

# Compte rendu stage Simulation parfaite

Antonin LAURENT

26 mai 2019

## Table des matières

<b>1</b>	<b>Chapitre 1 : Introduction</b>	<b>3</b>
1.1	Définitions et théorèmes . . . . .	3
1.2	Modèles étudiés . . . . .	6
1.2.1	Champs aléatoires de Markov . . . . .	6
1.2.2	Permutations . . . . .	8
1.3	Chaînes de Markov et simulation approchée . . . . .	9
1.4	Création de chaînes de Markov . . . . .	10
1.4.1	Échantillonnage de Gibbs . . . . .	11
1.4.2	Metropolis-Hastings . . . . .	17
1.4.3	Variables aléatoires auxiliaires / Auxiliary random variables . . . . .	22
1.4.4	Shift Chains . . . . .	23
<b>2</b>	<b>Chapitre 2 : Procédures de rejet</b>	<b>24</b>
2.1	Théorème et exemple simple . . . . .	24
2.2	Méthode de rejet pour variables aléatoires à densité . . . . .	24
2.2.1	Exemple : D'une Cauchy vers une Normale . . . . .	26
2.3	Union d'ensembles . . . . .	28
2.4	Simulation à l'aide des inégalités de Markov et Chernoff . . . . .	29
2.4.1	L'inégalité de Markov en tant que procédure de rejet . . . . .	29
2.4.2	Les inégalités de Chernoff en tant que procédure de rejet . . . . .	30
<b>3</b>	<b>Chapitre 3</b>	<b>32</b>

# 1 Chapitre 1 : Introduction

## 1.1 Définitions et théorèmes

### Définition 1. Temps d'arrêt

On dit que  $T$  est un temps d'arrêt pour une suite  $X_1, X_2, \dots$  si, connaissant les valeurs de  $X_1, X_2, \dots, X_n$  on peut déterminer si  $T \leq n$ .

### Définition 2. Fonction calculable

Une fonction est dite calculable (computable en anglais) si il existe un algorithme capable de retourner le résultat de la fonction.

### Définition 3. Algorithme probabiliste

Soient  $\mathcal{I}$  un ensemble d'indices tel que pour tout  $I \in \mathcal{I}$ , il existe une distribution  $\pi_I$  sur un espace d'états  $\Omega_I$ . On se donne une suite de variables aléatoires  $X_1, X_2, \dots$  où les  $X_I \in S_I$ .

Un algorithme probabiliste est une famille de temps d'arrêts  $\{T_I\}_{I \in \mathcal{I}}$  et de fonctions calculables  $\{f_{I,t}\}_{I \in \mathcal{I}, t \in \{1,2,\dots\}}$ . La sortie de l'algorithme est  $f_{I,T_I}(X_1, \dots, X_{T_I})$ .

### Définition 4. Algorithme de simulation parfaite

Un algorithme de simulation parfaite est un algorithme probabiliste dont la sortie est une variable aléatoire qui provient d'une distribution cible. Les algorithmes de simulation parfaite sont une sous-classe des algorithmes de simulation exacte, algorithmes qui tirent d'une distribution ciblée. Cependant les algorithmes dont le temps d'arrêt  $T_I$  est déterministe (algorithmes tournant selon un nombre fini de choix aléatoires) sont généralement considérés comme algorithmes de simulation exacte mais pas comme algorithmes de simulation parfaite.

### Théorème 1. Théorème fondamental de la simulation parfaite

On suppose que pour  $U_1, U_2, \dots$  iid tel que  $U_i \sim \text{Unif}([0,1])$ , il existe des fonctions calculables  $b, g$  et  $f$  telles que la fonction  $b$  aie pour image  $\{0, 1\}$  et  $\mathbb{P}(b(U) = 1) > 0$ .

Pour une variable aléatoire  $X$  qui vérifie :

$$X \sim b(U)g(U) + (1 - b(U))f(X, U), \quad (1)$$

soit  $T = \inf\{t : b(U_t) = 1\}$ . On a alors que :

$$Y = f(\dots f(f(g(U_T), U_{T-1}), U_{T-2}), \dots, U_1)$$

a la même distribution que  $X$  et on a  $\mathbb{E}[T] = \frac{1}{\mathbb{P}(b(U)=1)}$ .

*Démonstration.* Soient  $X_0, X_1, \dots$  des tirages indépendants chacun distribués selon  $X$ , et  $U_1, U_2, \dots$  iid tel que  $U_i \sim \text{Unif}([0,1])$ . Pour  $X_t$ , fixons  $X_{t,t} = X_t$  et récursivement, on pose :

$$X_{t,i} = b(U_{i+1})g(U_{i+1}) + (1 - b(U_{i+1}))f(X_{t,i+1}, U_{i+1}),$$

pour  $i \in \{0, \dots, t-1\}$ .

On a alors d'après la relation (1) :  $X_{t,0} \sim X$ . On montre ce résultat pour les premiers indices  $t$ , les suivants sont prouvés de la même manière.

**t=0**

$$X_{0,0} = X_0$$

comme on l'a posé précédemment, d'où

$$X_{0,0} \sim X.$$

**t=1**

$$X_{1,0} = b(U_1)g(U_1) + (1 - b(U_1))f(X_{1,1}, U_1)$$

Or,

$$X_{1,1} = X_1 \sim X.$$

En remplaçant dans l'équation précédente, on obtient :

$$X_{1,0} = b(U_1)g(U_1) + (1 - b(U_1))f(X_1, U_1)$$

D'après la relation (1), on obtient alors  $X_{1,0} = X_1 \sim X$ .

**t=2**

$$\begin{aligned} X_{2,0} &= b(U_1)g(U_1) + (1 - b(U_1))f(X_{2,1}, U_1) \\ X_{2,1} &= b(U_2)g(U_2) + (1 - b(U_2))f(X_{2,2}, U_2) \\ &= b(U_2)g(U_2) + (1 - b(U_2))f(X_2, U_2) \end{aligned}$$

À nouveau, par la relation (1), on a :  $X_{2,1} = X_2$  On remplace donc par cette valeur dans  $X_{2,0}$  et on obtient :

$$X_{2,0} = b(U_1)g(U_1) + (1 - b(U_1))f(X_2, U_1)$$

D'après la relation (1), on obtient alors  $X_{2,0} = X_2 \sim X$ .

On a donc  $X_{0,0}, X_{1,0}, X_{2,0}, \dots$  de même distribution que  $X$  mais pas forcément indépendants. On considère à présent la variable  $Y$  énoncée dans le théorème. On a alors la relation suivante :  $X_{t,0} = Y$  si  $t \geq T$ . On illustre ce résultat avec comme exemple  $T = 2$  et  $t = 2, 3$  :

On a donc :

$$Y = f(g(U_2), U_1)$$

Montrons que  $X_{2,0} = X_{3,0} = Y$ . Étant donné que  $T = 2$ , on a  $b(U_1) = 0$  et  $b(U_2) = 1$ , d'où :

$$\begin{aligned} X_{2,0} &= f(X_{2,1}, U_1) \\ X_{2,1} &= g(U_2) \end{aligned}$$

on remplace dans la première équation et on obtient

$$X_{2,0} = f(g(U_2), U_1)$$

D'où le résultat pour  $t=2$ . Voyons pour  $t=3$ .

De même, puisque  $T = 2$ , on a  $b(U_1) = 0$  et  $b(U_2) = 1$ , d'où :

$$\begin{aligned} X_{3,0} &= f(X_{3,1}, U_1) \\ X_{3,1} &= g(U_2) \end{aligned}$$

on remplace dans la première équation et on obtient

$$X_{3,0} = f(g(U_2), U_1)$$

D'où le résultat.

Ensuite, puisque les  $U_i$  sont indépendants, on a que :  $\mathbb{P}(T > t) = (1 - \mathbb{P}(b(U) = 1))^t$  et puisque, par hypothèse,  $\mathbb{P}(b(U) = 1) > 0$ , on a la relation

qui tend vers 0 pour  $t$  tendant vers l'infini. Il ne reste plus qu'à montrer  $Y \sim X$ . Pour tout  $t$ , pour tout ensemble  $C$  :

$$\mathbb{P}(Y \in C) = \mathbb{P}(Y \in C, t \geq T) + \mathbb{P}(Y \in C, t < T)$$

puisque  $X_{t,0} = Y$  si  $t \geq T$ , on a

$$\begin{aligned} &= \mathbb{P}(X_{t,0} \in C, t \geq T) + \mathbb{P}(Y \in C, t < T) \\ &= \mathbb{P}(X_{t,0} \in C) - \mathbb{P}(X_{t,0} \in C, t < T) + \mathbb{P}(Y \in C, t < T) \end{aligned}$$

puisque  $X_{t,0} \sim X$ , on a

$$= \mathbb{P}(X \in C) - \mathbb{P}(X_{t,0} \in C, t < T) + \mathbb{P}(Y \in C, t < T)$$

Les deux derniers termes sont bornés par  $\mathbb{P}(T > t) = (1 - \mathbb{P}(b(U) = 1))^t$ , et puisque l'équation est vraie pour n'importe quel  $t$ , on obtient :  $\mathbb{P}(Y \in C) = \mathbb{P}(X \in C)$  pour tout ensemble  $C$ , donc  $Y \sim X$ . Le fait que  $\mathbb{E}[T] = \frac{1}{\mathbb{P}(b(U)=1)}$  provient du fait que  $T$  suit une loi géométrique de paramètre  $\mathbb{P}(b(U) = 1)$

□

**Définition 5.** *Algorithme arretable*

Dans le théorème fondamental de la simulation parfaite, si  $X$  et  $T$  sont indépendants, on dit que l'algorithme est arretable, sinon il est non-arretable.

**Définition 6.** *Algorithme de simulation parfaite à lecture unique*

Un algorithme de simulation parfaite qui utilise  $X \sim b(U)g(U) + (1 - b(U))f(X, U)$  est un algorithme à lecture unique si  $f(X, u) = f(X, u')$  pour tout  $u, u'$ . Sinon, c'est un algorithme à lecture double.

En général, un algorithme arretable est préférable à un algorithme non-arretable, et un algorithme à lecture unique est préférable à un algorithme à lecture double.

## 1.2 Modèles étudiés

### 1.2.1 Champs aléatoires de Markov

On considère un graphe  $G = (V, E)$ .  $V$  est l'ensemble des nœuds et  $E$  est l'ensemble des arêtes. On notera par la suite  $\Delta$  le degré du graphe (égal au degré du nœud ayant un nombre de voisins maximal). On notera par  $\Omega$  l'ensemble des labels des nœuds.

**Définition 7. Ensemble séparant**

On dit qu'un sous-ensemble de nœuds  $S$  sépare les nœuds  $i$  et  $j$  si tout chemin du graphe menant  $i$  à  $j$  passe par  $S$ .

**Définition 8. Champ aléatoire de Markov**

Pour un graphe  $G = (V, E)$  et un ensemble de labels  $\Omega$ , on dit que la distribution de  $X$  sur  $\Omega$  est un champ aléatoire de Markov si, pour tous les nœuds  $i$  et  $j$ , pour tous les ensembles  $S$  séparant  $i$  de  $j$ , on a :  $[X(i)|X(j), X(S)] \sim [X(i)|X(S)]$ . Un état  $x \in \Omega$  est appelé une configuration.

**Définition 9. Clique**

Une clique est un sous-ensemble de nœuds du graphe tel que chaque paire de nœuds soit connectée par une arête.

**Théorème 2. Théorème de Hammersley Clifford**

Pour un graphe fini  $G = (V, E)$ , la distribution  $\pi$  est un champ aléatoire de Markov si elle a pour densité  $f_X$  et qu'il existe des fonctions  $\phi_C$  pour toute clique  $C$  telles que  $f_X$  puisse s'écrire :

$$f_X(x) = \frac{1}{Z} \prod_{C \in \text{cliques}(G)} \phi_C(x)$$

**Définition 10. Auto-modèle**

On dit qu'un champ aléatoire de Markov est un auto-modèle, si il existe des fonctions  $f_i$  et  $g_i$  telles que la densité de  $X \sim \pi$  peut-être écrite sous la forme :

$$f_X(x) = \frac{1}{Z} \left[ \prod_{i \in V} f_i(X(i)) \right] \left[ \prod_{\{i,j\} \in E} g_{\{i,j\}}(X(i), X(j)) \right]$$

Dans la suite, on définit des exemples bien connus d'auto-modèles que l'on étudiera par la suite.

**Exemple 1 : Modèle d'Ising**

Le modèle d'Ising est un auto-modèle de paramètres  $\mu$  et  $\beta$ , où  $\Omega = \{-1, 1\}^V$ , et  $f(c) = \exp(\mu c)$ ,  $g(c_1, c_2) = \exp(\beta \mathbf{1}(c_1 = c_2))$ . Dans la littérature, on appelle  $\mu$  magnétisation et  $\beta$  la température inverse. Lorsque  $\beta > 0$ , on dit que le modèle est ferromagnétique. Lorsque  $\mu = 0$ , on dit que le modèle est sans champ magnétique.

### Exemple 2 : Modèle Hard-core gas

Le modèle hard-core gas est un auto-modèle défini sur  $\{0, 1\}^V$  et de paramètre  $\lambda > 0$  où  $f(c) = \lambda^c$  et  $g(c_1, c_2) = 1 - c_1 c_2$ . Lorsqu'un nœud  $\nu$  a pour label 1, on dit que le nœud est occupé, sinon, il est inoccupé.

### Exemple 3 : Modèle de Strauss

Le modèle de Strauss est un auto-modèle sur  $\{0, 1\}^V$  de paramètres  $\lambda > 0$  et  $\gamma \in [0, 1]$  où  $f(c) = \lambda^c$  et  $g(c_1, c_2) = 1 + (\gamma - 1)c_1 c_2$ . Lorsqu'un nœud  $\nu$  a pour label 1, on dit que le nœud est occupé, sinon, il est inoccupé.

#### 1.2.2 Permutations

Un problème classique de distribution sur les permutations est lié à la recherche du permanent d'une matrice non-négative.

#### Exemple

Pour une matrice non-négative  $w(i, j)$ , notons :

$$w(\sigma) = \prod_{i=1}^n w(i, \sigma(i))$$

Tant qu'il existe au moins une permutation  $\sigma$  telle que  $w(i, \sigma(i)) > 0$  pour tout  $i$ , cela donne une densité non normalisée sur l'ensemble des permutations. La constante de normalisation pour cette densité est alors appelée le permanent de la matrice  $w(i, j)$ . Si aucun  $\sigma$  ne vérifie cette relation, le permanent est 0.

Une autre distribution importante que l'on va considérer est la distribution uniforme sur les permutations où certains objets doivent avoir une position plus faible que les autres.

#### Définition 11. Relation d'ordre partiel

Soit un ensemble  $P$ . Une relation d'ordre partiel sur  $P$  est une relation binaire  $\preceq$  telle que pour tout  $a, b, c \in P$ , la relation est :

1. (Réflexive)  $a \preceq a$
2. (Antisymétrique) Si  $a \preceq b$  et  $b \preceq a$ , alors  $a = b$
3. (Transitive) Si  $a \preceq b$  et  $b \preceq c$ , alors  $a \preceq c$

Un ensemble disposant d'une relation d'ordre partiel est parfois appelé poset d'après l'anglais partially ordered set.

#### Définition 12. Extension linéaire d'un ordre partiel



Une extension linéaire d'un ordre partiel sur  $1, \dots, n$  est une permutation pour laquelle si  $i$  et  $j$  sont tels que  $\sigma(i) \prec \sigma(j)$ , alors  $i < j$ .

### 1.3 Chaînes de Markov et simulation approchée

Jusqu'au développement des algorithmes de simulation parfaite/exacte, la manière principale d'obtenir une réalisation d'une distribution ciblée était une méthode d'approche. Plusieurs algorithmes et méthodes ont été mis en place et l'ensemble de ces méthodes porte le nom de Chaîne de Markov Monte Carlo (Markov Chain Monte Carlo, MCMC dans la littérature). On ne rappellera pas ici les principales définitions pour les chaînes de Markov mais d'autres seront nécessaires pour la suite. On s'intéressera notamment aux chaînes de Harris et au théorème ergodique associé.

#### Définition 13. Chaîne de Harris

Une chaîne de Markov  $\{X_t\}$  sur un espace d'état  $\Omega$  est une chaîne de Harris si il existe des ensembles mesurables  $A, B \in \Omega$  et  $\epsilon > 0$  pour  $x \in A$  et  $y \in B$ , et une mesure de probabilité  $\rho$  où  $\rho(B) = 1$  tels que l'on ait :

1. Pour  $T_A = \inf \{t \geq 0 : X_t \in A\}$ ,  $(\forall z \in \Omega)(\mathbb{P}(T_A < \infty | X_0 = z) > 0)$ .
2. Si  $x \in A$  et  $C \subseteq B$ , alors  $\mathbb{P}(X_1 \in C | X_0 = x) \geq \epsilon \rho(C)$

#### Définition 14. Chaîne récurrente

Soit  $R = \inf \{n > 0 : X_n \in A\}$ . On dit qu'une chaîne de Harris est une chaîne récurrente si pour tout  $x \in A$ ,  $\mathbb{P}(R < \infty | X_0 = x) = 1$ . Une chaîne qui n'est pas récurrente est dite transiente.

#### Définition 15. Chaîne apériodique

Une chaîne de Harris récurrente est apériodique si pour tout  $x \in \Omega$ , il existe  $n$  tel que pour tout  $n' \geq n$ ,  $\mathbb{P}(X_{n'} \in A | X_0 = x) > 0$

#### Théorème 3. Théorème ergodique pour les chaînes de Harris

Soit  $X_n$  une chaîne de Harris récurrente et apériodique de distribution stationnaire  $\pi$ . Si  $\mathbb{P}(R < \infty | X_0 = x) = 1$  pour tout  $x$ , alors, pour  $t \rightarrow \infty$ , pour tout ensemble mesurable  $C$  et pour tout  $x$  :

$$|\mathbb{P}(X_t \in C | X_0 = x) - \pi(C)| \rightarrow 0$$

Ce théorème est le cœur des méthodes MCMC, puisqu'il "suffit" de construire une chaîne de Harris ayant pour distribution stationnaire la distribution souhaitée et de la faire tourner pendant un nombre infini de pas. Cependant, n'ayant pas un temps infini, les utilisateurs font tourner leurs algorithmes

durant un grand nombre de pas et espèrent arriver dans la distribution stationnaire.

Nous pourrions cependant déterminer à quel point la chaîne est proche de la loi stationnaire à l'aide du concept de couplage (voir chapitre 3 pour une définition et application étendue).

**Définition 16.** *Couplage*

Supposons que  $\{X_t\} \sim \nu_X$  et  $\{Y_t\} \sim \nu_Y$ . Un couplage de  $\{X_t\}$  et  $\{Y_t\}$  est un processus bivarié  $\{(X'_t, Y'_t)\}$  tel que  $\{X'_t\} \sim \nu_X$  et  $\{Y'_t\} \sim \nu_Y$ .

**Théorème 4.** *Lemme de Couplage*

Soit  $Y_0 \sim \pi$  et  $X_0 = x_0$  tels que les deux variables évoluent de manière couplée. Alors, pour tout mesurable  $C$  :

$$|\mathbb{P}(X_t \in C | X_0 = x) - \pi(C)| \leq \mathbb{P}(X_t \neq Y_t).$$

## 1.4 Création de chaînes de Markov

Afin d'utiliser les méthodes MCMC, il faut créer des chaînes de Harris qui convergent vers la distribution  $\pi$  ciblée. Il est généralement mieux de créer des chaînes réversibles plutôt que des chaînes simplement stationnaires.

Nous utiliserons la notation suivante pour la définition de réversibilité :  $\pi(dx) = f(x)dx$ . Et donc, pour tout mesurable  $A$ ,  $\pi(A) = \int_{x \in A} \pi(dx) = \int_{x \in A} f(x)dx$ .

Nous utiliserons l'équation de balance détaillée pour en déduire la réversibilité :

**Définition 17.** *Équation de balance détaillée*

Une distribution  $\pi$  est réversible selon une chaîne de Markov  $\{X_t\}$  en particulier si :  $\pi(dx)\mathbb{P}(X_{t+1} \in dy | X_t = x) = \pi(dy)\mathbb{P}(X_{t+1} \in dx | X_t = y)$ .

**Lemme 1.**

Si  $\pi$  est réversible, alors  $\pi$  est stationnaire.

*Démonstration.* Soit  $\Omega$  l'espace d'état de la chaîne de Markov  $\{X_t\}$  considérée et  $\pi$  réversible pour cette chaîne. Pour  $X_t \sim \pi$ , et  $C$  un ensemble mesurable, alors on a :

$$\begin{aligned}
\mathbb{P}(X_{t+1} \in C) &= \mathbb{E}[\mathbf{1}(X_{t+1} \in C)] \\
&= \mathbb{E}[\mathbb{E}[\mathbf{1}(X_{t+1} \in C) | X_t]] \\
&= \int_{x \in \Omega} \mathbb{E}[\mathbf{1}(X_{t+1} \in C) | X_t = x] \pi(dx) \\
&= \int_{x \in \Omega} \mathbb{P}(X_{t+1} \in C | X_t = x) \pi(dx) \\
&= \int_{x \in \Omega} \int_{y \in C} \mathbb{P}(X_{t+1} \in dy | X_t = x) \pi(dx) \\
&= \int_{y \in C} \int_{x \in \Omega} \mathbb{P}(X_{t+1} \in dx | X_t = y) \pi(dy) \\
&= \int_{y \in C} \mathbb{P}(X_{t+1} \in \Omega | X_t = y) \pi(dy) \\
&= \int_{y \in C} \pi(dy) \\
&= \pi(C)
\end{aligned}$$

D'où la stationnarité de  $\pi$ .

□

Plusieurs types de chaînes réversibles existent telles que l'échantillonnage de Gibbs (Gibbs sampler), Metropolis-Hastings, etc. Ces chaînes sont présentées dans la suite.

#### 1.4.1 Échantillonnage de Gibbs

On présente ici un échantillonneur de Gibbs, qui agit sur un espace d'états de la forme  $C^V$ . On appelle  $\nu \in V$  une dimension du problème. Pour  $X_t = x$ , l'échantillonneur choisit une dimension  $\nu$  uniformément sur  $V$ . Soit  $L(x, \nu)$  l'ensemble des états qui sont exactement les nœuds de la configuration  $x$  sauf en  $\nu$ , écrit de la manière suivante :  $L(x, \nu) = \{y : (\forall w \in V \setminus \{\nu\})(y(w) = x(w))\}$ . Pour  $X_t = x$ , l'état suivant  $X_{t+1}$  est choisi selon  $\pi$  à conditionné à être dans  $L(x, \nu)$ .

On présente comme exemple le modèle d'Ising précédemment vu. Une dimension est alors un nœud de la configuration. On choisit donc un nœud uniformément dans  $V$ , et on considère les états qui sont exactement  $x$  en tous les autres nœuds autres que  $\nu$ . Pour le modèle d'Ising, la valeur en  $\nu$  de  $x$  est 1 ou  $-1$ . On note ces configurations  $x_{\nu \rightarrow 1}$  et  $x_{\nu \rightarrow -1}$ . On choisit alors l'état suivant entre  $x_{\nu \rightarrow 1}$  et  $x_{\nu \rightarrow -1}$ , où le choix est fait proportionnellement à  $\pi$ . On a donc :

$$\mathbb{P}(X_{t+1} = x_{\nu \rightarrow 1}) = \frac{\pi(\{x_{\nu \rightarrow 1}\})}{\pi(\{x_{\nu \rightarrow 1}\}) + \pi(\{x_{\nu \rightarrow -1}\})}$$

Or, pour le modèle d'Ising,

$$\pi(\{x\}) = \frac{1}{Z} \prod_{i \in V} \exp(\mu X(i)) \prod_{\{i,j\} \in E} \exp(\beta \mathbf{1}(x(i) = x(j))).$$

Après simplification, on obtient :

$$\begin{aligned} \mathbb{P}(X_{t+1} = x_{\nu \rightarrow 1}) &= \frac{\exp(\mu) \prod_{\{\nu,j\} \in E} \exp(\beta \mathbf{1}(x(j) = 1))}{\exp(\mu) \prod_{\{\nu,j\} \in E} \exp(\beta \mathbf{1}(x(j) = 1)) + \exp(-\mu) \prod_{\{\nu,j\} \in E} \exp(\beta \mathbf{1}(x(j) = -1))} \\ &= \frac{\exp(\beta n_1 + \mu)}{\exp(\beta n_1 + \mu) + \exp(\beta n_{-1} - \mu)} \end{aligned}$$

Où  $n_c$  est le nombre de voisins de  $\nu$  de label  $c$ . Par exemple, si  $\nu$  est adjacent à trois nœuds de label 1 et un nœud de label  $-1$ , alors la probabilité que  $\nu$  se voit labelliser 1 est  $\exp(3\beta + \mu) / (\exp(3\beta + \mu) + \exp(\beta - \mu))$ . Mettre ensuite en place cette méthode algorithmiquement est très simple.

Vérifions la réversibilité de la chaîne.

*Démonstration.* Les cas où la chaîne ne change pas d'état après un pas, (ie  $X_t = x_{\nu \rightarrow 1}$  et  $X_{t+1} = x_{\nu \rightarrow 1}$  ou  $X_t = x_{\nu \rightarrow -1}$  et  $X_{t+1} = x_{\nu \rightarrow -1}$ ) nous donnent immédiatement l'équation de balance détaillée, il ne reste alors qu'à traiter deux cas.

— Premier cas :  $X_t = x_{\nu \rightarrow 1}$  et  $X_{t+1} = x_{\nu \rightarrow -1}$ .

On a alors :

$$\pi(\{x_{\nu \rightarrow 1}\}) \mathbb{P}(X_{t+1} = x_{\nu \rightarrow -1} | X_t = x_{\nu \rightarrow 1}) = \pi(\{x_{\nu \rightarrow 1}\}) \frac{\exp(-\mu + \beta n_{-1})}{\exp(-\mu + \beta n_{-1}) + \exp(\mu + \beta n_1)}$$

D'autre part :

$$\pi(\{x_{\nu \rightarrow -1}\}) \mathbb{P}(X_{t+1} = x_{\nu \rightarrow 1} | X_t = x_{\nu \rightarrow -1}) = \pi(\{x_{\nu \rightarrow -1}\}) \frac{\exp(\mu + \beta n_1)}{\exp(-\mu + \beta n_{-1}) + \exp(\mu + \beta n_1)}$$

On aura donc l'égalité entre ces termes si et seulement si on a :

$$\pi(\{x_{\nu \rightarrow 1}\}) \exp(-\mu + \beta n_{-1}) = \pi(\{x_{\nu \rightarrow -1}\}) \exp(\mu + \beta n_1)$$

Les simplifications suivantes s'opèrent dans les termes en  $\pi$  :

- simplification des  $\frac{1}{Z}$
- simplification des  $\prod_{i \in V} \exp(\mu X(i))$  sauf au nœud  $\nu$
- simplification des  $\prod_{\{i,j\} \in E} \exp(\beta \mathbf{1}(x(i) = x(j)))$  sauf aux arêtes ayant  $\nu$  comme extrémité

On obtient alors :

$$\exp(\mu + \beta n_1) \exp(-\mu + \beta n_{-1}) = \exp(-\mu + \beta n_{-1}) \exp(\mu + \beta n_1)$$

On a donc bien vérifié l'égalité et donc l'équation de balance détaillée.

Le deuxième cas :  $X_t = x_{\nu \rightarrow -1}$  et  $X_{t+1} = x_{\nu \rightarrow 1}$  se vérifie de la même manière.

Ayant vérifié tous les cas possibles, on obtient donc l'équation de balance détaillée qui conduit bien à la réversibilité de la chaîne.

□

## Programmation sous R

```
1  ## X = graphe (rectangulaire ou carré), u = lancer uniforme sur
   [0,1], v = noeud de X ##
2  IsingGibbsUpdate <- function(X,u,v,beta,mu){ ## Permet de faire un
   pas de la chaîne de Markov associée selon le Gibbs Sampler##
3
4  ##On entoure la matrice de valeurs aberrantes pour la manipuler
   plus facilement##
5  ##On change les indices du point considéré pour se conformer à la
   nouvelle matrice##
6
7  i = v[1] + 1
8  j = v[2] + 1
9
10 X = rbind(rep(0,dim(X)[2]),X)
11 X = rbind(X,rep(0,dim(X)[2]))
12 X = cbind(rep(0,dim(X)[1]),X)
13 X = cbind(X,rep(0,dim(X)[1]))
14
15 ##Les indicatrices permettent de vérifier quelles valeurs ont les
   voisins du point considéré ##
16
17 if(u< exp(mu + beta*((X[i+1,j]==1)*1 + (X[i-1,j]==1)*1 +(X[i,j
   +1]==1)*1 + (X[i,j-1]==1)*1))/((exp(mu + beta*((X[i+1,j]==1)*1
   + (X[i-1,j]==1)*1 +(X[i,j+1]==1)*1 + (X[i,j-1]==1)*1))) + exp
   (-mu + beta*((X[i+1,j]==-1)*1 + (X[i-1,j]==-1)*1 +(X[i,j
   +1]==-1)*1 + (X[i,j-1]==-1)*1))))){
18   X[i,j] = 1
19 }
20 else{
21   X[i,j] = -1
22 }
23
24 ##On supprime les valeurs insérées##
25 X=X[,-1]
26 X=X[,-dim(X)[2]]
27 X=X[-1,]
28 X=X[-dim(X)[1],]
29
30 X
31 }
```

```

1  IsingGibbssteps <- function(X,beta,mu,n){ ##effectue n pas de la
      chaîne de Markov pour le sampler de Gibbs, pour un graphe
      rectangulaire X##
2
3  for (i in 1:n){
4
5      ##choix uniforme du noeud considéré##
6      k = sample(1:dim(X)[1],1)
7      l = sample(1:dim(X)[2],1)
8      ##lancer uniforme sur [0,1]##
9      u = runif(1)
10     ##mise à jour du point##
11     X=IsingGibbsUpdate(X,u,c(k,l),beta,mu)
12
13 }
14
15 X
16 }

```

```

1  IsingGibbsconditionnalstop <- function(X,beta,epsilon){ ##effectue
    un nombre aléatoire de pas de la chaîne de Markov pour le
    sampler de Gibbs, pour un graphe carré X##
2
3  ##On crée le vecteur Y qui contiendra les ratios de 1 pour chaque
    graphe##
4  Y=c(mean(X==1))
5  bol = TRUE
6  i=1
7
8  while (bol){
9
10     ##Choix uniforme du noeud considéré##
11     k = sample(1:dim(X)[1],1)
12     l = sample(1:dim(X)[2],1)
13
14     ##lancer uniforme sur [0,1]##
15     u = runif(1)
16
17     ##Mise à jour du noeud considéré##
18     X=IsingGibbsUpdate(X,u,c(k,l),beta)
19
20     ##Mise à jour du vecteur des ratios##
21     Y = c(Y,mean(X==1))
22
23     ##Si la ligne est décommentée, on obtient une représentation de
        la configuration X tout les (dim(X)[1]**2)/epsilon pas##
24     ##if ((i%((dim(X)[1]**2)/epsilon))==0){image(X)}
25
26     ##Si on a effectué un nombre de pas suffisant, et que le nombre de
        1 n'a pas varié énormément durant un certain nombre d'
        itérations, on s'arrête##
27     if ((i>=((log(dim(X)[1]**2)*dim(X)[1]**2)/epsilon)) && ((mean(Y[
        i+1-floor((dim(X)[1]**(1/2))/epsilon):i])-Y[i])<epsilon)){
28         break()
29     }
30     i=i+1
31 }
32
33 return(list(X,i))
34
35 }

```



### 1.4.2 Metropolis-Hastings

Pour la méthode de Metropolis-Hastings, pour chaque configuration  $x$  de la chaîne, on a une densité  $q_x$ . On propose ensuite un état suivant  $y$ , selon la densité  $q_x$ . L'état suivant sera donc  $x$  ou  $y$ . La probabilité avec laquelle est choisie la configuration  $y$  est donnée de telle sorte que la réversibilité s'applique. On la donne ci-dessous. Notez que, pour que cela fonctionne, il est nécessaire que si  $q_x(y) > 0$ , alors  $q_y(x) > 0$  aussi. On a alors le changement vers  $y$  avec probabilité :

$$\min \left\{ 1, \frac{f_\pi(y)q_y(x)}{f_\pi(x)q_x(y)} \right\}$$

On notera aussi que  $f_\pi$  peut-être normalisée ou non.

En exemple, on considère à nouveau le modèle d'Ising. A chaque pas, on choisit un nœud uniformément dans  $V$  puis un label candidat pour ce nœud est choisi uniformément dans  $\{-1, 1\}$ . On a alors  $q_x(y) = q_y(x) = 1/2$ . On calcule ensuite, pour  $n_c$  le nombre de voisins de  $\nu$  de label  $c$  (où  $c$  est le label proposé) et  $n_{x(\nu)}$  le nombre de voisins de  $\nu$  de même label que  $\nu$  :

$$\frac{f_\pi(y)}{f_\pi(x)} = \frac{\exp(n_c\beta + c\mu)}{\exp(n_{x(\nu)}\beta + x(\nu)\mu)}$$

L'algorithme est alors très simple à mettre en place. D'abord, vérifions la réversibilité de la chaîne.

*Démonstration.* Les cas où la chaîne ne change pas d'état après un pas, (ie  $X_t = x_{\nu \rightarrow 1}$  et  $X_{t+1} = x_{\nu \rightarrow 1}$  ou  $X_t = x_{\nu \rightarrow -1}$  et  $X_{t+1} = x_{\nu \rightarrow -1}$ ) nous donnent immédiatement l'équation de balance détaillée, il ne reste alors qu'à traiter deux cas.

— Premier cas :  $X_t = x_{\nu \rightarrow 1}$  et  $X_{t+1} = x_{\nu \rightarrow 1}$ .

On a alors :

$$\pi(\{x_{\nu \rightarrow 1}\})\mathbb{P}(X_{t+1} = x_{\nu \rightarrow 1} | X_t = x_{\nu \rightarrow 1}) = \pi(\{x_{\nu \rightarrow 1}\})\min \left\{ 1, \frac{\exp(-\mu + \beta n_{-1})}{\exp(\mu + \beta n_1)} \right\}$$

D'autre part :

$$\pi(\{x_{\nu \rightarrow -1}\})\mathbb{P}(X_{t+1} = x_{\nu \rightarrow 1} | X_t = x_{\nu \rightarrow -1}) = \pi(\{x_{\nu \rightarrow -1}\})\min \left\{ 1, \frac{\exp(\mu + \beta n_1)}{\exp(-\mu + \beta n_{-1})} \right\}$$

On aura donc l'égalité entre ces termes si et seulement si on a :

$$\pi(\{x_{\nu \rightarrow 1}\})\min \left\{ 1, \frac{\exp(-\mu + \beta n_{-1})}{\exp(\mu + \beta n_1)} \right\} = \pi(\{x_{\nu \rightarrow -1}\})\min \left\{ 1, \frac{\exp(\mu + \beta n_1)}{\exp(-\mu + \beta n_{-1})} \right\}$$

Les simplifications suivantes s'opèrent dans les termes en  $\pi$  :

- simplification des  $\frac{1}{Z}$
- simplification des  $\prod_{i \in V} \exp(\mu X(i))$  sauf au nœud  $\nu$
- simplification des  $\prod_{\{i,j\} \in E} \exp(\beta \mathbf{1}(x(i) = x(j)))$  sauf aux arêtes ayant  $\nu$  comme extrémité

On obtient alors :

$$\exp(\mu + \beta n_1) \min \left\{ 1, \frac{\exp(-\mu + \beta n_{-1})}{\exp(\mu + \beta n_1)} \right\} = \exp(-\mu + \beta n_{-1}) \min \left\{ 1, \frac{\exp(\mu + \beta n_1)}{\exp(-\mu + \beta n_{-1})} \right\}$$

Par propriété du min, on obtient alors l'égalité entre les termes, qui conduit à l'équation de balance détaillée.

Le deuxième cas :  $X_t = x_{\nu \rightarrow -1}$  et  $X_{t+1} = x_{\nu \rightarrow 1}$  se vérifie de la même manière.

Ayant vérifié tous les cas possibles, on obtient donc l'équation de balance détaillée qui conduit bien à la réversibilité de la chaîne.

□

## Programmation sous R

```
1  ## X = graphe (rectangulaire ou carré), u = lancer uniforme sur
   [0,1], v = noeud de X ##
2
3  MetropolisHastingsUpdate <- function(X,u,v,beta){ ## Permet de faire
   un pas de la chaîne de Markov associée selon Metropolis-
   Hastings
4
5  ##On entoure la matrice de valeurs aberrantes pour la manipuler
   plus facilement##
6  ##On change les indices du point considéré pour se conformer à la
   nouvelle matrice##
7
8  i = v[1] + 1
9  j = v[2] + 1
10
11 X = rbind(rep(0,dim(X)[2]),X)
12 X = rbind(X,rep(0,dim(X)[2]))
13 X = cbind(rep(0,dim(X)[1]),X)
14 X = cbind(X,rep(0,dim(X)[1]))
15
16 ##On choisit une couleur candidate pour le noeud v##
17
18 c=sample(c(-1,1),1,prob=c(1/2,1/2))
19
20 ##Les indicatrices permettent de vérifier quelles valeurs ont les
   voisins du point considéré ##
21
22 if(u< exp(beta*((X[i+1,j]==c)*1 + (X[i-1,j]==c)*1 +(X[i,j+1]==c)*1
   + (X[i,j-1]==c)*1))/exp(beta*((X[i+1,j]==X[i,j])*1 + (X[i-1,j]
   )==X[i,j])*1 +(X[i,j+1]==X[i,j])*1 + (X[i,j-1]==X[i,j])*1))) {
23   X[i,j] = c
24 }
25
26 ##On supprime les valeurs insérées##
27 X=X[, -1]
28 X=X[, -dim(X)[2]]
29 X=X[-1, ]
30 X=X[-dim(X)[1], ]
31
32 X
33 }
```

```

1  MetropolisHastingssteps <- function(X,beta,n){ ##effectue n pas de
    la chaîne de Markov pour Metropolis-Hastings, pour un graphe
    rectangulaire X##
2
3  for (i in 1:n){
4
5    ##choix uniforme du noeud##
6    k = sample(1:dim(X)[1],1)
7    l = sample(1:dim(X)[2],1)
8
9    ##lancer uniforme sur [0,1]##
10   u = runif(1)
11
12   ##mise à jour de la configuration X##
13   X=MetropolisHastingsUpdate(X,u,c(k,l),beta)
14 }
15
16 X
17
18 }

```

```

1
2 MetropolisHastingsconditionnalstop <- function(X,beta,epsilon){ ##
   effectue un nombre aléatoire de pas de la chaîne de Markov pour
   Metropolis-Hastings, pour un graphe carré X##
3
4 ##On créé le vecteur Y qui contiendra les ratios de 1 pour chaque
   graphe##
5 Y=c(mean(X==1))
6 bol = TRUE
7 i=1
8
9
10 while (bol){
11
12   ##Choix uniforme du noeud considéré##
13   k = sample(1:dim(X)[1],1)
14   l = sample(1:dim(X)[2],1)
15
16   ##lancer uniforme sur [0,1]##
17   u = runif(1)
18
19   ##Mise à jour du noeud considéré##
20   X=MetropolisHastingsUpdate(X,u,c(k,l),beta)
21
22   ##Mise à jour du vecteur des ratios##
23   Y = c(Y,mean(X==1))
24
25   ##Si la ligne est décommentée, on obtient une représentation de la
   configuration X tout les (dim(X)[1]**2)/epsilon pas##
26   ##if ((i%%((dim(X)[1]**2)/epsilon))==0){image(X)}##
27
28   ##Si on a effectué un nombre de pas suffisant, et que le nombre de
   1 n'a pas varié énormément durant un certain nombre d'
   itérations, on s'arrête##
29   if ((i>=((log(dim(X)[1]**2)*dim(X)[1]**2)/epsilon)) && ((mean(Y[
   i+1-floor((dim(X)[1]**(1/2)))/epsilon:i]-Y[i])<epsilon)){
30     break()
31   }
32   i=i+1
33 }
34
35 return(list(X,i))
36
37 }

```

## Discussion du critère d'arrêt dans "IsingGibbsconditionnalstop"

### et "MetropolisHastingsconditionnalstop"

Le premier critère pour l'arrêt du programme est si le nombre de pas  $i$  vérifie :  $i \geq ((\log(n^2) * n^2) / \epsilon)$ , où  $n^2$  est la taille du graphe carré considéré et  $\epsilon$  est la "précision" que l'utilisateur souhaite.

Le terme en  $n^2 * \log(n^2)$  provient du problème du collectionneur. En effet, on souhaite, pour atteindre un certain équilibre, avoir donné la chance à chacun des nœuds d'avoir subi un changement. Cela revient à avoir sélectionné chacun des nœuds au moins une fois, ce qui revient au problème du collectionneur, où en moyenne le temps d'obtention de tous les éléments de la collection se fait en  $n * \log(n)$  si  $n$  est la taille de la collection complète.

Le terme en  $\epsilon$  permet de s'assurer d'avoir obtenu la visite complète de tous les nœuds, mais servira plus amplement dans l'autre condition d'arrêt présentée ci-après.

Le second critère est si le vecteur des ratios de 1 vérifie :

$$|Y[i] - \frac{1}{\lfloor \frac{n}{\epsilon} \rfloor} \sum_{k=i+1-\lfloor \frac{n}{\epsilon} \rfloor}^i Y[k]| < \epsilon$$

C'est-à-dire, si le ratio de 1 n'a pas évolué de plus de  $\epsilon$  en moyenne depuis le temps  $i + 1 - \lfloor \frac{n}{\epsilon} \rfloor$ . On parle donc de "précision"  $\epsilon$  puisque plus cette valeur est petite, plus le temps requis pour que peu de changements s'effectuent en moyenne est grand.

Le terme en  $n$  (où  $n^2$  est la taille de la matrice) est ajouté pour obtenir une dépendance à la dimension dans la précision.

### **1.4.3 Variables aléatoires auxiliaires / Auxiliary random variables**

Dans la plupart des applications du MCMC, la densité ciblée  $X$  a une structure multiplicative. Pour ces densités, il est possible d'ajouter un vecteur  $Y$  de variables aléatoires supplémentaires telles que  $(X, Y)$  a une distribution jointe plus simple. Par exemple, on considère à nouveau le modèle d'Ising avec  $\mu = 0$ . Pour chaque arête  $\{i, j\}$ , on crée une variable aléatoire auxiliaire  $Y(\{i, j\})$  telle que sa distribution conditionnée sur  $X$  soit uniforme sur  $[0, 1]$  si  $X(i) \neq X(j)$  et uniforme sur  $[0, \exp(\beta)]$  si on a l'égalité.

La densité jointe est alors uniforme sur :

$$X \in \{0, 1\}^V \text{ et } Y \in \{[0, \infty) : (\forall \{i, j\})(y(\{i, j\}) \leq \min(\exp(\beta), 1))\}$$

La chaîne de Markov est alors la suivante : pour  $X$  donné, il suffit de choisir une nouvelle valeur de  $Y$  sachant  $X$ , puis choisir un nouveau  $X$  conditionné sur  $Y$ .

De par la construction de la chaîne, tirer  $Y$  sachant  $X$  est direct. Cependant,

tirer  $X$  sachant  $Y$  est une autre histoire : prenons pour exemple le modèle d'Ising avec  $\beta > 0$ . On a alors  $\exp(\beta) > 1$ . Lorsque  $y(\{i, j\}) \in [0, 1]$ , alors il est possible que  $x(i) = x(j)$  ou que  $x(i) \neq x(j)$ , mais lorsque  $y(\{i, j\}) > 1$ , alors on a forcément que  $x(i) = x(j)$ . Les arêtes avec  $y(\{i, j\}) > 1$  séparent le graphe en groupes de composants connectés tels que chacun des composants doivent avoir le même label.

Il suffit donc de séparer le graphe en groupes de composants connectés en se servant des arêtes  $y(\{i, j\}) > 1$ , puis il faut choisir un label uniformément sur  $\{-1, 1\}$  pour les composantes de ces groupes.

#### 1.4.4 Shift Chains

Un autre type de chaîne utile pour les modèles répulsifs est le shift. Pour illustrer le modèle, on considère le modèle hard-core où chaque nœud du graphe est soit occupé ( $x(\nu) = 1$ ) soit inoccupé ( $x(\nu) = 0$ ). Chaque nœud occupé contribue d'un facteur  $\lambda$  à la densité.

Dès lors qu'un des voisins de  $\nu$  est occupé, la chaîne ne peut changer la valeur de  $\nu$ .

Le shift autorise un échange de label entre  $\nu$  et son voisin occupé avec une certaine probabilité  $p_{shift}$ . Lorsque deux voisins de  $\nu$  ou plus sont occupés, aucun changement n'est possible et  $\nu$  reste inoccupé.

## 2 Chapitre 2 : Procédures de rejet

### 2.1 Théorème et exemple simple

#### Théorème 5.

Soit  $\nu$  une mesure finie sur un ensemble  $B$  et soit  $A$  un sous-ensemble de  $B$  tel que  $\nu(A) > 0$ . Soit  $X_1, X_2, \dots$  des variables aléatoires iid tirées selon  $\nu$  sur l'ensemble  $B$ . Soit enfin  $T = \inf \{t : X_t \in A\}$ . Alors :

$$X_T \sim [X_1 | X_1 \in A]$$

C'est à dire, si on tire successivement des valeurs dans  $B$  selon  $\nu$  et que l'on rend la première valeur dans  $A$ , alors on le résultat provient de la mesure  $\nu$  conditionnée à être dans  $A$ .

*Démonstration.* Cela suit du théorème fondamental de la simulation parfaite en ayant  $U_1, U_2, \dots$  des tirages dans  $B$  selon  $\nu$ ,  $b(U) = \mathbf{1}(U \in A)$ ,  $g(U) = U$  et  $f(X, U) = X$ .  $\square$

Le nombre moyen de pas nécessaire pour obtenir un tirage dans  $A$  dépend de la taille de  $A$  par rapport à la taille de  $B$ .

#### Théorème 6.

La probabilité qu'un tirage  $X$  dans  $B$  selon  $\nu$  soit dans  $A$  est  $\nu(A)/\nu(B)$ . Le nombre moyen de tirages nécessaires est  $\nu(B)/\nu(A)$ .

*Démonstration.* La première moitié du théorème découle simplement de la définition de mesure de probabilité. La seconde provient du fait que  $T$  soit une loi géométrique de paramètre  $\nu(A)/\nu(B)$ , et donc sa moyenne est  $\nu(B)/\nu(A)$ .  $\square$

#### Exemple : tirage uniforme sur le cercle unité

On choisit  $B = [-1, 1] \times [-1, 1]$  et  $A = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 \leq 1\}$ . L'aire de  $B$  est 4 tandis que l'aire du cercle unité est  $\pi$ . Le nombre moyen de lancers est alors  $2 * (4/\pi) = 2 * 1.273$ , où le facteur 2 provient du nombre d'uniformes nécessaires pour obtenir un tirage dans  $B$ .

### 2.2 Méthode de rejet pour variables aléatoires à densité

La méthode de rejet de base n'inclut pas les variables aléatoires à densité. Cependant, celles-ci sont traitées aisément à l'aide des variables aléatoires auxiliaires. Pour cela, on considère le théorème suivant, provenant de la théorie de la mesure.

#### Théorème 7. Théorème fondamental de la simulation de Monte Carlo



Soit  $X$  variable aléatoire de densité (possiblement non-normalisée)  $f_X$  par rapport à une mesure  $\nu$  sur  $\Omega$ . Si on a  $[Y|X] \sim \text{Unif}([0, f_X])$ , alors  $(X, Y)$  est un lancer provenant de la mesure produit  $\nu \times \text{Unif}$  sur l'ensemble  $\{(x, y) : x \in \Omega, 0 \leq y \leq f_x\}$ .

Essentiellement, ce théorème nous dit que pour obtenir un lancer selon une certaine densité, il suffit de tirer selon une loi uniforme sur un espace plus grand. De plus, la densité de  $X$  n'est pas forcément normalisée, ce qui sera essentiel dans les applications qui suivent.

Pour comprendre comment cela fonctionne, nous allons traiter un exemple en toute généralité, puis un exemple concret.

Soit  $\mu$  une mesure sur un ensemble  $\Omega$ . On suppose que, pour une densité  $g$ , il est possible de générer un lancer provenant de la densité produit  $\mu \times \text{Unif}$  sur  $\Omega_g = \{(x, y) : x \in \Omega, 0 \leq y \leq g\}$ .

Une fois que ce lancer est obtenu,  $X$  provient alors de la densité  $g$ . Mais supposons ici que le but est d'obtenir un lancer selon une densité  $f$ .

Notons d'abord que, si  $(X, Y)$  est de mesure  $\mu \times \text{Unif}$  sur  $\Omega_g$ , alors  $(X, 2Y)$  est de mesure  $\mu \times \text{Unif}$  sur  $\Omega_{2g} = \{(x, y) : x \in \Omega, 0 \leq y \leq 2g\}$ . On peut même généraliser : pour toute constante  $c > 0$ ,  $(X, cY)$  est de mesure  $\mu \times \text{Unif}$  sur  $\Omega_{cg}$ .

Supposons alors que  $c \geq \sup_{x \in \Omega} f(x)/g(x)$ . Alors :

$$\Omega_f \subseteq \Omega_{cg}$$

On peut alors utiliser une méthode de rejet pour obtenir une réalisation provenant de la densité  $f$ . Il suffit de générer  $X$  selon la densité  $g$ , puis de tirer uniformément  $Y$  sur  $[0, c * g(X)]$ . Si  $Y$  est aussi dans  $[0, f(X)]$  (de sorte que  $(X, Y) \in \Omega_f$ ) alors on accepte  $(X, y)$ , ce qui signifie que  $X$  est un lancer selon la densité  $f$ .

Pour tester si le lancer uniforme de  $[0, c * g(X)]$  est dans  $[0, f(X)]$ , il n'est pas nécessaire de vérifier si l'uniforme tombe effectivement dedans : on peut simplement tirer une variable aléatoire  $C$  suivant une loi de Bernoulli de paramètre  $f(X)/[c * g(X)]$ . Si  $C = 1$ , on considère que  $X$  est un lancer provenant de la densité  $f$ .

La probabilité d'acceptation varie inversement avec  $c$ , le meilleur choix possible pour  $c$  est alors  $\sup_{x \in \Omega} f(x)/g(x)$ . La valeur de  $c$  n'est pas toujours évidente à calculer. Dans ce cas, une majoration suffit : l'algorithme fera simplement plus de lancers en moyenne.

Nous allons maintenant calculer la probabilité d'acceptation d'un lancer.

## **Lemme 2.**

Soit  $Z_f = \int_{x \in \Omega} f(x) \nu(dx)$  et  $Z_g = \int_{x \in \Omega} g(x) \nu(dx)$ , les constantes de normalisation pour les densités (possiblement) non-normalisées  $f$  et  $g$ . Alors,

pour la méthode par rejet présentée précédemment, la probabilité d'acceptation est  $Z_f/[c * Z_g]$  et le nombre moyen de lancers est  $c * Z_g/Z_f$ .

*Démonstration.* Pour  $C$ , la loi de Bernoulli présentée précédemment, le problème est de calculer  $\mathbb{P}(C = 1)$ .

$$\begin{aligned}\mathbb{P}(C = 1) &= \mathbb{E}[\mathbb{E}[\mathbf{1}(C = 1)|X]] \\ &= \int_{x \in \Omega} \mathbb{P}(C = 1|X = x) \mathbb{P}(X \in dx) \\ &= \int_{x \in \Omega} f(x)/[cg(x)](g(x)/Z_g) \nu(dx) \\ &= [cZ_g]^{-1} \int_{x \in \Omega} f(x) \nu(dx) \\ &= Z_f/[c * Z_g]\end{aligned}$$

□

### 2.2.1 Exemple : D'une Cauchy vers une Normale

Un exemple simple est la situation suivante : on suppose qu'un utilisateur peut générer une réalisation d'une loi de Cauchy standard de densité  $g(x) = [\pi(1 + x^2)]^{-1}$ . Cet utilisateur souhaite générer une réalisation d'une loi Normale standard de densité  $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$ . D'abord, la normalisation n'est pas nécessaire, donc soit  $g_1(x) = (1 + x^2)^{-1}$  et  $f_1(x) = \exp(-x^2/2)$ . On a que  $f_1(x)/g_1(x) \leq 2, \forall x$ . On implémente alors sous R l'algorithme :

```

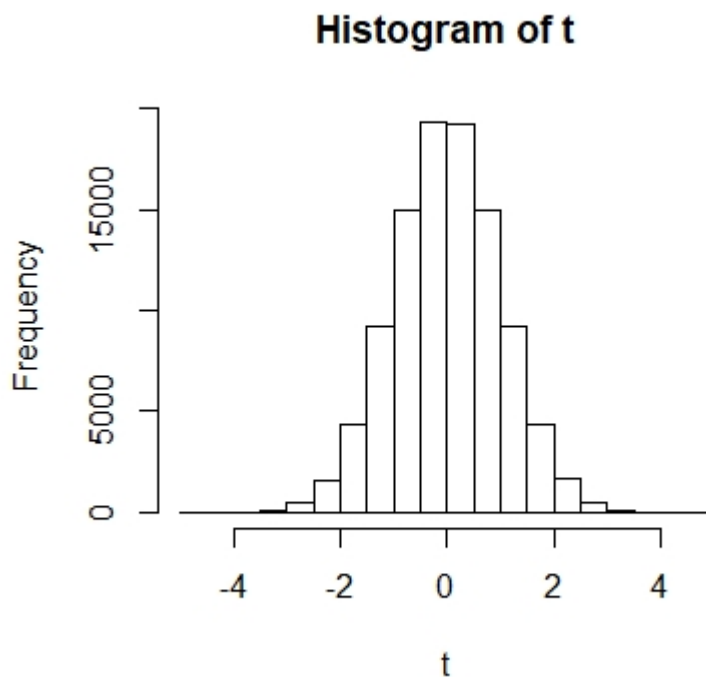
1  normalfromcauchy <- function(){ #changer le return pour obtenir une
    réalisation ou le temps mis à l'obtenir#
2  C=0
3  x=0
4  n=0
5  while(C!=1){
6
7  #lancer aléatoire selon la loi de Cauchy standard##
8    x = rcauchy(1)
9
10 #lancer aléatoire selon la loi de Bernoulli exprimée précédemment
    ##
11    C = rbinom(1,1,(1/2)*(1+x^2)*exp(-(1/2)*x^2))
12
13    n=n+1
14  }
15  return(x)
16 }
17
```

```

18 s=0
19 for (i in 1:10^6) { #changer le return pour obtenir le nombre moyen
    de lancers nécessaires
20   s=s+normalfromcauchy()
21 }
22
23 #s contiendra le nombre de lancers moyens
24 s = s/10^6
25
26 t = matrix(0,1,10^5)
27 for (i in 1:10^5) { #changer le return pour obtenir des réalisations
28   t[i]=normalfromcauchy()
29 }
30 }
31
32 #t contiendra tous les lancers effectués, on construit ensuite un
    histogramme##
33 hist(t)

```

L'histogramme rendu après un lancer du programme est le suivant :



On obtient bien la densité d'une loi normale standard.

On récupère aussi le nombre moyen de lancers requis afin d'obtenir une réalisation : 2.50588. On peut retrouver ce résultat de manière théorique en calculant la probabilité d'acceptation :

$$\begin{aligned}
\mathbb{P}(C = 1) &= \int_{x \in \mathbb{R}} \mathbb{P}(X \in dx, C = 1) \\
&= \int_{x \in \mathbb{R}} g(x) \frac{1}{2} (1 + x^2) \exp(-x^2/2) dx \\
&= (2\pi)^{-1/2}
\end{aligned}$$

Le nombre moyen de lancers est donc  $\sqrt{2\pi} = 2.506 \dots$

Il est possible d'accélérer l'algorithme : on a en fait  $\sup_x (1+x^2) \exp(-x^2/2) = 2/\sqrt{e}$ . En remplaçant dans l'algorithme (mais aussi en calculant exactement l'intégrale), le nombre moyen de lancers devient  $\sqrt{2\pi/e} = 1.520 \dots$

### 2.3 Union d'ensembles

Une autre manière d'utiliser les méthodes de rejet est de prendre en compte les multiples manières d'obtenir une réalisation. Par exemple, on considère le problème de génération selon une mesure  $\nu$  sur  $S_1 \cup \dots \cup S_n$ , où les ensembles  $S_i$  sont de mesure finie  $\nu(S_i)$ .

On considère l'algorithme suivant : on tire  $I$  dans  $\{1, 2, \dots, n\}$  de telle sorte que  $\mathbb{P}(I = i) \propto \nu(S_i)$ . Puis, conditionnellement sur  $I$ , on tire  $X$  dans  $S_I$  selon la mesure  $\nu$ .

Le problème est que  $X$  n'est pas tiré selon  $\nu$  sur  $S_1 \cup \dots \cup S_n$ , car il y a plusieurs manières pour lesquelles  $X$  aurait pu être choisi. Par exemple, si  $X \in S_1$  et  $X \in S_2$  mais  $X \notin S_3 \dots S_n$ , alors  $I$  aurait pu être 1 ou 2 lors du choix de  $X$ .

#### Lemme 3.

Pour l'algorithme précédent,  $X$  est un tirage selon  $\nu$  sur  $S_1 \cup \dots \cup S_n$  avec densité  $f(x) = \# \{i : x \in S_i\} / \nu(S_1 \cup \dots \cup S_n)$

*Démonstration.* Soit  $x \in S_1 \cup \dots \cup S_n$  et  $C = \nu(S_1 \cup \dots \cup S_n)$  Alors :

$$\begin{aligned}
\mathbb{P}(X \in dx) &= \sum_{i=1}^n \mathbb{P}(X \in dx, I = i) \\
&= \sum_{i=1}^n \mathbb{P}(I = i) \mathbb{P}(X \in dx | I = i) \\
&= \sum_{i=1}^n C^{-1} \nu(S_i) \mathbf{1}(x \in S_i) \nu(dx) \nu(S_i)^{-1} \\
&= C^{-1} \# \{i : x \in S_i\}
\end{aligned}$$

□

Par exemple, supposons que l'on veuille tirer uniformément sur les 3 cercles unités respectivement centrés en  $(0,0)$ ,  $(1/2,0)$  et  $(1/2,1/2)$ . On tire alors  $I \sim \text{Unif}(\{1, 2, 3\})$ . Si  $I = 1$ , on tire  $X$  sur le cercle unité centré en  $(0,0)$ . Pour  $I = 2$  ou  $I = 3$ , on tire  $X$  sur le cercle centré en  $(1/2,0)$  et  $(1/2,1/2)$  respectivement. Après, avoir tiré  $I$  puis  $X$  conditionnellement à  $I$ , on accepte  $X$  avec probabilité égale à l'inverse du nombre de cercles où se situe  $X$ .

## 2.4 Simulation à l'aide des inégalités de Markov et Chernoff

### 2.4.1 L'inégalité de Markov en tant que procédure de rejet

**Lemme 4.** *Inégalité de Markov*

On considère une variable aléatoire  $X$  non-négative avec probabilité 1 et d'espérance finie  $\mathbb{E}[X]$ . Alors, pour tout  $a > 0$  :

$$\mathbb{P}(X \geq a) \leq \mathbb{E}[X]/a$$

Du point de vue simulation, l'inégalité de Markov nous permet de tirer depuis  $[X|X \geq a]$ . C'est à dire, si  $X$  a pour densité  $f_X(x)$  alors le but est de tirer depuis la densité non-normalisée  $f_X(x)\mathbf{1}(x \geq a)$ . Pour pouvoir appliquer une méthode de rejet, on utilise la densité non-normalisée  $xf_X(x)$ . On a alors que  $f_X(x)\mathbf{1}(x \geq a)/[xf_X(x)]$  vaut 0 lorsque  $x < a$ , et  $1/x$  lorsque  $x \geq a$ . Le produit n'est donc jamais supérieur à  $1/a$  ce qui permet d'utiliser la méthode de rejet suivante :

---

**Méthode rejet Inégalité de Markov** Entrées :  $f, a$  Sortie :  $X \sim f$  sachant  $X > a$

---

**while**  $C \neq 1$  **do**

    Tirer  $X$  selon la densité non-normalisée  $xf(x)$

    Tirer  $C \sim \text{Bern}(a\mathbf{1}(X \geq a)/x)$

**end while**

---

On remarque que la constante de normalisation pour la densité  $xf(x)$  est simplement :  $\int xf(x)\nu(dx) = \mathbb{E}[X]$ . On a alors :

$$\mathbb{P}(C = 1) = \int_{x \geq 0} \frac{xf_X(x)}{\mathbb{E}[X]} \frac{a\mathbf{1}(X > a)}{x} = \int_{x \geq a} \frac{af_X(x)}{\mathbb{E}[X]} = \frac{a\mathbb{P}(X > a)}{\mathbb{E}[X]}$$

Puisque la chance d'accepter est au plus 1, cette fraction est aussi au plus 1, ce qui montre l'inégalité de Markov.

### 2.4.2 Les inégalités de Chernoff en tant que procédure de rejet

Les inégalités de Chernoff donnent des limites supérieures sur la probabilité qu'une somme de variables aléatoires est supérieure (ou inférieure) à une certaine valeur. Soit  $S_n = X_1 + \dots + X_n$ , où  $X_1, \dots, X_n$  sont iid ; alors le but est de générer un tirage selon  $S_n$  tel que  $S_n \geq a$  ou  $S_n \leq a$ .

L'inégalité de Markov peut aussi être appliquée à une somme de variables aléatoires, mais la borne donnée n'est pas aussi précise que celle obtenue avec les inégalités de Chernoff.

#### Lemme 5. Inégalités de Chernoff

Soit une variable aléatoire  $X$  telle que  $\mathbb{E}[e^{tX}]$  soit finie pour  $t \in [a, b]$ , où  $a$  est négatif et  $b$  positif. Alors pour un certain  $c \in \mathbb{R}$ , et en notant par  $mgf_X(t)$  la fonction génératrice des moments pour la variable aléatoire  $X$ ,

$$\begin{aligned}\mathbb{P}(X \geq c) &\leq mgf_X(t) / \exp(tc), \quad \forall t \in [0, b] \\ \mathbb{P}(X \leq c) &\leq mgf_X(t) / \exp(tc), \quad \forall t \in [a, 0]\end{aligned}$$

*Démonstration.* Tout d'abord :  $\mathbb{P}(X \geq c) = \mathbb{P}(tX \geq tc)$  pour tout  $t$  positif. De plus,  $\mathbb{P}(tX \geq tc) = \mathbb{P}(\exp(tX) \geq \exp(tc))$ . Enfin, on applique l'inégalité de Markov pour obtenir  $\mathbb{P}(\exp(tX) \geq \exp(tc)) \leq \mathbb{E}[\exp(tX)] / \exp(tc)$ . Le résultat pour l'autre inégalité est démontré de la même manière.  $\square$

Nous considérons à présent ce qu'il se passe lorsque l'on considère les inégalités de Chernoff pour une somme de variables aléatoires indépendante. On rappelle d'abord un résultat sur les fonctions génératrices des moments.

#### Lemme 6.

Soit  $S_n = X_1 + \dots + X_n$ , où les  $\{X_i\}$  sont iid de fonction génératrice des moments finie. Alors on a :  $\mathbb{E}[\exp(tS_n)] = [\mathbb{E}[\exp(tX_i)]]^n$

On applique ce lemme aux inégalité de Chernoff pour obtenir :

#### Lemme 7.

Soit  $S_n = X_1 + \dots + X_n$ , où les  $\{X_i\}$  sont iid de fonction génératrice des moments  $mgf_{X_i}(t)$  finie pour  $t \in [a, b]$ , où  $a$  est négatif et  $b$  positif. Alors on a :

$$\begin{aligned}\mathbb{P}(S_n \geq \alpha n) &\leq \left( \frac{mgf_{X_i}(t)}{\exp(t\alpha)} \right)^n, \quad \forall t \in [0, b] \\ \mathbb{P}(S_n \leq \alpha n) &\leq \left( \frac{mgf_{X_i}(t)}{\exp(t\alpha)} \right)^n, \quad \forall t \in [a, 0]\end{aligned}$$

Supposons que  $X_i \sim f_X$ . Le but est d'utiliser la fonction génératrice des moments pour obtenir un meilleur algorithme de rejet. Pour cela, il doit être possible de générer une réalisation selon la densité  $g_t(x) \propto e^{tx} f_X(x)$ . Lorsque  $t$  est grand, cette densité aura tendance à prendre de plus grandes valeurs. Lorsque  $t$  est grand dans les négatifs,  $g_t$  tendra à prendre des valeurs plus faibles.

Soit  $t > 0$ . Pour  $x \geq a$ , alors  $g_t(x) = e^{tx} f_X(x) \geq e^{ta} f_X(x)$ . Alors une réalisation de  $g_t$  peut être acceptée comme réalisation de  $f_X$  avec probabilité  $e^{ta}/e^{tx}$ . Si la probabilité que  $x$  soit bien plus grand que  $a$  est faible, alors la probabilité d'acceptation sera très proche de 1. Une méthode similaire s'applique lorsque  $t < 0$ . On en déduit alors l'algorithme suivant :

---

### Méthode rejet Inégalités de Chernoff

---

Entrées :  $f_X, a, t$

Sortie :  $S_n = X_1 + \dots + X_n$  sachant  $S_n \geq a$  (lorsque  $t > 0$ ) ou sachant  $S_n \leq a$  (lorsque  $t < 0$ )

---

Si  $t > 0$ , alors  $A \leftarrow [a, \infty)$ , sinon  $(-\infty, a]$

**while**  $C \neq 1$  **do**

    Tirer  $X_1, \dots, X_n$  iid selon la densité non-normalisée  $e^{tx} f(x)$

$S_n \leftarrow X_1 + \dots + X_n$

    Tirer  $C \sim \text{Bern}(\exp(t(a - S_n)) \mathbf{1}(S_n \in A))$

**end while**

---

### Lemme 8.

Supposons que  $mgf_{X_i}(t) \exp(-ta/n) < 1$ . Alors l'algorithme précédent génère une réalisation de  $[S_n | S_n > a]$  lorsque  $t > 0$  ou de  $[S_n | S_n < a]$  lorsque  $t < 0$ .

*Démonstration.* On considère tout vecteur  $(x_1, \dots, x_n)$  tel que  $\sum x_i = s$ . On considère  $X_1, \dots, X_n$  iid de densité  $f_X(x)$  et  $X'_1, \dots, X'_n$  iid de densité  $\exp(tx) f_X(x)$ . Alors la densité de  $S'_n = X'_1 + \dots + X'_n$  est simplement la densité de  $S_n = X_1 + \dots + X_n$  avec un facteur  $\exp(tx_1) \exp(tx_2) \dots \exp(tx_n) = \exp(ts)$ .

On génère dans l'algorithme une variable aléatoire de densité  $g(s) = \exp(ts) f_{S_n}(s)$  et la densité ciblée est  $f(s) = f_{S_n}(s) \mathbf{1}(s \in A)$ . On a alors  $f(s)/g(s)$  qui est soit 0 lorsque  $s \notin A$  soit  $\exp(-ts)$  lorsque  $s \in A$ .

Si  $A = [a, \infty)$ , alors  $t > 0$  et  $\exp(-ts) \leq \exp(-ta)$ . De même, lorsque  $A = (-\infty, a]$ , alors  $t < 0$  et  $\exp(-ts) \leq \exp(-ta)$ . On choisit donc  $c = \exp(-ta)$  et on obtient  $f(s)/[cg(s)] = \exp(ta) \exp(-ts) \mathbf{1}(s \in A)$  de même que dans l'algorithme.  $\square$

### 3 Chapitre 3

#### Définition : Fonction de mise à jour

On dit que  $\phi : \Omega \times [0, 1] \rightarrow \Omega$  est une fonction de mise à jour pour une chaîne de Markov  $\{X_t\}$  si, pour  $U \sim \text{Unif}[0, 1]$ ,  $[X_{t+1}|X_t] \sim \phi(X_t, U)$ .

La fonction  $\phi$  est déterministe : tout l'aléatoire est contenu dans la variable  $U$ .

Toute chaîne qui peut-être simulée sur ordinateur est un exemple de fonction de mise à jour.

Une même chaîne de Markov peut-être représentée par plusieurs fonctions de mise à jour : la fonction de mise à jour n'est pas forcément unique.

À l'aide d'une fonction de mise à jour  $\phi$ , on peut représenter la trajectoire d'une chaîne de Markov  $\{X_t\}$ . En effet, soient  $U_0, U_1, U_2, \dots$  iid  $\sim \text{Unif}([0, 1])$ . Pour un état initial  $x_0$ , on a :  $X_1 = \phi(x_0, U_0)$  puis pour  $i > 1$ ,  $X_i = \phi(X_{i-1}, U_{i-1})$ .

On notera alors la trajectoire jusqu'au temps  $t$  sous la forme :

$$\phi_t(x_0, U) = \phi(\phi(\phi(\dots(\phi(x_0, U_0), U_1), \dots, U_{t-1}))$$

Ensuite, pour n'importe quels états  $x_0$  et  $y_0$  dans  $\Omega$ , on définit pour une fonction de mise à jour  $\phi : X_t = \phi_t(x_0, U)$  et  $Y_t = \phi_t(y_0, U)$  (en utilisant les mêmes valeurs de  $U$ ). On appelle ce procédé un couplage. Notons qu'avec ce couplage, si il existe  $t \geq 0$  tel que  $X_t = Y_t$ , alors on dit que les processus ont fusionné (ou se sont rejoints, etc).

#### Définition : Couplage

Soit  $\mathcal{S}$  un ensemble de processus stochastiques définis sur un même ensemble  $\mathcal{I}$  et un même espace d'états  $\Omega$ . Si il existe un indice  $i \in \mathcal{I}$  et un état  $x \in \Omega$  tels que pour tout  $S \in \mathcal{S}$ , on aie  $S_i = x$ , alors on dit que les processus stochastiques ont fusionné (ou se sont rejoints, couplés, etc).

#### Exemple

Soit  $\Omega = \{0, 1, 2\}$ . Soit une fonction de mise à jour  $\phi$  telle que :

$$\phi(x, U) = x + \mathbf{1}(x < 2, U > \frac{1}{2}) - \mathbf{1}(x > 0, U \leq \frac{1}{2}).$$

Soient  $\{X_t\}$  et  $\{Y_t\}$  deux chaînes de Markov ayant  $\phi$  comme fonction de mise à jour et telles que  $X_0 = 0$  et  $Y_0 = 2$ . On suppose que  $U_0 = 0.64$ ,  $U_1 = 0.234$  et  $U_2 = 0.1$ . On a donc les trajectoires suivantes :  $(X_0, X_1, X_2, X_3) = (0, 1, 0, 0)$  et  $(Y_0, Y_1, Y_2, Y_3) = (2, 2, 1, 0)$ .

Les deux chaînes ont donc fusionné au temps  $t = 3$ .



### **Théorème CFTP (Coupling From The Past)**

On suppose que  $\phi$  est une fonction de mise à jour pour une chaîne de Markov définie sur  $\Omega$ , telle que pour  $U = (U_0, U_{-1}, \dots, U_{-t-1}) \sim \text{Unif}([0, 1]^t)$  on ait :

- Pour  $Y \sim \pi, \phi(Y, U_0) \sim \pi$
- Il existe un ensemble  $A \subseteq [0, 1]^t$  tel que  $\mathbb{P}(U \in A) > 0$  et  $\phi_t(\Omega, A) = \{x\}$  pour un certain  $x \in \Omega$

Posons alors, pour tout  $x \in \Omega$ ,

$$Y_0 = \phi_t(x_0, U_0)\mathbf{1}(U_0 \in A) + \phi_t(\phi_t(x_0, U_{-1}), U_0)\mathbf{1}(U_{-1} \in A) + \\ \phi_t(\phi_t(\phi_t(x_0, U_{-2}), U_{-1}), U_0)\mathbf{1}(U_{-2} \in A) + \dots$$

Alors  $Y_0 \sim \pi$ .

#### **Preuve**

Soit  $x_0 \in \Omega$ . Le résultat est immédiat en utilisant le théorème fondamental de la simulation parfaite, en posant  $g(U) = \phi_t(x_0, U), b(U) = \mathbf{1}(U \in A)$ , et  $f(X, U) = \phi_t(X, U)$ .