chi(t) = \int dt' [m(t)m(t' + t) - \langle m \rangle^2]$$

Pour une variable  $t$  discrète, on calculera explicitement :

$$\begin{aligned} \chi(t) &= \frac{1}{t_{max} - t} \sum_{t'=0}^{t_{max}-t} m(t')m(t' + t) \\ &- \frac{1}{t_{max} - t} \sum_{t'=0}^{t_{max}-t} m(t') \times \frac{1}{t_{max} - t} \sum_{t'=0}^{t_{max}-t} m(t' + t) \end{aligned} \quad (6.2)$$

qui correspond à une discrétisation de l'intégrale précédente. Dans les mêmes conditions que pour les figures précédentes ( $kT/J = 3$ ), on obtient le  $\chi(t)/\chi(0)$ , représenté sur la figure 6.4, en échelle semi logarithmique.

Cette fonction suit, au départ, une loi exponentielle décroissante :

$$\chi(t) = \chi(0) \exp(-t/\tau)$$

$\tau$  représente le temps de corrélation (ici  $\tau \approx 15 \times N$ )

Pour estimer la valeur moyenne de  $m$ , on pourrait n'additionner que des valeurs sélectionnées à des intervalles de temps supérieurs à  $\tau$ . On préfère additionner les valeurs obtenues 'à chaque temps  $t$ , mais alors le nombre de mesures indépendantes ne sera pas  $N = t_{max} - t_{eq}$  mais  $N' = (t_{max} - t_{eq})/\tau$  et l'intervalle de confiance pour la mesure de  $\langle m \rangle$  sera obtenu à partir de :

$$\sigma \sqrt{\frac{\tau}{(t_{max} - t_{eq})}}$$

## 6.4 Evolution en fonction de la température

On part de  $T = 0$  et d'un état complètement ordonné. On augmente la température par pas discrets et pour chaque température on applique l'algorithme précédent sur  $200 \times N$  pas en effectuant les valeurs moyennes sur les  $100 \times N$  derniers. Lorsqu'on varie la température, on prend pour état initial, l'état final obtenu à la température précédente. On obtient alors

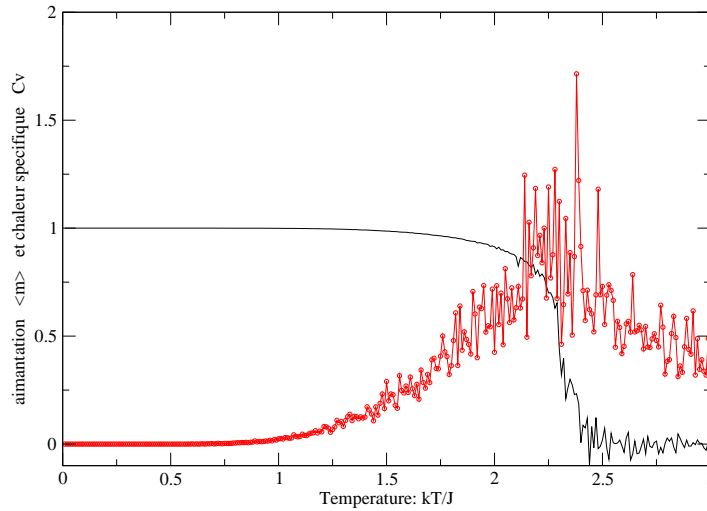


FIG. 6.5 – Aimantation moyenne par site et chaleur spécifique par site en fonction de la température

pour l'aimantation moyenne et pour la chaleur spécifique les courbes représentées sur la figure 6.5.

On observe une transition de phase vers un état complètement ordonné à  $kT/J \approx 2.3$ . On constate que l'erreur, en particulier sur la chaleur spécifique est d'autant plus grande que l'on s'approche de la transition. La raison est que lorsqu'on s'approche de la transitions, les "longueurs de corrélation" qui se traduisent par la taille des domaines ordonnés à courte distance divergent. Notre algorithme "local" n'est plus adéquat.

Ce phénomène est appelé "ralentissement critique"





# Utilisation de biais. Algorithme de Wolff

---

## 7.1 Introduction de biais

Dans l'algorithme utilisé précédemment pour la modèle d'Ising, la probabilité de choisir un spin parmi  $N$  pour le retourner est toujours  $1/N$ , quelle que soit la configuration de spins.

On peut envisager des algorithmes plus généraux, pour lesquels un certain mouvement de la configuration  $x$  vers la configuration  $y$  est choisi avec une probabilité  $P_{gen}(x \rightarrow y)$  puis, il est accepté avec une probabilité  $P_{acc}(x \rightarrow y)$ . Dans ce cas, la "loi de passage" de  $x$  à  $y$  est le produit de ces deux probabilités :

$$P(x \rightarrow y) = P_{gen}(x \rightarrow y)P_{acc}(x \rightarrow y)$$

En travaillant dans l'ensemble canonique, la relation du bilan détaillé s'écrit :

$$\begin{aligned} P_{gen}(x \rightarrow y)P_{acc}(x \rightarrow y) \exp(-E(x)/kT) = \\ P_{gen}(y \rightarrow x)P_{acc}(y \rightarrow x) \exp(-E(y)/kT) \end{aligned} \quad (7.1)$$

Si on utilise l'algorithme de Metropolis, la règle d'acceptation sera :

$$P_{acc}(x \rightarrow y) = \text{Min} \left\{ 1, \exp[-(E(y) - E(x))/kT] \frac{P_{gen}(y \rightarrow x)}{P_{gen}(x \rightarrow y)} \right\}$$

On retrouve la règle usuelle de Métropolis dans le cas où  $P_{gen}$  est symétrique :

$$P_{gen}(x \rightarrow y) = P_{gen}(y \rightarrow x)$$

Par contre rien n'empêche de créer une dissymétrie telle que la probabilité de générer  $x$  à partir de  $y$  soit supérieure à la probabilité de générer  $y$  à partir de  $x$ , d'où le nom de "biais". L'idéal serait de réaliser :

$$P_{gen}(x \rightarrow y) \exp(-E(x)/kT) = P_{gen}(y \rightarrow x) \exp(-E(y)/kT)$$

de manière à avoir une probabilité d'acceptation égale à 1. Ceci n'est possible que dans certains cas particulier : nous allons en étudier un exemple, l'algorithme de Wolff pour le modèle d'Ising. De manière plus générale on cherchera à augmenter la probabilité d'acceptation sans toujours pouvoir atteindre la limite idéale : 1.

## 7.2 Exemple : algorithme de Wolff pour le modèle d'Ising

On envisage maintenant, dans le cadre de la simulation du modèle d'Ising, le retournement global d'un petit ensemble de spins voisins de même orientation

### Définition de l'amas

1. On choisit un des  $N$  sites toujours suivant une loi uniforme comme premier site de l'amas.
2. On regarde successivement ses 4 voisins. Tout voisin qui a la même orientation de spin, est ajouté à l'amas avec une probabilité  $P$  ( $P$  est un nombre entre 0 et 1 fixé une fois pour toutes).
3. On regarde les voisins des nouveaux spins ajoutés à l'amas à l'étape précédente et qui n'appartiennent pas à l'amas. Tous ceux qui ont la même orientation que l'amas sont ajoutés à l'amas avec la probabilité  $P$ .
4. On réitère l'étape précédente.
5. On s'arrête lorsqu'à une certaine étape on n'a rien rajouté à l'amas.

### Retournement de l'amas

La configuration  $y$  est celle obtenue à partir de  $x$  en retournant en bloc tous les spins de l'amas (cf. exemple de la Fig. 7.1).

### Probabilité d'acceptation de la configuration retournée

1. On considère uniquement les spins de l'amas situés à la surface de l'amas. On compte le nombre  $\nu$  de leurs voisins n'appartenant pas à l'amas et qui pointent dans le même sens que ceux de l'amas. Le nombre  $\nu_x$  correspondant à la configuration  $x$  est en général différent de celui  $\nu_y$  correspondant à l'amas retourné de la configuration  $y$ . A titre d'exemple compter les nombres  $\nu_x$  et  $\nu_y$  correspondant à la Fig. 7.1.

Montrer que :

$$p_{choix}(x \rightarrow y) = A(1 - P)^{\nu_x}$$

où le prefacteur  $A$  ne dépend que de la topologie de l'amas

Montrer que :

$$p_{choix}(y \rightarrow x) = A(1 - P)^{\nu_y}$$

avec le même prefacteur  $A$ .

2. Montrer que :

$$\frac{p_{acc}(x \rightarrow y)}{p_{acc}(y \rightarrow x)}$$

s'exprime simplement en fonction de  $P$ ,  $\exp \left[ \frac{2J}{kT} \right]$  et de  $(\nu_x - \nu_y)$ .



3. Montrer qu'un choix judicieux de  $P$  conduit à

$$p_{acc}(x \rightarrow y) = p_{acc}(y \rightarrow x) = 1$$

c'est à dire à l'acceptation certaine du retournement à chaque étape.

4. Pourquoi cet algorithme est-il nettement supérieur au précédent ?

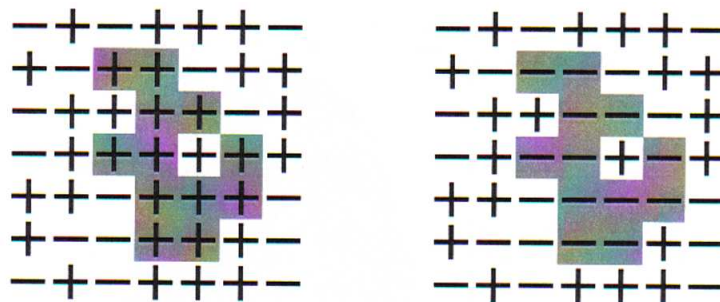


FIG. 7.1 – Choix d'un amas de la configuration de départ  $x$  (à gauche) et retournement des spins de l'amas pour une nouvelle configuration d'essai  $y$  (à droite).





# Applications à la physique médicale

---

## 8.1 Introduction

Ce chapitre est consacré à la simulation de processus stochastiques par essence : le transport de radiations (photons) ou de particules massives (neutrons, électrons ...) dans la matière. Il illustre l'une des applications les plus importantes de la méthode de Monte-Carlo. N'oublions pas que cette méthode a vu son essor, à partir des dernières années de la seconde guerre mondiale, à travers la simulation sur ordinateur du transport de neutrons dans la matière.

Une particule, massive (neutron, électron...) ou non massive (photon X, photon  $\gamma$ ) se déplaçant dans la matière a une certaine probabilité de rencontrer un atome ou une molécule avec lequel (ou laquelle) elle va interagir. Après interaction, sa trajectoire, son énergie vont être modifiées suivant certaines lois de probabilité. Entre deux interactions, la trajectoire de la particule est rectiligne.

Une particule, dans la matière suit un chemin aléatoire constitué de segments de droite. Les étapes (segments) successives de ce chemin constituent une chaîne de Markov, car la probabilité pour une particule, –ayant subi une interaction au point de coordonnée  $\mathbf{x}$  au temps  $t$ –, de subir une nouvelle interaction au point de coordonnée  $\mathbf{y}$  au temps  $t+1$ , après un parcours rectiligne ne dépend que de la position  $\mathbf{x}$  au temps  $t$  (indépendamment de tout ce qui s'est passé aux temps antérieurs à  $t$ ).

Nous nous intéresserons essentiellement ici au transport de photons (rayons X, rayons  $\gamma$ )

## 8.2 Modélisation d'un rayonnement

Un rayonnement X ou  $\gamma$  peut être décrit par un ensemble de photons. La simulation peut être effectuée en suivant l'évolution (trajectoire, énergie) d'un photon. On pourra aussi simuler en parallèle l'évolution d'un ensemble de photons, et dans ce cas l'usage d'ordinateurs massivement parallèles s'avérera particulièrement efficace (bien qu'il soit aussi possible de le faire sur un ordinateur monoprocesseur !)

A un instant donné, un photon individuel est caractérisé par sa position  $\mathbf{r}$  dans l'espace (trois variables), sa direction  $\hat{\mathbf{d}}$  (vecteur unitaire défini par deux variables indépendantes) et son énergie  $E$ .

Entre deux collisions, le photon conserve sa direction et son énergie. Le parcours du photon sera donc caractérisé par un ensemble discret d'états  $\{\mathbf{r}_n, \hat{\mathbf{d}}_n, E_n\}$  représentant sa position, sa direction son énergie juste après une interaction ("collision") avec une particule du milieu dans lequel il évolue.

### 8.3 Modélisation de la source

La première étape de la simulation est bien sûr de définir l'état initial  $\{\mathbf{r}_0, \hat{\mathbf{d}}_0, E_0\}$  du photon. Ces variables initiales seront sélectionnés suivant une densité de probabilité représentant la source de rayonnement.

$$S(\mathbf{r}_0, \hat{\mathbf{d}}_0, E_0)$$

### 8.4 Modélisation d'une trajectoire

Supposons que la trajectoire d'un photon ait été simulée jusqu'à l'état  $\{\mathbf{r}_n, \hat{\mathbf{d}}_n, E_n\}$  juste après la  $n^{ieme}$  collision. Le photon va continuer son chemin en ligne droite le long de la direction donnée par le vecteur unitaire  $\hat{\mathbf{d}}_n$  en parcourant une distance  $s$  puis subir une  $(n+1)^{ieme}$  collision au point :

$$\mathbf{r}_{n+1} = \mathbf{r}_n + s\hat{\mathbf{d}}_n$$

Comment choisir cette distance aléatoire  $s$  ? Nous allons montrer que  $s$  est une variable aléatoire de densité de probabilité :

$$p(s) = \frac{1}{\lambda} \exp(-s/\lambda)$$

Nous considérons un milieu homogène constitué de molécules de section efficace  $\sigma$ . Le nombre de molécules par unité de volume est  $\mathcal{N}$ .

La probabilité d'interaction, c'est à dire de rencontre d'une molécule lorsque le photon parcourt un petit chemin  $ds$  est indépendante de la position  $\mathbf{r}$  et de l'orientation  $\hat{\mathbf{d}}$  de sa trajectoire. Elle est simplement proportionnelle à la longueur du chemin élémentaire parcouru  $ds$  soit :

$$\frac{ds}{\lambda}$$

où  $\lambda$  est une constante —que nous préciserons par la suite— qui ne dépend que du milieu homogène considéré.

Soit  $p(s)$  la densité de probabilité correspondant à la loi de probabilité de la longueur  $s$  du chemin parcouru entre deux collisions.

La probabilité de parcourir un chemin de longueur  $s$  sans interagir est

$$\mathcal{F}(s) = 1 - \int_0^s p(u)du = \int_s^\infty p(u)du \quad (8.1)$$

La probabilité de parcourir un chemin de longueur  $s$  sans interagir, puis d'interagir entre  $s$  et  $s + ds$  est le produit des deux probabilités précédentes :

$$p(s)ds = \mathcal{F}(s)\frac{ds}{\lambda}$$

soit

$$p(s)ds = \frac{ds}{\lambda} \int_s^\infty p(u)du$$

$p(s)$  obéit donc à l'équation intégrale :

$$p(s) = \frac{1}{\lambda} \int_s^\infty p(u)du$$

En dérivant chaque membre de cette équation par rapport à  $s$ , on obtient :

$$\frac{dp(s)}{ds} = -\frac{1}{\lambda}p(s)$$

$p(s)$  est donc exponentielle :  $p(s) = A \exp(-s/\lambda)$ . La constante  $A$  est déterminée par la normalisation :

$$\int_0^\infty p(s)ds = 1$$

Ce qui impose :

$$p(s) = \frac{1}{\lambda} \exp(-s/\lambda)$$

Il est immédiat de vérifier que l'on a :

$$\int_0^\infty s p(s)ds = \lambda$$

$\lambda$  correspond à l'espérance ou valeur moyenne de la variable aléatoire associée au chemin parcouru entre deux collisions.  $\lambda$  correspond donc au libre parcours moyen du photon. En fonction de la densité  $\mathcal{N}$  de molécules et de leur section efficace, il s'exprime par la relation :

$$\lambda = 1/(\mathcal{N}\sigma)$$

La question est maintenant d'échantillonner des valeurs de  $s$  suivant la densité de probabilité exponentielle  $p(s)$ , ce qui est facile par la méthode de transformation de variables. La fonction de partition correspondant à  $p(s)$  est :

$$\eta = F(s) = \frac{1}{\lambda} \int_0^s \exp(-u/\lambda)du = 1 - \exp(-s/\lambda)$$

On inverse cette relation :

$$s = -\lambda \ln(1 - \eta)$$

Donc concrètement on tire des nombres  $\eta_i$  suivant la loi uniforme sur le segment  $[0,1]$ , et les nombres

$$s_i = -\lambda \ln(1 - \eta_i)$$

sont alors répartis suivant la densité de probabilité exponentielle souhaitée. Remarque : il est équivalent de choisir :

$$s_i = -\lambda \ln(\eta_i)$$

puisque si  $\eta$  est réparti uniformément sur le segment  $[0,1]$ , alors  $1 - \eta$  l'est aussi.

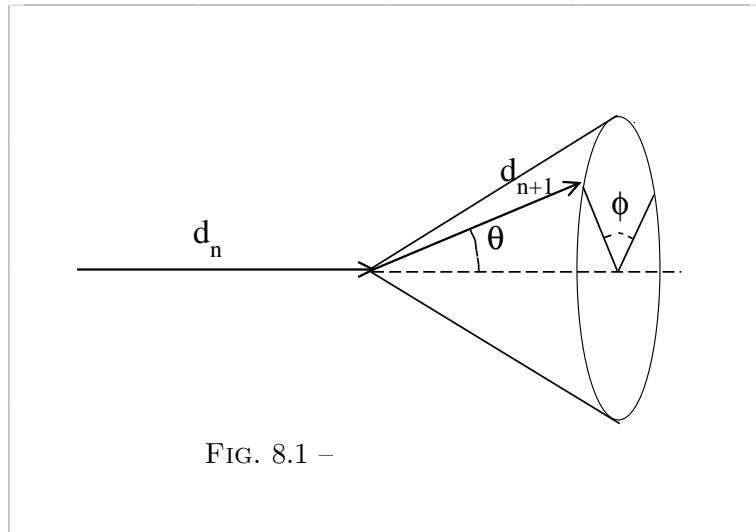
## 8.5 Interaction photon-matière

Il nous reste à échantillonner le nouveau vecteur directeur  $\hat{\mathbf{d}}_{n+1}$  après l'interaction suivante et la nouvelle énergie. Ceci va dépendre du type d'interaction.

### 8.5.1 Diffusion élastique de Rayleigh

Pour une diffusion élastique, il y a changement de la direction de propagation, mais conservation de l'énergie. La nouvelle direction de propagation  $\hat{\mathbf{d}}_{n+1}$  est définie par rapport à la précédente  $\hat{\mathbf{d}}_n$  par les deux angles  $\phi$  et  $\theta$  indiqués sur la figure précédente. L'angle azimuthal  $\phi$  obéit à une loi de probabilité de densité uniforme sur le segment  $[0, 2\pi]$ . Par contre, la loi de probabilité de l'angle  $\theta$  n'est pas uniforme.

Pour un photon d'énergie  $E$  subissant une diffusion élastique sur un atome de numéro atomique  $Z$ , la probabilité pour le photon d'être diffusé dans un petit angle solide  $d\Omega$



autour de la direction donnée par  $\{\theta, \phi\}$  est, dans le cadre de l'approximation de Born :

$$d\sigma_{Rayleigh} = r_e^2 \frac{1 + \cos^2 \theta}{2} F^2(q, Z) d\Omega$$

$q = |\vec{q}|$  est le module du transfert de moment :

$$\vec{q} = \vec{k}_{n+1} - \vec{k}_n$$

où  $\vec{k}_n$  et  $\vec{k}_{n+1}$  sont respectivement les vecteurs d'onde incident et diffusés de directions  $\hat{\mathbf{d}}_n$  et  $\hat{\mathbf{d}}_{n+1}$  et de même module :

$$|\vec{k}_n| = |\vec{k}_{n+1}| = E/c$$

( $c$  est la vitesse de la lumière). On a donc, en fonction de  $E$  et  $\theta$  :

$$q = 2(E/c) \sin(\theta/2)$$

$r_e$  est le rayon de Bohr de l'électron et  $F(q, Z)$  est le facteur de forme de l'atome, c'est à dire la transformée de fourier de sa densité électronique  $\rho(\vec{r})$ . Pour un atome de symétrie sphérique, on a :

$$F(q, Z) = 4\pi \int_0^\infty \rho(r) \frac{\sin(qr/\hbar)}{qr/\hbar} r^2 dr$$

Pour échantillonner  $\theta$  suivant la densité de probabilité correspondant à la “section efficace de diffusion”  $d\sigma_{Rayleigh}/d\Omega$  exprimée ci dessus, on aura en général recours à la méthode de réjection de Von Neumann.

### 8.5.2 Diffusion inélastique de Compton

Dans une diffusion de Compton, un photon d'énergie  $E_n$  interagit avec un électron d'un atome. Le photon transfère une partie de son énergie à l'électron qui est alors éjecté de l'atome. Le photon est alors diffusé suivant les angles  $\{\theta, \phi\}$  par rapport à la direction incidente  $\hat{\mathbf{d}}_n$  avec une énergie  $E_{n+1} < E_n$ . L'angle azimutal  $\phi$  est, ici aussi, distribué uniformément sur le segment  $[0, 2\pi]$ .

Nous nous restreindrons ici au cas où l'électron éjecté est, au départ, au repos. La conservation de l'énergie et de l'impulsion du système total {photon+électron} impose une relation entre l'angle  $\theta$  et les énergies :

$$\frac{m_e c^2}{E_{n+1}} - \frac{m_e c^2}{E_n} = 1 - \cos \theta$$

Les quantités  $m_e c^2/E = \lambda_{Compton}$  représentent les “longueurs d'onde de Compton” ( $m_e$  est la masse de l'électron au repos).

La section efficace de diffusion, c'est à dire la densité de probabilité correspondant à l'angle  $\theta$  est donnée par la relation de Klein-Nishina :

$$\frac{d\sigma_{Compton}}{d\Omega} = \frac{r_e^2}{2} \left( \frac{E_{n+1}}{E_n} \right)^2 \left( \frac{E_{n+1}}{E_n} + \frac{E_n}{E_{n+1}} - \sin^2 \theta \right)$$

### 8.5.3 Absorption photoélectrique

L'absorption photoélectrique est un processus où le photon est absorbé par un électron qui effectue une transition vers un état d'énergie plus élevée. La trajectoire du photon prend donc fin à cet endroit.

### 8.5.4 Choix d'un des événements

En fonction du milieu considéré, de l'énergie des photons etc... les probabilités de chacun des trois événements décrits ci-dessus :  $P_{Rayleigh}$ ,  $P_{Compton}$ ,  $P_{absorption}$  sont connues ( $P_{Rayleigh} + P_{Compton} + P_{absorption} = 1$ ). A chaque étape  $n$  il faudra tirer un de ces trois événements suivant ces probabilités.

### 8.5.5 Pour en savoir plus...

Il existe des logiciels programmés pour simuler toutes les interactions possibles de particules non massives (photons) ou massives (neutrons, électrons, positrons), chargées ou non chargées dans la matière. L'un d'entre eux, développé par un groupe de scientifique est nommé "PENELOPE". Le code source est payant mais les méthodes utilisées sont en libre accès sur internet.

<http://www.nea.fr/html/dbprog/penelope-2003.pdf>

## 8.6 Propagation dans un milieu inhomogène

En physique médicale, il est rare d'avoir un milieu homogène. Dans la plupart des cas, un photon traverse une succession de milieux de densités et sections efficaces différentes. Exemple : pour la simple radiographie d'une racine dentaire, les rayons X vont traverser la gencive, puis le tissu osseux, puis la pulpe etc...

Comme indiqué sur la figure 8.2, considérons un photon, qui après avoir subi une interaction en  $\mathbf{r}_t$  traverse successivement différents tissus  $T_1, T_2, \dots, T_{n-1}$  avant de subir l'interaction suivante en  $\mathbf{r}_{t+1}$  dans le tissu  $T_n$ . Nous noterons respectivement  $\lambda_1, \lambda_2, \dots, \lambda_n$  les libres parcours moyens du photon dans les tissus  $T_1, T_2, \dots, T_n$ . Comme l'indique la figure, le photon a parcouru une distance  $s_1$  dans le tissu  $T_1$ , une distance  $s_2$  dans le tissu  $T_2$  ... une distance  $s_{n-1}$  dans le tissu  $T_{n-1}$  et enfin une distance  $(s - s_{n-1} - s_{n-2} - \dots - s_2 - s_1)$  dans le dernier tissu  $T_n$ .

D'après ce qui précède, la probabilité de parcourir une distance  $s_i$  dans le tissu  $T_i$  sans interagir est (cf. Equation 8.1)

$$\mathcal{F}(s_i) = \frac{1}{\lambda_i} \int_{s_i}^{\infty} \exp\left(-\frac{u}{\lambda_i}\right) du = \exp\left(-\frac{s_i}{\lambda_i}\right)$$



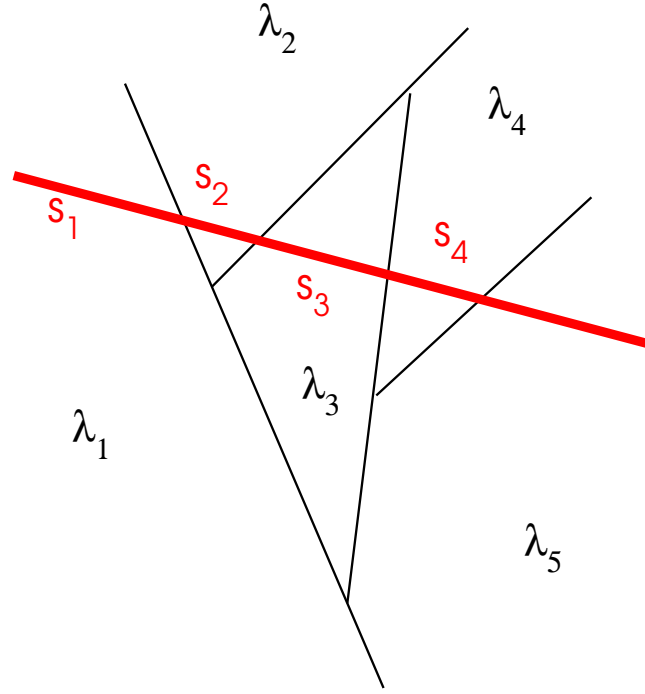


FIG. 8.2 –

Et la probabilité d'avoir interagi entre  $s$  et  $s + ds$  dans le milieu  $T_n$  est :

$$\frac{ds}{\lambda_n}$$

La probabilité d'avoir parcouru le chemin de longueur  $s$  à travers les divers tissus sans interagir, puis d'avoir interagi entre  $s$  et  $s + ds$  est donc :

$$p(s)ds = \exp\left(-\frac{s_1}{\lambda_1}\right) \exp\left(-\frac{s_2}{\lambda_2}\right) \cdots \exp\left(-\frac{s_{n-1}}{\lambda_{n-1}}\right) \exp\left(-\frac{[s - s_1 - s_2 - \cdots - s_{n-1}]}{\lambda_n}\right) \frac{ds}{\lambda_n}$$

En définissant, par simple translation la variable :

$$\tilde{s} = s_1 \frac{\lambda_n}{\lambda_1} + s_2 \frac{\lambda_n}{\lambda_2} + \cdots + s_{n-1} \frac{\lambda_n}{\lambda_{n-1}} + s - s_1 - s_2 - \cdots - s_{n-1}$$

on a

$$p(\tilde{s}) = \frac{1}{\lambda_n} \exp\left(-\frac{\tilde{s}}{\lambda_n}\right)$$

On peut donc générer  $\tilde{s}$ , comme précédemment suivant la loi exponentielle :

– On tire des nombres  $\eta$  suivant une loi uniforme sur le segment  $[0,1]$

- les nombres

$$\tilde{s} = -\lambda_n \ln \eta$$

sont générés suivant la loi exponentielle ci dessus

- on en déduit :

$$s = s_1 + s_2 + \cdots + s_{n-1} - \lambda_n \left[ \ln \eta + \frac{s_1}{\lambda_1} + \frac{s_2}{\lambda_2} + \cdots + \frac{s_{n-1}}{\lambda_{n-1}} \right] \quad (8.2)$$

On procédera donc, en général de la manière suivante. La direction du photon étant donnée, on connaît donc ses intersections avec les différentes frontières du milieu et les longueurs  $s_1, s_2, \dots$

- On tire un nombre  $\eta$  uniformément distribué sur le segment  $[0,1]$
- si  $-\ln \eta < \frac{s_1}{\lambda_1}$  alors l'interaction suivante a lieu dans le milieu  $T_1$ , il n'y a rien de nouveau par rapport aux paragraphes précédents.
- si  $\frac{s_1}{\lambda_1} < -\ln \eta < \frac{s_1}{\lambda_1} + \frac{s_2}{\lambda_2}$  alors l'interaction suivante a lieu dans le milieu  $T_2$  et on prend :

$$s = s_1 - \lambda_2 \left[ \ln \eta + \frac{s_1}{\lambda_1} \right]$$

- etc ...

- si

$$\frac{s_1}{\lambda_1} + \frac{s_2}{\lambda_2} + \cdots + \frac{s_{n-1}}{\lambda_{n-1}} < -\ln \eta < \frac{s_1}{\lambda_1} + \frac{s_2}{\lambda_2} + \cdots + \frac{s_{n-1}}{\lambda_{n-1}} + \frac{s_n}{\lambda_n}$$

alors l'interaction a lieu dans le milieu  $T_n$  et  $s$  est donné par la relation 8.2



# Méthode de Monte Carlo et Dynamique Moléculaire

## 9.1 Introduction

Bien que la méthode de “Dynamique Moléculaire” soit parfaitement déterministe, donc en dehors du cadre de ce cours, nous en donnons ici les principes, car dans les simulations numériques elle est souvent employée pour les mêmes problèmes, conjointement à la méthode de Monte-Carlo. Dans ce chapitre, nous soulignerons les avantages et les inconvénients de chacune de ces méthodes.

La “dynamique moléculaire” simule *ab initio* l’évolution d’un système physique –en particulier son évolution vers l’équilibre–, à partir des équations qui régissent ce système. Pour un système classique, ces équations sont simplement celles de la mécanique newtonienne.

En préambule, et pour rendre la comparaison entre les deux méthodes plus évidente, nous allons montrer qu’une variante de la Méthode de Monte-Carlo, appliquée à la simulation de la situation d’équilibre d’un système classique, peut se construire autour de la simulation d’une équation différentielle stochastique.

## 9.2 Equation stochastique

Nous allons montrer qu’une chaîne de Markov, satisfaisant le principe du bilan détaillé peut être obtenue à partir de l’intégration d’une *équation différentielle stochastique* (i.e. faisant intervenir des variables aléatoire).

Nous reprenons notre système  $\mathcal{S}$  caractérisé par les coordonnées :  $x = (\alpha_1, \alpha_2, \dots, \alpha_{3N})$  de ses  $N$  particules. Son énergie est  $E(x)$ , et nous introduisons “l’action”  $S(x) = \beta E(x)$ . Nous considérons l’équation de Langevin :

$$\frac{\partial \alpha_i}{\partial \tau} = -\Gamma_i \frac{\partial S(x)}{\partial \alpha_i} + \xi_i(\tau) \quad (9.1)$$

où les  $\alpha_i$  sont des fonctions d’une variable continue  $\tau$ .  $\Gamma_i$  est un paramètre d’échelle arbitraire et les  $\xi_i(\tau)$  représentent un ensemble de variables aléatoires indépendantes gaussiennes.

ennes centrées de même variance  $2\Gamma_i$

$$E[\xi_i(\tau)\xi_j(\tau')] = 2\Gamma_i\delta_{i,j}\delta_{\tau-\tau'} \quad (9.2)$$

Cette équation représente “un bruit blanc”. Remplaçons l’équation de Langevin continue par une équation aux différences finies :

$$\alpha_i(\tau_{n+1}) = \alpha_i(\tau_n) + \Delta\tau \left[ -\Gamma_i \frac{\partial S(x)}{\partial \alpha_i(\tau_n)} + \xi_i(\tau_n) \right] \quad (9.3)$$

de sorte que  $\alpha_i(\tau_n)$  correspond au  $n^{ieme}$  élément d’une chaîne de Markov.

La distribution de probabilité des variables aléatoires  $\xi_i$  correspondant à l’équation 9.2 est :

$$\prod_i \sqrt{\frac{\Delta\tau}{4\pi\Gamma_i}} e^{-\sum_i \int_{\tau_n}^{\tau_{n+1}} \frac{dt}{4\Gamma_i} \xi_i(t)^2}$$

soit

$$\prod_i \sqrt{\frac{\Delta\tau}{4\pi\Gamma_i}} e^{-\sum_i \frac{\Delta\tau}{4\Gamma_i} \xi_i(\tau_n)^2} \quad (9.4)$$

La règle  $p(x \rightarrow y)$  pour cette chaîne de Markov est que la probabilité de passer de  $x = (\alpha_1, \alpha_2, \dots, \alpha_{3N})$  à  $y = (\beta_1, \beta_2, \dots, \beta_{3N})$  est égale à la probabilité pour la variable aléatoire (bruit blanc)  $\xi_i$  d’être égale à

$$\frac{\beta_i - \alpha_i}{\Delta\tau} + \Gamma_i \frac{\partial S(x)}{\partial \alpha_i}$$

Il nous reste à vérifier que cette règle obéit à l’équation du bilan détaillé. On a

$$\frac{p(x \rightarrow y)}{p(y \rightarrow x)} = \frac{e^{-\frac{\Delta\tau}{4\Gamma_i} \sum_i \left( \frac{\beta_i - \alpha_i}{\Delta\tau} + \Gamma_i \frac{\partial S}{\partial \alpha_i} \right)^2}}{e^{-\frac{\Delta\tau}{4\Gamma_i} \sum_i \left( \frac{\alpha_i - \beta_i}{\Delta\tau} + \Gamma_i \frac{\partial S}{\partial \beta_i} \right)^2}}$$

soit lorsque  $\Delta\tau \rightarrow 0$

$$\frac{p(x \rightarrow y)}{p(y \rightarrow x)} = e^{-\sum_i (\beta_i - \alpha_i) \frac{\partial S}{\partial \alpha_i}}$$

ce qui peut encore s’écrire en identifiant  $S$  et son développement au premier ordre :

$$\frac{p(x \rightarrow y)}{p(y \rightarrow x)} \rightarrow \frac{e^{-S(y)}}{e^{-S(x)}} \quad \text{lorsque } \Delta\tau \rightarrow 0$$

CQFD

### 9.3 Comparaison avec la Dynamique Moléculaire

Nous avons jusqu'ici toujours considéré l'ensemble canonique : notre système  $\mathcal{S}$  est en contact avec un réservoir ou thermostat  $\mathcal{R}$ . La température  $T$  du système est fixée par le thermostat, mais son énergie fluctue autour d'une valeur moyenne  $\langle E \rangle$ . La valeur moyenne d'une observable  $\mathcal{O}(x)$  est :

$$\langle \mathcal{O} \rangle = \frac{\int \mathcal{O}(x) e^{-\beta H(p,x)} dp dx}{Z} \quad (9.5)$$

où  $Z$  représente la fonction de partition :

$$Z = \int e^{-\beta H(p,x)} dp dx \quad (9.6)$$

$x = (\alpha_1, \alpha_2, \dots, \alpha_{3N})$  représente les coordonnées des particules et  $p = (p_1, p_2, \dots, p_{3N})$  les variables conjuguées ou impulsions :  $p_i/m_i = d\alpha_i/dt$ .

$$H = \sum_i \frac{p_i^2}{2m_i} + E(x) \quad (9.7)$$

représente l'Hamiltonien du système. Le premier terme représente l'énergie cinétique et le second terme  $E(x)$  l'énergie potentielle.

Examinons la formulation de la méthode de dynamique moléculaire dans l'ensemble microcanonique.

On considère un système isolé, ayant donc une énergie  $\mathcal{E}$  fixée. Les trajectoires des particules sont déterminées par les équations de Hamilton :

$$\begin{aligned} d\alpha_i/dt &= p_i/m_i \\ dp_i/dt &= -\partial E(x)/\alpha_i \end{aligned} \quad (9.8)$$

Ces équations de mouvement sont intégrées numériquement. En suivant l'évolution d'une observable  $\mathcal{O}(x(t))$  pendant un temps très long, on peut calculer sa valeur moyenne dans le temps :

$$\langle \mathcal{O} \rangle_t = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathcal{O}(x(t)) dt \quad (9.9)$$

Si le système est *ergodique*, alors on montre que, dans la limite thermodynamique, la moyenne d'ensemble  $\langle \mathcal{O} \rangle$  [relation 9.5] calculée dans l'ensemble canonique est égale à la moyenne dans le temps  $\langle \mathcal{O} \rangle_t$  calculée dans l'ensemble microcanonique, pour une valeur fixée  $\mathcal{E}$  de l'énergie égale à la valeur moyenne  $\langle E \rangle$  de l'énergie dans l'ensemble canonique.

Dans la limite thermodynamique, on passe de l'ensemble canonique à l'ensemble micro-canonique en remarquant que, dans l'ensemble canonique, l'énergie fluctue de moins en moins autour de sa valeur moyenne  $\langle E \rangle$ . On peut alors appliquer *l'approximation du col* à l'intégrale représentée par la relation 9.5. Dans ce cas, la contribution essentielle à l'intégrale provient de l'hypersurface d'énergie constante définie par :

$$H(p, x) = \langle E \rangle$$

En général, les méthodes de dynamique moléculaire procèdent à l'intégration des équation du mouvement par discrétisation du temps. On suit l'évolution du système à des temps  $\tau_k = k\Delta\tau$ , espacés régulièrement de  $\Delta\tau$ . Les équations du mouvement :

$$m_i \frac{d^2 \alpha_i}{dt^2} = - \frac{\partial E(x)}{\partial \alpha_i} \quad (9.10)$$

conduisent aux équations d'évolution en temps discret :

$$\alpha_i(\tau_{n+1}) = 2\alpha_i(\tau_n) - \alpha_i(\tau_{n-1}) - \frac{1}{m_i}(\Delta\tau)^2 \frac{\partial E(x)}{\partial \alpha_i} \quad (9.11)$$

Ces équations constituent le fondement de l'algorithme de Verlet[10] le plus couramment utilisé en dynamique moléculaire[11]. On peut remanier les termes de cette équation pour les écrire de manière plus suggestive :

$$\alpha_i(\tau_{n+1}) = \alpha_i(\tau_n) - \frac{1}{2m_i}(\Delta\tau)^2 \frac{\partial E}{\partial \alpha_i} + \frac{1}{2}[\alpha_i(\tau_{n+1}) - \alpha_i(\tau_{n-1})] \quad (9.12)$$

Il est intéressant de comparer cette équation à celle obtenue après discrétisation de l'équation de Langevin pour un échantillonnage suivant la méthode de Monte-Carlo. Dans l'équation 9.3, le paramètre  $\Gamma_i$  est arbitraire. Prenons

$$\Gamma_i = \frac{\Delta\tau}{2\beta m_i}$$

alors, on obtient :

$$\alpha_i(\tau_{n+1}) = \alpha_i(\tau_n) - \frac{1}{2m_i}(\Delta\tau)^2 \frac{\partial E}{\partial \alpha_i} + \Delta\tau \xi_i(\tau_n) \quad (9.13)$$

En comparant les équations 9.12 et 9.13, on constate que les vitesses des particules :

$$v_i = \frac{1}{2\Delta\tau}[\alpha_i(\tau_{n+1}) - \alpha_i(\tau_{n-1})]$$

qui correspondent au dernier terme de l'équation d'évolution 9.13 en dynamique moléculaire sont remplacées, dans la méthode de Monte-Carlo par des variables aléatoires gaussiennes indépendantes  $\xi_i$  (bruit stochastique)[8, 12].

La méthode de dynamique moléculaire a l'avantage de conduire directement dans les régions de l'espace des phases  $\{p, x\}$  qui contribuent le plus à l'intégrale donnant la valeur moyenne d'une observable, mais dans certains cas non triviaux où un temps très long est nécessaire pour parcourir l'espace des phases, on peut "manquer" certaines régions importantes de cet espace. La méthode de Monte-Carlo explore tout l'espace des phases, mais on peut perdre du temps à stagner dans une région de l'espace des phases qui contribue peu à l'intégrale.

La comparaison ci-dessus des deux méthodes ouvre la possibilité d'algorithmes hybrides[12, 13] où l'évolution du système aux temps discrets successifs correspond à une chaîne de Markov déterminée par l'équation :

$$\alpha_i(\tau_{n+1}) = \alpha_i(\tau_n) - \frac{1}{2m_i}(\Delta\tau)^2 \frac{\partial E}{\partial \alpha_i} + \Delta\tau \zeta_i(\tau_n) \quad (9.14)$$

$\zeta_i$  est soit la vitesse (deterministe!) des particules  $v_i$ , soit un bruit stochastique  $\xi_i$ . La manière d'alterner dans la chaîne de Markov entre ces deux possibilités est propre à chaque problème. Cette méthode est à la fois plus rapide que la méthode de Monte-Carlo conventionnelle et plus sûre que la dynamique moléculaire pure.







# Optimisation

---

## 10.1 Méthode du “Recuit Simulé” (Simulated Annealing)

Lorsqu'on souhaite minimiser une fonction à grand nombre de variables, les méthodes classiques n'ont aucune difficulté pour trouver des minima locaux (en général le plus proche d'un point défini comme point de départ de la méthode). Il est par contre pratiquement impossible de trouver le minimum absolu. Seule, la méthode de Monte-Carlo permet de résoudre ce problème.

La fonction  $E(x_1, x_2, \dots, x_N)$  à minimiser est considérée comme l'énergie d'un système fictif ayant pour espace de configuration  $R^N$ . Ce système est supposé en interaction avec un thermostat à température  $T$ . On simule l'évolution du système dans l'ensemble Canonique par la Méthode de Monte-Carlo.

Comme en métallurgie, le “recuit” consiste à “chauffer le système” à une température élevée puis à diminuer très doucement la température de manière à obtenir l'équilibre thermodynamique à chaque température. Lorsque la température s'approche de zéro, le système atteint son énergie minimale, c'est à dire le minimum absolu de la fonction  $E$ .

## 10.2 “Recuit Parallèle” (Parallel Tempering)

Lorsque la fonction  $E(x_1, x_2, \dots, x_N)$  est particulièrement compliquée avec un très grand nombre de minima relatifs, la méthode simple précédente ne permet pas forcément à basse température d'atteindre l'équilibre en un temps raisonnable, donc de converger vers le minimum absolu à température nulle. On peut facilement resté “piégé” à basse température dans une région de l'espace multidimensionnel, dont on ne s'échappe plus.

Dans ce cas, une méthode dite “Parallel Tempering” en anglais, —que nous traduirons par “Recuit parallèle”— a été proposée, il y a une vingtaine d'années par les physiciens des “verres de spin”. Elle s'applique aussi bien à l'obtention de l'équilibre d'un système thermodynamique comprenant un très grand nombre d'états d'énergies voisines —c'est le cas des solides désordonnés ou “verres”, ou d'assemblée de spins avec fort désordre dans les interactions entre spins (“verres de spins”)— qu'à un problème de minimisation de fonction

présentant les mêmes caractéristiques.

Il s'agit d'effectuer parallèlement plusieurs simulations du même système à des températures différentes, avec des échanges de configurations, selon des règles que nous allons définir.

Considérons le cas (schématisé à une dimension sur la figure 10.1 où deux régions de l'espace A et B sont séparées par une importante barrière d'énergie  $\Delta E$ . A haute température  $kT_H \gg \Delta E$ , il est évident que l'algorithme de Métropolis permet d'explorer les deux régions A et B. Par contre, à basse température  $kT_B \ll \Delta E$ , on va resté piégé dans une des deux régions A ou B.

L'idée est de simuler en parallèle le même système à deux températures différentes  $T_H$  et  $T_B$  et d'échanger "de temps en temps" les configurations des deux systèmes, avec une probabilité que nous allons définir. Cet "échange" va permettre au système à basse température de visiter les deux régions A et B.

La condition d'ergodicité (visiter tout l'espace) sera ainsi réalisée. Il nous reste à définir la probabilité d'échange des configurations de manière à satisfaire la "microréversibilité".

Soit  $E_H$  L'énergie de la configuration du système à haute température  $T_H$  et  $E_B$  celle du système à basse température  $T_B$ . Nous allons montrer que si on accepte l'échange des configurations avec la probabilité :

$$\exp[-(\beta_B - \beta_H)(E_H - E_B)] \quad \text{si} \quad (E_H - E_B) > 0 \quad (10.1)$$

1 sinon

alors la condition de microréversibilité est satisfaite (on note  $\beta_H = 1/kT_H$  et  $\beta_B = 1/kT_B$ )

Considérons la probabilité conjointe  $p_{\mu,\nu}$  pour que le système à basse température soit dans l'état  $\mu$  et que le système à haute température soit dans l'état  $\nu$ . A l'équilibre les deux systèmes doivent satisfaire la distribution canonique de Boltzmann :

$$p_{\mu,\nu} = \frac{1}{Z_B Z_H} \exp[-\beta_B E(\mu)] \exp[-\beta_H E(\nu)] \quad (10.2)$$

Toutes les M itérations, nous tentons un échange des configurations  $\mu$  et  $\nu$  avec la probabilité  $P(\{\mu, \nu\} \rightarrow \{\nu, \mu\})$  En appliquant la loi 10.1, on a :

$$\frac{P(\{\mu, \nu\} \rightarrow \{\nu, \mu\})}{P(\{\nu, \mu\} \rightarrow \{\mu, \nu\})} = \exp[-(\beta_B - \beta_H)(E(\nu) - E(\mu))]$$

Mais la loi de Boltzmann 10.2 conduit à :

$$\frac{p_{\mu,\nu}}{p_{\nu,\mu}} = \frac{\exp[-\beta_B E(\mu)] \exp[-\beta_H E(\nu)]}{\exp[-\beta_B E(\nu)] \exp[-\beta_H E(\mu)]} = \exp[-(\beta_B - \beta_H)(E(\mu) - E(\nu))]$$

On a donc :

$$\frac{P(\{\mu, \nu\} \rightarrow \{\nu, \mu\})}{P(\{\nu, \mu\} \rightarrow \{\mu, \nu\})} = \frac{p_{\nu,\mu}}{p_{\mu,\nu}}$$

et la condition de microréversibilité est bien vérifiée.

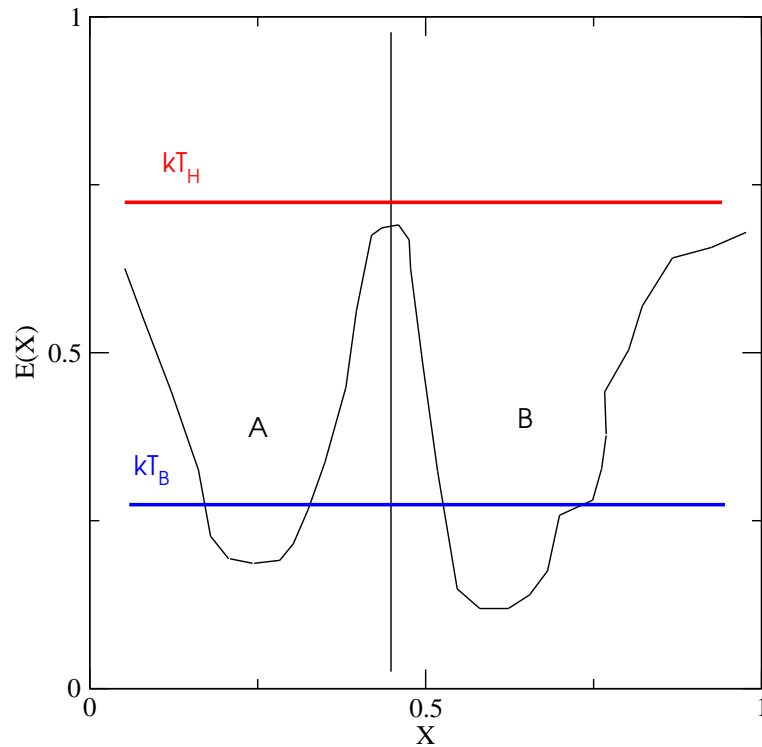


FIG. 10.1 –

En ce qui concerne les  $M-1$  autres itérations, nous tentons d’abord une transition  $\mu \rightarrow \mu'$  du système à basse température, suivant l’algorithme habituel de Métropolis, et il est facile de vérifier à partir de 10.2 que la condition de microréversibilité est vérifiée. Puis nous tentons une transition  $\nu \rightarrow \nu'$  du système à haute température, et il en est de même.

*Comment choisir  $M$  ?*

Entre deux tentatives d’échange, il convient que le système à haute température soit passé à un état décorrélé du précédent.  $M$  sera donc de l’ordre du temps d’autocorrélation de l’énergie.

*Généralisation à plus de deux systèmes*

Le schéma que nous venons d’établir pour deux systèmes se généralise de manière triviale à  $n$  systèmes placés à  $n$  températures différentes. Il se prête particulièrement bien à l’usage d’ordinateurs massivement parallèles, d’autant plus qu’il est absolument équivalent d’échanger les températures plutôt que les états, et ceci sera beaucoup moins coûteux en temps de communication entre processeurs. On peut découper un intervalle de températures en segments relativement petits, avec une probabilité importante d’échanger les états de deux systèmes de températures voisines.





# Polymères

---

## 11.1 Introduction

Dés les années 50, les pionniers de la méthode Monte-Carlo, en particulier les époux Rosenbluth, ont simulé des systèmes de polymères.

Nous allons étudier quelques algorithmes spécifiques pour la simulation de longues chaînes de molécules.

## 11.2 Modélisation de polymères

### 11.2.1 Modèle du “collier de perles”

C’est le modèle le plus simple. Chaque monomère est modélisé par une sphère dure de rayon  $a$ . Chaque sphère dure est accolée à ses deux voisines. Les mouvements de l’une par rapport à l’autre de ces sphères sont libres.

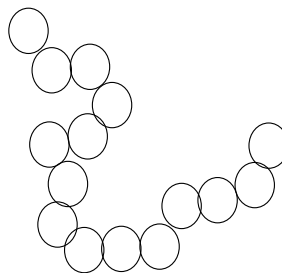


FIG. 11.1 – Modélisation suivant le “collier de perles”

### 11.2.2 Modèle du “Chapelet de ressorts” (String beads)

1. La cohésion de la chaînes est modélisée par un potentiel attractif de “ressort” entre deux molécules consécutives :

$$E_{el} = K|R_{i+1} - R_i|^2 \quad (11.1)$$

2. On prend en compte un potentiel entre paires de monomères, fortement répulsif à courte distance  $a$ , —de l'ordre de grandeur du diamètre de sphère dures du modèle plus simple du collier de perles—, et faiblement attractif à longue distance (pour tenir compte des forces de Van-der-Waals). On prendra par exemple le potentiel de “Lennard-Jones” :

$$V_{i,j} = \epsilon \left[ \frac{a}{|R_i - R_j|^{12}} - \frac{a}{|R_i - R_j|^6} \right] \quad (11.2)$$

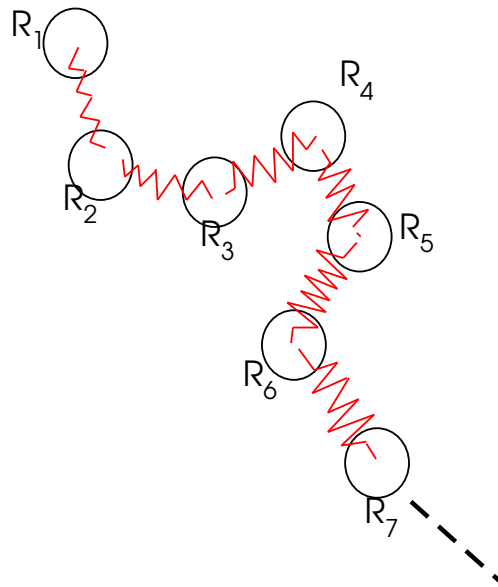


FIG. 11.2 – Modélisation suivant le “chapelet de ressorts”

## 11.3 Mouvements dans un algorithme de Metropolis

### 11.3.1 Reptation

On supprime le premier monomère d'une chaîne puis on le reconstruit au bout de la chaîne

### 11.3.2 Mouvements de pivot

1. On choisit, au hasard un monomère  $P$  de la chaîne
2. On choisit suivant une répartition uniforme un vecteur unitaire  $\hat{u}$  dans un angle solide centré autour du segment défini par le monomère  $P$  et le suivant ( $P + 1$ )
3. On choisit un angle  $\phi$  suivant une loi uniforme sur le segment  $[0, 2\pi]$
4. On effectue une rotation en bloc d'un angle  $\phi$  autour de  $\hat{u}$  de la partie du polymère comprise entre le monomère ( $P + 1$ ) et le dernier monomère.

## 11.4 “Reconstruction biaisée”

La méthode du “pivot” que nous avons décrite au paragraphe précédent est efficace pour un ensemble de polymères “dilués i.e. pas trop proches les uns des autres. Pour un système

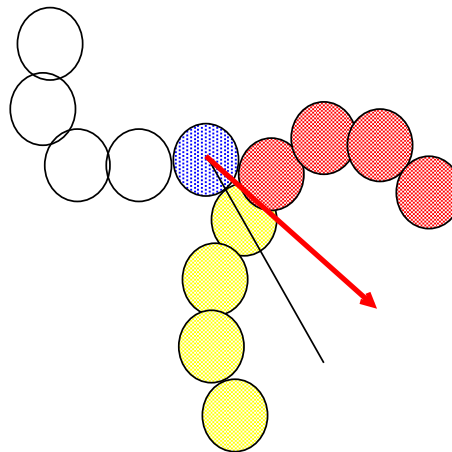


FIG. 11.3 – Mouvement de Pivot

dense, lorsqu'on tente de bouger un grand nombre de monomères par la méthode du pivot, par exemple, on a de grandes chances que dans la nouvelle configuration d'essai certains "coeurs durs" se recouvrent, auquel cas le mouvement est rejeté.

En 1954, les époux Rosenbluth[14] ont proposé une technique astucieuse consistant à supprimer une chaîne de polymères pour la reconstruire monomère par monomère à un autre endroit en utilisant un "biais" pour éviter le recouvrement de coeurs durs. Cette méthode a été perfectionnée[15] et est maintenant couramment utilisée sous le nom anglais de "configurational-bias Monte-Carlo algorithm"

1. Dans la configuration initiale  $x$ , on choisit une chaîne  $\Gamma^a$  comprenant  $N$  monomères  $\Gamma_i^a$
2. On supprime cette chaîne pour la reconstruire ailleurs, pas à pas de la manière suivante :
  - On choisit le premier monomère  $\Gamma_1^b$  de manière uniforme.
  - On suppose ici, pour simplifier que les segments joignant deux monomères ne peuvent prendre que  $k$  orientations discrètes, ce qui est le cas si on travaille sur un espace discrétisé, mais la méthode s'étend facilement au cas continu[15].
 On choisit pour le second monomère  $\Gamma_2^b$  de la chaîne une des  $k$  orientations possibles avec la probabilité :

$$P_2^b = \frac{\exp(-u_2^b/kT)}{Z_2^b}$$

avec

$$Z_2^b = \sum_{j=1}^k \exp(-u_{2,j}^b/kT)$$

$u_{2,j}^b$  représentant l'énergie d'interaction du monomère  $\Gamma_2^b$  avec toutes les autres particules du système.

- On construit le second monomère de la même manière, etc.. jusqu'au  $N^{ieme}$  monomère.
- A la fin du processus de construction, la conformation finale  $\Gamma^b$  du polymère a été générée avec une probabilité :

$$P_{gen}^b = \prod_{i=2}^N \frac{\exp(-u_i^b/kT)}{Z_i^b} = \frac{\exp[-(u_1^b + u_2^b + u_3^b + \dots + u_N^b)/kT]}{W^b}$$

où

$$W^b = \exp(-u_1^b/kT) \prod_{i=2}^N Z_i^b$$

Mais  $(u_1^b + u_2^b + u_3^b + \dots + u_N^b)$  n'est autre que la contribution totale  $U^b$  de la "nouvelle chaîne", dans sa nouvelle conformation  $\Gamma^b$  à l'énergie totale du système. Et on peut écrire :

$$P_{gen}^b = \frac{\exp(-U^b/kT)}{W^b}$$

$W^b$  est appelé le "facteur de Rosenbluth"



- Pour satisfaire la condition de microreversibilité, il faut calculer le facteur de Rosenbluth :

$$W^a = \exp(-u_1^a/kT) \prod_{i=2}^N Z_i^a$$

de la chaîne dans sa conformation initiale  $\Gamma^a$  et on a :

$$P_{gen}^a = \frac{\exp(-U^a/kT)}{W^a}$$

$U^a$  représentant la contribution de la chaîne dans son ancienne conformation  $\Gamma^a$  à l'énergie totale du système.

3. On accepte la nouvelle conformation  $\Gamma^b$  avec la probabilité :

$$\text{Min} \left[ 1, \exp[-(U^b - U^a)/kT] \frac{P_{gen}^a}{P_{gen}^b} \right] = \text{Min}[1, W^b/W^a]$$





# Monte-Carlo quantique

## 12.1 Monte-Carlo Variationnel

Pour calculer l'énergie fondamentale  $E_0$  (à température nulle) d'un système quantique décrit par un Hamiltonien  $H$ , on utilise couramment l'approximation variationnelle. Cette approximation consiste à :

- Partir d'une fonction d'onde  $|\psi_{\lambda,\mu,\dots,\nu}\rangle$  censée représenter une bonne approximation de l'état fondamental du système (on n'exigera pas forcément que la norme de cette fonction soit 1), et dépendant d'un certain nombre de paramètres ajustables  $\lambda, \mu, \dots, \nu$
- Calculer :

$$E = \frac{\langle \psi_{\lambda,\mu,\dots,\nu} | H | \psi_{\lambda,\mu,\dots,\nu} \rangle}{\langle \psi_{\lambda,\mu,\dots,\nu} | \psi_{\lambda,\mu,\dots,\nu} \rangle}$$

On montre (théorème de Rietz ou principe variationnel) l'inégalité  $E_0 < E$

- Le principe variationnel nous dit que la meilleure fonction d'essai est celle qui minimise l'énergie  $E$ . On ajustera donc les paramètres  $\lambda, \mu, \dots, \nu$  pour obtenir l'énergie minimale. Que l'on parte d'une représentation de  $|\psi_{\lambda,\mu,\dots,\nu}\rangle$  dans l'espace réel à  $3N$  dimensions,  $N$  représentant le nombre de particules, ou d'une représentation dans l'espace réciproque (vecteurs d'ondes  $K$ ), chacune des quantités  $\langle \psi_{\lambda,\mu,\dots,\nu}^* | H | \psi_{\lambda,\mu,\dots,\nu} \rangle$  et  $\langle \psi_{\lambda,\mu,\dots,\nu}^* | \psi_{\lambda,\mu,\dots,\nu} \rangle$  représente une intégrale multidimensionnelle. Dans l'espace réel :

$$E = \frac{\int \psi_{\lambda,\mu,\dots,\nu}^*(r_1, r_2, \dots, r_N) H \psi_{\lambda,\mu,\dots,\nu}(r_1, r_2, \dots, r_N) dr_1 dr_2 \dots dr_N}{\int \psi_{\lambda,\mu,\dots,\nu}^*(r_1, r_2, \dots, r_N) \psi_{\lambda,\mu,\dots,\nu}(r_1, r_2, \dots, r_N) dr_1 dr_2 \dots dr_N}$$

D'où l'idée de calculer directement ce rapport de deux intégrales par la méthode de Monte-Carlo.

Il est alors naturel de choisir pour densité de probabilité :

$$p(r_1, r_2, \dots, r_N) = \frac{\psi_{\lambda,\mu,\dots,\nu}^*(r_1, r_2, \dots, r_N) \psi_{\lambda,\mu,\dots,\nu}(r_1, r_2, \dots, r_N) dr_1 dr_2 \dots dr_N}{\int \psi_{\lambda,\mu,\dots,\nu}^*(r_1, r_2, \dots, r_N) \psi_{\lambda,\mu,\dots,\nu}(r_1, r_2, \dots, r_N) dr_1 dr_2 \dots dr_N}$$

On échantillonne cette densité de probabilité par la méthode de Metropolis, par exemple. Une estimation de  $E$  est donnée par l'espérance de :

$$\frac{\psi_{\lambda,\mu,\dots,\nu}^*(r_1, r_2, \dots, r_N) H \psi_{\lambda,\mu,\dots,\nu}(r_1, r_2, \dots, r_N)}{\psi_{\lambda,\mu,\dots,\nu}^*(r_1, r_2, \dots, r_N) \psi_{\lambda,\mu,\dots,\nu}(r_1, r_2, \dots, r_N)}$$

## 12.2 “Diffusion Monte-Carlo”

Nous allons décrire un algorithme qui permet de trouver, sans approximation, l'énergie et l'état fondamental de l'équation de Schrödinger :

$$H|\Psi\rangle = E|\Psi\rangle$$

Par souci de simplicité, nous considérons une particule de masse  $m$ , soumise à un potentiel  $V(x)$ , se déplaçant sur un axe  $x$ . La méthode est bien-sûr généralisable à un système à plusieurs dimensions.

L'Hamiltonien s'écrit donc simplement :

$$H = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x)$$

et l'évolution de la fonction d'onde  $\Psi(x, t)$  en fonction du temps est gouvernée par l'équation de Schrödinger :

$$i\hbar \frac{\partial |\psi\rangle}{\partial t} = H|\psi\rangle$$

Soit  $\{|\phi\rangle_n\}$  une base orthonormée de fonctions propres associées aux énergies  $E_n$ , c'est à dire :

$$H|\phi_n\rangle = E_n|\phi_n\rangle$$

avec la condition :

$$\langle \phi_n | \phi_m \rangle = \int_{-\infty}^{\infty} \phi_n^*(x) \phi_m(x) dx = \delta_{nm}$$

Alors on peut écrire en développant  $|\Psi(x, t)\rangle$  sur cette base :

$$|\Psi(x, 0)\rangle = \sum_{n=0}^{\infty} c_n \Phi_n(x)$$

avec

$$c_n = \langle \phi_n | \Psi(x_0, 0) \rangle = \int_{-\infty}^{\infty} \phi_n(x_0) \Psi(x_0, 0) dx_0$$

on a donc

$$|\Psi(x, t)\rangle = \sum_{n=0}^{\infty} c_n \phi_n(x) e^{-i \frac{E_n}{\hbar} t}$$

En combinant ces deux dernières équations, on obtient après permutation des signes somme et intégrale :

$$|\Psi(x, t)\rangle = \int_{-\infty}^{\infty} dx_0 \left[ \sum_{n=0}^{\infty} \phi_n^*(x) e^{-i \frac{E_n}{\hbar} t} \phi_n(x_0) \right] \Psi(x_0, 0)$$

On obtient donc  $|\Psi(x, t)\rangle$  sous forme d’une équation intégrale :

$$|\Psi(x, t)\rangle = \int_{-\infty}^{\infty} dx_0 K(x, t|x_0, 0) \Psi(x_0, 0)$$

où le noyau :

$$K(x, t|x_0, 0) = \sum_{n=0}^{\infty} \phi_n^*(x) e^{-i \frac{E_n}{\hbar} t} \phi_n(x_0)$$

appelé *opérateur d’évolution* représente la probabilité pour une particule se trouvant en  $x_0$  à l’instant  $t = 0$  de passer en  $x$  à l’instant  $t$ .

On peut écrire formellement :

$$K(x, t|x_0, 0) = \sum_{n=0}^{\infty} \phi_n^*(x) e^{-i \frac{t}{\hbar} H} \phi_n(x_0)$$

Pour résoudre l’équation intégrale par Monte-Carlo, on a besoin de fonctions réelles positives, et on va se débarrasser du facteur imaginaire par le changement de variable :

$$\tau = it$$

D’autre part on va effectuer une translation arbitraire de l’origine des énergies :

$$E_n \rightarrow E_n - E_R$$

$$V(x) \rightarrow V(x) - E_R$$

L’équation de Schrödinger devient :

$$\hbar \frac{\partial \Psi(x, \tau)}{\partial \tau} = \frac{\hbar^2}{2m} \frac{\partial^2 \Psi(x, \tau)}{\partial x^2} - [V(x) - E_R] \Psi(x, \tau)$$

Le premier terme du second membre correspond à une équation de diffusion où  $\Psi$  représente la densité de particules qui diffusent. Une telle équation peut être simulée avec une marche aléatoire de particules dans l’espace de configuration.

Le second terme est semblable à une équation décrivant un processus de mort et de naissance d’individus dans une population.

L’équation globale peut être simulée par la combinaison d’un processus de diffusion et d’un processus de branchement, dans lequel le nombre de particules diffusant augmente ou diminue de façon à réduire la densité de probabilité dans les régions où  $V(x)$  est grand et à l’augmenter dans les zones d’énergie potentielle favorables.

On a :

$$|\Psi(x, \tau)\rangle = \sum_{n=0}^{\infty} c_n \phi_n(x) e^{-\frac{(E_n - E_R)}{\hbar} \tau}$$

et le comportement asymptotique de  $|\Psi(x, \tau)\rangle$  pour  $\tau \rightarrow \infty$  ne dépend que du signe de  $(E_0 - E_R)$ ,  $E_0$  représentant l'énergie fondamentale :

- $E_R > E_0$  :  $|\Psi(x, \tau)\rangle \rightarrow 0$  pour  $\tau \rightarrow \infty$
- $E_R = E_0$  :  $|\Psi(x, \tau)\rangle \rightarrow c_0 |\phi_0(x)\rangle$  pour  $\tau \rightarrow \infty$
- $E_R < E_0$  :  $|\Psi(x, \tau)\rangle \rightarrow \infty$  pour  $\tau \rightarrow \infty$

Sous forme intégrale,  $\Psi(x, \tau)$  obéit à l'équation :

$$|\Psi(x, \tau)\rangle = \int_{-\infty}^{\infty} dx_0 K(x, \tau | x_0, 0) \Psi(x_0, 0) \quad (12.1)$$

avec

$$K(x, t | x_0, 0) = \sum_{n=0}^{\infty} \phi_n^*(x) e^{-\frac{(E_n - E_R)}{\hbar} \tau} \phi_n(x_0)$$

que l'on peut écrire formellement :

$$K(x, t | x_0, 0) = \sum_{n=0}^{\infty} \phi_n^*(x) e^{-\frac{(\hat{T} + \hat{V})}{\hbar} \tau} \phi_n(x_0)$$

avec

$$\hat{T} = \frac{\hbar^2}{2m} \frac{\partial^2 \Psi(x, \tau)}{\partial x^2}$$

et

$$\hat{V} = -[V(x) - E_R]$$

Les opérateurs  $\hat{T}$  et  $\hat{V}$  ne commutent pas, donc en général :

$$e^{-\frac{(\hat{T} + \hat{V})}{\hbar} \tau} \neq e^{-\frac{\hat{T}}{\hbar} \tau} e^{-\frac{\hat{V}}{\hbar} \tau}$$

Cependant, pour  $\delta\tau$  petit, on a [formule de Trotter-Suzuki] :

$$e^{-\frac{(\hat{T} + \hat{V})}{\hbar} \delta\tau} = e^{-\frac{\hat{V}}{\hbar} \frac{\delta\tau}{2}} e^{-\frac{\hat{T}}{\hbar} \tau} e^{-\frac{\hat{V}}{\hbar} \frac{\delta\tau}{2}} + \mathcal{O}(\delta\tau^3) \quad (12.2)$$

L'opérateur :

$$K_{diff}(x, t | x_0, 0) = \sum_{n=0}^{\infty} \phi_n^*(x) e^{-\frac{\hat{V}}{\hbar} \tau} \phi_n(x_0)$$

est connu exactement car il correspond à la propagation d'une particule libre (potentiel constant).

Considérons une particule de masse  $m$  dans une boîte unidimensionnelle de longueur  $L$ . Les fonctions propres sur l'intervalle  $[-L/2, L/2]$  sont

$$\phi_n = \frac{1}{\sqrt{L}} e^{ik_n x}, \quad \text{avec} \quad k_n = \frac{2\pi n}{L} \quad \text{et} \quad E_n = \frac{\hbar^2}{2m} k_n^2$$

Ce qui donne :

$$K_{diff}(x, t|x_0, 0) = \sum_{n=0}^{\infty} \frac{1}{\sqrt{L}} e^{-ik_n(x-x_0)} e^{-it\hbar \frac{k_n^2}{2m}}$$

soit en passant à la limite continue, avec  $L \rightarrow \infty$  :

$$K_{diff}(x, t|x_0, 0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ik(x-x_0)} e^{-it\hbar \frac{k^2}{2m}} dk$$

soit après intégration et substitution de  $it$  par  $\tau$  :

$$K_{diff}(x, \tau|x_0, 0) = \left[ \frac{m}{2\pi\hbar\tau} \right]^{1/2} e^{-\frac{m}{2\hbar\tau}(x-x_0)^2} \quad (12.3)$$

On va donc diviser  $\tau$  en  $N$  petits intervalles  $\delta\tau = \tau/N$  et pour  $\delta\tau$  petit approximer  $K(x, t|x_0, 0)$  en utilisant la relation de Trotter-Suzuki :

$$K(x, \delta\tau|x_0, 0) = \left[ \frac{m}{2\pi\hbar\delta\tau} \right]^{1/2} e^{-\frac{\delta\tau}{\hbar} \frac{(V(x)-E_R)}{2}} e^{-\frac{m}{2\hbar\delta\tau}(x-x_0)^2} e^{-\frac{\delta\tau}{\hbar} \frac{(V(x_0)-E_R)}{2}} \quad (12.4)$$

Et en itérant  $N$  fois la relation intégrale 12.1 on obtient :

$$\Psi(x, \tau) = \left[ \prod_{j=0}^{N-1} \int_{-\infty}^{\infty} dx_j \right] \left\{ \prod_{n=1}^N W(x_n) P(x_n, x_{n-1}) \right\} \Psi(x_0, 0)$$

avec

$$P(x_n, x_{n-1}) = \left( \frac{m}{2\pi\hbar\delta\tau} \right)^{\frac{1}{2}} \exp \left[ -\frac{m(x_n - x_{n-1})^2}{2\hbar\delta\tau} \right]$$

et

$$W(x_n) = \exp \left[ -\frac{(V(x_n) - E_R)\delta\tau}{\hbar} \right]$$

La fonction :  $P(x_n, x_{n-1})$  représente une densité de probabilité car :

$$\int_{-\infty}^{\infty} P(x, y) dy = 1 \quad \forall x$$

ceci va nous permettre de construire une chaîne de Markov ayant pour loi de transition  $P(x, y)$ .

Les poids  $W(x_n)$  dépendant uniquement du potentiel  $V(x)$  et de l'énergie de référence  $E_R$  vont être traduits par un processus de naissance-mort de particules. C'est à dire qu'au lieu de raisonner sur une seule trajectoire de la particules, on va suivre en parallèle l'évolution de  $\mathcal{N}$  trajectoires.

- On part de  $\mathcal{N}$  points distribués suivant une probabilité proportionnelle à  $|\Psi(x_0, 0)|^2$
- A chaque étape  $\tau_n = n\delta\tau$  on tire  $\mathcal{N}$  nouvelles abscisses suivant la loi de probabilité  $P(x_n, x_{n-1})$

- On définit  $m_n = \min[\text{int}(W(x_n) + \eta), \nu]$  où  $\eta$  est un nombre aléatoire uniformément distribué sur  $[0,1]$ ,  $\nu$  est un petit nombre arbitraire (en pratique  $\nu \approx 3$ ) Si  $m_n = 0$ , on arrête le processus; si  $m_n = 1$  la particule continue normalement son chemin; enfin si  $m_n > 1$  on crée  $(m_n - 1)$  nouvelles particules fictives au même endroit.

Au cours de l'évolution, on ajuste  $E_R$  pour obtenir une "population" stable, on a alors

$$E_0 = E_R = \langle V \rangle$$

Donc, en pratique on ajuste à chaque étape  $(E_R)_n$  à  $\langle V \rangle_n$ . Lorsque le processus a convergé, la densité de points sur l'axe  $x$  représente le module au carré de la fonction d'onde correspondant à l'état fondamental.

Le rôle de  $\nu$  est d'éviter au début un trop grand nombre de branchements. Il n'intervient pas dans le résultat final, car lorsque l'équilibre est atteint, la probabilité pour que  $W(x)$  soit supérieur à  $\nu$  est faible.

## 12.3 "Path Integral Monte-Carlo"

### 12.3.1 Valeur moyenne d'une grandeur physique

En mécanique quantique, un système est décrit par un hamiltonien  $H$ . Soit  $|\Phi_i\rangle$  un ensemble de vecteurs propres orthonormés associés respectivement aux valeurs propres  $E_i$  :

$$H|\Phi_i\rangle = E_i|\Phi_i\rangle$$

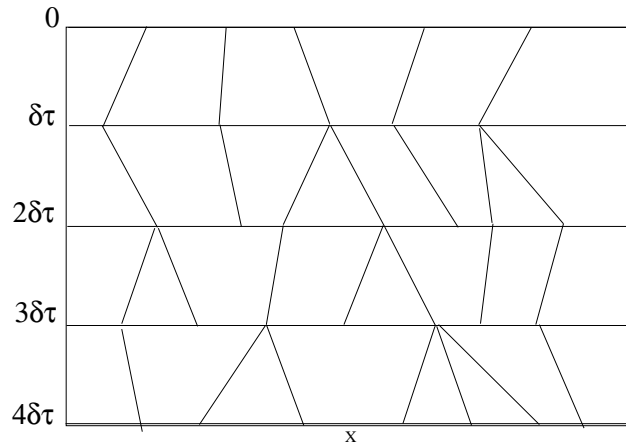


FIG. 12.1 – Processus de diffusion + naissance-mortalité



Une grandeur physique est représentée par un opérateur  $\mathcal{O}$ . Dans l'ensemble canonique la probabilité du système de se trouver dans l'état  $|\Phi_i\rangle$  est proportionnelle à  $\exp(-E_i/kT)$  et la valeur moyenne de la grandeur physique correspondante est :

$$\langle \mathcal{O} \rangle = \frac{\sum_i \langle \Phi_i | \mathcal{O} | \Phi_i \rangle \exp[-\beta E_i]}{Z}$$

$$Z = \sum_i \exp[-\beta E_i]$$

est la fonction de partition et  $\beta = 1/kT$

Cette valeur moyenne peut encore s'écrire sous la forme plus générale, indépendante de la représentation (base orthonormée de l'espace de Hilbert) choisie :

$$\langle \mathcal{O} \rangle = \frac{\text{Tr } \mathcal{O} \exp[-\beta H]}{Z} \quad (12.5)$$

avec  $Z = \text{Tr } \exp[-\beta H]$ . La trace “Tr” d'un opérateurs étant invariante, quelle que soit la représentation choisie..

### 12.3.2 La Matrice Densité

On choisit, en général pour représentation non pas une base de vecteurs propres telle que  $\{|\Phi_i\rangle\}$  mais une base d'états orthonormés  $\{|R\rangle\}$  décrivant les positions dans l'espace des particules. Cette base est continue et les sommes discrètes relatives à la base discrète  $\{|\Phi_i\rangle\}$  considérée précédemment sont remplacées par des intégrales. En particulier, la trace d'un opérateur, somme de ses valeurs propres de vient une intégrale :

$$\text{Tr } A = \int dR' \langle R' | A | R' \rangle$$

Alors la relation générale 12.5 s'écrit :

$$\langle \mathcal{O} \rangle = \frac{\int dR' \langle R' | \mathcal{O} \exp(-\beta H) | R' \rangle}{Z} \quad (12.6)$$

En utilisant l'orthonormalité de la base  $\{|R\rangle\}$  qui se traduit par la relation “de fermeture”

$$\int dR |R\rangle \langle R| = 1 \quad (12.7)$$

on peut rendre plus explicite le produit des deux matrices  $\mathcal{O}$  et  $\exp(-\beta H)$  pour arriver à l'expression :

$$\langle \mathcal{O} \rangle = \frac{\int dR dR' \langle R' | \mathcal{O} | R \rangle \langle R | \exp(-\beta H) | R' \rangle}{Z} \quad (12.8)$$

En introduisant la “Matrice Densité :

$$\rho(R, R'; \beta) = \langle R | \exp(-\beta H) | R' \rangle \quad (12.9)$$

on a

$$\langle \mathcal{O} \rangle = \frac{\int dR dR' \langle R' | \mathcal{O} | R \rangle \rho(R, R'; \beta)}{Z} \quad (12.10)$$

avec

$$Z = \int dR \rho(R, R, \beta)$$

Pour calculer l'intégrale ci-dessus par la méthode de Monte-Carlo, nous allons donc générer de configurations  $\{R, R'\}$  suivant une probabilité proportionnelle à  $\rho(R, R'; \beta)$  et la moyenne des éléments de matrice  $\langle R' | \mathcal{O} | R \rangle$  correspondant à ces configurations nous donnera accès à une approximation de  $\langle \mathcal{O} \rangle$

### 12.3.3 Echantillonnage de $\rho(R, R'; \beta)$

#### Une particule à une dimension

Nous illustrons d'abord la méthode sur le cas simple utilisé précédemment pour l'algorithme de "Diffusion Monte-Carlo" : une particule à une dimension, de masse  $m$ , soumise à un potentiel  $V(x)$ . L'hamiltonien s'écrit :

$$H = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x)$$

soit

$$H = \hat{T} + \hat{V}$$

en séparant la partie cinétique :

$$\hat{T} = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2}$$

de la partie potentielle

$$\hat{V} = V(x)$$

La matrice densité s'écrit :

$$\rho(x, x'; \beta) = \langle x | \exp[-\beta(\hat{T} + \hat{V})] | x' \rangle$$

Cette quantité est très difficile à calculer, alors que les quantités  $\langle x | \exp[-\beta\hat{T}] | x'' \rangle$  et  $\langle x'' | \exp[-\beta\hat{V}] | x' \rangle$  sont relativement simples. Malheureusement, les opérateurs  $\hat{T}$  et  $\hat{V}$  ne commutent pas et

$$\exp[-\beta(\hat{T} + \hat{V})] \neq \exp[-\beta\hat{T}] \exp[-\beta\hat{V}]$$

Par contre, en divisant  $\beta$  en  $M$  intervalles  $\delta\tau = \beta/M$ , on pourra recourir à la formule de Trotter-Suzuki (cf. Equation 12.2)

$$\exp[-\beta(\hat{T} + \hat{V})] = \lim_{M \rightarrow \infty} \left[ \exp(-\delta\tau\hat{T}) \exp(-\delta\tau\hat{V}) \right]^M \quad (12.11)$$

Enfin, en intercalant des relations de fermeture :

$$\int dx_i |x_i\rangle\langle x_i| = 1$$

entre chaque produit d'opérateurs, on aboutit à la relation :

$$\begin{aligned} \rho(x_0, x_M; \beta) = & \int dx_1 \int dx_2 \cdots \int dx_{M-1} \langle x_0 | \exp(-\delta\tau\hat{T}) \exp(-\delta\tau\hat{V}) | x_1 \rangle \langle x_1 | \exp(-\delta\tau\hat{T}) \exp(-\delta\tau\hat{V}) | x_2 \rangle \\ & \cdots \langle x_{M-1} | \exp(-\delta\tau\hat{T}) \exp(-\delta\tau\hat{V}) | x_M \rangle \end{aligned}$$

Calcul de  $\langle x_{i-1} | \exp(-\delta\tau\hat{T}) \exp(-\delta\tau\hat{V}) | x_i \rangle$

L'opérateur  $\hat{V}$  est diagonal, on a donc :

$$\langle x_{i-1} | \exp(-\delta\tau\hat{T}) \exp(-\delta\tau\hat{V}) | x_i \rangle = \langle x_{i-1} | \exp(-\delta\tau\hat{T}) | x_i \rangle \exp[-\delta\tau\hat{V}(x_i)]$$

Mais l'élément de matrice :

$$\langle x_{i-1} | \exp(-\delta\tau\hat{T}) | x_i \rangle = \left[ \frac{m}{2\pi\hbar^2\delta\tau} \right]^{1/2} \exp\left[-\frac{m}{2\hbar^2\delta\tau}(x_i - x_{i-1})^2\right]$$

pour une particule libre a été calculé au paragraphe précédent (cf. Relation 12.3)

En reportant dans l'expression précédente, le calcul de  $\rho(x_0, x_M; \beta)$  se présente comme une intégrale à  $M - 1$  dimensions :

$$\rho(x_0, x_M; \beta) = C \int dx_1 \int dx_2 \cdots \int dx_{M-1} \exp\left(-\sum_{i=1}^M \left[ \frac{m}{2\hbar^2\delta\tau}(x_{i-1} - x_i)^2 + \delta\tau V(x_i) \right]\right) \quad (12.12)$$

avec

$$C = \left[ \frac{m}{2\pi\hbar^2\delta\tau} \right]^{M/2}$$

La somme figurant en exposant du facteur exponentiel correspond à l'énergie d'une chaîne de particules de coordonnées  $x_0, x_1, \dots, x_M$ , chaque particule étant reliée à la précédente par un ressort d'énergie proportionnelle au carré de la distance qui les sépare. Chaque particule  $i$  étant soumise à un potentiel  $V(x_i)$ . On obtient l'image d'un polymère (cf section 11.2.2)

### N particules discernables à 3 dimensions

Ces calculs se généralisent facilement pour une particule à deux ou trois dimensions... et pour un ensemble de N particules à trois dimensions soumises à un potentiel d'interaction de paires

$$V(|R_i - R_j|)$$

ne dépendant que de la distance entre paires de particules. Dans ce cas, chaque particule  $p$  est représentée par une chaîne de  $M$  monomères de coordonnées  $R_{p,i}$ . La matrice densité s'exprime alors comme une intégrale à  $3NM$  dimensions :

$$\rho(R_{1,0}, R_{2,0}, \dots, R_{N,0}, R_{1,M}, R_{2,M}, \dots, R_{N,M}; \beta) = \left[ \frac{m}{2\pi\hbar^2\delta\tau} \right]^{3NM/2} \int dR_{1,1} \cdots \int dR_{N,M} \exp \left[ -\frac{m}{2\hbar^2\delta\tau} \sum_{p=1}^N \sum_{i=1}^M (R_{p,i-1} - R_{p,i})^2 \right] \exp \left[ -\delta\tau \sum_{p=1}^N \sum_{k>p}^N \sum_{i=1}^M V(|R_{p,i} - R_{k,i}|) \right]$$

### N Bosons

Au paragraphe précédent, nous avons raisonné sur des particules discernables. Or, en mécanique quantique, les particules sont indiscernables.

Pour des bosons, l'expression de la matrice densité  $\rho_B$  doit être complètement symétrique par rapport à une permutation quelconque des particules.

On doit donc sommer sur toutes les permutations possibles de l'expression d'une matrice densité de particules discernables :

$$\rho_B(R_{1,0}, \dots, R_{N,0}, R_{1,M}, \dots, R_{N,M}; \beta) = \frac{1}{N!} \sum_P \rho_B(R_{1,0}, \dots, R_{N,0}, R_{P(1),M}, \dots, R_{P(N),M}; \beta) =$$

où  $P(1), P(2), \dots, P(N)$ , représente une permutation quelconque des  $N$  particules.

Les permutations seront elles-mêmes échantillonnées par un algorithme de Monte-Carlo.

Nous n'aborderons pas le problème des Fermions pour lesquels l'antisymétrisation conduit à une alternance de signe qui constitue un des problèmes les plus difficiles pour l'application de la méthode de Monte-Carlo.

### Pour en savoir plus...

Le livre de Feynman [16] est essentiel pour approfondir la formulation de la mécanique quantique sous forme "d'intégrales de chemin", que nous avons brièvement introduite ci-dessus.

Pour une approche plus complète des méthodes de Monte-Carlo quantique, nous conseillons les articles de revue et cours de Ceperley [17]



## **Annexes : Probabilités et Physique Statistique**

---





# Rappels de théorie des probabilités

---

*Nous nous contenterons de rappeler brièvement les notions indispensables à la justification des méthodes de Monte-Carlo. Pour un exposé complet et didactique sur la théorie des probabilités, nous conseillons le livre (niveau second cycle) de J. Bass[18] ou le cours de l'Ecole Polytechnique de J. Neveu[19].*

## A.1 Expérience, ensemble des résultats possibles.

Une expérience aléatoire est une expérience dont le résultat est soumis au hasard. Elle se décrit par la donnée de l'ensemble des résultats possibles. Nous noterons  $\omega$  un tel résultat et  $\Omega$  l'espace formé par tous les résultats possibles.

Exemple : l'espace  $\Omega$  associé au jet aléatoire de deux dés est composé des 36 couples d'entiers  $\omega = (x, y)$  tels que  $1 \leq x, y \leq 6$

## A.2 Notion d'événement aléatoire.

Un événement aléatoire est représenté par l'ensemble des résultats  $\omega$  de l'expérience qui le réalisent. Dans l'exemple précédent du jet de deux dés, on peut définir comme événement : “obtenir un total de points supérieur à 8”. Cet événement est représenté par le sous ensemble de résultats  $A \subset \Omega$  :

$$A = \{(3, 6), (4, 5), (5, 4), (6, 3), (4, 6), (5, 5), (6, 4), (5, 6), (6, 5), (6, 6)\}$$

Nous devons maintenant envisager des opérations logiques sur ces événements.

- A tout événement  $A$  est associé son contraire  $A^c$  représenté par l'ensemble complémentaire de  $A$  dans  $\Omega$ . (Le contraire de  $A$  dans l'exemple précédent est “obtenir un total de points inférieur ou égal à 8”).
- Pour tout couple d'événements  $(A_1, A_2)$ , l'événement “ $A_1$  et  $A_2$ ” est par définition celui qui est réalisé si les événements  $A_1$  et  $A_2$  sont réalisés à la fois. En langage de théorie des

ensembles, il correspond à l'intersection  $A_1 \cap A_2$ . En prenant toujours le même exemple de jet de deux dés, considérons,  $A_1$  : “obtenir un total de points supérieur à 8” et  $A_2$  : “chacun des dés est tombé sur un nombre pair”, l'événement “ $A_1$  et  $A_2$ ” est donc représenté par :

$$A_3 = A_1 \cap A_2 = \{(4, 6), (6, 4), (6, 6)\}$$

- L'événement impossible sera représenté par l'ensemble vide  $\emptyset$ . L'équation  $A_1 \cap A_2 = \emptyset$  signifie que les événements  $A_1$  et  $A_2$  sont incompatibles, c'est à dire que les parties de  $\Omega$  qui les représentent sont disjointes
- Pour tout couple d'événements  $(A_1, A_2)$ , l'événement “ $A_1$  ou  $A_2$ ” est par définition celui qui est réalisé si l'un au moins des deux événements  $A_1$  ou  $A_2$  est réalisé (“ou” non exclusif). En langage de théorie des ensembles, il correspond à l'union  $A_1 \cup A_2$ .
- L'événement certain sera noté  $\Omega$  puisqu'il est réalisé quel que soit le résultat  $\omega$  de l'expérience.

En conclusion, une expérience aléatoire est décrite mathématiquement par la donnée d'un espace  $\Omega$  et d'une classe  $\mathcal{A}$  de parties de  $\Omega$ . Chaque élément  $A$  de  $\mathcal{A}$  représente un événement que l'on considère relativement à l'expérience. Nous exigeons de la classe  $\mathcal{A}$  les propriétés suivantes :

- Elle contient l'événement certain : l'ensemble  $\Omega$  lui-même.
- Si  $A$  appartient à  $\mathcal{A}$ , l'ensemble  $A^c$  complémentaire de  $A$  par rapport à  $\Omega$  (i.e. “le contraire de  $A$ ”) est aussi un élément de la classe  $\mathcal{A}$
- Si  $A_1$  et  $A_2$  appartiennent à  $\mathcal{A}$ , la réunion et l'intersection de  $A_1$  et  $A_2$  appartiennent à  $\mathcal{A}$ .
- Lorsque l'espace  $\Omega$  n'est pas fini, il est encore indispensable d'exiger que la réunion d'une infinité dénombrable d'ensembles de  $\mathcal{A}$  appartienne à  $\mathcal{A}$ .

*On dit que  $\mathcal{A}$  est une “tribu de parties de  $\Omega$ ”, ou encore une “ $\sigma$ -algèbre construite sur  $\Omega$ .”*

Par exemple, dans  $\mathbb{R}$ , la tribu engendrée par les segments ouverts [i.e la plus petite tribu contenant les segments ouverts] est nommée “tribu Borélienne”. D'après les propriétés énoncées ci-dessus, elle contient aussi les segments fermés (complémentaires d'ouverts), l'intersection et la réunion de nombres infinis dénombrables d'entre eux.

## A.3 Notion de probabilité.

### A.3.1 Définition heuristique

Une propriété essentielle d'une expérience est de pouvoir être répétée indéfiniment. Si nous comptons, au cours de  $N$  répétitions d'une expérience, le nombre de fois  $N_A$  où un événement  $A$  est réalisé, nous observons que la fréquence statistique  $N_A/N$  tend vers une limite  $p \leq 1$  lorsque  $N \rightarrow \infty$ . D'où l'idée d'associer un nombre positif ou nul  $P(A) = p$ , inférieur à 1, à tout événement  $A$ .

Si deux événements sont incompatibles ( $A_1 \cap A_2 = \emptyset$ ), la fréquence statistique de l'événement



$A_1 \cup A_2$  au cours de  $N$  répétitions d'une expérience aléatoire est égale à la somme des fréquences statistiques des événements  $A_1$  et  $A_2$  puisque un seul des événements, au plus, peut être réalisé.

$$\frac{N_{A_1 \cup A_2}}{N} = \frac{N_{A_1}}{N} + \frac{N_{A_2}}{N}$$

donc

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) \quad \text{si} \quad A_1 \cap A_2 = \emptyset$$

### A.3.2 Définition mathématique

- A chaque événement  $A \in \mathcal{A}$  est associé un nombre réel  $P(A) \in [0, 1]$  appelé probabilité de  $A$
- La probabilité 1 est attribuée à l'événement certain :  $P(\Omega) = 1$
- Nous imposons la propriété d'additivité pour toute suite (finie ou infinie-dénombrable) d'événements deux à deux indépendants :

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n) \quad (\text{A.1})$$

si

$$A_i \cap A_j = \emptyset \quad \forall (i, j), \quad i \neq j$$

En langage mathématique,  $P$  est une “mesure positive, de masse totale  $P(\Omega) = 1$ , définie sur la tribu  $\mathcal{A}$  de parties de  $\Omega$ ”.

On note “espace probabilisé”  $\{\Omega, \mathcal{A}, P\}$  le triplet formé par l'ensemble des résultats possibles de l'expérience  $\Omega$ , la tribu de parties de cet ensemble  $\mathcal{A}$  et la mesure de probabilités  $P$ .

## A.4 Probabilités conditionnelles. Événements indépendants

Considérons une expérience qui se réalise en deux temps. En reprenant toujours l'exemple du jet de deux dés, on peut lancer successivement un premier dé puis le second. On peut définir des événements relatifs aux résultats du premier temps (exemple  $B$  : “le résultat du premier dé est pair”) et des événements relatifs au résultat global après le second temps (exemple  $A$  : “le total des points est 8”). Ces deux événements, bien sûr, appartiennent à une même tribu  $\mathcal{A}$  de parties de  $\Omega$

Nous devons alors distinguer heuristiquement entre la probabilité, dans l'absolu,  $P(A)$  de l'événement  $A$ , et la probabilité de l'événement  $A$ , sachant que  $B$  s'est produit, que nous appellerons “probabilité de l'événement  $A$  conditionnelle en l'événement  $B$ ”. Nous la noterons  $P(A/B)$ .

De même que la fréquence  $N_A/N$  des réalisations de l'événement  $A$  au cours de  $N$  répétitions de l'expérience aléatoire, donne une estimation statistique de la probabilité  $P(A)$  de l'événement  $A$ , nous pouvons penser que le rapport  $N_{A \cap B}/N_B$  de réalisations de l'événement  $A$

parmi les  $N_B$  répétitions de l'expérience qui ont donné  $B$  fournisse une estimation statistique de la probabilité de l'événement  $A$  conditionnelle en l'événement  $B$ .

Après avoir remarqué que :

$$\frac{N_{A \cap B}}{N_B} = \frac{\frac{N_{A \cap B}}{N}}{\frac{N_B}{N}}$$

il devient naturel de poser comme définition :

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad (\text{A.2})$$

Deux événement  $A$  et  $B$  seront dits indépendants si la probabilité de l'événement  $A$  conditionnelle en l'événement  $B$  ne dépend pas de l'événement  $B$ , c'est à dire si  $P(A/B) = P(A)$ , ce qui d'après la relation précédente correspond à l'égalité

$$P(A \cap B) = P(A)P(B) \quad (\text{A.3})$$

## A.5 Variable aléatoire.

Une variable aléatoire est définie par référence à une expérience aléatoire comme un nombre réel, ou plus généralement un vecteur  $X = \phi(\omega)$  dont la valeur dépend du résultat  $\omega$  de l'expérience. Par exemple, dans l'expérience du jet de deux dés, la somme  $S$  des points amenés par les deux dés est une variable aléatoire réelle. Le vecteur  $V = (m, n)$  où  $m$  est le nombre de points amené par le premier dé et  $n$  le nombre de points amenés par le second dé est un vecteur aléatoire qui prend ses valeurs dans  $R^2$ .

Une variable aléatoire est donc représentée par une application  $\phi$  de  $\Omega$  dans  $R^d$ . Elle est "aléatoire" du fait du caractère aléatoire de l'expérience, dont chaque résultat possible est un point  $\omega$  de l'ensemble  $\Omega$ .

Nous exigerons une condition essentielle sur cette application :

L'ensemble des résultats  $\omega$  de l'expérience, tels que :

$$a < X = \phi(\omega) < b \quad (\text{A.4})$$

appartient à la tribu  $\mathcal{A}$

Si  $X$  est une variable aléatoire réelle,  $a$  et  $b$  représentent deux réels quelconques tels que  $a < b$ .

Si  $X$  est une variable aléatoire vectorielle à valeurs dans  $R^d$ ,  $a = (a_1, a_2, \dots, a_d)$  et  $b = (b_1, b_2, \dots, b_d)$  représentent deux vecteurs de  $R^d$  et les inégalités sont vérifiées composante par composante. Dans ce cas, l'inégalité

$$a < x < b \quad \text{i.e.} \quad a_i < x_i < b_i \quad \forall i$$

définit un pavé  $B$  de  $R^d$ .

Donc, l'image réciproque par l'application  $X = \phi(\omega)$  d'un segment  $B$  de  $R$ , pour une variable aléatoire réelle, (ou plus généralement d'un "pavé"  $B$  de  $R^d$ , pour une variable aléatoire vectorielle) est un élément (i.e "un événement") de la tribu  $\mathcal{A}$  de sous ensembles de  $\Omega$ .

Cette condition est indispensable pour que nous puissions parler de la probabilité de "l'événement :  $\{a < X = \phi(\omega) < b\}$ . Elle correspond à exiger que l'application  $X = \phi(\omega)$  soit une "application mesurable".

Remarque importante : la classe  $\mathcal{B}$  de parties de  $R^d$  formée par les pavés de  $R^d$ , leurs intersections et la réunion d'un nombre fini ou infini-dénombrable d'entre eux constitue une tribu (cf. ci-dessus) appelée "tribu Borélienne"

## A.6 Loi de probabilité d'une variable aléatoire

Soit  $A_B$  l'image réciproque dans  $\Omega$  d'un pavé  $B$  de  $R^d$  ( $B$  est un élément de la tribu Borélienne  $\mathcal{B}$ ) par l'application  $X = \phi(\omega)$ , c'est à dire :

$$A_B = \{\omega \text{ tels que } X = \phi(\omega) \in B\}$$

alors les sous ensembles "événements"  $A_B$  forment une sous tribu de la tribu  $\mathcal{A}$  appelée "la tribu engendrée par la variable aléatoire  $X$ ".

La loi :

$$P_{(X)}(B) = P(A_B)$$

définit une probabilité  $P_{(X)}$  sur  $R^d$  par rapport à la tribu  $\mathcal{B}$  "Borélienne" formée par les pavés ouverts ou fermés de  $R^d$  et la réunion d'un nombre fini, ou infini-dénombrable d'entre eux.

Cette probabilité  $P_{(X)}$  sur  $R^d$ , image de la probabilité  $P$  par l'application  $X = \phi(\omega)$  de  $\Omega$  dans  $R^d$  est appelée "la loi de probabilité de la variable aléatoire  $X$ "

L'intérêt de l'introduction de ces "lois de probabilité" tient à ce qu'il est plus aisé de manipuler des "mesures" sur  $R^d$  que sur l'espace  $\Omega$ . Pour les applications en physique, nous utiliserons le plus souvent la mesure de Lebesgue ou la mesure de Riemann sur  $R^d$ .

## A.7 Fonction de répartition d'une variable aléatoire réelle

Considérons le segment semi infini  $B = ]-\infty, z]$  alors

$$F(z) = P_X(B) = P(A_B) = P(\{-\infty < X(\omega) \leq z\})$$

est la fonction de partition relative à la loi de probabilité de la variable aléatoire  $X$ . C'est une fonction positive, croissante qui atteint sa valeur maximale pour  $z \rightarrow \infty$ .

$$\lim_{z \rightarrow \infty} F[z] = P(\Omega) = 1$$

C'est une fonction à valeurs dans le segment  $[0, 1]$

## A.8 Densité de probabilité.

Si la loi de probabilité  $P_{(X)}$  de la variable aléatoire  $X$  peut s'écrire :

$$P_{(X)}(B) = P(\{X = \phi(\omega) \in B\}) = \int_B f(x) dx \quad (\text{A.5})$$

où  $dx$  désigne la mesure de Lebesgue ou la mesure de Rieman.

Alors nous dirons que la variable aléatoire  $X$  possède la densité de probabilité  $f(x)$

( $f$  est la densité de la “mesure de probabilité”  $P_{(X)}$ )

$f$  est une fonction positive, et il résulte de l'égalité  $P_{(X)}(R^d) = 1$  que :

$$\int_{R^d} f(x) dx = 1 \quad (\text{A.6})$$

Si une variable aléatoire  $X$  possède une densité de probabilité  $f(x)$ , on peut écrire si  $X$  est définie dans  $R$  ( $d = 1$ ) :

$$P(\{x < X = \phi(\omega) < x + dx\}) = f(x) dx \quad \text{lorsque} \quad dx \rightarrow 0 \quad (\text{A.7})$$

et plus généralement pour une variable aléatoire définie dans  $R^d$ , en notant  $x = (\alpha_1, \alpha_2, \dots, \alpha_d)$  un élément de  $R^d$  :

$$P(\{\alpha_k < (X)_k = \phi(\omega)_k < \alpha_k + d\alpha_k\}) = f(x) d\alpha_1 d\alpha_2 \dots d\alpha_d \quad (\text{A.8})$$

Etant donné que, dans les applications physiques que nous allons étudier, nous nous restreignons au cas où les variables aléatoires considérées possèdent une densité de probabilité  $f$ , le lecteur, quelque peu dérouté par le formalisme général peut se “raccrocher” aux relations ci-dessus en tant que définition d'une densité de probabilité.

la fonction de répartition d'une variable aléatoire réelle admettant une densité de probabilité  $f(x)$  est

$$F(z) = \int_{-\infty}^z f(x) dx$$

Quelques exemples de densités de probabilité dans  $R$

- La loi uniforme sur un segment :

$$f(x) = 1/(b-a) \quad \text{si} \quad a < b < x \quad \text{et} \quad 0 \quad \text{sinon}$$

- La loi exponentielle

$$f(x) = \lambda e^{-\lambda x} \quad x \geq 0 \quad \text{et} \quad 0 \quad \text{sinon}$$

qui décrit un certain nombre de phénomènes physiques, dont le temps de désintégration d'un élément radioactif.

- La loi de Gauss (ou loi “normale”)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad (\text{A.9})$$

très importante en physique et statistique

## A.9 Espérance, moments d'ordre $n$ , variance

### A.9.1 Valeur moyenne ou “Espérance” d'une variable aléatoire réelle

Considérons tout d'abord une variable aléatoire discrète  $X$ , c'est à dire à valeurs dans l'ensemble des entiers relatifs  $Z \subset R$ . Au cours de  $N$  répétitions de l'expérience aléatoire, chaque valeur  $\{X = x_i\}$  “sort”  $k_i$  fois. Si nous voulons évaluer empiriquement une valeur moyenne  $m$  de la variable aléatoire  $X$ , nous prendrons :

$$m = \sum_i k_i x_i / N = \sum_i \left( \frac{k_i}{N} \right) x_i$$

Lorsque  $N \rightarrow \infty$ , on associe la probabilité  $p_i$  à la fréquence  $k_i/N$ , d'où la définition de la valeur moyenne ou espérance dans le cas discret :

$$E[X] = m = \sum_Z x_i p_i \quad (\text{A.10})$$

Dans le cas continu, la somme discrète est remplacée par une intégrale :

$$E[X] = \int_{\Omega} X dP = \int_R x dP_{(X)} \quad (\text{A.11})$$

où les intégrales sont prises au sens général de la théorie de l'intégration par rapport aux mesures abstraites  $dP$ ,  $dP_{(X)}$ .

Dans le cas où la variable aléatoire  $X$  possède la densité de probabilité  $f$ , alors, on peut écrire plus simplement :

$$E(X) = m = \int_R x f(x) dx \quad (\text{A.12})$$

où  $dx$  représente la mesure de Lebesgue ou la mesure de Rieman sur  $R$

### A.9.2 Moment d'ordre $n$ . Variance d'une variable aléatoire réelle

Toute fonction  $Y = g(X) = g[\phi(\omega)] = g \circ \phi(\omega)$  définit une nouvelle variable aléatoire. Son espérance est :

$$E[Y] = E[g(X)] = \int_R g(x)f(x)dx \quad (\text{A.13})$$

Considérons en particulier  $g(X) = X^n$ , alors

$$\mu_n = E[X^n] = \int_R x^n f(x)dx \quad (\text{A.14})$$

est le moment d'ordre  $n$  de la variable aléatoire réelle  $X$

Le moment d'ordre deux :

$$E[X^2] = \int_R x^2 f(x)dx \quad (\text{A.15})$$

joue un rôle particulièrement important.

On appelle variance le moment d'ordre deux de  $(X - m) = (X - E[X])$  et écart type  $\sigma$  la racine carrée de cette quantité

$$\sigma = \sqrt{E[(X - E[X])^2]} \quad (\text{A.16})$$

La variable aléatoire  $(X - E[X])/\sigma$  a pour espérance 0 et pour écart type 1. On dit qu'elle est réduite.

## A.10 Fonction Caractéristique

### A.10.1 Définition

La fonction caractéristique d'une variable aléatoire réelle  $X$  est la transformée de Fourier de la loi de probabilité  $P_{(X)}$  de la variable aléatoire  $X$  :

$$\mathcal{F}(t) = E[\exp(itX)]$$

Son développement de Taylor à l'origine s'exprime en fonction des moments d'ordre  $n$  successifs de la variable aléatoire  $X$

$$\mathcal{F}(t) = 1 + itE(X) - t^2 E[X^2]/2 - it^3 E[X^3]/3! \dots$$

Si  $X$  admet une densité de probabilité  $f(x)$ , alors sa fonction caractéristique est la transformée de Fourier de la fonction  $f(x)$ .

## A.10.2 Exemple :

*Loi gaussienne :*

Nous choisissons pour exemple la loi gaussienne de densité de probabilité :

$$f_G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

car sa fonction caractéristique est très simple, et le résultat nous servira à la démonstration du théorème de la limite centrale.

$$\mathcal{F}_G(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp(itx) \exp\left(-\frac{x^2}{2\sigma^2}\right) dx$$

soit

$$\mathcal{F}_G(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{\sigma^2 t^2}{2}\right) \exp\left[-\frac{(x - \sigma^2 it)^2}{2\sigma^2}\right] dx$$

Nous allons montrer que l'intégrale :

$$I = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left[-\frac{(x - \sigma^2 it)^2}{2\sigma^2}\right] dx$$

est simplement 1. Pour cela nous considérons dans le plan complexe la fonction générale :

$$G(Z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{Z^2}{2\sigma^2}\right]$$

et nous l'intégrons sur le contour fermé représenté sur la figure A.1, Puisqu'il n'y a aucun pôle à l'intérieur de ce contour, le théorème des résidus nous dit que l'intégrale sur le contour est nulle :

$$\int_{C_1} G(Z) dZ + \int_{C_2} G(Z) dZ + \int_{C_3} G(Z) dZ + \int_{C_4} G(Z) dZ = 0$$

Lorsque  $A \rightarrow \infty$  les deux derniers termes tendent vers 0 car  $|Z|^2 \rightarrow \infty$ . Le second terme n'est autre que  $I$  et le premier terme représente l'intégrale sur l'axe réel :

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left[-\frac{x^2}{2\sigma^2}\right] dx = 1$$

Donc  $I = 1$  et la fonction caractéristique de la loi gaussienne est :

$$\mathcal{F}_G(t) = \exp\left(-\frac{\sigma^2 t^2}{2}\right) \quad (\text{A.17})$$

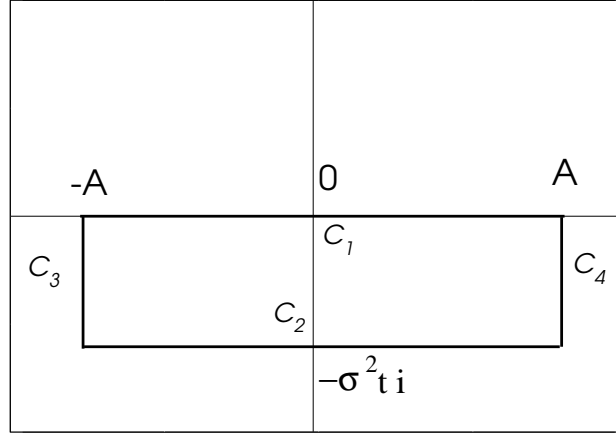


FIG. A.1 –

### A.11 Inégalité de Bienaymé-Tchebichef

Nous allons démontrer une inégalité triviale, mais néanmoins importante introduite par Bienaymé et Tchebichef.

Pour toute variable aléatoire réelle  $X$ , de carré intégrable (i.e  $\mu_2 = E[X^2]$  est fini), et pour tout nombre réel  $a$  strictement positif :

$$P(|X| \geq a) \leq \frac{\mu_2}{a^2} \quad (\text{A.18})$$

Démonstration : Désignons par

$$1_{\{|x| \geq a\}}$$

la fonction qui vaut 1 si  $|x| \geq a$  et 0 sinon, alors on a

$$1_{\{|x| \geq a\}} \leq \frac{1}{a^2} x^2$$

et on a aussi, entre les variables aléatoires définies par les fonctions correspondant à chacun des deux membres :

$$1_{\{|X| \geq a\}} \leq \frac{1}{a^2} X^2$$

En prenant l'intégrale par rapport à la mesure de probabilité  $P$  de chacun des deux membres, on obtient :

$$P(|X| \geq a) \leq E \left[ \frac{X^2}{a^2} \right]$$



Et comme

$$E\left[\frac{X^2}{a^2}\right] = E[X^2]/a^2$$

l'inégalité de Bienaymé-Tchebichev est ainsi démontrée

## A.12 Couples de variables aléatoires

Soient deux variables aléatoires  $X_1$  et  $X_2$  de lois de probabilité  $P_{X_1}$  et  $P_{X_2}$ . On peut définir la loi de probabilité conjointe du couple (ou vecteur)  $U = (X_1, X_2)$  :

$$P_{(U)}(A \times B) = P(\{X_1 \in A\} \cap \{X_2 \in B\})$$

### A.12.1 Covariance

On appelle covariance des deux variables aléatoires :

$$\text{cov}(X_1, X_2) = E[(X_1 - E[X_1])(X_2 - E[X_2])] = E[X_1 X_2] - E[X_1]E[X_2]$$

### A.12.2 Indépendance

Deux variables aléatoires sont indépendantes si quels que soient  $A$  et  $B$  :

$$P_{(U)}(A \times B) = P_{(X_1)}(A)P_{(X_2)}(B)$$

soit

$$P(\{X_1 \in A\} \cap \{X_2 \in B\}) = P(\{X_1 \in A\})P(\{X_2 \in B\})$$

On en déduit :

$$E(X_1 X_2) = E(X_1)E(X_2) \quad \text{donc} \quad \sigma(X_1, X_2) = 0$$

Si les variables aléatoires  $X_1$  et  $X_2$  admettent des densités de probabilité  $f_1(x_1)$ ,  $f_2(x_2)$  alors la loi de probabilité conjointe du couple admet pour densité de probabilité :

$$h(x_1, x_2) = f_1(x_1)f_2(x_2)$$

### A.12.3 Somme de 2 variables aléatoires indépendantes

La fonction caractéristique de la somme  $S = X_1 + X_2$  de deux variables aléatoires  $X_1$  et  $X_2$  s'écrit :

$$\mathcal{F}(t) = E[\exp it(X_1 + X_2)] = E[\exp(itX_1) \exp(itX_2)]$$

Si deux variables aléatoires  $X_1$  et  $X_2$  sont indépendantes, les variables aléatoires  $\exp(itX_1)$  et  $\exp(itX_2)$  sont indépendantes et l'espérance du produit de ces dernières est égale au produit des espérances. Donc

$$\mathcal{F}(t) = E[\exp(itX_1)]E[\exp(itX_2)]$$

si  $X_1$  et  $X_2$  admettent pour densités de probabilité  $f_1(x_1)$  et  $f_1(x_1)$  alors  $\mathcal{F}(t)$ , transformée de fourier de la densité de probabilité de la somme  $X_1 + X_2$  est le produit des transformées de Fourier des densités de probabilité  $f_1$  et  $f_2$ . On en déduit que la densité de probabilité de  $X_1 + X_2$  est le produit de convolution :

$$f_1 * f_2(x) = \int f_1(y)f_2(x - y)dy$$

Ce résultat peut aussi est obtenu de manière élémentaire : L'ensemble des valeurs du couple

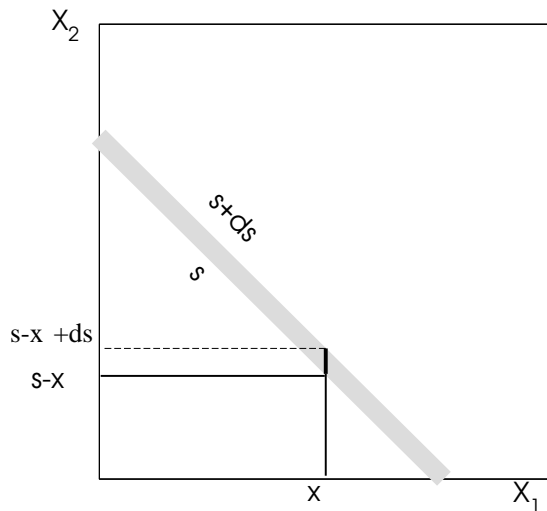


FIG. A.2 –

$U = (X_1, X_2)$  étant représenté par le plan  $(x_1, x_2)$  de la figure A.2 on a :

$$P(\{x_1 < X_1 < x_1 + dx_1\} \cap \{x_2 < X_2 < x_2 + dx_1\}) = h(x_1, x_2)dx_1dx_2$$

Puisque les variables sont indépendantes, la densité de probabilité  $h(x_1, x_2)$  du couple  $(X_1, X_2)$  est le produit des densités de probabilité des variables  $X_1$  et  $X_2$  soit :

$$P(\{x_1 < X_1 < x_1 + dx_1\} \cap \{x_2 < X_2 < x_2 + dx_1\}) = f_1(x_1)f_2(x_2)dx_1dx_2$$

On a alors :

$$P(\{s < S < s + ds\}) = \int_{[\text{aire en gris}]} f_1(x_1) f_2(x_2) dx_1 dx_2$$

soit

$$P(\{s < S < s + ds\}) = \int_{-\infty}^{\infty} f_1(x) dx \int_{s-x}^{s-x+ds} f_2(x_2) dx_2$$

Et dans la limite où  $ds \rightarrow 0$  :

$$P(\{s < S < s + ds\}) = \left[ \int_{-\infty}^{\infty} f_1(x) f_2(s-x) dx \right] ds$$

## A.13 Généralisation à N variables aléatoires

Tous les raisonnements précédents pour un couple de deux variables aléatoires se généralisent à un n-uplet  $U = (X_1, X_2, \dots, X_n)$  (ou vecteur à  $n$  dimensions) de variables aléatoires.

Nous dirons que les  $n$  variables  $X_1, X_2, \dots, X_n$  sont indépendantes si pour tout ensemble  $A_1, A_2, \dots, A_n$  de segments ouverts de  $R$ , la loi conjointe :

$$P_{(U)}(A_1 \times A_2 \times \dots \times A_n) = P_{(X_1)}(A_1) P_{(X_2)}(A_2) \dots P_{(X_n)}(A_n)$$

soit

$$\begin{aligned} P(\{X_1 \in A_1\} \cap \{X_2 \in A_2\} \cap \dots \cap \{X_n \in A_n\}) = \\ P(\{X_1 \in A_1\}) P(\{X_2 \in A_2\}) \dots P(\{X_n \in A_n\}) \end{aligned}$$

On a donc :

$$E[X_1 X_2 \dots X_n] = E[X_1] E[X_2] \dots E[X_n]$$

Et si chacune de ces lois a une densité de probabilité  $f_i(x_i)$ , alors le “n-uplet”

$$(X_1, X_2, \dots, X_n)(x_1, x_2, \dots, x_n)$$

admet pour densité de probabilité le produit :

$$h(x_1, x_2, \dots, x_n) = f(x_1) f(x_2) \dots f(x_n)$$

## A.14 Somme de variables aléatoires indépendantes

### A.14.1 Lois “faible” des grands nombres.

Soit  $X_1, X_2, \dots, X_N, \dots$  une suite de variables aléatoires réelles deux à deux indépendantes et ayant toutes la même moyenne  $E[X_i] = m$  et le même écart type fini :  $\sigma = E[X_i^2]$ . Alors la moyenne arithmétique :

$$\frac{X_1 + X_2 + \dots + X_N}{N}$$

converge “en probabilité” vers la moyenne stochastique  $m$  lorsque  $n \rightarrow \infty$ .

C’est à dire, quel que soit le nombre positif  $\epsilon$  donné (si petit soit-il) :

$$\lim_{N \rightarrow \infty} P \left( \left| \frac{X_1 + X_2 + \cdots + X_N}{N} - m \right| > \epsilon \right) = 0$$

Démonstration :

Considérons la variable aléatoire :

$$Z_N = X_1 + X_2 + \cdots + X_N$$

On a :

$$E[Z_N] = Nm, \quad \text{ou} \quad E \left[ \frac{Z_N}{N} \right] = m$$

On peut se ramener à une somme de variables aléatoires centrées :

$$Z_N - E[Z_N] = (X_1 - m) + (X_2 - m) + \cdots + (X_N - m)$$

En utilisant l’indépendance deux à deux des  $X_i$ , i.e

$$E[X_i X_j] = E[X_i] E[X_j] \quad \forall i \neq j$$

On peut écrire :

$$E[(Z_N - E[Z_N])^2] = \sum_{i=1}^N E[(X_i - m)^2] = N\sigma^2$$

Soit, en divisant par  $N^2$  :

$$E \left[ \left( \frac{Z_N}{N} - m \right)^2 \right] = \frac{\sigma^2}{N} \quad (\text{A.19})$$

En appliquant à la variable aléatoire  $(Z_N/N - m)$  l’inégalité de Bienaymé-Tchebichef, démontrée au paragraphe précédent :

$$P \left( \left| \frac{Z_N}{N} - m \right| > \epsilon \right) < \frac{\sigma^2}{N\epsilon^2} \quad (\text{A.20})$$

Quel que soit  $\epsilon$  fixé, si petit soit-il, le second membre tend vers 0 lorsque  $N \rightarrow \infty$  C.Q.F.D.

Remarque importante :

La relation (A.18) montre que l’écart type sur la variable aléatoire  $Z_N/N$  est  $\sqrt{N}$  fois plus faible que l’écart type  $\sigma$  sur la variable aléatoire  $X$

Plus généralement soit une suite de variables aléatoires  $X_1, X_2, \dots, X_N, \dots$  deux à deux indépendantes, à valeurs dans  $R^d$ , ayant toutes la même loi de probabilité qu'une variable aléatoire donnée  $X$ . Pour toute fonction  $g$  de  $R^d$  dans  $R$ , la somme

$$\frac{S_N}{N} = \frac{g(X_1) + g(X_2) + \dots + g(X_N)}{N}$$

converge “en probabilité” vers la moyenne stochastique  $\bar{g} = E[g(X)]$  lorsque  $N \rightarrow \infty$  et l'écart type sur la variable aléatoire  $S_N/N$  est  $\sqrt{N}$  fois plus faible que l'écart type :  $\sigma_g = \sqrt{E[(g(X) - \bar{g})^2]}$  sur la variable aléatoire  $Y = g(X)$

Cette généralisation est évidente :

Considérons les variables aléatoires  $Y_n = g(X_n)$ , si les  $X_n$  sont deux à deux indépendantes, alors les  $Y_n$  le sont aussi. Nous pouvons donc appliquer la loi forte des grands nombres à la suite de variables aléatoires réelles  $Y_n$  et le théorème énoncé ci-dessus est démontré.

Le type de convergence (convergence en probabilité) que nous venons de démontrer, à partir d'une inégalité triviale (de Bienaymé-Tchebichef) est “faible”.

On peut prouver une loi beaucoup plus forte, dite “loi forte des grands nombres” que nous ne redémontrons pas dans ce cours :

Pour toute suite de variables aléatoires indépendantes équidistribuées, la somme :

$$(X_1 + X_2 + \dots + X_n)/N$$

converge presque sûrement vers la moyenne  $m = E[X_1]$ .

*La convergence presque sûre de la variable aléatoire  $S_n$  vers  $S$  a lieu si et seulement si  $S_i = \phi_i(\omega)$  tend vers  $S = \phi(\omega)$  pour tout résultat  $\omega \in \Omega$ , sauf peut-être pour certains résultats  $\omega$  particuliers appartenant à un ensemble de mesure nulle (i.e. de probabilité nulle). [En théorie générale de la mesure on emploie le terme “convergence presque partout”]*

### A.14.2 Théorème de la limite centrale

La “loi des grands nombres” démontrée au paragraphe précédent nous dit que la variable aléatoire  $Z_N/N$ , somme de  $N$  variables aléatoires indépendantes tend vers la moyenne stochastique  $E[Z_N/N]$ . Par contre, l'inégalité triviale de Bienaymé-Tchebichef utilisée pour la démonstration ne donne pas une estimation précise de “l'erreur” :  $\epsilon_N = (Z_N/N - E[Z_N/N])$ . Nous pouvons le faire dans le cas où les variables aléatoires  $X_n$  sont indépendantes dans leur ensemble (la loi de probabilité de  $(X_1, X_2, \dots, X_N)$  est le produit des lois de probabilités de  $X_1, X_2, \dots, X_N$ ).

Soit une suite de variables aléatoire indépendantes dans leur ensemble :

$$X_1, X_2, \dots, X_N$$

ayant toutes la même moyenne  $m$  et un même écart type  $\sigma$  fini.

Considérons la variable aléatoire :

$$\frac{\sqrt{N}}{\sigma}\epsilon_N = \frac{Z_N}{N} - E\left[\frac{Z_N}{N}\right] = \frac{(X_1 - m) + (X_2 - m) + \dots + (X_N - m)}{\sigma\sqrt{N}}$$

Cette variable aléatoire a une moyenne nulle son écart type est l'unité. Elle est donc réduite.

Nous allons montrer que lorsque  $N \rightarrow \infty$  la loi de probabilité de  $\frac{\sqrt{N}}{\sigma}\epsilon_N$  converge “en loi” vers la “loi normale” (loi Gaussienne) réduite ayant pour densité de probabilité :

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

Nous pouvons réécrire  $\frac{\sqrt{N}}{\sigma}\epsilon_N$  sous la forme :

$$\mathcal{S}_N = \frac{\sqrt{N}}{\sigma}\epsilon_N = \frac{X_1 - m}{\sigma\sqrt{N}} + \frac{X_2 - m}{\sigma\sqrt{N}} + \dots + \frac{X_N - m}{\sigma\sqrt{N}} = Y_1 + Y_2 + \dots + Y_N$$

Les variables :

$$Y_i = \frac{X_i - m}{\sigma\sqrt{N}}$$

sont indépendantes, de moyenne nulle et de même écart type  $1/\sqrt{N}$ . A l'ordre deux leur fonctions caractéristiques ont le même développement en série de Taylor :

$$\mathcal{F}_{Y_i}(t) = E[\exp(itY_i)] = 1 + E(Y_i)it - E[Y_i^2]\frac{t^2}{2} + \dots = 1 - \frac{t^2}{2N} + \frac{1}{N}O(1/\sqrt{N})$$

la fonction caractéristique de la somme s'écrit :

$$\mathcal{F}_{\mathcal{S}_N}(t) = E[\exp(it(Y_1 + Y_2 + \dots + Y_N))] = E[\exp(itY_1)\exp(itY_2)\dots\exp(itY_N)]$$

et puisque les variables aléatoires  $\exp(itY_i)$  sont indépendantes, l'espérance du produit est le produit des espérances :

$$\begin{aligned}\mathcal{F}_{\mathcal{S}_N}(t) &= E[\exp(itY_1)]E[\exp(itY_2)]\dots E[\exp(itY_N)] = \\ &= \left[1 - \frac{t^2}{2N} + \frac{1}{N}O(1/\sqrt{N})\right]^N\end{aligned}$$

En prenant le logarithme :

$$\text{Ln}\mathcal{F}_{\mathcal{S}_N}(t) = N\text{Ln}\left[1 - \frac{t^2}{2N} + \frac{1}{N}O(1/\sqrt{N})\right] = -\frac{t^2}{2} + O(1/\sqrt{N})$$

Donc  $\text{Ln}\mathcal{F}_{\mathcal{S}_N}(t)$  tend vers  $-t^2/2$  lorsque  $N \rightarrow \infty$  et :

$$\lim_{t \rightarrow \infty} \mathcal{F}_{\mathcal{S}_N}(t) = \exp\left(-\frac{t^2}{2}\right)$$

Mais  $\exp(-\frac{t^2}{2})$  est la fonction caractéristique de la loi normale réduite. On en déduit que  $\mathcal{S}_N$  converge en loi vers la loi normale réduite.

*La convergence “en loi” est équivalente à la convergence des fonctions caractéristiques.*

## A.15 Divers modes de convergence

A travers les divers exemples que nous avons rencontrés, nous avons vu qu'en théorie des probabilités, plusieurs types de convergence, plus ou moins fortes, pour des suites  $X_n$  de variables aléatoires peuvent avoir lieu :

- Nous venons de voir la convergence “en loi” qui a lieu si les fonctions caractéristiques convergent simplement pour tout réel  $t$ .
- Nous avons vu lors de la démonstration de la loi faible des grands nombres “la convergence en probabilités” :

$$P(\{|X_n - X| > \epsilon\}) \rightarrow 0$$

- Nous avons cité lors de l'énoncé de la loi forte des grands nombres la convergence “presque sûre” (ou “presque partout”) :

$$X_n(\omega) \rightarrow X(\omega)$$

pour tout événement  $\omega$  de  $\Omega$  sauf peut-être pour  $\omega$  appartenant à un sous ensemble de  $\Omega$  de probabilité (mesure) nulle

- Citons encore la convergence en moyenne d'ordre  $p$  qui correspond à

$$\lim_{n \rightarrow \infty} \int |X_n - X|^p dP = 0$$

En particulier pour  $p = 2$  la convergence en moyenne quadratique qui correspond à :

$$\lim_{n \rightarrow \infty} E[(X_n - X)^2] = 0$$

La convergence presque sûre entraîne la convergence en probabilités, qui entraîne la convergence en loi. La convergence en moyenne d'ordre  $p$  entraîne la convergence en probabilités.







# Chaînes de Markov

## B.1 Définition

Une suite infinie *ordonnée*  $(X_1, X_2, \dots, X_t, \dots)$  de variables aléatoires constitue une chaîne de Markov si la loi de probabilité conditionnelle de  $X_{t+1}$  lorsqu'on se donne les valeurs de  $X_1, X_2, \dots, X_t$ , se réduit à la loi de probabilité conditionnelle de  $X_{t+1}$  lorsqu'on se donne seulement la valeur de  $X_t$ .

Remarque importante : pour une suite ainsi définie, la connaissance du passé influe sur l'avenir ; mais, d'un ensemble d'événements passés, seule subsiste l'influence du plus récent. Si l'on se donne  $X_t$ , la connaissance de  $X_{t'}$  pour  $t' > t$  n'est en rien précisée par la donnée des valeurs de  $X_{t-1}, X_{t-2}, \dots, X_1$ .

*Ici  $t$  désigne un entier, nous avons choisi cette notation (plutôt que  $n$ ) parce qu'en physique cette variable représente souvent un temps discret. Nous verrons, en fin de paragraphe la généralisation à une variable  $t$  continue.*

Nous nous restreindrons aux Chaînes de Markov “homogènes” où la loi de probabilité conditionnelle de  $X_{t+1}$  lorsqu'on se donne la valeur de  $X_t$  ne dépend pas de  $t$ . Alors la chaîne de Markov est entièrement déterminée par cette loi de probabilité conditionnelle appelée “*loi de transition*” et par la loi de probabilité initiale de la variable aléatoire  $X_1$ .

Nous supposons aussi que la loi de probabilité conditionnelle de  $X_{t+1}$  lorsqu'on se donne la valeur de  $X_t$  possède une densité de probabilité que nous noterons  $p(y/x)$  ou  $p(x \rightarrow y)$ , c'est à dire :

$$P(\{y < X_{t+1} < y + dy\} / \{X_t = x\}) = p(y/x)dy = p(x \rightarrow y)dy \quad (\text{B.1})$$

Nous nous contenterons d'énoncer quelques résultats sur les chaînes de Markov qui sont indispensables à la compréhension de l'algorithme de Métropolis de la Méthode de Monte-Carlo. Nous démontrerons ces résultats dans la cas discret et nous admettrons leur généralisation pour des variables continues.

## B.2 Chaînes dans un ensemble discret

### B.2.1 Propriétés

Nous nous restreignons ici au cas où les variables aléatoires  $X_t$  ne peuvent prendre que  $r$  valeurs discrètes :

$$X_t(\omega) \in \{a_1, a_2, \dots, a_r\}$$

A “l’instant”  $t$  (*discret, lui-aussi*), la loi de probabilité de la variable aléatoire  $X_t$  est définie par un vecteur ligne  $\pi^t$  dont le  $k^{ieme}$  élément est  $P(\{X_t = a_k\})$

$$\pi^t = (P(\{X_t = a_1\}), P(\{X_t = a_2\}), P(\{X_t = a_3\}), \dots, P(\{X_t = a_k\}), \dots)$$

Puisque  $\pi^t$  définit une loi de probabilité, la somme de ses composantes est 1 :

$$\sum_k (\pi^t)_k = 1$$

A la loi de probabilité conditionnelle (probabilité de transition)  $P(X_{t+1}/X_t)$  on peut associer une matrice :  $\mathbf{p}$  définie par :

$$(\mathbf{p})_{i,j} = P(\{X_{t+1} = a_j\} / \{X_t = a_i\})$$

Où chaque élément  $(\mathbf{p})_{i,j}$  représente la “probabilité de transition de l’état  $i$  à l’état  $j$ ”. Cette matrice est “non négative” (tous ses éléments sont positifs ou nuls) et elle satisfait la condition :

$$\sum_j (\mathbf{p})_{i,j} = 1 \quad \forall i$$

C’est à dire que la somme de chaque ligne est 1. On peut traduire cette propriété par la relation :

$$\mathbf{p} \cdot \mathbf{1} = \mathbf{1}$$

où  $\mathbf{1}$  est le vecteur colonne dont tous les éléments sont égaux à 1. Cette relation signifie que le nombre  $\lambda = 1$  est valeur propre de la matrice  $\mathbf{p}$  avec pour vecteur propre  $\mathbf{1}$ .

Connaissant la loi de probabilité  $\pi^t$  de la variable  $X_t$  à l’instant  $t$ , la loi de passage, donc ici la matrice  $\mathbf{p}$  permet de déterminer la loi de probabilité  $\pi^{t+1}$  de  $X_{t+1}$  à l’instant  $t + 1$  :

L’axiome des probabilités conditionnelles permet écrire :

$$P(\{X_{t+1} = a_k\} \cap \{X_t = a_j\}) = P(\{X_t = a_j\})P(\{X_{t+1} = a_k\} / \{X_t = a_j\})$$

En sommant chacun des deux membres sur tous les  $j$ , on obtient :

$$\sum_j P(\{X_{t+1} = a_k\} \cap \{X_t = a_j\}) = \sum_j P(\{X_t = a_j\})P(\{X_{t+1} = a_k\} / \{X_t = a_j\})$$

Puisque pour  $i \neq j$  les événements  $X_t = a_i$  et  $X_t = a_j$  sont incompatibles, le premier membre peut s'écrire, en vertu de la propriété d'additivité A.1

$$P(\{X_{t+1} = a_k\} \cap [\{X_t = a_1\} \cup \{X_t = a_2\} \cup \dots \cup \{X_t = a_n\} \cup \dots])$$

soit encore :

$$P(\{X_{t+1} = a_k\} \cap \Omega) = P(\{X_{t+1} = a_k\})$$

On obtient donc

$$(\pi^{t+1})_k = P(\{X_{t+1} = a_k\}) = \sum_j P(\{X_t = a_j\})P(\{X_{t+1} = a_k\} / \{X_t = a_j\}) \quad (\text{B.2})$$

soit

$$(\pi^{t+1})_k = \sum_j (\pi^t)_j (\mathbf{p})_{jk}$$

c'est à dire au sens de la multiplication d'un vecteur ligne par une matrice :

$$\pi^{t+1} = \pi^t \mathbf{p}$$

On en déduit

$$\pi^{t+\alpha} = \pi^t \mathbf{p}^\alpha$$

C'est l'équation de Chapman-Kolmogorov

### B.2.2 Rayon spectral de la matrice $\mathbf{p}$

.

Nous savons que  $\lambda = 1$  est valeur propre de la matrice de passage  $\mathbf{p}$ . Nous allons montrer que c'est la plus grande valeur propre en module de cette matrice, c'est à dire son "rayon spectral".

Si  $\lambda$  est valeur propre de la matrice  $\mathbf{p}$ , alors il existe un vecteur colonne  $\mathbf{v}$  tel que :

$$\mathbf{p} \cdot \mathbf{v} = \lambda \mathbf{v}$$

et il existe un vecteur ligne  $\mathbf{u}$  associé à cette même valeur propre :

$$\lambda \mathbf{u} = \mathbf{u} \cdot \mathbf{p}$$

soit :

$$\lambda \mathbf{u}_j = \sum_i \mathbf{u}_i \cdot \mathbf{p}_{ij} \quad \forall j$$

En prenant la valeur absolue de chaque membre, on a :

$$|\lambda| |\mathbf{u}_j| = \left| \sum_i \mathbf{u}_i \cdot \mathbf{p}_{ij} \right| \quad \forall j$$

et puisque la valeur absolue d'une somme est inférieure ou égale à la somme des valeurs absolues :

$$|\lambda| |\mathbf{u}_j| \leq \sum_i |\mathbf{u}_i| \cdot |\mathbf{p}_{ij}| \quad \forall j$$

Si maintenant nous sommions chaque membre sur  $j$ , nous obtenons

$$|\lambda| \sum_j |\mathbf{u}_j| \leq \sum_j \sum_i |\mathbf{u}_i| \cdot |\mathbf{p}_{ij}| = \sum_i \sum_j |\mathbf{u}_i| \cdot |\mathbf{p}_{ij}|$$

Mais puisque la matrice  $\mathbf{p}$  satisfait la condition :

$$\sum_j (\mathbf{p})_{i,j} = 1 \quad \forall i$$

(cf paragraphe précédent), on a alors :

$$|\lambda| \sum_j |\mathbf{u}_j| \leq \sum_i |\mathbf{u}_i|$$

donc

$$\lambda \leq 1$$

CQFD

### B.2.3 Convergence vers une loi de distribution stationnaire

Si pour un temps “suffisamment long” la loi de distribution  $\pi^t$  de la chaîne de Markov tend vers une distribution  $\pi^L$  qui devient invariante, c'est à dire indépendante du temps, on dit que la distribution  $\pi^L$  est stationnaire et on a :

$$\pi^L = \pi^L \cdot \mathbf{p}$$

C'est à dire que  $\pi^L$  est un vecteur propre (vecteur ligne) associé à la valeur propre  $\lambda = 1$

### B.2.4 Chaîne de Markov ergodique

D'après les paragraphes précédents, il existe au moins une distribution invariante  $\pi^L$ , mais elle n'est pas forcément unique. Pour cela on doit imposer une condition supplémentaire à la matrice  $\mathbf{p}$ .

#### Matrice régulière

Une matrice est non régulière s'il existe une permutation des indices qui permette de la mettre sous la forme :

$$\begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix}$$

dans ce cas :

$$(0, U) \begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix} = (0, U A_2)$$

et en partant d'un état correspondant à la partie droite  $U$  du vecteur ligne  $(0, U)$ , on reste dans ce même sous espace, sans jamais atteindre le sous espace correspondant à la partie gauche.

Une matrice est régulière si elle n'est pas décomposable sous la forme ci-dessus.

### Ergodicité

Si la matrice de passage  $\mathbf{p}$  d'une chaîne de Markov est régulière alors quel que soit le couple d'états  $\{a_i, a_j\}$  il existe un entier  $n$  tel que la probabilité de passage de  $a_i$  à  $a_j$  au bout d'un temps  $n$  soit strictement positive (i.e non nulle). Dans ce cas, tous les états sont visités au cours du temps, ce qui correspond à la propriété physique "d'ergodicité".

Le théorème mathématique de Perron-Fröbenius nous dit que pour une matrice non-négative, régulière la valeur propre correspondant à son rayon spectral (i.e la valeur propre de plus grand module) est non dégénérée, c'est à dire que le vecteur propre correspondant est unique.

Pour une chaîne de Markov ergodique, la distribution stationnaire  $\pi^L$  (i.e vecteur propre correspondant à la valeur propre maximale  $\lambda = 1$ ) est unique.

#### B.2.5 Généralisation à un ensemble d'états continus

Au lieu de raisonner sur une matrice de passage  $\mathbf{p}$  à deux dimensions, on considère la densité de probabilité  $p(x \rightarrow y)$  de la loi de transition qui est une fonction de deux variables continues, et dans les relations précédentes, démontrées dans le cas discret, on remplace les sommes discrètes par des intégrales.

On peut, en particulier faire un passage à l'intégrale pour l'équation de Chapman-Kolmogorov B.2 démontée pour un ensemble discret.

Dans le cas continu, si  $\pi^t(x)$  représente la densité de probabilité correspondant à la loi de probabilité de la variable  $X_t$ , au temps  $t + 1$ , la densité de probabilité associée à la loi de probabilité de la variable aléatoire  $X_{t+1}$  est :

$$\pi^{t+1}(x) = \int \pi^t(u) p(u \rightarrow x) du \quad (\text{B.3})$$

Cette relation représente l'équation de Chapman-Kolmogorov, pour un ensemble d'états continu.





# Rappels de Physique statistique

---

*Les rappels de ce chapitre sont tirés de l'excellent livre de Diu et al. [20]*

## C.1 Entropie statistique

### C.1.1 Définition de Shannon

Soit  $X$  une variable aléatoire discrète, sur un ensemble fini, avec pour loi de probabilité :

$$p(x_i) = P(\{X = x_i\})$$

l'entropie est définie par :

$$S[X] = -E[\log P(X)] = -\sum_{i=1}^M p(x_i) \log[p(x_i)]$$

avec la convention :  $0 \log(0) = 0$ . En théorie de l'information, on utilise en général le logarithme en base 2.

Elle mesure le “manque d'information” et possède les propriétés suivantes :

1. Elle est positive, complètement symétrique
2. Les événements de probabilité nulle ne contribuent pas
3. Elle atteint son minimum  $S = 0$  si l'un des événements est réalisé avec une probabilité 1.
4. Elle atteint son maximum lorsque tous les événements sont équiprobables
5. Additivité : considérons deux variables aléatoire indépendantes  $X, Y$  et la loi de probabilité du couple :

$$p_{x_m, y_n} = P(X = x_m, Y = y_n)$$

Alors l'entropie du couple est :

$$S = -\sum_{m,n} p_{x_m, y_n} \log[p_{x_m, y_n}]$$

Si  $X$  et  $Y$  sont indépendantes, alors  $p_{x_m, y_n} = p(x_m)p(y_n)$  et il est facile de montrer que :

$$S = -\left\{\sum_m p(x_m) \log[p(x_m)] + \sum_n p(y_n) \log[p(y_n)]\right\}$$

c'est à dire :

$$S[(X, Y)] = S[X] + S[Y]$$

Le manque d'information sur le couple est la somme des manques d'information sur chaque variable.

Plus généralement, si  $X$  et  $Y$  ne sont pas indépendantes, on peut montrer la propriété

$$S[(X, Y)] = S[X] + S[\{Y/X\}]$$

en faisant intervenir la probabilité conditionnelle  $P(\{Y/X\})$

### C.1.2 Théorème de Khinchin (1957).

Khinchin a montré que la seule fonction qui satisfasse aux propriétés 1-5 est :

$$S = -k \sum_{i=1}^M p(x_i) \ln[p(x_i)]$$

On a le choix de la constante  $k$  : en thermodynamique,  $k$  est la constante de Boltzman et le logarithme est en base e (logarithme Népérien)

## C.2 Systèmes macroscopiques à l'équilibre

### C.2.1 Ensemble micro-Canonique

On considère un système isolé.

Son état est déterminé par un ensemble de paramètres extérieurs :

- Son énergie  $E$  définie à  $\delta E$  près
- Son volume  $V$
- Son nombre de particules  $N$
- etc ...

Le système est caractérisé par ses états microscopiques. A chaque état  $l$  correspond une probabilité  $P_l$ . Pour un système isolé à l'équilibre, puisqu'on ne connaît aucune information plus précise sur ce système, on postule que tous les états microscopiques accessibles sont équiprobables (—entropie maximale—) :

$$P_l = 1/\Omega \quad \text{si} \quad E < E_l < E + \delta E \quad (\text{C.1})$$

$$P_l = 0 \quad \text{sinon}$$



$\Omega$  représente le nombre d'états du système et on a :

$$\sum_l P_l = 1$$

Son entropie est donc maximale :

$$S^{\mu C}(E, V, N, \dots) = k \ln(\Omega) \quad (\text{C.2})$$

Cette relation est l'effigie inscrite sur la tombe de Boltzman Cette hypothèse d'équipartition correspond donc au manque d'information maximal sur le système !

On peut considérer diverses dérivées par rapport aux paramètres extérieurs :

– La température microcanonique  $T^{\mu C}$  :

$$\frac{1}{T^{\mu C}} = \partial S^{\mu C}(E, N, V, \dots) / \partial E \quad (\text{C.3})$$

– Le potential chimique microcanonique :

$$\frac{\mu^{\mu C}}{T^{\mu C}} = -\partial S^{\mu C}(E, N, V, \dots) / \partial N \quad (\text{C.4})$$

– La pression microcanonique

$$\frac{P^{\mu C}}{T^{\mu C}} = \partial S^{\mu C}(E, N, V, \dots) / \partial V \quad (\text{C.5})$$

### C.2.2 Ensemble Canonique

On considère un système  $\mathcal{S}$  en contact avec un système beaucoup plus grand : réservoir de chaleur ou thermostat  $\mathcal{T}$ .

L'union de ces deux systèmes constitue un système isolé et son énergie totale :

$$E_{\mathcal{S}} + E_{\mathcal{T}} = E_{tot}$$

est fixée à  $\delta E$  près

$\mathcal{T}$  étant beaucoup plus grand que  $\mathcal{S}$  on a toujours  $E_{\mathcal{S}} \ll E_{\mathcal{T}}$ , et on peut considérer la température microcanonique du thermostat  $T_{\mathcal{T}}^{\mu C}$  comme indépendante de l'énergie du petit système  $\mathcal{S}$  et donc constante.

La température canonique  $T^c$  du petit système  $\mathcal{S}$  sera par définition la température microcanonique du thermostat.

$$T^c = T_{\mathcal{T}}^{\mu C}$$

Le système  $\mathcal{SUT}$  étant isolé, on peut lui appliquer les postulats de l'ensemble microcanonique. L'union des deux systèmes est caractérisée par l'ensemble des couples états  $(l, L)$  d'énergie

$$E_{l,L} = E_l + E_L$$

tels que :

$$E_{tot} \leq E_{l,L} \leq E_{tot} + \delta E$$

Tous les états microcanoniques sont équiprobables, et la probabilité  $P_l^C$  de trouver le système dans l'état d'énergie  $l$  est égale au rapport du nombre d'états  $\Omega_{\mathcal{T}}(E_{\mathcal{T}} = E_{total} - E_l)$  du thermostat d'énergie  $E_{\mathcal{T}} = E_{tot} - E_l$  sur le nombre total d'états de  $\mathcal{SUT}$  :  $\Omega_{\mathcal{SUT}}(E_{tot}) = \sum_l \Omega_{\mathcal{T}}(E_{\mathcal{T}} = E_{total} - E_l)$

$$P_l^C = \Omega_{\mathcal{T}}(E_{tot} - E_l) / \sum_i \Omega_{\mathcal{T}}(E_{tot} - E_i)$$

$\Omega_{\mathcal{T}}(E_{tot} - E_l)$  s'exprime en fonction de l'entropie microcanonique du Thermostat :

$$\Omega_{\mathcal{T}}(E_{tot} - E_l) = \exp[S_{\mathcal{T}}^{\mu C}(E_{tot} - E_l)/k]$$

Mais l'énergie  $E_l$  est toujours faible devant l'énergie du grand système  $T$ , et un développement au premier ordre donne :

$$S_{\mathcal{T}}^{\mu C}(E_{total} - E_l) = S_{\mathcal{T}}^{\mu C}(E_{total}) - E_l \frac{\partial S}{\partial E}(E_{tot})$$

soit

$$S_{\mathcal{T}}^{\mu C}(E_{total} - E_l) = S_{\mathcal{T}}^{\mu C}(E_{total}) - E_l \frac{1}{kT}$$

Ce qui conduit à

$$P_l^C = \exp[-E_l/kT]/Z \quad (\text{C.6})$$

avec

$$Z(T, N, V, \dots) = \sum_i \exp[-E_i/kT] \quad (\text{C.7})$$

On définit le potentiel thermodynamique, ou énergie libre :

$$F = -kT \ln Z(T, N, V, \dots) \quad (\text{C.8})$$

– La valeur moyenne de l'énergie est :

$$\bar{E} = \sum_i E_i \exp[-E_i/kT]/Z = -\frac{\partial}{\partial \beta} \ln Z(T, N, V, \dots) \quad (\text{C.9})$$

avec  $\beta = 1/T$

– La chaleur spécifique à volume constant est :

$$C_v = \frac{\partial \bar{E}}{\partial T} \quad (\text{C.10})$$

– L'entropie Canonique est :

$$S^C(T, V, N, \dots) = -k \sum_i P_i \ln[P_i] = -k(\ln Z - \bar{E}/kT) \quad (\text{C.11})$$

Soit

$$S^c(T, V, N, \dots) = -k \sum_l \frac{\exp(-E_l/kT)}{Z} \left( -\frac{E_l}{kT} - \ln(Z) \right)$$

d'où

$$TS^c(T, V, N, \dots) = \bar{E} + kT \ln(Z)$$

ce qui permet d'écrire l'énergie libre sous la forme habituelle :

$$F = \bar{E} - TS^C \quad (\text{C.12})$$

– Le potentiel chimique canonique est :

$$\mu^C = \frac{\partial F}{\partial N} \quad (\text{C.13})$$

– La pression canonique est

$$P^C = \frac{\partial F}{\partial V} \quad (\text{C.14})$$

### C.2.3 Ensemble Grand Canonique

On considère un petit système  $\mathcal{S}$  en contact avec un réservoir  $\mathcal{R}$  de chaleur et de particules. L'union des deux systèmes  $\mathcal{S}$  et  $\mathcal{R}$  constitue un système isolé auquel on peut appliquer les principes microcanoniques.

On procède comme pour l'ensemble Canonique, mais ici un état du système  $\mathcal{S}$  est caractérisé par son énergie  $E_l$  et son nombre de particules  $N_l$ .

La probabilité de trouver le système  $\mathcal{S}$  dans un état d'énergie  $E_l$ , avec un nombre de particules  $N_l$  est proportionnelle au nombre d'état du réservoir d'énergie  $E_{\mathcal{R}} = E_{tot} - E_l$  et de nombre de particules  $N_{\mathcal{R}} = N_{tot} - N_l$  soit :

$$P_l^{GC} = c \Omega(E_{\mathcal{R}} = E_{tot} - E_l, N_{\mathcal{R}} = N_{tot} - N_l)$$

$$P_l^{GC} = c \Omega \exp[S_{\mathcal{R}}^{\mu C}(E_{\mathcal{R}} = E_{tot} - E_l, N_{\mathcal{R}} = N_{tot} - N_l)/k]$$

Et en développant  $S_{\mathcal{R}}^{\mu C}$  au premier ordre en  $E_l$  et  $N_l$  :

$$S_{\mathcal{R}}^{\mu C} = S_{\mathcal{R}}^{\mu C}(E_{\mathcal{R}}, N_{\mathcal{R}}) - \frac{\partial S}{\partial E} E_l - \frac{\partial S}{\partial N} N_l$$

$$S_{\mathcal{R}}^{\mu C} = S_{\mathcal{R}}^{\mu C}(E_{\mathcal{R}}, N_{\mathcal{R}}) - \frac{1}{T} E_l + \frac{\mu}{T} N_l$$

On deduit :

$$P_l^{GC} = \exp\left[-\frac{E_l - \mu N_l}{kT}\right] / \mathcal{Z} \quad (\text{C.15})$$

avec :

$$\mathcal{Z} = \sum_l \exp\left[-\frac{E_l - \mu N_l}{kT}\right] \quad (\text{C.16})$$

On définit le Grand Potentiel thermodynamique :

$$G = -kT \ln \mathcal{Z} \quad (\text{C.17})$$

En regroupant les termes correspondant à un même nombre de particules  $N_l = n$ , on peut écrire :

$$\mathcal{Z} = \sum_{n=1}^N [\exp(\mu/kT)]^n \sum_{l \in \{N_l=n\}} \exp(-E_l/kT)$$

où la seconde somme porte sur le sous ensemble d'états d'énergie  $E_l$  correspondant à un même nombre de particules  $N_l = n$ . Ce n'est autre que la fonction de partition canonique  $Z_n$  d'un système à  $n$  particules.

$$\mathcal{Z} = \sum_{n=1}^N [\exp(\mu/kT)]^n Z_n$$

La valeur moyenne du nombre de particules est

$$\bar{N} = \sum_l N_l \exp[-\frac{E_l - \mu N_l}{kT}] / \mathcal{Z}$$

soit

$$\bar{N} = kT \frac{\partial \ln(\mathcal{Z})}{\partial \mu} = -\partial G / \partial \mu$$

## C.3 Evolution vers l'équilibre

### C.3.1 Probabilité de transition

On postule que la probabilité pour qu'un système qui se trouve à l'état  $l$  à l'instant  $t_0$  se retrouve à l'état  $m$  à l'instant  $t_0 + dt$  ( $dt$  long devant le temps caractéristique des transitions microscopiques, mais court à l'échelle macroscopiques est :

- Proportionnel à  $dt$
- indépendant de  $t_0$

$$\mathcal{P}_{ml}(t_0, t_0 + dt) = a_{ml} dt \quad (\text{C.18})$$

### C.3.2 Equation maîtresse

A un instant  $t$ , état macroscopique d'un système est caractérisé par l'ensemble des probabilités  $P_l(t)$  pour qu'il se trouve dans chacun de ses états microscopiques  $l$ .

Soit  $dP_l$  l'évolution de  $P_l$  entre les instants  $t$  et  $t + dt$ . On peut écrire intuitivement :

$$dP_l = \sum_m a_{lm} P_m(t) dt - \sum_m a_{ml} P_l(t) dt \quad (\text{C.19})$$

On postule "l'équation maîtresse suivante :

$$\frac{dP_l}{dt} = \sum_m [a_{lm} P_m(t) - a_{ml} P_l(t)] \quad (\text{C.20})$$

### C.3.3 Système isolé

– Les  $a_{lm}$  ne relient que des états de même énergie :

$$a_{lm}^{\mu C} = 0 \quad \text{si} \quad |E_l - E_m| > \delta E$$

– Les  $a_{lm}$  sont symétriques

$$a_{lm}^{\mu C} = a_{ml}^{\mu C}$$

### C.3.4 Système $\mathcal{S}$ en contact avec un thermostat $\mathcal{T}$

On considère l'union  $\mathcal{SUT}$  comme système isolé.

La probabilité  $P_{l,L}(t)$  pour que  $\mathcal{S}$  soit dans l'état  $l$  et  $\mathcal{T}$  dans l'état  $L$  obéit à l'équation maîtresse :

$$\frac{dP_{l,L}}{dt} = \sum_{m,M} [a_{(l,L)(m,M)}^{\mu C} P_{m,M}(t) - a_{(m,M)(l,L)}^{\mu C} P_{l,L}(t)]$$

La probabilité  $P_l$  pour que le système  $\mathcal{S}$  soit dans l'état  $l$  est

$$P_l = \sum_L P_{l,L}$$

On a

$$P_{l,L} = P_l P_L^{\mu C}(E_{tot} - E_l)$$

En substituant dans la relation :

$$\frac{dP_l}{dt} = \sum_L \frac{dP_{l,L}}{dt}$$

on obtient l'équation maîtresse :

$$\frac{dP_l}{dt} = \sum_m [a_{lm}^C P_m(t) - a_{m,l}^C P_l(t)] \quad (\text{C.21})$$

avec

$$a_{lm}^C = \sum_{M,L} a_{(l,L)(m,M)}^{\mu C} P_M^{\mu C}(E_{tot} - E_m)$$

$$a_{ml}^C = \sum_{M,L} a_{(m,M)(l,L)}^{\mu C} P_M^{\mu C}(E_{tot} - E_l)$$

on a

$$P_M^{\mu C}(E_{tot} - E_l) = \frac{1}{\Omega(E_{tot} - E_l)} = \exp[-S^{\mu C}(E_{tot} - E_l)]$$

et en développant  $S^{\mu C}$  au premier ordre :

$$S^{\mu C}(E_{tot} - E_l) = S^{\mu C}(E_{tot}) - \frac{\partial S^{\mu C}}{\partial E} E_l$$

on en déduit :

$$a_{lm}^C \exp(-E_m/kT) = a_{ml}^C \exp(-E_l/kT) \quad (C.22)$$

### C.3.5 Propriétés à l'équilibre. Relation du bilan détaillé.

A l'équilibre :

$$dP_l/dt = 0 \quad \forall l$$

L'équation maîtresse s'écrit alors :

$$\sum_m [a_{lm} P_m^e - a_{ml} P_l^e] = 0$$

L'équation ci dessus, avec les contraintes :

$$P_l^e \geq 0 \quad \forall l$$

et

$$\sum_l P_l^e = 1$$

admet pour seule solution :

- La distribution microcanonique pour un système isolé
- La distribution canonique pour un système en contact avec un thermostat

1. Pour un système isolé :

$$P^{\mu C}(E) = 1/\Omega(E) \quad E \leq E_l \leq E + \delta E$$

$$P^{\mu C}(E) = 0 \quad \text{sinon}$$

On a donc :

$$P_l^{\mu C}(E) = P_m^{\mu C}(E)$$

Puisque  $a_{lm}^{\mu C} = a_{ml}^{\mu C}$ , on a une relation terme à terme beaucoup plus forte, dite **relation du bilan détaillé** :

$$[a_{lm} P_m - a_{ml} P_l] = 0 \quad \forall l, \quad \forall m \quad (C.23)$$

On montre que c'est la seule solution d'équilibre.

2. Pour un système en contact avec un thermostat :

$$P_l^C = \exp(-E_l/kT)$$

et

$$a_{lm}^C \exp(-E_m/kT) = a_{ml}^C \exp(-E_l/kT)$$

Et on a aussi la relation du bilan détaillé :

$$[a_{lm}P_m - a_{ml}P_l] = 0 \quad \forall l, \quad \forall m$$







# Valeurs moyennes d'observables

---

## D.1 Définition

### Système quantique

Un système quantique est décrit par son Hamiltonien  $H$ . Ses différents états sont les vecteurs propres  $|n\rangle$  de  $H$ , avec pour énergies  $E_n$  les valeurs propres correspondantes :

$$H|n\rangle = E_n|n\rangle$$

Une grandeur physique, ou “observable” est caractérisée par un opérateur  $\mathcal{O}$  et la mesure de cette grandeur, lorsque le système est dans un état  $|n\rangle$  est

$$\langle n|\mathcal{O}|n\rangle$$

Dans l'ensemble Canonique, la fonction de partition est, d'après les rappels précédents :

$$Z = \sum_n \exp(-\beta E_n)$$

avec  $\beta = 1/kT$ .

On peut écrire en considérant une base d'états propres  $|n\rangle$

$$Z = \sum_n \langle n|\exp(-\beta H)|n\rangle$$

$Z$  apparaît donc comme la trace de l'opérateur  $\exp(-\beta H)$  et on sait que la trace d'un opérateur ne dépend pas de la base choisie. Donc on écrit de manière plus générale :

$$Z = \text{Tr}[\exp(-\beta H)] \quad (\text{D.1})$$

La valeur moyenne ou “espérance d'un opérateur  $\mathcal{O}$  décrit, dans l'ensemble canonique (cf. rappels précédents) :

$$\langle \mathcal{O} \rangle = \frac{\sum_n \langle n|\mathcal{O}|n\rangle \exp(-\beta E_n)}{Z}$$

soit

$$\langle \mathcal{O} \rangle = \frac{\sum_n \langle n | \mathcal{O} \exp(-\beta E_n) | n \rangle}{Z}$$

ou encore :

$$\langle \mathcal{O} \rangle = \frac{\sum_n \langle n | \mathcal{O} \exp(-\beta H) | n \rangle}{Z}$$

On voit apparaître la trace de l'opérateur  $\mathcal{O} \exp(-\beta H)$ , indépendante du choix de la base.

On écrira :

$$\langle \mathcal{O} \rangle = \frac{\text{Tr}[\mathcal{O} \exp(-\beta H)]}{Z} \quad (\text{D.2})$$

### Système classique

Pour un système classique, décrit par un Hamiltonien classique :  $H(p, x)$ , la valeur moyenne d'une grandeur physique, ou "observable"  $\mathcal{O}$  (par exemple : énergie, aimantation pour un système magnétique, etc...) est, dans l'Ensemble Canonique[21, 22, 23, 20] :

$$\langle \mathcal{O} \rangle = \frac{\int \mathcal{O}(x) e^{-\beta H(p, x)} dp dx}{Z} \quad (\text{D.3})$$

où  $Z$  représente la fonction de partition :

$$Z = \int e^{-\beta H(p, x)} dp dx \quad (\text{D.4})$$

On note  $\beta = 1/k_B T$ ,  $x = (\alpha_1, \alpha_2, \dots, \alpha_{3N})$  représente les coordonnées des particules et  $p = (p_1, p_2, \dots, p_{3N})$  les variables conjuguées ou impulsions :  $p_i = d\alpha_i/dt$ . L'Hamiltonien  $H$  du système s'écrit :

$$H = \sum_i \frac{p_i^2}{2m_i} + E(x) \quad (\text{D.5})$$

Le premier terme représente l'énergie cinétique du système et le second terme  $E(x)$  l'énergie potentielle.

Dans ce cas, les intégrales sur les  $p_i$  se séparent et se simplifient entre numérateur et dénominateur et on peut écrire plus simplement :

$$\langle \mathcal{O} \rangle = \frac{\int \mathcal{O}(x) e^{-\beta E(x)} dx}{Z} \quad (\text{D.6})$$

avec :

$$Z = \int e^{-\beta E(x)} dx \quad (\text{D.7})$$

La fonction :

$$f(x) = e^{-\beta E(x)} / Z \quad (\text{D.8})$$

est positive et a pour intégrale :  $\int f(x)dx = 1$ , elle peut donc représenter la densité de probabilité d'une variable aléatoire. Elle est grande dans les portions de l'espace qui nous intéressent pour le calcul de la valeur moyenne de toute observable  $\mathcal{O}$  et décroît rapidement en dehors. Elle constitue donc le meilleur choix pour un "échantillonnage suivant l'importance".

Il subsiste toutefois un problème important. Dans la définition de  $f$ , nous ne connaissons pas la constante de normalisation  $Z$ , c'est à dire la fonction de partition. Il est donc impossible d'échantillonner directement  $f$ .

Nous devons donc avoir recours à un autre procédé d'échantillonnage. Nous utiliserons pour outil des *chaînes de Markov*.





# Bibliographie

---

- [1] D.E. Knuth, Seminumerical Algorithms, 2nd. ed. vol. 2 of The Art of Computer Programming (Reading, Mass. : Addison-Wesley), (1981).
- [2] D.H. Lehmer, Proc. 2nd Symp, Large-Scale Digital Calculating machinery (Harvard Univ. Press, Cambridge, MA, 1951) pages : 141-146
- [3] W. H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes (Cambridge University Press, New-York) (1986).
- [4] G.J. Mitchell and D.P. Moore (1958) unpublished.
- [5] Heringa, Blöte, and Compagner, Int. J. Mod. Phys. C3, 561 (1992) et références citées.
- [6] R. C. Taushworthe, Math. Comput. 19, 201 (1965).
- [7] J.M. Hammersley et D.C Handscomb, Les Méthodes de Monte-Carlo (Ed. Dunod, Paris), (1967).
- [8] J.W Negele et H. Orland, Quantum Many-Particle System, Chapitre 8 : “Stochastic methods”, (Ed. Addison-Wesley, New-York) (1988).
- [9] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller et E. Teller, J. Chem. Phys. **21**, 1087 (1953).
- [10] L. Verlet, Phys. Rev. **159**, 98 (1967).
- [11] M.P. Allen et D.J. Tildesley, Computer Simulation of Liquids, Clarendon Press, Oxford (1987).
- [12] J.B. Kogut, J. of Stat. Phys. **43**, 771 (1986).
- [13] A.D. Kennedy, Nuclear Physics B (Proc. Suppl.) 4, 576 (1988).
- [14] M. Rosenbluth, A Rosenbluth, J. Chem. Phys. **23**, 356 (1954).
- [15] D. Frenkel, G. Mooij, B. Smith, J. Phys. Cond. Mat. **3**, 3053 (1991).
- [16] R.P. Feynman and A.R. Hibbs, Quantum Mechanics and Path Integrals, Ed. Mc Graw-Hill (1965).
- [17] D.M. Ceperley, Rev. Mod. Phys. **67**, 279 (1995).  
Voir aussi les cours de Ceperley, accessibles gratuitement sur internet :  
[http ://people.physics.uiuc.edu/Ceperley/papers/163.pdf](http://people.physics.uiuc.edu/Ceperley/papers/163.pdf)  
[http ://www.phys.uri.edu/nigh/QMC-NATO/webpage/abstracts/ceperley.ps](http://www.phys.uri.edu/nigh/QMC-NATO/webpage/abstracts/ceperley.ps)

- [18] J. Bass, *Eléments de calcul des probabilités ; théorique et appliqué*. Ed. Masson [Paris] (1967).
- [19] J. Neveu, *Cours de probabilités : Ecole Polytechnique* (1974) ; même auteur : *Probabilités*, Edition 1995, Ecole Polytechnique [Palaiseau, France].
- [20] B. Diu, C. Guthmann, D. Lederer, B. Roulet, *Physique Statistique*, Hermann Paris (1989).
- [21] R. Balian, *Du microscopique au macroscopique*, Cours de Physique statistique de l'Ecole polytechnique, tomes 1 et 2. Ellipses (1982).
- [22] L. Landau et E. Lifschitz, *Physique Statistique*, Editions Mir, Moscou (1967).
- [23] F. Reif, *Fundamentals of Statistical and Thermal Physics*, McGraw-Hill, New York (1965).