



HAL
open science

Méthodes de Monte-Carlo

Sylvain Rubenthaler

► **To cite this version:**

| Sylvain Rubenthaler. Méthodes de Monte-Carlo. Master. France. 2019. hal-02429146

HAL Id: hal-02429146

<https://cel.hal.science/hal-02429146>

Submitted on 6 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Méthodes de Monte-Carlo
(Cours et exercices)
M1 IM, 2018-2019



Sylvain Rubenthaler

Table des matières

Préface	iii
Chapitre 1. Introduction	1
1.1. Description de la méthode	1
1.2. Théorèmes de convergence	2
1.3. Exemples	3
1.4. Comparaison avec les méthodes déterministes	4
1.5. Exercices	5
Chapitre 2. Simulation de variables aléatoires	9
2.1. Inversion de la fonction de répartition	9
2.2. Simulation d'une loi exponentielle	11
2.3. Simulation de variables gaussiennes (algorithme de Box-Müller)	12
2.4. Simulation d'une variable aléatoire poissonnienne	13
2.5. Méthode de rejet	14
2.6. Vérification par histogramme	16
2.7. Exercices	18
Chapitre 3. Réduction de variance	21
3.1. Échantillonnage préférentiel (« importance sampling » en anglais)	21
3.2. Variable de contrôle	22
3.3. Variables antithétiques	23
3.4. Méthode de stratification	24
3.5. Valeur moyenne ou conditionnement	26
3.6. Exercices	27
Chapitre 4. Méthodes de Monte-Carlo par chaînes de Markov	29
4.1. Rappels sur les chaînes de Markov	29
4.2. Algorithme de Hastings-Metropolis	30
4.3. Algorithme de Metropolis simple	32
4.4. Le modèle d'Ising	33
4.5. Analyse bayésienne d'image	35
4.6. Cryptographie	37
4.7. Exercices	38
Annexe A. Table de la loi normale	41
Annexe B. Fonctions, intégrales et sommes usuelles	43
Bibliographie	45
Liste des symboles	47
Index	49

Préface

Ce polycopié est une introduction aux méthodes de Monte-Carlo. Les prérequis sont : cours de L3 MASS de probabilités¹, cours de M1 IM sur les chaînes de Markov, notions de \mathbf{R} (acquises au premier semestre). Les sources d'inspiration de ce document sont les suivantes : [DB01, Par08], le polycopié de Laure Élie et Bernard Lapeyre², le polycopié de Bernard Ycart³, des exercices fournis par François Delarue, un exposé de Persi Diaconis sur la cryptographie. Les TP se feront en \mathbf{R} , les exemples de programmes seront aussi donnés en \mathbf{R} .

Important : les fichiers sources et les corrigés des exercices sont disponibles sur : <http://math.unice.fr/~rubentha/cours.html>. J'encourage toute personne enseignant les méthodes de Monte-Carlo à utiliser ces fichiers et à ajouter son nom à la liste d'auteurs de ce polycopié.

Remerciements : Mlle Otgonbaatar (M1 MPA 2018-2019) m'a aidé à corriger des erreurs dans la section sur le modèle d'Ising.

1. <http://math.unice.fr/~rubentha/enseignement/poly-integration-probas.pdf>
2. <http://cermics.enpc.fr/~bl/PS/SIMULATION-X/poly-monte-carlo-x.pdf>
3. <http://ljk.imag.fr/membres/Bernard.Ycart/polys/montec.pdf>

Introduction

1.1. Description de la méthode

Supposons que l'on veuille calculer une quantité I . La première étape est de la mettre sous forme d'une espérance $I = \mathbb{E}(X)$ avec X une variable aléatoire. Si on sait simuler des variables X_1, X_2, \dots indépendantes et identiquement distribuées (ce que nous noterons i.i.d. dans la suite), alors nous pouvons approcher I par

$$(1.1) \quad I \approx \frac{X_1 + X_2 + \dots + X_N}{N}$$

avec N « grand », sous réserve d'application de la loi des grands nombres. C'est ce type d'approximation que l'on appelle « méthode de Monte-Carlo » (c'est une méthode pour faire des calculs numériques).

EXEMPLE 1.1. Supposons que l'on cherche à calculer

$$I = \int_{[0;1]^d} f(u_1, \dots, u_d) du_1 \dots du_d,$$

avec f bornée. Nous posons $X = f(U_1, \dots, U_d)$ où les U_1, \dots, U_d sont des variables aléatoires indépendantes suivant toutes une loi uniforme sur $[0; 1]$. Alors :

$$I = \mathbb{E}(f(U_1, \dots, U_d)) = \mathbb{E}(X).$$

Nous avons donc réalisé la première étape. Tout logiciel de programmation nous permet de simuler des variables uniformes sur $[0; 1]$. De plus, des appels successifs de variables uniformes renvoient des variables indépendantes. Nous pouvons donc facilement simuler $U_1^{(1)}, \dots, U_d^{(1)}, U_1^{(2)}, \dots, U_d^{(2)}, \dots$ des variables i.i.d. de loi uniforme sur $[0; 1]$ (nous noterons cette loi $\mathcal{U}([0; 1])$) et donc $X_1 = (U_1^{(1)}, \dots, U_d^{(1)})$, $X_2 = (U_1^{(2)}, \dots, U_d^{(2)})$, ... sont i.i.d. Le programme 1.1 fournit un exemple de programme approchant I par une méthode de Monte-Carlo (avec $d = 3$, $n = 1000$ itérations, $f(u_1, u_2, u_3) = \sin(u_1) \sin(u_2) \sin(u_3)$).

Programme 1.1 approximation de Monte-Carlo

```
d=3
n=1000
s=0
for (i in 1:n)
{
  u=runif(d,0,1) #simulation de d variables uniformes indépendantes
  s=s+sin(u[1])*sin(u[2])*sin(u[3])
}
print(s/n)
```

Si nous cherchons à évaluer une intégrale de la forme

$$I = \int_{\mathbb{R}^n} g(x)f(x)dx$$

avec f une densité de probabilité (c'est à dire f positive et $\int_{\mathbb{R}^n} f(x)dx = 1$), alors nous pouvons écrire $I = \mathbb{E}(g(X))$ avec X de loi de densité f . Nous sommes encore dans la situation où nous

voulons calculer l'espérance d'une variable aléatoire (si X est une variable aléatoire, alors $g(X)$ est une variable aléatoire, à condition que g soit mesurable).

1.2. Théorèmes de convergence

1.2.1. Convergence.

THÉORÈME 1.2 (Loi forte des grands nombres.). *Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires i.i.d., à valeurs dans \mathbb{R}^d ($d \in \mathbb{N}^*$). On suppose que $\mathbb{E}(|X_1|) < +\infty$. Alors*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n \rightarrow +\infty]{p.s.} \mathbb{E}(X).$$

Rappel : « p.s. » signifie « presque sûrement ».

Ce théorème nous dit pourquoi l'approximation de Monte-Carlo (1.1) est valide (et sous quelles hypothèses).

1.2.2. Vitesse de convergence.

THÉORÈME 1.3 (Théorème de la limite centrale.). *Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires i.i.d., à valeurs dans \mathbb{R} . On suppose que $\mathbb{E}(X_1^2) < +\infty$. Soit $\sigma^2 = \mathbb{V}(X)$ (la variance de X). Alors*

$$\frac{\sqrt{n}}{\sigma} \left(\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}(X_1) \right) \xrightarrow[n \rightarrow +\infty]{loi} \mathcal{N}(0, 1).$$

Rappels : Le symbole $\mathcal{N}(\mu, \sigma^2)$ se rapporte à la loi normale (on dit aussi gaussienne) de moyenne μ et de variance σ^2 . La convergence en loi d'une suite de variables réelles $Z_n \xrightarrow[n \rightarrow +\infty]{loi} Z$ (Z de loi de densité ϕ) signifie que pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ mesurable, bornée,

$$\mathbb{E}(f(Z_n)) \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}(f(Z)) = \int_{\mathbb{R}} f(x)\phi(x)dx.$$

Ceci implique en particulier, sous les hypothèses du théorème ci-dessus, pour tout $a < b$,

$$\begin{aligned} \mathbb{P} \left(a \leq \frac{\sqrt{n}}{\sigma} \left(\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}(X_1) \right) \leq b \right) \\ = \mathbb{E} \left(\mathbb{1}_{[a;b]} \left(\frac{\sqrt{n}}{\sigma} \left(\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}(X_1) \right) \right) \right) \\ \xrightarrow[n \rightarrow +\infty]{} \int_a^b 1 \times \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx. \end{aligned}$$

Si nous cherchons à approcher $\mathbb{E}(X)$ par une méthode de Monte-Carlo ($\mathbb{E}(X) \approx \frac{X_1 + \cdots + X_n}{n}$), nous pouvons donc dire que l'erreur $\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}(X)$ est d'ordre $\frac{\sigma}{\sqrt{n}}$. Mais cette erreur est aléatoire (le théorème ci-dessus nous dit que $\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}(X)$ est à peu près de loi $\mathcal{N}(0, \sigma^2/n)$), nous ne pouvons donc pas la borner. En revanche, nous pouvons donner un intervalle de confiance pour le résultat.

Toujours sous les hypothèse du théorème ci-dessus, notons $\epsilon_n = \frac{X_1 + \cdots + X_n}{n} - \mathbb{E}(X_1)$. On cherche à approcher $\mathbb{E}(X_1)$ à 0,01 près avec une confiance de 95% (on cherche à construire un intervalle de confiance). C'est à dire que l'on veut

$$\mathbb{P}(|\epsilon_n| \geq 0,01) \leq 0,05,$$

ce qui est équivalent à

$$\mathbb{P}(|\epsilon_n| \leq 0,01) \geq 0,95.$$

Calculons :

$$\begin{aligned} \mathbb{P}(|\epsilon_n| \leq 0,01) &= \mathbb{P} \left(-0,01 \leq \frac{X_1 + \cdots + X_n}{n} - \mathbb{E}(X_1) \leq 0,01 \right) \\ &= \mathbb{P} \left(-0,01 \frac{\sqrt{n}}{\sigma} \leq \frac{\sqrt{n}}{\sigma} \left(\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}(X_1) \right) \leq 0,01 \frac{\sqrt{n}}{\sigma} \right) \end{aligned}$$

$$\begin{aligned}
(\text{pour } n \text{ "assez grand"}) &\approx \int_{-0,01\frac{\sqrt{n}}{\sigma}}^{+0,01\frac{\sqrt{n}}{\sigma}} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \\
(\text{par symétrie de } \mathcal{N}(0,1)) &= 2 \int_{-\infty}^{+0,01\frac{\sqrt{n}}{\sigma}} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt - 1.
\end{aligned}$$

On veut donc n tel que $\int_{-\infty}^{+0,01\frac{\sqrt{n}}{\sigma}} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \geq 0,975$. On regarde dans une table de la loi normale (voir par exemple l'annexe A) et on voit qu'il suffit pour cela d'avoir $0,01\sqrt{n}/\sigma \geq 1,96$, c'est à dire

$$n \geq \frac{(1,96 \times \sigma)^2}{0,01^2}.$$

On remarque que le nombre de tirages nécessaire n pour atteindre un certain niveau d'erreur avec une certaine confiance est une fonction linéaire de σ^2 .

Dans la pratique, σ^2 pourrait ne pas être connu, mais on peut l'estimer par une méthode de Monte-Carlo.

On dit que σ^2 est la variance de la méthode. Sous les hypothèses du théorème 1.3, la variance de l'erreur $\frac{X_1 + \dots + X_n}{n} - \mathbb{E}(X_1)$ est asymptotiquement σ^2/n . La variance σ^2 est aussi la variance des variables i.i.d. dont on fait la moyenne empirique.

LEMME 1.4 (Estimateur de la variance). *Si X_1, X_2, \dots sont i.i.d. avec $\mathbb{E}(X_1^2) < +\infty$ et $\sigma^2 = \mathbb{V}(X)$, alors*

$$\frac{(X_1 - X_2)^2 + (X_3 - X_4)^2 + \dots + (X_{2n-1} - X_{2n})^2}{2n} \xrightarrow[n \rightarrow +\infty]{p.s.} \sigma^2.$$

DÉMONSTRATION. Les variables $(X_1 - X_2)^2, (X_3 - X_4)^2, \dots$ sont i.i.d. avec $\mathbb{E}((X_1 - X_2)^2) = \mathbb{E}(X_1^2 + X_2^2 - 2X_1X_2) = 2\mathbb{E}(X_1^2) - 2\mathbb{E}(X_1)^2 < +\infty$ (car $\mathbb{E}(X_1^2) < +\infty$ et $\mathbb{E}(X_1)^2 < +\infty$). Donc le résultat découle de la loi des grands nombres. \square

1.3. Exemples

1.3.1. Cas facile. Soit U une variables aléatoire dans \mathbb{R}^d ($d \in \mathbb{N}^*$) et $f : \mathbb{R}^d \rightarrow \mathbb{R}$ mesurable. On cherche à calculer $p = \mathbb{P}(f(U) \leq \lambda)$ pour un certain λ (p est donc dans $[0; 1]$). Soit $X = \mathbb{1}_{f(U) \leq \lambda}$. Alors $p = \mathbb{E}(X)$. On peut donc approcher p par (en faisant des tirages i.i.d. X_1, X_2, \dots de même loi que X)

$$p_n = \frac{X_1 + \dots + X_n}{n} \approx p.$$

Nous calculons $\sigma^2 = \mathbb{V}(X) = p(1-p)$. Si nous voulons une erreur $|p - p_n|$ inférieure à 0,01 avec une confiance à 0,95% (comme dans l'exemple ci-dessus), il suffit de prendre

$$n \geq \frac{(1,96 \times \sigma)^2}{0,01^2}.$$

Pour tout $p \in [0; 1]$, $p(1-p) \leq 1/4$, donc il suffit de prendre

$$n \geq \frac{1,96^2}{0,01^2} \times \frac{1}{4}.$$

1.3.2. Cas difficile. On cherche à calculer $\mathbb{E}(e^{\beta Z})$ avec $Z \sim \mathcal{N}(0, 1)$. Nous savons que

$$\begin{aligned}
\mathbb{E}(e^{\beta Z}) &= \int_{-\infty}^{+\infty} e^{\beta x} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x - \beta)^2 + \frac{\beta^2}{2}\right) dx \\
&= e^{\beta^2/2}.
\end{aligned}$$

Si nous tirons Z_1, Z_2, \dots i.i.d. de même loi que Z , nous aurons au bout de n tirages :

$$\frac{e^{\beta Z_1} + \dots + e^{\beta Z_n}}{n} \approx \mathbb{E}(e^{\beta Z}) + \frac{\sigma}{\sqrt{n}} Y$$

avec $Y \sim \mathcal{N}(0, 1)$, $\sigma^2 = \mathbb{V}(e^{\beta Z})$ (d'après le théorème de la limite-centrale). Un calcul similaire à celui que nous venons de faire nous donne : $\sigma^2 = e^{2\beta^2} - e^{\beta^2}$. En ce qui concerne l'erreur relative

$$\frac{1}{\mathbb{E}(e^{\beta Z})} \left| \frac{e^{\beta Z_1} + \dots + e^{\beta Z_n}}{n} - \mathbb{E}(e^{\beta Z}) \right| \approx \frac{\sigma}{\mathbb{E}(e^{\beta Z})\sqrt{n}} \mathbb{E}(|Y|) = \frac{\sigma}{\mathbb{E}(e^{\beta Z})\sqrt{n}} \sqrt{\frac{2}{\pi}}.$$

Nous avons $\frac{\sigma}{\mathbb{E}(e^{\beta Z})} = \sqrt{e^{\beta^2} - 1}$. Par exemple, si $\beta = 5$, si nous voulons une erreur relative de l'ordre de 1, il nous faut prendre n tel que

$$\frac{\sqrt{n}}{\sqrt{e^{\beta^2} - 1} \sqrt{\frac{2}{\pi}}} \geq 1,96,$$

c'est à dire n de l'ordre de 4×10^{11} , ce qui n'est pas réalisable dans la pratique. C'est pourquoi il est important de réduire la variance (voir le chapitre 3).

1.3.3. Les aiguilles de Buffon. On jette une grande quantité d'aiguilles de même longueur sur un parquet. Le parquet est constitué de planches parallèles de même largeur (et de longueur supposée infinie). On calcule la proportion d'aiguilles tombant sur une rainure du parquet (voir une animation sur un site de l'INRIA ¹). On en déduit une approximation de π (voir la formule sur le site de l'INRIA). Voir une variation avec lancer de saucisses ².

1.4. Comparaison avec les méthodes déterministes

Replaçons-nous dans le cadre de l'exemple 1.1. Nous pouvons également approcher l'intégrale en question par des sommes de Riemann :

$$\int_{[0;1]^d} f(x) dx = \lim_{n \rightarrow +\infty} \frac{1}{n^d} \sum_{1 \leq i_1, \dots, i_d \leq n} f\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right).$$

Pour comparer cette approximation avec l'approximation de Monte-Carlo, il convient de comparer les temps de calcul nécessaires pour atteindre une précision fixée, ou de comparer les précisions atteintes pour des temps de calcul égaux.

LEMME 1.5. *Si f est lipschitz alors*

$$\left| \int_{[0;1]^d} f(x) dx - \frac{1}{n^d} \sum_{1 \leq i_1, \dots, i_d \leq n} f\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right) \right| \leq \frac{C}{n}$$

pour une certaine constante C .

DÉMONSTRATION. La propriété lipschitz veut dire : il existe une constante C' telle que, pour tout $x, y \in [0;1]^d$, $|f(x) - f(y)| \leq C' \|x - y\|$. Nous avons donc

$$\begin{aligned} & \left| \int_{[0;1]^d} f(x) dx - \frac{1}{n^d} \sum_{1 \leq i_1, \dots, i_d \leq n} f\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right) \right| \\ &= \left| \sum_{1 \leq i_1, \dots, i_d \leq n} \int_{u_1 \in [\frac{i_1-1}{n}; \frac{i_1}{n}]} \dots \int_{u_d \in [\frac{i_d-1}{n}; \frac{i_d}{n}]} f(x) - f\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right) du_1 \dots du_d \right| \\ &\leq \sum_{1 \leq i_1, \dots, i_d \leq n} \int_{u_1 \in [\frac{i_1-1}{n}; \frac{i_1}{n}]} \dots \int_{u_d \in [\frac{i_d-1}{n}; \frac{i_d}{n}]} \left| f(x) - f\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right) \right| du_1 \dots du_d \\ &\leq \sum_{1 \leq i_1, \dots, i_d \leq n} \int_{u_1 \in [\frac{i_1-1}{n}; \frac{i_1}{n}]} \dots \int_{u_d \in [\frac{i_d-1}{n}; \frac{i_d}{n}]} C' \frac{\sqrt{d}}{n} du_1 \dots du_d = C' \frac{\sqrt{d}}{n}. \end{aligned}$$

□

1. <http://www-sop.inria.fr/mefisto/java/tutorial11/node14.html#SECTION00033110000000000000>
2. <https://www.youtube.com/watch?v=Q3jVk6k6CGY>

Pour calculer la somme $\sum_{1 \leq i_1, \dots, i_d \leq n} \dots$, nous devons exécuter d boucles emboîtées de longueur n (et nous sommes donc n^d termes). Nous dirons donc que le coût du calcul est d'ordre n^d . La précision est elle d'ordre n^{-1} . Si nous notons $N = n^d$ le coût du calcul, nous avons donc une précision en $N^{-1/d}$.

Dans le programme 1.1, le coût du calcul est d'ordre n (il y a n boucles) pour une précision d'ordre $n^{-1/2}$ (d'après le théorème 1.3, si $\mathbb{E}(X^2) < \infty$). Si nous notons $N = n$ le coût du calcul, nous avons donc une précision en $N^{-1/2}$.

- REMARQUE 1.6. — On voit donc que la méthode de Monte-Carlo n'est avantageuse numériquement qu'à partir de la dimension $d = 3$. Même si la plupart des exercices et exemples de ce livre sont donnés en dimension 1 par souci de simplicité, il convient de ne pas oublier ce fait dans la pratique
- Les hypothèses permettant d'appliquer l'une ou l'autre des méthodes sont différentes (moins contraignantes dans le cas Monte-Carlo).
 - La méthode de Monte-Carlo est facile à implémenter, c'est ce qui fait son succès.
 - Il existe des méthodes déterministes dites « quasi Monte-Carlo » qui sont plus rapides que les méthodes de Monte-Carlo simples (mais qui demandent plus d'hypothèses et qui sont plus difficiles à implémenter). Le lecteur pourra se reporter au polycopié de Laure Élie et Bernard Lapeyre cité dans l'introduction.

1.5. Exercices

EXERCICE 1.1. Donner au moins deux méthodes de Monte-Carlo différentes d'approximation de l'intégrale

$$I_1 = \int_0^1 \cos(x^3) \exp(-x) dx,$$

à l'aide d'une moyenne empirique impliquant des variables aléatoires de loi connue, puis coder chaque méthode en R.

Mêmes questions avec

$$I_2 = \int_0^{+\infty} \sin(x^4) \exp(-2x) \exp\left(-\frac{x^2}{2}\right) dx,$$

et

$$I_3 = \int_0^1 \ln(1+x^2) \exp(-x^2) dx.$$

EXERCICE 1.2. Soient X, X_1, X_2, \dots des variables i.i.d uniformes dans $C = [-1; 1] \times [-1; 1]$ (sous-entendu : $X = (U_1, U_2)$ avec U_1 et U_2 indépendantes et de loi uniforme sur $[-1; 1]$ toutes les deux). Soit D le disque (dans \mathbb{R}^2) de centre $(0, 0)$ et de rayon 1.

- (1) Calculer par une méthode de Monte-Carlo la valeur de $\mathbb{E}(\mathbb{1}_{X \in D})$ (coder en R).
- (2) Estimer la variance de la méthode (coder en R).
- (3) Trouver n un nombre de boucles à effectuer en (1) pour que l'estimation de l'espérance soit précise à 0,01 près avec une probabilité $\geq 0,95$.

EXERCICE 1.3. On s'intéresse à

$$I = \int_0^{+\infty} \sqrt{x} e^{-x} dx.$$

- (1) Donner deux méthodes de Monte-Carlo différentes pour calculer I (de manière approchée).
- (2) Écrire un programme en R qui calcule I (de manière approchée).

EXERCICE 1.4. On s'intéresse à $I = \int_0^1 \sin(\sqrt{x}) dx$.

- (1) Proposer une méthode de Monte-Carlo pour calculer I (de manière approchée). Dans la suite, on notera σ^2 la variance de cette méthode.
- (2) Écrire un programme en R qui calcule I (de manière approchée).

- (3) Trouver un nombre de boucle n (qui s'écrit comme une fonction de σ^2) telle que la méthode ci-dessus approche I à 0,01 près avec une probabilité $\geq 0,99$ (il n'est pas nécessaire de simplifier l'expression trouvée)

EXERCICE 1.5. Aiguilles de Buffon. On jette des aiguilles de longueur 2 dm sur un parquet dont les lattes sont espacées de 2 dm. Voir la figure .

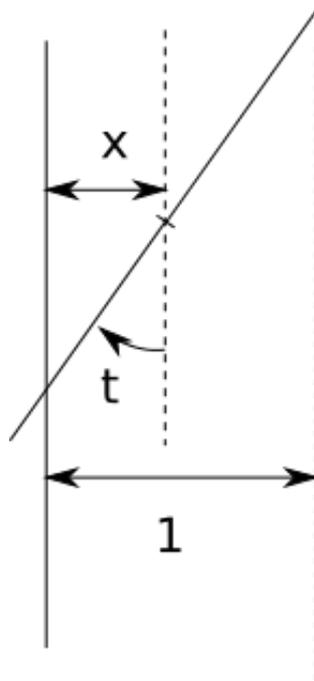


FIGURE 1.5.1. Aiguille de Buffon

- (1) Calculer la probabilité qu'une aiguille tombe sur une rainure du parquet. On pourra remarquer qu'en utilisant les symétries du problème, on peut se ramener au problème suivant avec une seule rainure : la distance du centre de l'aiguille à la rainure est une variable aléatoire X de loi uniforme dans $[0; 1]$, l'angle Θ (t sur la figure) que forme l'aiguille avec la direction des rainures est une variable uniforme dans $[0; \pi/2]$.
- (2) En déduire un programme qui calcule π de manière approchée par une méthode de Monte-Carlo.

EXERCICE 1.6. Soit $(U_n^1, \dots, U_n^3)_{n \geq 1}$ une famille de v.a. indépendantes et identiquement distribuées de loi uniforme sur $]0; 1[^3$. Donner la limite presque-sûre de

$$n^{-1} \sum_{i=1}^n \mathbf{1}_{(U_i^1)^2 + (U_i^2)^2 + (U_i^3)^2 < 1}.$$

Donner une méthode similaire pour retrouver le volume d'une boule de rayon 2.

EXERCICE 1.7 (Loi des grands nombres.). Étant donnée une suite de variables aléatoires $(X_n)_{n \geq 1}$ identiquement distribuées et indépendantes, telles que $\mathbb{E}[|X_1|^4] < +\infty$ et $\mathbb{E}(X_1) = 0$, on pose

$$S_n = \sum_{k=1}^n X_k.$$

- (1) Montrer qu'il existe une constante $C > 0$ telle que, pour tout $n \geq 1$,

$$\mathbb{E}[S_n^4] \leq Cn^2.$$

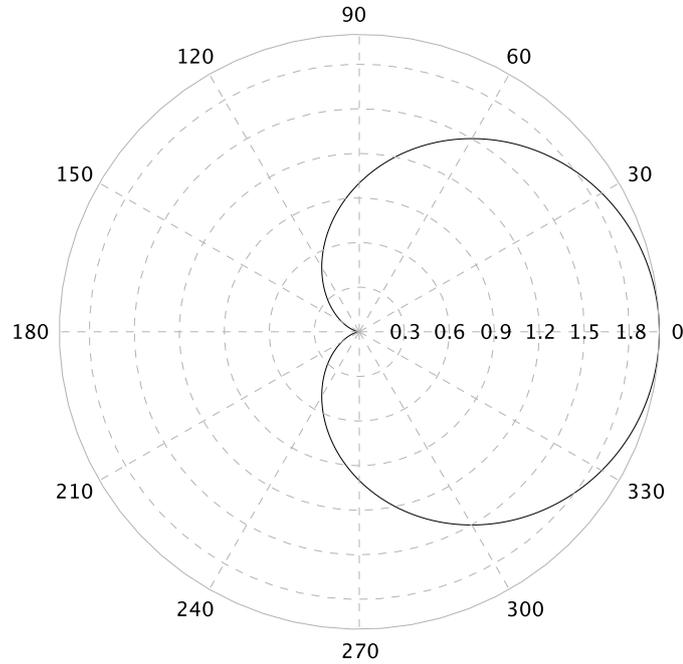


FIGURE 1.5.2. Cardioïde

- (2) En utilisant l'inégalité de Markov à l'ordre 4, en déduire qu'il existe une constante $c > 0$ telle que, pour tout $n \geq 1$,

$$\mathbb{P}\{|S_n| \geq n^{5/6}\} \leq cn^{-4/3}.$$

- (3) En utilisant le lemme de Borel-Cantelli, en déduire que, presque-sûrement, $S_n/n \rightarrow 0$.

EXERCICE 1.8 (Densité des polynômes dans $\mathcal{C}([0; 1])$). Étant donnée une fonction f continue sur le segment $[0, 1]$, on définit, pour tout rang $n \geq 1$, la fonction polynomiale

$$\mathcal{P}_n(x) = \sum_{k=0}^n [C_n^k x^k (1-x)^{n-k} f(\frac{k}{n})].$$

- (1) En introduisant, pour tout $x \in [0, 1]$, $(\varepsilon_n(x))_{n \geq 1}$, suite de variables aléatoires indépendantes de même loi de Bernoulli de paramètre x , et en posant $S_n(x) = \varepsilon_1(x) + \dots + \varepsilon_n(x)$, interpréter $\mathcal{P}_n(x)$ comme une espérance.
- (2) En utilisant l'inégalité de Bienaymé-Chebychev, montrer que

$$\forall \varepsilon > 0, \quad \mathbb{P}\left\{\left|\frac{S_n(x)}{n} - x\right| \geq \varepsilon\right\} \leq \frac{1}{4n\varepsilon^2}.$$

- (3) En déduire que

$$\lim_{n \rightarrow +\infty} \sup_{x \in [0, 1]} |f(x) - \mathcal{P}_n(x)| = 0.$$

EXERCISE 1.9. La figure 1.5.2 représente une cardioïde (section de pomme) d'équation $(x^2 + y^2 - x)^2 = (x^2 + y^2)$.

- (1) L'intérieur de la cardioïde est donnée par l'inéquation $(x^2 + y^2 - x)^2 < (x^2 + y^2)$.
En déduire une méthode probabiliste d'approximation de la surface intérieure de la cardioïde.
- (2) On obtient un solide de révolution en faisant tourner l'intérieur de la cardioïde autour de l'axe des x (on obtient quelque chose qui ressemble à une pomme). Proposer une méthode probabiliste d'approximation du volume de ce solide.
- (3) En fait, la cardioïde admet aussi une description paramétrique en coordonnées polaires : $\rho = 1 + \cos(\theta)$. En déduire que le volume d'une pomme est $8\pi/3$.

EXERCISE 1.10 (Plus difficile). Soient X_1, \dots, X_n des v.a. i.i.d. suivant une loi exponentielle de paramètre 1, de somme notée S_n .

- (1) En utilisant le TCL, calculer la limite quand n tend vers l'infini de

$$\mathbb{P}\left\{a \leq \sqrt{n}\left(\frac{S_n}{n} - 1\right) \leq b\right\},$$

pour a et b deux réels.

- (2) On rappelle que S_n suit une loi $\Gamma(n, 1)$, i.e. a pour densité la fonction

$$x \mapsto \mathbf{1}_{[0, +\infty)}(x) \frac{x^{n-1}}{\Gamma(n)} \exp(-x),$$

calculer directement la même quantité que ci-dessus.

- (3) Retrouver la formule de Stirling

$$n! \underset{n \rightarrow +\infty}{\sim} \sqrt{2\pi n} \left(\frac{n}{e}\right)^n.$$

(On rappelle que $\Gamma(n) = (n-1)!$.)

Simulation de variables aléatoires

Comme nous l'avons vu dans le chapitre précédent, pour appliquer une méthode de Monte-Carlo, il faut savoir simuler suivant une loi donnée. Nous supposons ici que nous disposons d'un générateur de nombres aléatoires de loi $\mathcal{U}([0; 1])$. Le langage **R** contient les générateurs d'un grand nombre de lois classiques.

REMARQUE 2.1. Les ordinateurs ne disposent pas de composant au comportement aléatoire¹. Quand on demande à un programme de générer des nombres aléatoires de loi $\mathcal{U}([0; 1])$, il renvoie des nombres pris dans une liste à la suite les uns des autres. En général, cette suite est suffisamment désordonnée et longue pour avoir un comportement similaire à celui d'une suite aléatoire. Dans le cas d'un programme écrit dans un langage de bas niveau (type **C++**), l'utilisateur aura intérêt à construire lui-même une procédure de génération de nombres aléatoires (la liste utilisée par le langage étant trop peu désordonnée).

2.1. Inversion de la fonction de répartition

LEMME 2.2. Soit X une v. a. r. (variable aléatoire réelle) de fonction de répartition F . Posons, pour $0 \leq t \leq 1$,

$$F^{-1}(t) = \inf\{x, F(x) \geq t\}.$$

Alors, si $U \sim \mathcal{U}([0; 1])$ (veut dire : « U suit la loi $\mathcal{U}([0; 1])$ »), alors $F^{-1}(U)$ a même loi que X .

DÉMONSTRATION. Attention, si F est continue, F^{-1} est l'inverse de F mais ce n'est pas le cas si F n'est pas continue (puisque F n'a alors pas d'inverse). La fonction F^{-1} ainsi définie s'appelle le pseudo-inverse. Voir les figures 2.1.1 et 2.1.2 pour un exemple (pour $t \in [1/3; 2/3]$, $\{x, F(x) \geq t\} = [1; +\infty[$ et donc $F^{-1}(t) = 1$). Soit maintenant $u \in [0; 1]$ et $t \in \mathbb{R}$.

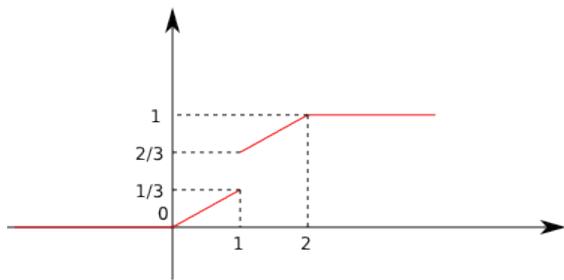


FIGURE 2.1.1. Fonction de répartition F .

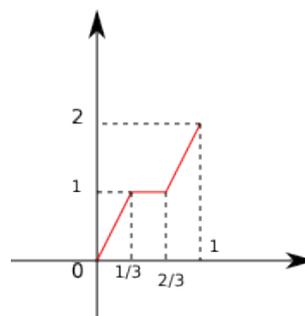


FIGURE 2.1.2. Fonction F^{-1} .

— Si $u \leq F(t)$ alors $t \in \{x, F(x) \geq u\}$ donc $F^{-1}(u) \leq t$.

1. Même un jeu de pile ou face ne se comporte pas aussi « aléatoirement » que nous le voudrions, voir <https://www.npr.org/templates/story/story.php?storyId=1697475> sur la machine de Persi Diaconis.

- Supposons que $F^{-1}(u) \leq t$. Soit $\epsilon > 0$. Il existe $x \leq t + \epsilon$ tel que $F(x) \geq u$. La fonction F est croissante donc $F(x) \leq F(t + \epsilon)$, donc $u \leq F(t + \epsilon)$. Ceci est vrai pour tout $\epsilon > 0$ et la fonction F est continue à droite (parce que c'est une fonction de répartition) donc $u \leq F(t)$.

Nous avons donc

$$[u \leq F(t)] \Leftrightarrow [F^{-1}(u) \leq t].$$

Nous avons alors, pour tout $t \in \mathbb{R}$,

$$\mathbb{P}(F^{-1}(U) \leq t) = \mathbb{P}(U \leq F(t)) = F(t).$$

Donc $F^{-1}(U)$ a la même fonction de répartition que X (et donc ils ont la même loi). \square

EXEMPLE 2.3. Dans le cas de la fonction F donnée dans les figures 2.1.1, 2.1.2, nous avons

$$F(x) = \begin{cases} \frac{x}{3} & \text{si } x \in [0; 1[, \\ \frac{x}{3} + \frac{1}{3} & \text{si } x \in [1; 2], \\ 1 & \text{si } x \geq 2, \\ 0 & \text{si } x \leq 0, \end{cases}$$

$$F^{-1}(t) = \begin{cases} 3t & \text{si } t \in [0; \frac{1}{3}], \\ 1 & \text{si } t \in [\frac{1}{3}; \frac{2}{3}], \\ 3t - 1 & \text{si } t \in [\frac{2}{3}; 1]. \end{cases}$$

Le programme 2.1 renvoie une variable de fonction de répartition F (on parle ici de la variable affichée à la fin du programme).

Programme 2.1 Simulation par inversion de la fonction de répartition

```

finv<-function(u)
{
  if (u<1/3)
    { z=3*u }
  else if (u<2/3)
    { z=1 }
  else
    { z=3*u-1 }
  return(z)
}
u=runif(1,0,1) #simulation d'une variable uniforme (dans [0;1])
print(finv(u))

```

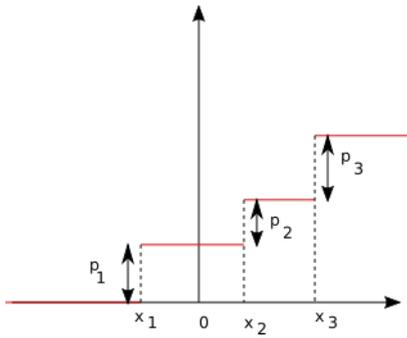
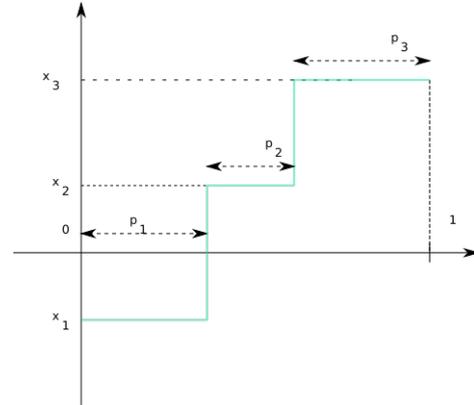
EXEMPLE 2.4. Soit $n \in \mathbb{N}^*$. Soient $x_1 < x_2 < \dots < x_n$ dans \mathbb{R} . Soient $p_1, p_2, \dots, p_n \in [0; 1]$ tels que $p_1 + p_2 + \dots + p_n = 1$. Soit X une variable aléatoire telle que $\mathbb{P}(X_i = x_i)$ pour tout $i \in \{1, 2, \dots, n\}$. La fonction de répartition de X est

$$F : t \in \mathbb{R} \mapsto \sum_{i=1}^n p_i \mathbb{1}_{t \geq x_i},$$

(voir un exemple de dessin avec $n = 3$ dans la Figure 2.1.3). Son pseudo-inverse est

$$F^{-1} : u \in [0; 1] \mapsto \sum_{j=1}^n x_j \mathbb{1}_{p_1 + \dots + p_{j-1} \leq u < p_1 + \dots + p_j},$$

(voir un exemple de dessin avec $n = 3$ dans la Figure 2.1.4). Le programme 2.2 renvoie une variable de fonction de répartition F (le tableau x contient les valeurs x_i et le tableau p contient les valeurs p_i).

FIGURE 2.1.3. Fonction de répartition F .FIGURE 2.1.4. Fonction F^{-1} .

Programme 2.2 Simulation par inversion de la fonction de répartition.

```

i=0
t=0
b=0
u=runif(1,0,1)
while (b==0)
{
  i=i+1
  t=t+p[i]
  if (t>u)
  { b=1 }
}
print(x[i])

```

2.2. Simulation d'une loi exponentielle

On rappelle que X suit une loi exponentielle de paramètre $\lambda (> 0)$ si pour tout $t \in \mathbb{R}_+$,

$$\mathbb{P}(X > t) = \exp(-\lambda t).$$

Nous noterons cette loi $\mathcal{E}(\lambda)$. Si F est la fonction de répartition de X , nous avons alors $F(t) = 1 - e^{-\lambda t}$ (pour $t \geq 0$) et

$$F^{-1}(x) = -\frac{\log(1-x)}{\lambda}.$$

Si $U \sim \mathcal{U}([0; 1])$, nous avons, d'après le résultat ci-dessus : $-\log(1-U)/\lambda \sim \mathcal{E}(\lambda)$. Remarquons que $1-U$ a même loi que U et donc

$$-\frac{\log(U)}{\lambda} \sim \mathcal{E}(\lambda).$$

Attention : nous écrivons log pour le logarithme népérien.

2.3. Simulation de variables gaussiennes (algorithme de Box-Müller)

LEMME 2.5. Si U, V sont de loi $\mathcal{U}([0; 1])$ et indépendantes alors

$$(\sqrt{-2 \log(U)} \cos(2\pi V), \sqrt{-2 \log(U)} \sin(2\pi V))$$

est un vecteur gaussien centré de matrice de covariance $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (donc les deux composantes sont indépendantes et de loi $\mathcal{N}(0, 1)$).

DÉMONSTRATION. On se rappelle que pour identifier la loi d'une variable aléatoire, il suffit de savoir calculer les espérances de fonctions de cette variables aléatoires (pour toute fonction dans un ensemble de fonctions tests, voir proposition 6.2.10 dans le polycopié de L3²). Nous prenons donc $f \in \mathcal{C}_b^+(\mathbb{R}^2)$ ($\mathcal{C}_b^+(\mathbb{R}^2)$ est l'ensemble des fonctions positives, continues, bornées de \mathbb{R}^2 dans \mathbb{R}). Notons X et Y deux variables indépendantes de loi $\mathcal{N}(0, 1)$. Nous avons :

$$\mathbb{E}(f(X, Y)) = \int_{\mathbb{R}^2} f(x, y) \frac{e^{-(x^2+y^2)/2}}{2\pi} dx dy.$$

Nous voulons effectuer le changement de variable $x = r \cos(\theta)$, $y = r \sin(\theta)$ ($r \in \mathbb{R}_+$, $\theta \in [0; 2\pi]$) (c'est le passage usuel en coordonnées polaires). Nous obtenons la matrice jacobienne

$$J_1(x, y) = \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -r \sin(\theta) & r \cos(\theta) \end{bmatrix},$$

de déterminant $\det(J_1) = r \cos^2(\theta) + r \sin^2(\theta) = r$. Donc

$$\mathbb{E}(f(X, Y)) = \int_{(r, \theta) \in \mathbb{R}_+ \times [0; 2\pi]} f(r \cos(\theta), r \sin(\theta)) \frac{e^{-r^2/2}}{2\pi} |r| dr d\theta.$$

Nous effectuons maintenant un nouveau changement de variable : $r = \sqrt{-2 \ln(u)}$, $\theta = 2\pi v$ ($(u, v) \in [0; 1]^2$). Nous obtenons la matrice jacobienne

$$J_2(r, \theta) = \begin{bmatrix} -\frac{2}{u} \frac{1}{2\sqrt{-2 \ln(u)}} & 0 \\ 0 & 2\pi \end{bmatrix},$$

de déterminant $\det(J_2) = \frac{2\pi}{u\sqrt{-2 \ln(u)}}$. Donc

$$\begin{aligned} \mathbb{E}(f(X, Y)) &= \int_{(u, v) \in [0; 1]^2} f(\sqrt{-2 \ln(u)} \cos(2\pi v), \sqrt{-2 \ln(u)} \sin(2\pi v)) \frac{u}{2\pi} \sqrt{-2 \ln(u)} \\ &\quad \times \frac{2\pi}{u\sqrt{-2 \ln(u)}} du dv \\ &= \int_{(u, v) \in [0; 1]^2} f(\sqrt{-2 \ln(u)} \cos(2\pi v), \sqrt{-2 \ln(u)} \sin(2\pi v)) du dv \\ &= \mathbb{E}(f(\sqrt{-2 \ln(U)} \cos(2\pi V), \sqrt{-2 \ln(U)} \sin(2\pi V))). \end{aligned}$$

Ceci est vrai pour toute fonction $f \in \mathcal{C}_b^+(\mathbb{R}^2)$ donc (X, Y) et

$$(\sqrt{-2 \ln(U)} \cos(2\pi V), \sqrt{-2 \ln(U)} \sin(2\pi V))$$

ont même loi. □

Pour simuler une variable X de loi $\mathcal{N}(0, 1)$, il suffit donc de prendre $U, V \sim \mathcal{U}([0; 1])$ indépendantes et poser $X = \sqrt{-2 \log(U)} \cos(2\pi V)$. Pour simuler une variable X de loi $\mathcal{N}(\mu, \sigma^2)$, il suffit de prendre $X = \mu + \sigma Y$ avec $Y \sim \mathcal{N}(0, 1)$.

2. <http://math.unice.fr/~rubentha/enseignement/poly-integration-probas.pdf>

2.4. Simulation d'une variable aléatoire poissonnienne

Une variable de loi de Poisson de paramètre $\lambda > 0$ est une variable à valeurs dans \mathbb{N} telle que

$$\mathbb{P}(X = n) = \frac{e^{-\lambda} \lambda^n}{n!}$$

(ici la variable s'appelle X).

LEMME 2.6. Soit $t > 0$. Soit $(T_i)_{i \geq 1}$ une suite de variables i.i.d. de loi $\mathcal{E}(\lambda)$ ($\lambda > 0$). Alors

$$N_t = \sum_{n \geq 1} n \mathbb{1}_{\{T_1 + \dots + T_n \leq t < T_1 + \dots + T_{n+1}\}}$$

est une variable de Poisson de paramètre λt .

DÉMONSTRATION. Commençons par vérifier que la variable N_t est bien définie. Pour tout $\omega \in \Omega$, il existe un $k \in \mathbb{N}^*$ et un seul tel que $T_1 + \dots + T_k \leq t < T_1 + \dots + T_{k+1}$ et donc $N_t(\omega) = k \mathbb{1}_{\{T_1 + \dots + T_k \leq t < T_1 + \dots + T_{k+1}\}}$ (tous les autres termes de la somme sont nuls). Nous avons, pour tout $n \in \mathbb{N}^*$,

$$\begin{aligned} \mathbb{P}(N_t = n) &= \mathbb{P}(T_1 + \dots + T_n \leq t < T_1 + \dots + T_{n+1}) \\ &= \int_{0 \leq t_1 \leq t} \dots \int_{t_{n+1} \geq 0} \mathbb{1}_{\{t_1 + \dots + t_n \leq t < t_1 + \dots + t_{n+1}\}} \lambda e^{-\lambda t_1} \dots \lambda e^{-\lambda t_{n+1}} dt_1 \dots dt_{n+1} \\ &= \int_{0 \leq t_1 \leq t} \int_{0 \leq t_2 \leq t - t_1} \dots \int_{0 \leq t_n \leq t - (t_1 + \dots + t_{n-1})} \int_{t - (t_1 + \dots + t_n) \leq t_{n+1}} \lambda e^{-\lambda t_1} \dots \lambda e^{-\lambda t_{n+1}} dt_{n+1} dt_n \dots dt_1 \\ &= \int_{0 \leq t_1 \leq t} \int_{0 \leq t_2 \leq t - t_1} \dots \int_{0 \leq t_n \leq t - (t_1 + \dots + t_{n-1})} \lambda e^{-\lambda t_1} \dots \lambda e^{-\lambda t_n} [-e^{-\lambda t_{n+1}}]_{t - (t_1 + \dots + t_n)}^{+\infty} dt_n \dots dt_1 \\ &= \int_{0 \leq t_1 \leq t} \int_{0 \leq t_2 \leq t - t_1} \dots \int_{0 \leq t_n \leq t - (t_1 + \dots + t_{n-1})} \lambda^n e^{-\lambda t} dt_1 \dots dt_n. \end{aligned}$$

Montrons par récurrence sur $n \in \mathbb{N}^*$ que pour tout $x \geq 0$,

$$\int_{0 \leq t_1 \leq x} \int_{0 \leq t_2 \leq x - t_1} \int_{0 \leq t_3 \leq x - (t_1 + t_2)} \dots \int_{0 \leq t_n \leq x - (t_1 + \dots + t_{n-1})} 1 dt_n \dots dt_1 = \frac{x^n}{n!}.$$

— Pour $n = 1$, $\int_{0 \leq t_1 \leq x} 1 dt_1 = x = \frac{x^1}{1!}$. Pour $n = 2$,

$$\int_{0 \leq t_1 \leq x} \int_{0 \leq t_2 \leq x - t_1} 1 dt_2 dt_1 = \int_{0 \leq t_1 \leq x} x - t_1 dt_1 = \left[-\frac{(x - t_1)^2}{2!} \right]_0^x = \frac{x^2}{2!}.$$

— Si la propriété est vraie jusqu'au rang n , calculons

$$\begin{aligned} &\int_{0 \leq t_1 \leq x} \int_{0 \leq t_2 \leq x - t_1} \int_{0 \leq t_3 \leq x - (t_1 + t_2)} \dots \int_{0 \leq t_{n+1} \leq x - (t_1 + \dots + t_n)} 1 dt_{n+1} \dots dt_1 \\ &= \int_{0 \leq t_1 \leq x} \left(\int_{0 \leq t_2 \leq x - t_1} \int_{0 \leq t_3 \leq x - (t_1 + t_2)} \dots \int_{0 \leq t_{n+1} \leq x - (t_1 + \dots + t_n)} 1 dt_{n+1} \dots dt_2 \right) dt_1 \\ &\quad \text{(par la propriété au rang } n) \\ &= \int_{0 \leq t_1 \leq x} \frac{(x - t_1)^n}{n!} dt_1 = \left[-\frac{(x - t_1)^{n+1}}{(n+1)!} \right]_0^x = \frac{x^{n+1}}{(n+1)!}. \end{aligned}$$

Nous avons donc

$$\mathbb{P}(N_t = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}.$$

□

Nous pouvons donc simuler une variable de Poisson à l'aide d'une boucle « tant que » (et en utilisant le résultat de la partie 2.2) (voir le programme 2.3)

Programme 2.3 Simulation d'une variable aléatoire poissonnienne.

```

lambda=2 #on fixe une valeur arbitraire de lambda
t=1 #on fixe une valeur arbitraire de t
b=0
s=0
n=-1
while (b==0)
{
  n=n+1
  s=s-log(runif(1,0,1))/lambda
  if (s>t)
  {
    b=1
  }
}
print(n)

```

2.5. Méthode de rejet

On veut simuler une variable de densité f (dans \mathbb{R}^d) et on suppose qu'il existe une loi de densité g facile à simuler telle que

$$(2.1) \quad \forall x \in \mathbb{R}^d, f(x) \leq kg(x).$$

avec k une constante. Posons $\alpha(x) = \frac{f(x)}{kg(x)}$ sur l'ensemble $\{x, g(x) > 0\}$.

PROPOSITION 2.7. *Soit $(X_n, U_n)_{n \geq 1}$ une suite de variable i.i.d. avec pour tout n , X_n est de loi de densité g et est indépendante de U_n , qui est de loi uniforme sur $[0; 1]$. Soit $T = \inf\{n, U_n \leq \alpha(X_n)\}$ (T est un entier aléatoire). Soit $X = X_T$.*

Alors la v.a. X ainsi simulée est de loi de densité f .

DÉMONSTRATION. Remarquons que $\alpha(X_n)$ est défini pour presque tout ω . En effet, $\mathbb{P}(X_n \in \{x : g(x) = 0\}) = 0$.

Soit $\varphi \in C_b^+(\mathbb{R}^d)$. Nous avons $\varphi(X_T) = \sum_{n=1}^{+\infty} \varphi(X_n) \mathbb{1}_{n=T}$ (pour tout ω , cette somme contient au plus un terme non nul).

$$\begin{aligned}
 \mathbb{E}(\varphi(X)) &= \sum_{n=1}^{+\infty} \mathbb{E}(\varphi(X_n) \mathbb{1}_{n=T}) \\
 &= \sum_{n=1}^{+\infty} \mathbb{E}(\varphi(X_n) \prod_{i=1}^{n-1} \mathbb{1}_{U_i > \alpha(X_i)} \times \mathbb{1}_{U_n \leq \alpha(X_n)}) \\
 (\text{par indépendance des } (X_i, U_i)) &= \sum_{n=1}^{+\infty} \mathbb{E}(\varphi(X_n) \mathbb{1}_{U_n \leq \alpha(X_n)}) \prod_{i=1}^{n-1} \mathbb{E}(\mathbb{1}_{U_i > \alpha(X_i)}).
 \end{aligned}$$

Calculons, pour tout n ,

$$\begin{aligned}
 \mathbb{E}(\varphi(X_n) \mathbb{1}_{U_n \leq \alpha(X_n)}) &= \int_{x \in \mathbb{R}^d} \left(\int_{u \in [0; 1]} \mathbb{1}_{u \leq \alpha(x)} du \right) \varphi(x) g(x) dx \\
 &= \int_{x \in \mathbb{R}^d} \frac{f(x)}{kg(x)} \varphi(x) g(x) dx \\
 &= \frac{1}{k} \int_{x \in \mathbb{R}^d} \varphi(x) f(x) dx.
 \end{aligned}$$

Nous en déduisons, pour tout n ,

$$\mathbb{E}(\mathbb{1}_{U_n > \alpha(X_n)}) = 1 - \mathbb{E}(\mathbb{1}_{U_n \leq \alpha(X_n)})$$

$$= 1 - \frac{1}{k}.$$

Remarquons que l'inégalité $f \leq kg$ implique

$$\int_{\mathbb{R}^d} f(x)dx \leq k \int_{\mathbb{R}^d} g(x)dx,$$

donc $k \geq 1$. Nous avons donc

$$\begin{aligned} \mathbb{E}(\varphi(X)) &= \sum_{n=1}^{+\infty} \int_{\mathbb{R}^d} \varphi(x) f(x) dx \times \frac{1}{k} \left(1 - \frac{1}{k}\right)^{n-1} \\ (\text{somme géométrique}) &= \int_{\mathbb{R}^d} \varphi(x) f(x) dx \times \frac{1}{k} \times \frac{1}{1 - (1 - \frac{1}{k})} \\ &= \int_{\mathbb{R}^d} \varphi(x) f(x) dx. \end{aligned}$$

D'où le résultat. \square

EXEMPLE 2.8. Soit $f(x) = \frac{e^{-x^3}}{Z} \mathbb{1}_{x \geq 1}$ avec $Z = \int_1^{+\infty} e^{-x^3} dx$ (la constante Z n'est pas connue mais nous n'en aurons pas besoin). Soit $g(x) = \frac{\mathbb{1}_{x \geq 1}}{x^2}$. Les fonctions f et g sont des densités de probabilité. Il est facile de simuler suivant la loi de densité g en utilisant la technique d'inversion de la fonction de répartition. Nous commençons par étudier la fonction $h : x \geq 1 \mapsto x^2 e^{-x^3}$. Nous avons $h'(x) = (2x - 3x^4)e^{-x^3} \leq 0$ pour tout $x \geq 1$. Donc, pour tout $x \geq 1$, $h(x) \leq h(1) = e^{-1}$. Nous en déduisons (pour $x \geq 1$)

$$f(x) \leq \frac{1}{eZ} g(x).$$

Nous posons $k = 1/(eZ)$. Soit

$$\alpha(x) = \frac{f(x)}{kg(x)} = x^2 e^{-x^3+1}.$$

Nous ne connaissons pas Z mais nous pouvons calculer $\alpha(x)$ pour tout $x \geq 1$ donc nous pouvons utiliser l'algorithme de simulation par rejet pour simuler suivant la densité f (voir programme 2.4). Dans ce programme, nous utilisons les résultats suivants pour simuler suivant la densité g

Programme 2.4 Simulation par la méthode de rejet.

```

b=0
while (b==0)
{
  u=runif(1,0,1)
  v=runif(1,0,1)
  y=1/(1-v)
  if (u<y^2*exp(-y^3)*exp(1))
    { b=1 }
}
print(y)

```

par inversion de la fonction de répartition :

$$\text{pour } x \geq 1, G(x) = \int_1^x g(u)du = \left[-\frac{1}{u}\right]_1^x = 1 - \frac{1}{x},$$

$$\text{pour } u \in [0; 1], G^{-1}(u) = \frac{1}{1-u}.$$

EXEMPLE 2.9 (Loi conditionnelle). Soit X une variable aléatoire à valeur dans \mathbb{R} , de densité g . Soit A un sous-ensemble (mesurable) de \mathbb{R} tel que $\mathbb{P}(X \in A) \neq 0$. La loi de X conditionné à tomber dans A a pour densité

$$x \mapsto f(x) = \frac{g(x) \mathbb{1}_A(x)}{\int_{u \in A} g(u) du}.$$

Nous remarquons que pour tout x , $f(x) \leq Cg(x)$ avec $C = (\int_{u \in A} f(u)du)^{-1}$. Donc, si nous tirons $(X_n, U_n)_{n \geq 0}$ i.i.d. avec $X_n \perp\!\!\!\perp U_n$, X_n de loi de densité g , $U_n \sim \mathcal{U}([0; 1])$, et si nous posons $T = \inf\{n : U_n \leq f(X_n)/(Cg(X_n))\}$, alors X_T est de loi de densité f . Ici, le ratio $f/(Cg)$ se simplifie en

$$\frac{f(x)}{Cg(x)} = \mathbb{1}_A(x).$$

Donc l'algorithme de rejet consiste simplement à tirer des X_n jusqu'à ce qu'il y en ait un dans A .

REMARQUE 2.10 (Temps de calcul). Le nombre de boucles effectuées dans la méthode du rejet est aléatoire. Sous les hypothèses de la proposition ci-dessus, ce nombre de boucles est T et nous pouvons calculer son espérance

$$\begin{aligned} \mathbb{E}(T) &= \sum_{n=1}^{+\infty} n\mathbb{P}(T = n) \\ &= \sum_{n=1}^{+\infty} n\mathbb{P}(U_1 > \alpha(X_1), \dots, U_{n-1} > \alpha(X_{n-1}), U_n \leq \alpha(X_n)) \\ (\text{par indépendance des } (X_i, U_i)) &= \sum_{n=1}^{+\infty} n(\mathbb{P}(U_1 > \alpha(X_1)))^{n-1} \mathbb{P}(U_1 \leq \alpha(X_n)) \\ (\text{d'après la dém. de la prop.}) &= \sum_{n=1}^{+\infty} n \left(1 - \frac{1}{k}\right)^{n-1} \frac{1}{k} \\ &= \frac{1}{k} \phi\left(1 - \frac{1}{k}\right), \end{aligned}$$

avec $\phi(z)$ la série entière $\phi(z) = \sum_{n \geq 1} nz^{n-1}$ (de rayon de convergence 1, c'est à dire que cette série converge pour $|z| < 1$). Nous avons $\phi(z) = \Phi'(z)$ avec $\Phi(z) = \sum_{n \geq 0} z^n = \frac{1}{1-z}$. Donc $\phi(z) = \frac{1}{(1-z)^2}$. Donc

$$\mathbb{E}(T) = \frac{1}{k} \frac{1}{\left(1 - \left(1 - \frac{1}{k}\right)\right)^2} = k.$$

Donc le temps de calcul moyen est proportionnel à k . C'est pourquoi on cherche une constante k satisfaisant (2.1) qui soit la plus petite possible.

REMARQUE 2.11. L'algorithme ci-dessus est encore valable si la v.a. X à simuler a une densité f par rapport à une mesure μ quelconque et que cette densité est majorée par kg où g est la densité par rapport à μ d'une variable Y facile à simuler. Ceci se traduit par (pour tout A)

$$\mathbb{P}(X \in A) = \int_A f(x)\mu(dx) \leq \int_A kg(x)\mu(dx) = k\mathbb{P}(Y \in A).$$

Si la v.a. X a une loi portée par un ensemble discret E , on peut choisir pour μ la mesure de comptage sur E et la méthode de rejet est aussi valable pour les lois discrètes, $f(x)$ étant dans ce cas égale à $\mathbb{P}(X = x)$.

2.6. Vérification par histogramme

Quand on simule une variable aléatoire réelle à l'aide d'une des méthodes ci-dessus, on peut vérifier empiriquement que la loi simulée est bien celle que l'on voulait. Soient X_1, X_2, \dots des v.a.r. i.i.d. de loi de densité f (avec une dérivée f' vérifiant $\|f'\|_\infty < C$ pour une certaine constante C). Pour tout $a < b$,

$$\frac{1}{n(b-a)} \sum_{i=1}^n \mathbb{1}_{[a;b]}(X_i) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \frac{\mathbb{P}(X_1 \in [a;b])}{b-a},$$

et $\frac{\mathbb{P}(X_1 \in [a;b])}{b-a} = \frac{1}{b-a} \int_a^b f(t)dt$ donc

$$\left| \frac{\mathbb{P}(X_1 \in [a;b])}{b-a} - f(a) \right| = \left| \frac{1}{b-a} \int_a^b f(t)dt - f(a) \right|$$

$$\begin{aligned}
&\leq \frac{1}{b-a} \int_a^b |f(t) - f(a)| dt \\
&\leq \frac{1}{b-a} \int_a^b C|t-a| dt \\
&= \frac{C(b-a)}{2} \xrightarrow{b \searrow a} 0.
\end{aligned}$$

Donc l'histogramme (correctement renormalisé) des n valeurs simulées X_1, X_2, \dots, X_n est proche de la densité f . Dans le programme 2.5, nous simulons des variables de densité $f : x \in \mathbb{R} \mapsto \mathbb{1}_{x \geq 1} e^{-x^3} / Z$ (comme dans l'exemple 2.8) et nous traçons l'histogramme empirique des variables et le graphe de f dans le même repère (voir figure 2.6.1, dans laquelle il y a aussi le graphe de g en bleu).

Programme 2.5 Tracé d'un histogramme

#Nous déclarons une fonction qui simule une variable de densité f .

#Remarque : l'argument t n'est pas utilisé.

```

simu<-function(t)
{
  b=0
  while (b==0)
  {
    u=runif(1,0,1)
    v=runif(1,0,1)
    y=1/(1-v)
    if (u<y^2*exp(-y^3)*exp(1))
      { b=1 }
  }
  return(y)
}
f1<-function(x)
{
  if (x>1)
    { z=exp(-x^3) }
  else
    { z=0 }
return(z)
}
Zi=integrate(f1,1,5)
Z=Zi$value #calcule la constante Z
n=5000
tab=c()
for (i in 1:n)
{
  x=simu(1)
  tab=c(tab,x)
}
hist(tab,freq=FALSE)
abs=seq(1.01,5,0.01)
y=exp(-abs*abs*abs)/Z
lines(abs,y,type='l',col='red')

```

Pour un exemple de construction d'histogramme pour la loi uniforme, voir : <http://www-sop.inria.fr/mefisto/java/tutorial1/node7.html#SECTION00031010000000000000>.

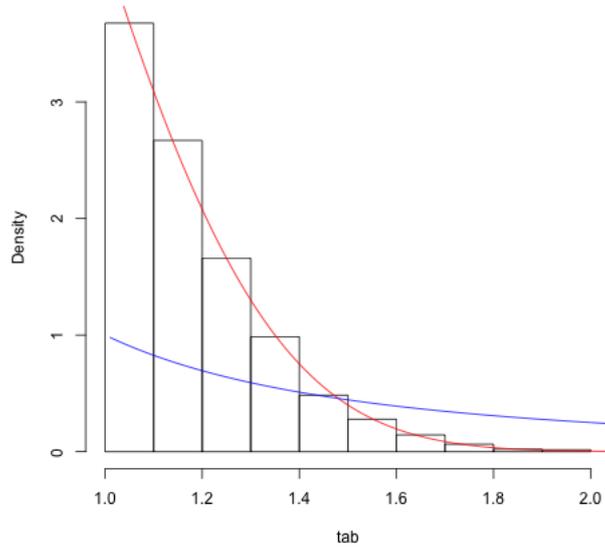


FIGURE 2.6.1. Histogramme et densité

2.7. Exercices

Préliminaire. Pour tous les exercices dans lesquels vous proposerez une méthode de simulation, il est conseillé de faire une vérification par histogramme.

- EXERCICE 2.1. (1) Simuler des variables exponentielles par la méthode du cours.
 (2) Tracer un histogramme des variables simulées par cette méthode. Comparer avec un tracé de la densité de la loi exponentielle.

EXERCICE 2.2. Montrer que l'algorithme suivant permet de simuler une réalisation de loi uniforme sur le disque unité de \mathbb{R}^2 .

- (1) Simuler (U, V) couple de deux v.a. i.i.d. de loi uniforme sur $[-1, 1]$.
- (2) Tant que $U^2 + V^2 > 1$, répéter (1).
- (3) Renvoyer la valeur de (U, V) en fin de boucle.

Coder l'algorithme en \mathbf{R} .

EXERCICE 2.3. Montrer que l'algorithme suivant permet de simuler un couple de v.a. i.i.d. de loi gaussiennes centrées réduites.

- (1) Simuler (U, V) couple de deux v.a. i.i.d. de loi uniforme sur $[-1, 1]$.
- (2) Tant que $U^2 + V^2 > 1$, répéter (1).
- (3) Renvoyer la valeur de (U, V) et de $R^2 = U^2 + V^2$ en fin de boucle.
- (4) Poser $Z = [-2 \ln(R^2)/R^2]^{1/2}$.
- (5) Poser $X = ZU$ et $Y = ZV$.

Coder l'algorithme en \mathbf{R} . Vérifier à l'aide d'un histogramme que la variable X simulée suit bien une loi gaussienne centrée réduite.

EXERCICE 2.4. Soient X et Y deux v.a. i.i.d. de loi exponentielle de paramètre 1.

- (1) Montrer que la loi conditionnelle de X sachant $2Y > (1 - X)^2$ a pour densité

$$x \mapsto \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \mathbf{1}_{[0, +\infty[}(x).$$

- (2) Soit S une v.a. de loi Bernoulli de paramètre $1/2$, indépendante du couple (X, Y) . Montrer que la loi conditionnelle de $(2S - 1)X$ sachant $2Y > (1 - X)^2$ suit une loi normale centrée réduite.
- (3) En déduire un algorithme de simulation de la loi $\mathcal{N}(0, 1)$. Le coder en R.

EXERCICE 2.5. Calcul de VaR. On suppose que le gain obtenu par gestion d'un portefeuille est une variable X dont on connaît la densité. On se fixe une probabilité p (que l'on prendra égale à 0.05). La VaR (« value at risk ») pour la probabilité p est le nombre x tel que $\mathbb{P}(X \leq -x) = p$ (c'est à dire que la probabilité que la perte soit plus grande que x vaut p). Nous supposons dans cet exercice que X est de loi $\mathcal{N}(1; 1)$. Nous voulons calculer la VaR.

Pour tout entier n dans \mathbb{N}^* , nous tirons X_1, X_2, \dots, X_n i.i.d. de loi $\mathcal{N}(1; 1)$ et nous construisons la fonction de répartition empirique

$$F_n : t \in \mathbb{R} \mapsto F_n(t) = \frac{\#\{i \in \{1, 2, \dots, n\} : X_i \leq t\}}{n}.$$

- (1) Montrer que pour tout t dans \mathbb{R} , $F_n(t) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} F(t)$, où F est la fonction de répartition de X .
- (2) Notons, pour tout n ,

$$Z_n = \inf\{t : F_n(-t) \leq p\}.$$

Montrer que $Z_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} x$.

- (3) Proposer un programme calculant x de manière approchée. Indice : pour ordonner un vecteur de nombres \mathbf{x} , exécuter $\mathbf{y} = \text{order}(\mathbf{x})$; $\mathbf{z} = \mathbf{x}[\mathbf{y}]$ (\mathbf{z} contient les valeurs de \mathbf{x} , triées par ordre croissant).

EXERCICE 2.6. Soient f et g ($\mathbb{R} \mapsto \mathbb{R}$) des densités. Soit h la fonction :

$$h(x) = \frac{\sup(f(x), g(x))}{\int_{\mathbb{R}} \sup(f(t), g(t)) dt}.$$

- (1) On simule une v.a. Z suivant une méthode d'acceptation/rejet. On tire X, Y indépendantes respectivement de lois de densité f, g et U, V indépendantes uniformes sur $[0, 1]$ (et indépendantes de X, Y) jusqu'à ce que
- $Uf(X) \leq g(X)$, auquel cas on prend $Z = Y$
 - ou $Vg(Y) \leq f(Y)$ et $Uf(X) > g(X)$, auquel cas on prend $Z = X$.
- Montrer que Z est de loi de densité h .
- (2) On suppose ici que f densité de la loi $\mathcal{N}(0, 1)$ et g densité de la loi $\mathcal{N}(3/2, 1)$.
- (a) Écrire un programme qui simule des variables aléatoires suivant h .
- (b) Écrire un programme qui dessine un histogramme empirique de h .

EXERCICE 2.7. On désigne par f la densité

$$f(x) = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{x^2}{2}\right) \mathbb{1}_{\{x > 0\}}, \quad x \in \mathbb{R}.$$

- (1) Pour $\lambda > 0$ fixé, trouver une constante $c_\lambda > 1$ telle que
- $$f(x) \leq c_\lambda \lambda \exp(-\lambda x), \quad x \in \mathbb{R}_+.$$
- (2) En déduire une méthode de simulation de la loi de densité f
- (3) Trouver λ tel que le temps moyen de calcul dans la méthode proposée soit le plus petit possible.

EXERCICE 2.8. Pour $a > 0$ donné, on désigne par f la fonction

$$f(x) = \mathbb{1}_{[0, a]}(x) \exp(-x), \quad x \in \mathbb{R}.$$

- (1) Trouver une constante C telle que Cf soit une densité.

- (2) Trouver une constante $c_1 > 1$ telle que

$$Cf(x) \leq c_1 \frac{\mathbb{1}_{[0;a]}(x)}{a}, x \in \mathbb{R}.$$

- (3) Trouver une constante $c_2 > 1$ telle que

$$Cf(x) \leq c_2 \mathbb{1}_{[0;+\infty[}(x) \exp(-x), x \in \mathbb{R}.$$

- (4) On veut mettre en place une méthode de rejet pour simuler la loi densité f en utilisant la loi uniforme sur $[0; a]$ ou la loi exponentielle de paramètre 1. Laquelle vaut-il mieux choisir ? Coder la méthode.

EXERCICE 2.9. On souhaite simuler suivant la loi de densité

$$f : x \in \mathbb{R} \mapsto \frac{1}{Z} \exp(-x^2) \mathbb{1}_{x \geq 1},$$

avec $Z = \int_1^{+\infty} e^{-x^2} dx$.

- (1) Trouver une densité g (suivant laquelle on sait simuler) et une constante C telles que $f \leq Cg$.
- (2) Soit $X \sim \mathcal{N}(0, 1/2)$. Montrer que f est la densité de la loi de X sachant $X \geq 1$ (on pourra calculer l'espérance sur une fonction test).
- (3) Coder la simulation de la loi de densité f (suivant la méthode du rejet basée sur g, C).

EXERCICE 2.10. Soit X une variable aléatoire admettant pour fonction de répartition

$$F(x) = (1 - \exp(-\alpha x^\beta)) \mathbb{1}_{[0;+\infty[}(x),$$

avec α et β des paramètres > 0 . Expliciter la densité de la loi de X et proposer une méthode de simulation de cette loi.

EXERCICE 2.11. Soit X une loi géométrique de paramètre p :

$$\mathbb{P}(X = k) = p(1 - p)^{k-1}, \text{ pour } k \in \mathbb{N}^*.$$

- (1) Rappeler la méthode classique de simulation de X à l'aide de tirages à pile ou face.
- (2) Donner une autre méthode de simulation de cette loi utilisant la fonction de répartition.
- (3) Comment comparer l'efficacité des 2 méthodes ?

EXERCICE 2.12. Soit $n \in \mathbb{N}^*$. Soient U_1, U_2, \dots, U_n des variables i.i.d. de loi $\mathcal{U}([0; 1])$. On note (V_1, \dots, V_n) la variable aléatoire telle que, pour tout ω , $\{V_1(\omega), \dots, V_n(\omega)\} = \{U_1(\omega), \dots, U_n(\omega)\}$ et $V_1(\omega) \leq V_2(\omega) \leq \dots \leq V_n(\omega)$ (c'est une remise en ordre des variables U_1, \dots, U_n). Le vecteur aléatoire (V_1, \dots, V_n) est appelé statistique d'ordre uniforme sur $[0; 1]$.

- (1) Soit $C_n = \{(v_1, \dots, v_n) \in [0; 1]^n : v_1 \leq v_2 \leq \dots \leq v_n\}$. Montrer que la densité de la loi de (V_1, \dots, V_n) est

$$(v_1, \dots, v_n) \in \mathbb{R}^n \mapsto n! \times \mathbb{1}_{C_n}(v_1, \dots, v_n).$$

- (2) Soient $X_1, X_2, \dots, X_n, X_{n+1}$ des variables i.i.d. de loi $\mathcal{E}(1)$. Montrer que

$$\left(\frac{X_1}{X_1 + X_2 + \dots + X_{n+1}}, \frac{X_1 + X_2}{X_1 + X_2 + \dots + X_{n+1}}, \dots, \frac{X_1 + \dots + X_n}{X_1 + X_2 + \dots + X_{n+1}} \right)$$

a même loi que (V_1, \dots, V_n) .

Réduction de variance

Nous avons vu au chapitre 1 (section 1.2.2) que si nous calculons une quantité $\mathbb{E}(X)$ par la méthode de Monte-Carlo, c'est à dire si nous approchons $\mathbb{E}(X)$ par une moyenne empirique $\frac{X_1 + \dots + X_n}{n}$, X_1, X_2, \dots i.i.d. de même loi que X , alors l'erreur $|\mathbb{E}(X) - \frac{X_1 + \dots + X_n}{n}|$ est d'ordre σ/\sqrt{n} , où $\sigma^2 = \mathbb{V}(X)$. Nous allons présenter ici des méthodes qui permettent d'écrire $\mathbb{E}(X) = \mathbb{E}(Y)$ avec $\mathbb{V}(Y) \leq \mathbb{V}(X)$. Ces méthodes sont dites « de réduction de variance ».

3.1. Échantillonnage préférentiel (« importance sampling » en anglais)

Nous cherchons à calculer $\mathbb{E}(g(X))$ avec X variable à valeurs dans \mathbb{R}^d , de densité f . Pour toute densité $\tilde{f} > 0$, nous pouvons écrire

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}^d} g(x)f(x)dx = \int_{\mathbb{R}^d} \frac{g(x)f(x)}{\tilde{f}(x)}\tilde{f}(x)dx = \mathbb{E}\left(\frac{g(Y)f(Y)}{\tilde{f}(Y)}\right),$$

avec Y de densité \tilde{f} . Nous avons donc deux méthodes de Monte-Carlo pour approcher $\mathbb{E}(g(X))$:

$$\mathbb{E}(g(X)) \approx \frac{g(X_1) + \dots + g(X_n)}{n},$$

$$\mathbb{E}(g(X)) \approx \frac{1}{n} \left(\frac{g(Y_1)f(Y_1)}{\tilde{f}(Y_1)} + \dots + \frac{g(Y_n)f(Y_n)}{\tilde{f}(Y_n)} \right),$$

avec des X_1, X_2, \dots i.i.d. de même loi que X , des Y_1, Y_2, \dots i.i.d. de même loi que Y . La deuxième méthode est plus intéressante que la première si $\mathbb{V}\left(\frac{g(Y)f(Y)}{\tilde{f}(Y)}\right) \leq \mathbb{V}(g(X))$.

Supposons $g \geq 0$ et choisissons $\tilde{f} : x \mapsto \frac{g(x)f(x)}{\mathbb{E}(g(X))}$ (c'est bien une densité de probabilité), nous avons alors

$$\begin{aligned} \mathbb{V}\left(\frac{g(Y)f(Y)}{\tilde{f}(Y)}\right) &= \mathbb{E}\left(\left(\frac{g(Y)f(Y)}{\tilde{f}(Y)}\right)^2\right) - \left(\mathbb{E}\left(\frac{g(Y)f(Y)}{\tilde{f}(Y)}\right)\right)^2 \\ &= \int_{\mathbb{R}^d} \frac{g(u)^2 f(u)^2}{g(u)^2 f(u)^2} (\mathbb{E}(g(X)))^2 \tilde{f}(u) du - \left(\int_{\mathbb{R}^d} \mathbb{E}(g(X)) \tilde{f}(u) du\right)^2 \\ &= (\mathbb{E}(g(X)))^2 - (\mathbb{E}(g(X)))^2 = 0. \end{aligned}$$

Nous avons donc ici une méthode de Monte-Carlo de variance nulle, ce qui semble n'avoir aucun sens. En fait, la variance de cette méthode n'a pas d'intérêt car cette méthode n'est pas implémentable. En effet, il faudrait pour cela savoir simuler suivant la densité \tilde{f} , mais l'expression de \tilde{f} contient la constante $\mathbb{E}(g(X))$, que nous ne connaissons pas.

La discussion ci-dessus nous donne une idée d'une démarche à suivre pour réduire la variance.

- (1) Trouver une fonction \tilde{f}_1 proche de $|f \times g|$ et telle que l'on sache simuler suivant la densité $\tilde{f} = \tilde{f}_1 / \int_{\mathbb{R}^d} \tilde{f}_1(x) dx$.
- (2) Soit Y variable aléatoire de densité \tilde{f} . Comparer $\mathbb{V}(g(X))$ et $\mathbb{V}(g(Y)f(Y)/\tilde{f}(Y))$ (il faut en général approcher ces variables par un estimateur, voir par exemple le lemme 1.4 à ce sujet).

La fonction \tilde{f} s'appelle la fonction d'importance.

EXEMPLE 3.1. On cherche à calculer $I = \int_0^1 \cos(\pi x/2) dx$ par une méthode de Monte-Carlo. On peut approcher I par

$$(3.1) \quad I \approx \frac{\cos(\pi U_1/2) + \dots + \cos(\pi U_n/2)}{n},$$

avec U_1, U_2, \dots i.i.d. de loi $\mathcal{U}([0; 1])$ (toujours pour « n grand »). Si nous reprenons les notations ci-dessus, nous avons $g(x) = \cos(\pi x/2)$ et $f(x) = 1$ (pour $x \in [0; 1]$). Soit $\tilde{f}_1(x) = 1 - x$,

$$\tilde{f}(x) = \frac{1 - x}{(1/2)}.$$

La fonction \tilde{f}_1 est (grossièrement) proche de $g \times f$. Si $U \sim \mathcal{U}([0; 1])$ alors $1 - \sqrt{U}$ est de loi de densité \tilde{f} (nous trouvons ce résultat par la méthode d'inversion de la fonction de répartition, voir chapitre 2). Donc nous savons simuler suivant la loi de densité \tilde{f} . Dans le programme 3.1, nous estimons les variances des deux méthodes. Nous trouvons $\approx 1.10^{-1}$ pour la méthode de l'équation (3.1) ci-dessus et 7.10^{-3} pour la méthode d'échantillonnage d'importance par \tilde{f} . La variance est donc réduite.

Programme 3.1 Comparaison de variances

```
simu<-function(t)
{
  u=runif(1,0,1)
  return(1-sqrt(u))
}
s=0
for (i in 1:n)
{
  u=runif(1,0,1)
  v=runif(1,0,1)
  u1=cos(pi*u/2)
  v1=cos(pi*v/2)
  s=s+(u1-v1)^2
}
s=s/(2*n)
cat("Variance de la première méthode : ",s)
s=0
for (i in 1:n)
{
  v=simu(1)
  v1=cos(pi*v/2)/(2*(1-v))
  w=simu(1)
  w1=cos(pi*w/2)/(2*(1-w))
  s=s+(v1-w1)^2
}
s=s/(2*n)
cat("Variance de la méthode d'échantillonnage préférentiel : ",s)
```

3.2. Variable de contrôle

Supposons que l'on veuille calculer $\mathbb{E}(f(X))$ avec X une variable aléatoire. La première méthode consiste à faire l'approximation

$$\mathbb{E}(f(X)) \approx \frac{f(X_1) + \dots + f(X_n)}{n}$$

avec X_1, X_2, \dots i.i.d. de même loi que X (toujours pour « n grand »). Si on sait calculer $\mathbb{E}(h(X))$ pour une certaine fonction h , alors on peut aussi faire l'approximation

$$\mathbb{E}(f(X)) \approx \frac{(f(X_1) - h(X_1)) + \dots + (f(X_n) - h(X_n))}{n} + \mathbb{E}(h(X)).$$

Il faut ensuite comparer les variances pour savoir quelle est la méthode la plus avantageuse. Dans le cas de la deuxième méthode, l'erreur est

$$\begin{aligned} \mathbb{E}(f(X)) - \left(\frac{(f(X_1) - h(X_1)) + \dots + (f(X_n) - h(X_n))}{n} + \mathbb{E}(h(X)) \right) \\ = \mathbb{E}(f(X) - h(X)) - \frac{(f(X_1) - h(X_1)) + \dots + (f(X_n) - h(X_n))}{n}. \end{aligned}$$

Elle est d'ordre de grandeur $\mathbb{V}(f(X) - h(X))^{1/2}/\sqrt{n}$.

La méthode est donc la suivante.

- (1) Trouver une fonction h proche de f et telle que l'on sache calculer $\mathbb{E}(h(X))$ (le fait que h soit proche de f devrait faire en sorte que $\mathbb{V}(f(X) - h(X))$ est petite).
- (2) Estimer les variances $\mathbb{V}(f(X))$ et $\mathbb{V}(f(X) - h(X))$, et les comparer.

EXEMPLE 3.2. On cherche à calculer $I = \int_0^1 \exp(x^2) dx$ par une méthode de Monte-Carlo. Nous avons $I = \mathbb{E}(e^{U^2})$ avec $U \sim \mathcal{U}([0; 1])$. Donc nous pouvons faire l'approximation

$$I \approx \frac{e^{U_1^2} + \dots + e^{U_n^2}}{n},$$

avec U_1, U_2, \dots i.i.d., $\sim \mathcal{U}([0; 1])$ (toujours pour « n grand »). Nous avons donc une première méthode de Monte-Carlo. Avec les notations ci-dessus, nous avons $f(x) = \exp(x^2)$. Remarquons que $h : x \mapsto 1 + x^2$ est proche de f sur $[0; 1]$ (c'est le début du développement limité de f en 0). Nous savons calculer

$$\mathbb{E}(h(U)) = \int_0^1 1 + x^2 dx = 1 + \frac{1}{3} = \frac{4}{3}.$$

Nous pouvons faire l'approximation

$$I \approx \frac{(e^{U_1^2} - 1 - U_1^2) + \dots + (e^{U_n^2} - 1 - U_n^2)}{n} + \mathbb{E}(h(U)).$$

Nous estimons les deux variances dans le programme 3.2. La variance est réduite d'un facteur 10 grâce à cette méthode.

3.3. Variables antithétiques

Supposons que l'on cherche à calculer $I = \int_{[0;1]^d} f(u_1, \dots, u_d) du_1 \dots du_d$ (une intégrale sur l'hypercube de dimension d). Nous avons $I = \mathbb{E}(f(U))$ avec $U = (U_1, \dots, U_d)$ variable aléatoire dont les composantes sont indépendantes et de loi $\mathcal{U}([0; 1])$. Nous pouvons utiliser l'approximation

$$I \approx \frac{f(X_1) + \dots + f(X_n)}{n},$$

avec X_1, X_2, \dots i.i.d. de même loi que U . Nous avons $(1, \dots, 1) - U \stackrel{\text{loi}}{=} U$. Alors

$$I = \mathbb{E} \left(\frac{f((1, \dots, 1) - U) + f(U)}{2} \right),$$

ce qui nous donne l'idée d'une deuxième approximation

$$I \approx \frac{1}{n} \sum_{i=1}^n \frac{f((1, \dots, 1) - X_i) + f(X_i)}{2}.$$

LEMME 3.3. *Sous les hypothèses ci-dessus, $\mathbb{V} \left(\frac{f((1, \dots, 1) - U) + f(U)}{2} \right) \leq \mathbb{V}(f(U))$. (La variance est donc toujours réduite.)*

Programme 3.2 Variable de contrôle.

```

n=5000
s=0
for (i in 1:n)
{
  u=runif(1,0,1)
  v=runif(1,0,1)
  s=s+(exp(u*u)-exp(v*v))^2
}
s=s/(2*n)
cat("Variance de la première méthode : ",s)
s=0
for (i in 1:n)
{
  u=runif(1,0,1)
  v=runif(1,0,1)
  s=s+(exp(u*u)-1-u*u-exp(v*v)+1+v*v)^2
}
s=s/(2*n)
cat("Variance de la méthode de variable de contrôle : ",s)

```

DÉMONSTRATION.

$$\begin{aligned}
& \mathbb{V} \left(\frac{f((1, \dots, 1) - U) + f(U)}{2} \right) \\
&= \mathbb{E} \left(\left(\frac{f((1, \dots, 1) - U) + f(U)}{2} \right)^2 \right) - \mathbb{E} \left(\frac{f((1, \dots, 1) - U) + f(U)}{2} \right)^2 \\
&= \frac{1}{4} \mathbb{E}(f((1, \dots, 1) - U)^2) + \frac{1}{4} \mathbb{E}(f(U)^2) + \frac{1}{2} \mathbb{E}(f((1, \dots, 1) - U)f(U)) - \mathbb{E}(f(U))^2 \\
&\quad \text{(Cauchy-Schwarz)} \\
&\leq \frac{1}{4} \mathbb{E}(f((1, \dots, 1) - U)^2) + \frac{1}{4} \mathbb{E}(f(U)^2) + \frac{1}{2} \mathbb{E}(f((1, \dots, 1) - U)^2)^{1/2} \mathbb{E}(f(U)^2)^{1/2} - \mathbb{E}(f(U))^2 \\
&\quad = \mathbb{V}(f(U)).
\end{aligned}$$

□

REMARQUE 3.4. Le même résultat est encore valable si on remplace $[0; 1]^d$ par un domaine quelconque A de \mathbb{R}^d , si on remplace $(u_1, \dots, u_d) \in [0; 1]^d \mapsto (1 - u_1, \dots, 1 - u_d) \in [0; 1]^d$ par une transformation bijective $\varphi : A \rightarrow A$ telle que $|\det(\text{Jac } \varphi)(x)| = 1$ ($\forall x$) et si $f(\varphi(x)) = f(x)$ pour tout $x \in A$.

3.4. Méthode de stratification

On veut calculer une intégrale de la forme $I = \mathbb{E}(g(X))$ avec X variable de densité f sur un ensemble D (par exemple par rapport à la mesure de Lebesgue). Cette écriture nous donne l'idée d'une première méthode de Monte-Carlo :

$$I \approx \frac{g(X_1) + \dots + g(X_n)}{n},$$

avec X_1, X_2, \dots i.i.d. de même loi que X . L'erreur commise $I - \frac{g(X_1) + \dots + g(X_n)}{n}$ est approximativement de loi $\mathcal{N}(0, \sigma^2/n)$ avec $\sigma = \mathbb{V}(g(X))$.

Supposons que D est partitionné en D_1, D_2, \dots, D_m (ce que nous noterons $D = \bigsqcup_{1 \leq i \leq m} D_i$). Nous décomposons I en :

$$I = \sum_{i=1}^m p_i I_i$$

avec pour tout i , $I_i = \mathbb{E}(g(X)|X \in D_i)$, $p_i = \mathbb{P}(X \in D_i)$ (nous avons $\sum_{i=1}^m p_i = 1$). Nous supposons que nous savons simuler suivant la loi $\mathcal{L}(X|X \in D_i)$ pour tout i et que tous les p_i sont connus. Pour chaque i , nous pouvons approcher I_i par une méthode de Monte-Carlo à n_i tirages :

$$I_i \approx \widehat{I}_i = \frac{g(X_1^{(i)}) + \dots + g(X_{n_i}^{(i)})}{n_i},$$

avec des variables $X_1^{(i)}, X_2^{(i)}, \dots$ i.i.d. $\sim \mathcal{L}(X|X \in D_i)$. Chacune de ces approximations a un coût numérique n_i (puisque l'on fait n_i boucles pour calculer la somme ci-dessus). Nous en déduisons une deuxième approximation :

$$I \approx p_1 \widehat{I}_1 + \dots + p_m \widehat{I}_m.$$

L'erreur commise se décompose dans ce cas en une somme d'erreur :

$$I - (p_1 \widehat{I}_1 + \dots + p_m \widehat{I}_m) = \sum_{i=1}^m p_i (I_i - \widehat{I}_i).$$

Dans cette somme, chacun des termes est (approximativement) de loi $\mathcal{N}(0, p_i^2 \sigma_i^2 / n_i)$ avec

$$\begin{aligned} \sigma_i^2 &= \mathbb{V}(g(X)|X \in D_i) \\ &= \mathbb{E}((g(X) - \mathbb{E}(g(X)|X \in D_i))^2 | X \in D_i) \\ &= \frac{\mathbb{E}((g(X) - \mathbb{E}(g(X)|X \in D_i))^2 \mathbb{1}_{\{X \in D_i\}})}{p_i} \\ &= p_i^{-1} \mathbb{E}(g(X)^2 \mathbb{1}_{\{X \in D_i\}}) + p_i^{-1} \mathbb{E}(g(X)|X \in D_i)^2 \mathbb{E}(\mathbb{1}_{\{X \in D_i\}}) \\ &\quad - 2p_i^{-1} \mathbb{E}(g(X)|X \in D_i) \mathbb{E}(g(X) \mathbb{1}_{\{X \in D_i\}}) \\ &= p_i^{-1} \mathbb{E}(g(X)^2 \mathbb{1}_{\{X \in D_i\}}) - p_i^{-2} \mathbb{E}(g(X) \mathbb{1}_{\{X \in D_i\}})^2. \end{aligned}$$

Donc l'erreur est (approximativement) de loi $\mathcal{N}(0, \sum_{i=1}^m p_i^2 \sigma_i^2 / n_i)$. Nous voulons que cette erreur soit la plus petite possible donc nous voulons minimiser la variance $\sum_{i=1}^m p_i^2 \sigma_i^2 / n_i$. Comme nous voulons pouvoir faire une comparaison avec la première méthode, nous ajoutons la contrainte $\sum_{i=1}^m n_i = n$.

LEMME 3.5. Soit $S = \{(x_1, \dots, x_m) \in \mathbb{R}_+^m : x_1 + \dots + x_m = n\}$. Soit

$$f : x = (x_1, \dots, x_m) \in S \mapsto \sum_{i=1}^m \frac{p_i^2 \sigma_i^2}{x_i}.$$

Alors f atteint son minimum en $n \left(\frac{p_1 \sigma_1}{p_1 \sigma_1 + \dots + p_m \sigma_m}, \dots, \frac{p_m \sigma_m}{p_1 \sigma_1 + \dots + p_m \sigma_m} \right)$.

DÉMONSTRATION. L'ensemble S est la surface $\{x \in \mathbb{R}^m : g(x) = n\}$ avec $g : x = (x_1, \dots, x_m) \in \mathbb{R}^m \mapsto x_1 + \dots + x_m$ (plus précisément, l'intersection de cette surface avec \mathbb{R}_+^m). Nous commençons par chercher les points critiques de f (voir le cours d'optimisation pour les détails de la démonstration qui va suivre). Ce sont les points x de S tels que $\nabla f(x)$ est colinéaire à $\nabla g(x)$ (nous notons ∇ pour le gradient). Nous avons

$$\begin{aligned} \nabla f(x) &= \left(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_m}(x) \right) = \left(\frac{-p_1^2 \sigma_1^2}{x_1^2}, \dots, \frac{-p_m^2 \sigma_m^2}{x_m^2} \right), \\ \nabla g(x) &= (1, 1, \dots, 1). \end{aligned}$$

Donc nous cherchons $x \in S$ et $\lambda \in \mathbb{R}$ tels que, pour tout $i \in \{1, \dots, m\}$, $\frac{-p_i^2 \sigma_i^2}{x_i^2} = \lambda$. L'unique solution est $x_i = n \frac{p_i \sigma_i}{p_1 \sigma_1 + \dots + p_m \sigma_m}$, pour tout i . La fonction f est convexe (sur \mathbb{R}_+^m) donc ce point critique est un minimum. Comme il est l'unique point critique alors, c'est un minimum absolu. \square

Comme nous voulons des n_i entiers, nous prenons $n_i = \left\lfloor n \frac{p_i \sigma_i}{p_1 \sigma_1 + \dots + p_m \sigma_m} \right\rfloor$, pour tout i (rappel : $\lfloor x \rfloor = \inf\{n \in \mathbb{Z} : n \leq x\}$). Dans la pratique, il faudra donc estimer les σ_i et les p_i (par une méthode de Monte-Carlo). Si on oublie les parties entières, la deuxième méthode est de variance

$$\sum_{i=1}^m \frac{p_i^2 \sigma_i^2}{n_i} = \sum_{i=1}^m p_i^2 \sigma_i^2 \frac{(p_1 \sigma_1 + \dots + p_m \sigma_m)}{n p_i \sigma_i} = \frac{(p_1 \sigma_1 + \dots + p_m \sigma_m)^2}{n}.$$

LEMME 3.6. *Nous avons*

$$\sigma^2 \geq (p_1 \sigma_1 + \dots + p_m \sigma_m)^2.$$

(La variance est donc bien réduite par la stratification.)

DÉMONSTRATION. Nous avons

$$\begin{aligned} \mathbb{V}(g(X)) &= \mathbb{E} \left(\left(\sum_{i=1}^m g(X) \mathbb{1}_{\{X \in D_i\}} \right)^2 \right) - \mathbb{E}(g(X))^2 \\ (\text{les } D_i \text{ partitionnent } D) &= \mathbb{E} \left(\sum_{i=1}^m g(X)^2 \mathbb{1}_{\{X \in D_i\}} \right) - \left(\sum_{i=1}^m p_i \mathbb{E}(g(X)|X \in D_i) \right)^2 \\ &= \sum_{i=1}^m \{p_i \mathbb{E}(g(X)^2|X \in D_i) - p_i \mathbb{E}(g(X)|X \in D_i)^2\} \\ &\quad + \sum_{i=1}^m \{p_i \mathbb{E}(g(X)|X \in D_i)^2\} - \left(\sum_{i=1}^m p_i \mathbb{E}(g(X)|X \in D_i) \right)^2 \end{aligned}$$

Rappelons que pour $x_1, \dots, x_m \in \mathbb{R}$ et $\lambda_1, \dots, \lambda_m \in \mathbb{R}_+$ tels que $\sum_{i=1}^m \lambda_i = 1$, nous avons

$$(3.2) \quad \sum_{i=1}^m \lambda_i x_i^2 \geq \left(\sum_{i=1}^m \lambda_i x_i \right)^2$$

(c'est l'inégalité de convexité pour la fonction carré). Donc

$$\begin{aligned} \mathbb{V}(g(X)) &\geq \sum_{i=1}^m \{p_i \mathbb{E}(g(X)^2|X \in D_i) - p_i \mathbb{E}(g(X)|X \in D_i)^2\} \\ (\text{encore (3.2)}) &\geq \left(\sum_{i=1}^m p_i (\mathbb{E}(g(X)^2|X \in D_i) - \mathbb{E}(g(X)|X \in D_i)^2)^{1/2} \right)^2 \\ &= \left(\sum_{i=1}^m p_i \sigma_i \right)^2. \end{aligned}$$

□

3.5. Valeur moyenne ou conditionnement

On cherche à calculer une intégrale de la forme

$$I = \mathbb{E}(g(X, Y)) = \int g(x, y) f(x, y) dx dy$$

où f est la densité du couple (X, Y) . Cette écriture nous donne l'idée d'une première approximation

$$I \approx \frac{1}{n} \sum_{i=1}^n g(X_i, Y_i)$$

avec des (X_i, Y_i) i.i.d. de même loi que (X, Y) . Soit $h(x) = \frac{1}{m(x)} \int g(x, y) f(x, y) dy$ avec $m(x) = \int f(x, y) dy$. Nous avons

$$\mathbb{E}(h(X)) = \int_{x,y} h(x) f(x, y) dx dy$$

$$\begin{aligned}
&= \int_{x,y} \frac{1}{\int_u f(x,u)du} \left(\int_v g(x,v)f(x,v)dv \right) f(x,y)dy \\
\text{(Fubini)} &= \int_{v,x} g(x,v)f(x,v) \frac{\int_y f(x,y)dy}{\int_u f(x,u)du} dx dv \\
&= \int_{v,x} g(x,v)f(x,v) dx dv \\
&= \mathbb{E}(g(X,Y)).
\end{aligned}$$

Supposons que nous savons calculer $h(x)$ pour tout x . Nous avons alors l'idée d'une deuxième approximation

$$I \approx \frac{h(X_1) + \dots + h(X_n)}{n}.$$

LEMME 3.7. *Nous avons*

$$\mathbb{V}(h(X)) \leq \mathbb{V}(g(X,Y)).$$

(Donc la méthode de la valeur moyenne réduit la variance.)

DÉMONSTRATION. Nous savons que $\mathbb{E}(h(X)) = \mathbb{E}(g(X,Y))$ donc pour comparer $\mathbb{V}(h(X))$ et $\mathbb{V}(g(X,Y))$, il suffit de comparer $\mathbb{E}(h(X)^2)$ et $\mathbb{E}(g(X,Y)^2)$. Nous avons :

$$\begin{aligned}
\mathbb{E}(h(X)^2) &= \int_{x,y} h(x)^2 f(x,y) dx dy \\
&= \int_{x,y} \left(\frac{\int_v g(x,v)f(x,v)dv}{\int_u f(x,u)du} \right)^2 f(x,y) dx dy \\
\text{(Cauchy-Schwarz)} &\leq \int_{x,y} \frac{\int_v g(x,v)^2 f(x,v)dx}{\int_u f(x,u)du} f(x,y) dx dy \\
\text{(Fubini)} &= \int_{x,v} \frac{g(x,v)^2 f(x,v)}{\int_u f(x,u)du} \left(\int_y f(x,y)dy \right) dx dv \\
&= \int_{x,v} g(x,v)^2 f(x,v) dx dv \\
&= \mathbb{E}(g(X,Y)^2).
\end{aligned}$$

Donc $\mathbb{V}(h(X)) \leq \mathbb{V}(g(X,Y))$. □

3.6. Exercices

EXERCICE 3.1. Cet exercice est inspiré par l'exercice 1.7 de [Par08]. On veut calculer

$$I = \mathbb{E}(\mathbf{1}_{X>0} e^{\beta X}),$$

où $X \sim \mathcal{N}(0,1)$ et $\beta = 5$. On estimera la variance à chaque étape de l'exercice.

- (1) Calculer (par Monte-Carlo) la variance de la méthode « naïve » (quand on tire des X_1, X_2, \dots i.i.d. de loi $\mathcal{N}(0,1)$ et que l'on approche I par $\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i>0} e^{\beta X_i}$).
- (2) Proposer une méthode d'échantillonnage préférentiel.
- (3) Proposer une méthode de variable de contrôle.
- (4) Améliorer la méthode à l'aide d'une technique de variable antithétique.

EXERCICE 3.2. (1) On s'intéresse au programme suivant.

Programme 3.3 Boucle de Monte-Carlo

```

n=1000
s=0
for (i in 1:n)
{
  u=rnorm(1,0,1) #tire une variable N(0,1)
  s=s+sqrt(abs(u))
}
print(s/n)

```

On note Z le résultat affiché par l'algorithme. Quelle est la quantité I approchée par Z ?

- (2) Écrire un programme qui estime la variance de la méthode ci-dessus.
- (3) Proposer une méthode de réduction de variance pour le calcul de I .
- (4) Écrire un programme qui implémente cette nouvelle méthode. Écrire un programme qui estime la nouvelle variance (elle soit être plus petite que la variance de la question 2).

EXERCICE 3.3. On veut calculer $p_l = \mathbb{P}(X \in [l; l+1])$ pour X une variable exponentielle de paramètre 1 et $l \geq 0$.

- (1) (a) Proposer une méthode de Monte-Carlo pour calculer p_l .
(b) Calculer la variance de cette méthode.
- (2) (a) Proposer une méthode d'échantillonnage préférentiel (par exemple : telle que les nouveaux tirages soient tous dans $[l; l+1]$).
(b) Calculer la variance de cette méthode.
- (3) Comparer les variances des deux méthodes quand $l \rightarrow +\infty$.

Recommencer avec $X \sim \mathcal{N}(0; 1)$. On pourra comparer les variances entre les cas suivants : densité \tilde{f} uniforme sur $[0, 1]$ et densité \tilde{f} de la loi exponentielle conditionnée à tomber dans $[l; l+1]$.

EXERCICE 3.4 (Recyclage dans la méthode du rejet). On veut évaluer par une méthode de Monte-Carlo

$$\int_{\mathbb{R}} f(x)p(x)dx$$

(où p est une densité de probabilité). On fixe un entier n . On simule X_1, \dots, X_n de densité p par la méthode du rejet (basée sur une inégalité $p \leq Mq$). On doit donc simuler un nombre aléatoire N de variables aléatoires Y_i de densité q . Parmi celles-ci, notons Z_1, \dots, Z_{N-n} les variables Y qui sont rejetées, c'est à dire telles que :

$$p(Y_i) \leq MU_i q(Y_i).$$

- (1) Quelle est la loi de N ?
- (2) Quelle est la loi de Z_1, \dots, Z_{N-n} conditionnellement à N et n ?
- (3) Montrez que

$$\frac{1}{N} \left(\sum_{i=1}^n f(X_i) + \sum_{i=1}^{N-n} \frac{(M-1)p(Z_i)}{(Mq-p)(Z_i)} f(Z_i) \right)$$

est un estimateur sans biais de $E(f(X_i))$.

Méthodes de Monte-Carlo par chaînes de Markov

Dans tout ce chapitre, on suppose que l'on est dans un espace E fini. Les algorithmes décrits fonctionnent aussi dans le cas E dénombrable.

4.1. Rappels sur les chaînes de Markov

THÉORÈME 4.1 (Théorème ergodique). *Soit Q un noyau de Markov irréductible (sur E). Alors, il existe une probabilité Q -invariante π_0 et de plus :*

- (1) π_0 est l'unique probabilité Q -invariante.
- (2) Tous les états sont récurrents (pour Q).
- (3) Si $(\bar{X}_n)_{n \geq 0}$ est une chaîne de Markov de loi initiale π_0 et de transition Q alors

$$\frac{1}{n+1} \sum_{p=0}^n f(\bar{X}_p) \xrightarrow[n \rightarrow +\infty]{p.s.} \int f(x) \pi_0(dx),$$

pour toute fonction f qui est π_0 -intégrable.

- (4) Si $(X_n)_{n \geq 0}$ est une chaîne de Markov de loi initiale π quelconque et de transition Q alors

$$\frac{1}{n+1} \sum_{p=0}^n f(X_p) \xrightarrow[n \rightarrow +\infty]{p.s.} \int f(x) \pi_0(dx),$$

pour toute fonction f qui est π_0 -intégrable.

Ce théorème nous permet d'approcher une intégrale par une moyenne empirique. Ici, la suite des (X_p) n'est pas i.i.d. mais est une chaîne de Markov. On parle alors de méthode de Monte-Carlo par chaîne de Markov (« MCMC » pour « Monte Carlo Markov Chain » en anglais).

DÉFINITION 4.2. Si π une probabilité sur E vérifie $\pi(x)Q(x, y) = \pi(y)Q(y, x)$ pour tous x, y , alors π est dite symétrique (par rapport à Q) ou Q -symétrique.

LEMME 4.3. *Si π est Q -symétrique, alors elle est Q -invariante.*

DÉMONSTRATION. Pour tout y ,

$$\begin{aligned} (\pi Q)(y) &= \sum_{x \in E} \pi(x)Q(x, y) \\ (\text{symétrie}) &= \sum_{x \in E} \pi(y)Q(y, x) \\ &= \pi(y). \end{aligned}$$

□

DÉFINITION 4.4. Soit Q un noyau de Markov. Pour n dans \mathbb{N}^* , on note

$$Q^n(x, y) = \sum_{x_1, \dots, x_{n-1}} Q(x, x_1)Q(x_1, x_2) \dots Q(x_{n-1}, y).$$

Pour $x \in E$, on note

$$d(x) := \text{PGCD}\{n : Q^n(x, x) > 0\}.$$

Un élément x de E est dit apériodique si $d(x) = 1$.

EXEMPLE 4.5 (Marche simple dans \mathbb{Z}). Soit Q noyau de Markov de \mathbb{Z} dans \mathbb{Z} donné par

$$Q(x, x+1) = \frac{1}{2}, \quad Q(x, x-1) = \frac{1}{2}, \quad \forall x.$$

Pour tout x , $Q^2(x, x) = \mathbb{P}(X_2 = x | X_0 = x)$ (où (X_n) chaîne de Markov de transition Q). Nous avons donc

$$Q^2(x, x) \geq \mathbb{P}(X_2 = x, X_1 = x+1 | X_0 = x) = \left(\frac{1}{2}\right)^2 > 0.$$

Par ailleurs, si $Q^n(x, x) > 0$ alors n est pair (on fait forcément un nombre pair de pas pour aller de x à x). Donc $d(x) = 2$. Donc x n'est pas apériodique.

DÉFINITION 4.6. Soit Q un noyau de Markov. Si tous les x de E sont apériodique alors Q est dit apériodique.

LEMME 4.7. Si Q est irréductible, $[\exists x \in E \text{ qui est apériodique}] \Rightarrow [Q \text{ est apériodique}]$.

THÉORÈME 4.8. Supposons que Q est une matrice de transition irréductible, apériodique et admettant une probabilité invariante π_0 .

(1) Il existe deux réels $\alpha \in [0; 1[$, $M \in \mathbb{R}_+$ tels que pour tout $x, y \in E$,

$$|Q^n(x, y) - \pi_0(y)| \leq M\alpha^n.$$

(2) Pour toute fonction $f : E \rightarrow \mathbb{R}$, si on appelle $(X_n)_{n \geq 0}$ une chaîne de Markov de loi initiale π quelconque et de transition Q

$$\sqrt{n+1} \left(\frac{1}{n+1} \sum_{p=0}^n f(X_p) - \sum_{y \in E} \pi_0(y) f(y) \right) \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathcal{N}(0, \bar{\sigma}^2),$$

avec une variance $\bar{\sigma}^2 < +\infty$.

REMARQUE 4.9. Dans le cas d'un espace fini ou dénombrable, $\int_{x \in E} f(x) \pi(dx) = \sum_{x \in E} f(x) \pi(x)$ pour toute probabilité π et toute fonction f .

REMARQUE 4.10. Pour toute chaîne $(X_n)_{n \geq 0}$ de transition Q vérifiant les hypothèses du théorème ci-dessus, $X_n \xrightarrow[n \rightarrow +\infty]{\text{loi}} \pi_0$ (c'est le point 1). Le point 2 du théorème ci-dessus nous donne une vitesse de convergence pour la convergence du théorème ergodique (comme le théorème de la limite centrale nous donne la vitesse de convergence de la loi des grands nombres).

4.2. Algorithme de Hastings-Metropolis

On se fixe une loi π suivant laquelle on aimerait simuler ou dont on voudrait calculer une intégrale $\int_E f(x) \pi(dx)$. Nous appellerons π la *loi cible*. On se donne un noyau de Markov Q . Nous appellerons Q le *noyau de proposition*. Nous allons construire une chaîne de Markov $(X_n)_{n \geq 0}$.

- Nous prenons $X_0 = x_0$ tel que $\pi(x_0) > 0$.
- Si $X_n = x$, nous simulons Y_{n+1} et U_{n+1} indépendants (et indépendants des simulations passées) avec

$$Y_{n+1} \sim Q(x, \cdot), \quad U_{n+1} \sim \mathcal{U}([0; 1]).$$

La variable Y_{n+1} s'appelle une *proposition*. Posons, pour tout x, y ,

$$\alpha(x, y) = \min \left(1, \frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)} \right).$$

Alors :

$$X_{n+1} = \begin{cases} Y_{n+1} & \text{si } U_{n+1} \leq \alpha(X_n, Y_{n+1}), \\ X_n & \text{sinon.} \end{cases}$$

Dans le cas où $U_{n+1} \leq \alpha(X_n, Y_{n+1})$, on dit qu'on accepte la proposition, et dans le cas contraire on dit qu'on refuse la proposition.

PROPOSITION 4.11. *La suite aléatoire $(X_n)_{n \geq 0}$ construite ci-dessus est une chaîne de Markov de transition P avec*

$$\begin{cases} P(x, y) = Q(x, y)\alpha(x, y) & \text{si } x \neq y, \\ P(x, x) = 1 - \sum_{y \neq x} P(x, y). \end{cases}$$

La loi π est P -invariante

DÉMONSTRATION. Pour tout x , $\mathbb{P}(X_{n+1} = x | X_0, \dots, X_n) = \mathbb{P}(X_{n+1} = x | X_n)$ (d'après la construction ci-dessus). Donc $(X_n)_{n \geq 0}$ est bien une chaîne de Markov. Calculons, pour tout $x \neq y$

$$\begin{aligned} \mathbb{P}(X_{n+1} = y | X_n = x) &= \mathbb{P}(Y_{n+1} = y, U_{n+1} \leq \alpha(X_n, Y_{n+1}) | X_n = x) \\ &= \mathbb{P}(Y_{n+1} = y, U_{n+1} \leq \alpha(x, Y_{n+1}) | X_n = x) \\ &= \mathbb{P}(Y_{n+1} = y, U_{n+1} \leq \alpha(x, y) | X_n = x) \\ (Y_{n+1} \perp\!\!\!\perp U_{n+1}) &= Q(x, y)\alpha(x, y). \end{aligned}$$

Nous en déduisons

$$\mathbb{P}(X_{n+1} = x | X_n = x) = 1 - \sum_{y \neq x} P(x, y) = 1 - \sum_{y \neq x} Q(x, y)\alpha(x, y).$$

Pour tout $x \neq y$,

$$\begin{aligned} \pi(x)P(x, y) &= \pi(x)Q(x, y)\alpha(x, y) \\ &= \pi(x)Q(x, y) \min\left(1, \frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)}\right) \\ &= \min(\pi(x)Q(x, y), \pi(y)Q(y, x)) \\ &= \min(\pi(y)Q(y, x), \pi(x)Q(x, y)) \end{aligned}$$

$$(\text{on refait le calcul en inversant } x \text{ et } y) = \pi(y)P(y, x)$$

Donc π est symétrique par rapport à P . D'après le lemme 4.3, nous avons donc $\pi P = \pi$. \square

EXEMPLE 4.12. Soit Q défini dans l'exemple 4.5 et π une loi sur \mathbb{Z} telle que $\pi(x) = C \exp(-\sqrt{|x|})$ pour tout x (C est alors la constante telle que $\sum_{x \in \mathbb{Z}} \pi(x) = 1$). Le programme 4.1 fait une simulation d'une marche de Metropolis de noyau de proposition Q et de loi-cible π . On remarque qu'il n'est pas nécessaire de connaître C pour implémenter l'algorithme.

Programme 4.1 Chaîne de Metropolis

```
n=100
f<-function(k)
{
  return(exp(-sqrt(abs(k))))
}
liste=c()
x=0
for (k in 1:n)
{
  liste=c(liste,x)
  v=runif(1,0,1)
  if (v<0.5)
  { y=x+1 }
  else
  { y=x-1 }
  u=runif(1,0,1)
  alpha=min(1,f(y)*0.5/(f(x)*0.5))
  if (u<alpha)
  { x=y }
}

```

LEMME 4.13. Avec les notations définies ci-dessus, si $Q(x, y)$ est irréductible et vérifie : $\forall x, y, Q(x, y) \neq 0 \Rightarrow Q(y, x) \neq 0$, et $\pi(x) > 0$ pour tout x, y alors P est irréductible.

DÉMONSTRATION. Soient x, y dans $E, x \neq y$. Il existe n dans \mathbb{N} et x_1, \dots, x_n dans E tels que

$$Q(x, x_1)Q(x_1, x_2) \dots Q(x_n, y) > 0.$$

On peut toujours supposer $x \neq x_1, \dots, x_n \neq y$. On déduit des hypothèses que

$$P(x, x_1)P(x_1, x_2) \dots P(x_n, y) > 0.$$

□

4.3. Algorithme de Metropolis simple

Si $Q(x, y) = Q(y, x)$ pour tout x, y (on dit que le noyau Q est *symétrique*). Dans la version originale de l'algorithme de Metropolis, le noyau Q est symétrique. Dans ce cas, α se simplifie en $\alpha(x, y) = \min\left(1, \frac{\pi(y)}{\pi(x)}\right)$ pour tout x, y .

PROPOSITION 4.14. Si Q est un noyau de Markov irréductible, symétrique et si π est une probabilité non constante telle que $\pi(x) > 0$ pour tout x , alors la chaîne de Markov de Metropolis de noyau de proposition Q et de loi cible π est irréductible, apériodique, de loi invariante π .

DÉMONSTRATION. Soit P le noyau de Metropolis de noyau de proposition Q et de loi cible π . Pour tout $x, y, \alpha(x, y) > 0$. De plus, pour tout x, y , il existe $p \in \mathbb{N}$ et une suite x_1, \dots, x_p de E telle que $Q(x, x_1)Q(x_1, x_2) \dots Q(x_{p-1}, x_p)Q(x_p, y) > 0$. Donc :

$$P(x, x_1)P(x_1, x_2) \dots P(x_p, y) = Q(x, x_1)\alpha(x, x_1) \dots Q(x_p, y)\alpha(x_p, y) > 0.$$

Pour prouver l'apériodicité de P , il suffit de montrer qu'il existe x tel que $P(x, x) > 0$. Supposons que $P(x, x) = 0$ pour tout x .

$$\begin{aligned} P(x, x) &= 1 - \sum_{y:y \neq x} P(x, y) \\ &= \sum_y Q(x, y) - \sum_{y:y \neq x} Q(x, y)\alpha(x, y) \\ &= Q(x, x) + \sum_{y:y \neq x} Q(x, y)(1 - \alpha(x, y)) \\ &= Q(x, x) + \sum_{y:x \neq y} Q(x, y) \left(1 - \frac{\pi(y)}{\pi(x)}\right)_+, \end{aligned}$$

donc

$$P(x, x) - Q(x, x) = \sum_{y:x \neq y} Q(x, y) \left(1 - \frac{\pi(y)}{\pi(x)}\right)_+.$$

Fixons $x \in E$. Les deux termes de l'équation ci-dessus sont de signes opposés, ils valent donc 0. Nous avons donc, pour tout y tel que $y \neq x$ et $Q(x, y) \neq 0$ (il en existe car Q est irréductible),

$$(4.1) \quad \left(1 - \frac{\pi(y)}{\pi(x)}\right)_+ = 0,$$

c'est à dire $\pi(y) \geq \pi(x)$.

Nous avons donc montré, pour tout x , que si y est tel que $Q(x, y) \neq 0$, alors $\pi(y) \geq \pi(x)$. Nous fixons maintenant x et y tels que $Q(x, y) \neq 0$. Comme Q est symétrique, nous avons $Q(y, x) = Q(x, y)$ et donc $\pi(x) \geq \pi(y)$. Donc $\pi(x) = \pi(y)$.

Nous avons maintenant montré que pour tous x, y tels que $Q(x, y) \neq 0, \pi(x) = \pi(y)$. Comme Q est irréductible, cela implique que π est constante, ce qui est en contradiction avec les hypothèses. Conclusion : Il existe x tel que $P(x, x) > 0$. Donc le noyau P est apériodique. □

Sous les hypothèse de cette proposition, nous pouvons appliquer le théorème 4.8 sur la vitesse de convergence.

4.4. Le modèle d'Ising

C'est un modèle très populaire en physique statistique. Soit $N \in \mathbb{N}$. Soit

$$\Lambda = \{-N, -N+1, \dots, -1, 0, 1, \dots, N\}^2 \subset \mathbb{Z}^2$$

(on pourrait aussi regarder ce modèle en dimension quelconque en prenant l'exposant $d \in \mathbb{N}^*$ à la place de 2). Nous définissons l'espace des configurations par

$$E = \{-1; 1\}^\Lambda$$

(il s'agit donc des fonctions $f : \Lambda \rightarrow \{-1; 1\}$). Le cardinal de E est 2^{2N+1} . Pour $x \in E$, nous posons

$$H(x) = \frac{1}{2} \sum_{\substack{m, m' \in \Lambda \\ |m-m'|=1}} |x(m) - x(m')|^2,$$

où la somme porte sur toutes les paires de sites m, m' de E qui sont voisins. Remarquons que $H(x)$ est nulle si x est constante. Soit $\beta > 0$, nous définissons la probabilité sur E :

$$\pi(x) = \frac{1}{Z(\beta)} e^{-\beta H(x)},$$

avec $Z(\beta) = \sum_{x \in E} e^{-\beta H(x)}$. La constante $Z(\beta)$ n'est pas connue.

Nous aimerions simuler suivant la loi π ou calculer des intégrales contre la loi π . Pour cela, nous allons utiliser l'algorithme de Metropolis. Nous allons utiliser un noyau de proposition Q particulier (appelé échantillonneur de Gibbs).

Nous définissons une partition de Λ en

$$\Lambda^+ = \{(m_1, m_2) \in \Lambda; m_1 + m_2 \text{ pair}\},$$

$$\Lambda^- = \{(m_1, m_2) \in \Lambda; m_1 + m_2 \text{ impair}\}.$$

Remarquons que si m appartient à Λ^+ alors tous ses voisins sont dans Λ^- (et réciproquement).

Pour $x \in E$, nous notons

$$x^+ = (x(m), m \in \Lambda^+),$$

$$x^- = (x(m), m \in \Lambda^-).$$

Les composantes x^+ et x^- sont des restrictions de x (respectivement à Λ^+ et Λ^-) et nous écrivons $x = (x^+, x^-)$. Nous cherchons maintenant à calculer $\pi(x^+|x^-)$ ($\pi(x^+|x^-) = \mathbb{P}(X^+ = x^+ | X^- = x^-)$) avec X variable aléatoire de loi π). Nous avons :

$$\begin{aligned} \pi(x^+|x^-) &= \frac{\pi(x^+, x^-)}{\pi(x^-)} \\ &= \frac{\pi(x^+, x^-)}{\sum_{\substack{y \in E \\ y^- = x^-}} \pi(y)} \\ &= \frac{\exp\left(-\beta \sum_{m \in \Lambda^+} \sum_{\substack{m' \in \Lambda^- \\ |m-m'|=1}} (x^+(m) - x^-(m'))^2\right)}{\sum_{\substack{y \in E \\ y^- = x^-}} \exp\left(-\beta \sum_{m \in \Lambda^+} \sum_{\substack{m' \in \Lambda^- \\ |m-m'|=1}} (y^+(m) - y^-(m'))^2\right)} \\ &= \frac{\exp\left(-\beta \sum_{m \in \Lambda^+} \sum_{\substack{m' \in \Lambda^- \\ |m-m'|=1}} (1 + 1 - 2x^+(m)x^-(m'))\right)}{\sum_{\substack{y \in E \\ y^- = x^-}} \exp\left(-\beta \sum_{m \in \Lambda^+} \sum_{\substack{m' \in \Lambda^- \\ |m-m'|=1}} (1 + 1 - 2y^+(m)y^-(m'))\right)} \end{aligned}$$

$$\begin{aligned}
&= \frac{\exp\left(2\beta \sum_{m \in \Lambda^+} \sum_{\substack{m' \in \Lambda^- \\ |m-m'|=1}} x^+(m)x^-(m')\right)}{\sum_{\substack{y \in E \\ y^- = x^-}} \exp\left(2\beta \sum_{m \in \Lambda^+} \sum_{\substack{m' \in \Lambda^- \\ |m-m'|=1}} y^+(m)x^-(m')\right)} \\
&= \frac{\prod_{m \in \Lambda^+} \exp\left(2\beta x^+(m) \sum_{\substack{m' \in \Lambda^- \\ |m-m'|=1}} x^-(m')\right)}{\sum_{\substack{y \in E \\ y^- = x^-}} \exp\left(2\beta \sum_{m \in \Lambda^+} \sum_{\substack{m' \in \Lambda^- \\ |m-m'|=1}} y^+(m)x^-(m')\right)}.
\end{aligned}$$

Cette expression est un produit de fonctions de $x^+(m)$ sur $m \in \Lambda^+$. Donc, sous la loi $\pi(\cdot|x^-)$, les composantes $X^+(m)$, $m \in \Lambda$ sont indépendantes et de loi

$$\mathbb{P}(X^+(m) = x^+(m) | X^- = x^-) \propto \exp\left(2\beta x^+(m) \sum_{\substack{m' \in \Lambda^- \\ |m-m'|=1}} x^-(m')\right).$$

Donc, en notant $M = \sum_{\substack{m' \in \Lambda^- \\ |m-m'|=1}} x^-(m')$

$$\begin{aligned}
\mathbb{P}(X^+(m) = 1 | X^- = x^-) &= \frac{e^{2\beta M}}{e^{2\beta M} + e^{-2\beta M}}, \\
\mathbb{P}(X^+(m) = -1 | X^- = x^-) &= \frac{e^{-2\beta M}}{e^{2\beta M} + e^{-2\beta M}}.
\end{aligned}$$

Il est donc très facile de simuler suivant $\pi(\cdot|x^-)$. Nous définissons maintenant un noyau Q en décrivant comment simuler un saut selon Q . Soit x dans E . Nous tirons $U \sim \mathcal{B}(1/2)$. Si $U \leq 1/2$, nous tirons Y^+ de loi $\pi(\cdot|x^-)$ et nous sautons en (Y^+, x^-) . Si $U > 1/2$, nous tirons Y^- de loi $\pi(\cdot|x^+)$ et nous sautons en (x^+, Y^-) . Nous pouvons alors écrire :

$$Q(x, y) = \begin{cases} \frac{1}{2}\pi(y^+|x^-) & \text{si } y^- = x^- \text{ et } y^+ \neq x^+, \\ \frac{1}{2}\pi(y^-|x^+) & \text{si } y^+ = x^+ \text{ et } y^- \neq x^-, \\ \frac{\pi(x^+|x^-) + \pi(x^-|x^+)}{2} & \text{si } x = y, \\ 0 & \text{dans les autres cas.} \end{cases}$$

Soit

$$\alpha(x, y) = \min\left(1, \frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)}\right).$$

Calculons dans le cas $x^- = y^-$ (les autres cas étant similaires) :

$$\begin{aligned}
\frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)} &= \frac{\pi(y^+|y^-)\mathbb{P}(X^- = y^-)(1/2)\pi(x^+|y^-)}{\pi(x^+|x^-)\mathbb{P}(X^- = x^-)(1/2)\pi(y^+|x^-)} \\
&= 1.
\end{aligned}$$

Donc $\alpha(x, y) = 1$ tout le temps¹. En particulier, pour x, y dans E , nous n'avons pas besoin de $Z(\beta)$ pour calculer $\alpha(x, y)$. Nous pouvons donc simuler la chaîne de Metropolis de noyau de proposition Q et de loi cible π . Nous noterons $(X_n)_{n \geq 0}$ cette chaîne. Le noyau Q est irréductible et symétrique. La loi π n'est pas constante. Donc par la proposition 4.14 et le théorème 4.8, la loi de X_n est proche de π (pour $n \rightarrow +\infty$) et, pour une fonction f , nous pouvons faire l'approximation

$$\int_E f(x)\pi(dx) \approx \frac{1}{n+1} \sum_{n=0}^{n+1} f(X_n),$$

quand $n \rightarrow +\infty$.

1. Nous sommes donc en présence d'un algorithme de Metropolis un peu dégénéré, qui s'appelle en fait un échantillonneur de Gibbsz. Voir [Par08] pour plus de précisions sur cette classe d'algorithmes.

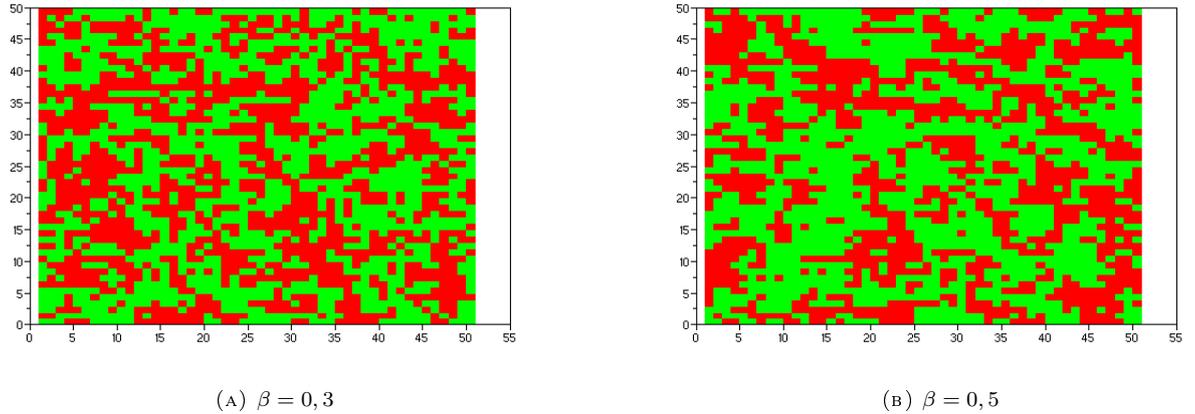


FIGURE 4.4.1

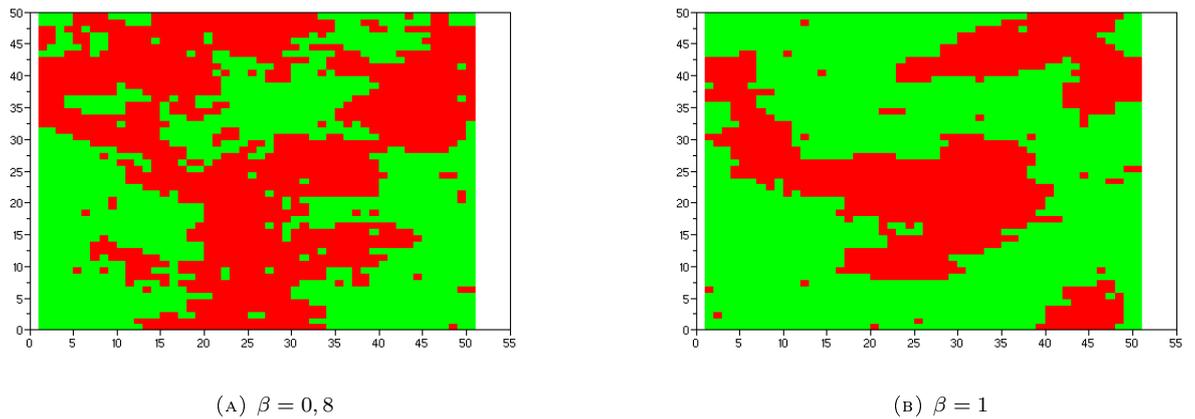


FIGURE 4.4.2

Dans les figures (4.4.1) et (4.4.2), nous avons simulé des tirages suivant la loi π à l'aide de l'algorithme de Metropolis (les résultats sont donc approximativement de loi π) pour $N = 25$. Les sites en rouge (ou gris foncé) représentent des $+1$ et les sites en vert (ou gris clair) représentent des -1 . On s'attend à ce que les variables tirées soient représentatives de la loi π . On remarque que plus β est grand, plus la loi π privilégie les configurations x tels que le nombre de couples de voisins en désaccord est petit (on dit que $m, m' \in \Lambda$ sont en désaccord si $x(m) \neq x(m')$). Et en effet, dans les tirages effectués, les deux couleurs sont plus mélangées pour β petit que pour β grand.

4.5. Analyse bayésienne d'image

Reprenons le modèle précédent et associons à chaque x de E une image : $m \in \Lambda$ représente un pixel, le pixel est noir si $x(m) = +1$ et blanc si $x(m) = -1$. Nous observons les couleurs des pixels et l'observation nous donne la couleur exacte de chaque pixel avec probabilité p ($p \in]0; 1[$). Alors, la loi a posteriori, ou encore la loi conditionnelle de $X = x$ sachant que l'on a observé y (l'image de départ est modélisée par une variable aléatoire X parce que nous ne la connaissons pas) est

$$\pi(x|y) \propto e^{-\beta H(x)} p^{a(x,y)} (1-p)^{d(x,y)},$$

où $a(x, y) = \#\{m \in \Lambda : x(m) = y(m)\}$, $d(x, y) = \#\{m \in \Lambda : x(m) \neq y(m)\}$. Nous cherchons alors à simuler une variable de loi $\pi(\cdot|y)$ (pour avoir une idée de l'image de départ).

REMARQUE 4.15. Le paramètre p représente l'erreur de mesure. Plus p est petit, plus l'image observée est bruitée par rapport à l'image de départ. Le paramètre β dépend de l'idée que l'on se fait du contraste de l'image de départ. Plus β est grand, plus on privilégie les images de fort contraste.

Nous allons construire une chaîne de Metropolis semblable à celle de la partie précédente. L'observation y est fixée (c'est l'image dont nous disposons) et la loi cible est $\mu = (\mu(x), x \in E)$ avec $\mu(x) = \pi(x|y)$ pour tout x . En nous inspirant de la partie précédente, nous obtenons :

$$\begin{aligned} \mu(x^+|x^-) &= \frac{\mu(x^+, x^-)}{\sum_{\substack{z \in E \\ z^- = x^-}} \mu(z)} \\ &= \frac{\exp\left(-\beta \sum_{m \in \Lambda^+} \sum_{\substack{m' \in \Lambda^- \\ |m-m'|=1}} (x^+(m) - x^-(m'))^2\right) p^{a(x,y)} (1-p)^{d(x,y)}}{\sum_{\substack{z \in E \\ z^- = x^-}} \exp\left(-\beta \sum_{m \in \Lambda^+} \sum_{\substack{m' \in \Lambda^- \\ |m-m'|=1}} (z^+(m) - z^-(m'))^2\right) p^{a(z,y)} (1-p)^{d(z,y)}} \\ &= \frac{\prod_{m \in \Lambda^+} \exp\left(2\beta x^+(m) \sum_{\substack{m' \in \Lambda^- \\ |m-m'|=1}} x(m')\right)}{\sum_{\substack{y \in E \\ y^- = x^-}} \exp\left(2\beta \sum_{m \in \Lambda^+} \sum_{\substack{m' \in \Lambda^- \\ |m-m'|=1}} y^+(m) x^-(m')\right) p^{a(z,y)} (1-p)^{d(z,y)}} \\ &\quad \times \prod_{m \in \Lambda^+} p^{1-|x^+(m)-y^+(m)|/2} (1-p)^{|x^+(m)-y^+(m)|/2} \\ &\quad \times \prod_{m \in \Lambda^-} p^{1-|x^-(m)-y^-(m)|/2} (1-p)^{|x^-(m)-y^-(m)|/2} \\ &\propto \prod_{m \in \Lambda^+} \exp\left(2\beta x^+(m) \sum_{\substack{m' \in \Lambda^- \\ |m-m'|=1}} x(m')\right) p^{1-|x^+(m)-y^+(m)|/2} (1-p)^{|x^+(m)-y^+(m)|/2}. \end{aligned}$$

Cette expression est un produit de fonctions de $x^+(m)$ sur $m \in \Lambda^+$. Donc, sous la loi $\mu(\cdot|x^-)$, les composantes $X^+(m)$, $m \in \Lambda$ sont indépendantes et de loi

$$\mu(x^+(m)|x^-) \propto \exp\left(2\beta x^+(m) \sum_{\substack{m' \in \Lambda^- \\ |m-m'|=1}} x(m')\right) p^{1-|x^+(m)-y^+(m)|/2} (1-p)^{|x^+(m)-y^+(m)|/2}.$$

Donc, en notant $M = \sum_{\substack{m' \in \Lambda^- \\ |m-m'|=1}} x^-(m')$, si $y^+(m) = 1$,

$$\begin{aligned} \mu(x^+(m) = 1|x^-) &= \frac{e^{2\beta M p}}{e^{2\beta M p} + e^{-2\beta M}(1-p)}, \\ \mu(x^+(m) = -1|x^-) &= \frac{e^{-2\beta M}(1-p)}{e^{2\beta M p} + e^{-2\beta M}(1-p)}; \end{aligned}$$

et si $y^+(m) = -1$,

$$\begin{aligned} \mu(x^+(m) = 1|x^-) &= \frac{e^{2\beta M}(1-p)}{e^{2\beta M}(1-p) + e^{-2\beta M p}}, \\ \mu(x^+(m) = -1|x^-) &= \frac{e^{-2\beta M p}}{e^{2\beta M}(1-p) + e^{-2\beta M p}}. \end{aligned}$$

Il est donc très facile de simuler suivant $\mu(\cdot|x^-)$. Nous définissons maintenant un noyau Q en décrivant comment simuler un saut selon Q . Soit x dans E . Nous tirons $U \sim \mathcal{B}(1/2)$. Si $U \leq 1/2$, nous tirons Z^+ de loi $\mu(\cdot|x^-)$ et nous sautons en (Z^+, x^-) . Si $U > 1/2$, nous tirons Z^- de loi $\mu(\cdot|x^+)$ et nous sautons en (x^+, Z^-) .

Comme dans la section 4.4, pour tous x, y ,

$$\alpha(x, y) = \min \left(1, \frac{\mu(y)Q(y, x)}{\mu(x)Q(x, y)} \right) = 1.$$

Nous pouvons donc simuler la chaîne de Metropolis de noyau de proposition Q et de loi cible μ . Nous noterons $(X_n)_{n \geq 0}$ cette chaîne. Le noyau Q est irréductible et symétrique. La loi μ n'est pas constante. Donc par la proposition 4.14 et le théorème 4.8, la loi de X_n est proche de μ (pour $n \rightarrow +\infty$).

Pour manipuler des images en R, nous utilisons le package PNG². Nous partons de l'image de gauche dans la Figure 4.5.1 (cette image est bruitée) et après quelques itérations, nous arrivons à l'image de droite ($\beta = 4, p = 0.9$).

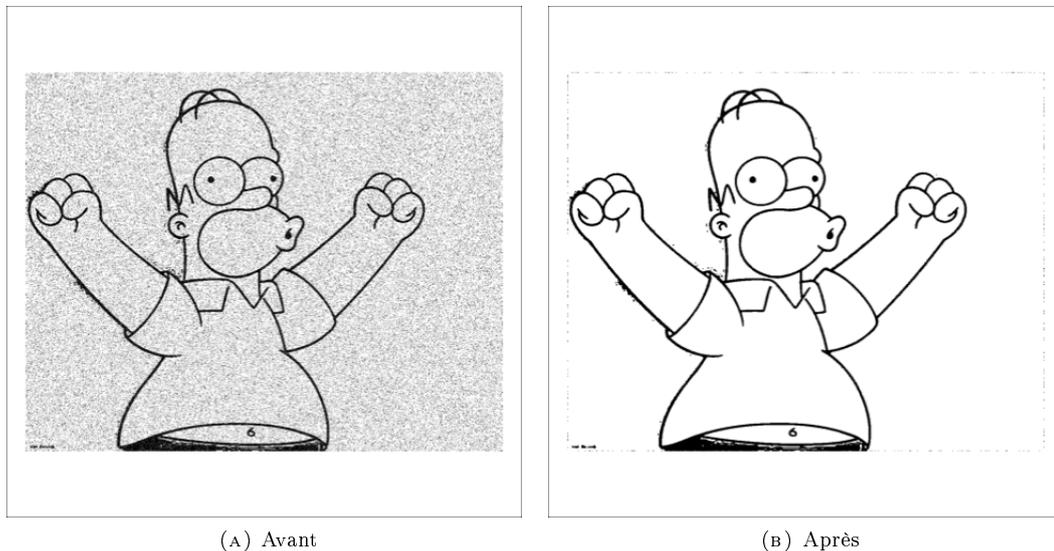


FIGURE 4.5.1

4.6. Cryptographie

Une des techniques de cryptographie les plus simples consiste à remplacer les caractères du texte à coder au moyen d'une bijection. Supposons par exemple que l'ensemble des caractères utilisés soit $\mathcal{A} = \{ 'a', 'b', 'c', \dots, 'x', 'y', 'z', ' ', ' ' \}$ (espace), ' ', ' ' (guillemet simple), ' ' (guillemet double à l'anglaise), ' ', ' ', ' ', ' ', ' ', ' ': }. Cet ensemble est de cardinal 36, on l'identifiera dans la suite à $[36] = \{1, 2, \dots, 36\}$ (chaque caractère est identifié par son numéro). Le codage consiste en une permutation des caractères, c'est à dire en une application g bijective $[36] \rightarrow [36]$. Voici une phrase codée tirée d'une œuvre littéraire bien connue

```
g,hpobv ,gpochb :chbp ;hgpo'og,hpo"b ? :sbyo"bos,gpbymbkohgoabhgbox,eebroch
ogn'm' vo?'po :n' yovy bpo gvb : : lbgvro?'y :ochb :chbpo gpv'gvp'o'mbsoh-
goe,gp bhych opbovy,hm' vo'os,vbo"bo :h ro? h
po:o' : :opn'ppb,yko"bhuxbhybpo ? :hpov'y"roabo :boybgs,gvy' o"bog,hmb'hoqo :obv'
vobgos ,e?'lg bo"nhgos'e'y"bobvo?'y : ' vosx!! ,gp
```

2. voir <https://iut-info.univ-reims.fr/users/nocent/R/#dimensions-de-limage> pour un tutoriel sur ce package

Chaque lettre 'a' de la phrase d'origine a été remplacée par la lettre 'g(a)', etc. Nous notons $E = \{f : [36] \rightarrow [36], \text{ bijective}\}$.

On se donne une matrice de Markov M de taille 36×36 qui représente les probabilités de transition d'un caractère à l'autre quand on regarde les caractères les uns après les autres dans un texte écrit en français. Par exemple, si on calibre cette matrice sur un volume des *Trois Mousquetaires*, $M(2, 1) = 0,14$ parce qu'un 'a' est suivi d'un 'b' dans environ 14% des cas. Notons c_1, c_2, \dots, c_{313} la suite des numéros des caractères apparaissant dans la phrase codée. On s'intéresse à la loi π sur E telle que, pour toute f :

$$\pi(f) \propto \prod_{i=1}^{313} M(f(c_i), f(c_{i+1})).$$

On remarque que plus $\pi(f)$ est grande, plus f est un décodage plausible. Nous voulons construire une chaîne de Markov $(F_n)_{n \geq 0}$ telle que F_n converge en loi vers π quand n tend vers l'infini. Soit un noyau de Markov de proposition Q sur E décrit par :

- Quand on est en $f \in E$. On tire $\{X, Y\}$ suivant la loi uniforme sur $\mathcal{P}_2([36])$ (les parties à deux éléments de $[36]$).
- On saute en $f^{(X,Y)}$ définie par

$$f^{(X,Y)}(z) = \begin{cases} f(z) & \text{si } z \notin \{X, Y\}, \\ f(X) & \text{si } z = Y, \\ f(Y) & \text{si } z = X. \end{cases}$$

Nous remarquons que le noyau Q est symétrique. Ce noyau est irréductible (parce que les transitions engendrent le groupe symétrique). La loi π est non constante. Soit $(F_n)_{n \geq 0}$ une chaîne de Metropolis de noyau de proposition Q et de loi cible π . La proposition 4.14 et le théorème 4.8 nous disent que pour n grand, la loi de F_n est proche de π . Voici ce que l'on peut obtenir après 10000 itérations

nous etions quelquesshuns a nous declamer de monserve, un jeune boppe. qui n'avait cas l'air tres in telligent. carla quelques instants avem un ponsieur qui se trouvait a mote de lui. cuis il all a s'asseoir, deux beures clus tard. je le renmontrai de nouveau - il etait en mopcagnie d'un m aparade et carlait mbiffons

Donc le texte d'origine a de fortes chances d'être :

Nous étions quelques-uns à nous déplacer de conserve. Un jeune homme, qui n'avait pas l'air très intelligent, parla quelques instants avec un monsieur qui se trouvait à côté de lui, puis il alla s'asseoir. Deux heures plus tard, je le rencontrai de nouveau ; il était en compagnie d'un camarade et parlait chiffons.

Raymond Queneau, *Exercices de style (Litotes)*, 1947.

Pour coder cet algorithme de Metropolis, on pourra consulter les archives 2012-2013³ de la page <http://www.math.unice.fr/~rubentha/enseignement> (le devoir no 2 contient des indications sur comment s'y prendre en *scilab*, la matrice M ...) et aussi les archives 2014-2015⁴, ligne « TP sur la cryptographie ... » (pour savoir comment s'y prendre en \mathbb{R}).

4.7. Exercices

EXERCICE 4.1. Soit E, F des espaces finis. Soit Q noyau de transition de E dans E (pour tout $x \in E$, $Q(x, \cdot)$ est une loi de probabilité sur E). Le noyau Q est supposé irréductible et apériodique. On suppose que $Q(x, y) > 0 \Rightarrow Q(y, x) > 0$. Soit K noyau de transition de E dans F (pour tout $x \in E$, $K(x, \cdot)$ est une loi de probabilité sur F). On suppose que pour tout $(x, y) \in E \times F$, $K(x, y) > 0$. On dispose de $Z : E \times F \rightarrow \mathbb{R}^{+*}$ et on note, pour tout $x \in E$, $\sum_{u \in F} Z(x, u)K(x, u) = z(x)$. On s'intéresse à la chaîne de Markov suivante dans $E \times F$.

- $X_0 = x_0 \in E$ (quelconque), $U_0 \sim K(x_0, \cdot)$.

3. <http://math.unice.fr/~rubentha/archives-12-13.html>

4. <http://math.unice.fr/~rubentha/archives-14-15.html>

- Si on a $(X_n, U_n) \in E \times F$.
- On tire $X_{n+1}^* \sim Q(X_n, \cdot)$ et $V_{n+1} \sim \mathcal{U}([0, 1])$ (indépendant de tout le reste). On tire $U_{n+1}^* \sim K(X_{n+1}^*, \cdot)$.
- On calcule

$$\alpha(X_n, U_n, X_{n+1}^*, U_{n+1}^*) = \inf \left(1, \frac{Z(X_{n+1}^*, U_{n+1}^*)Q(X_{n+1}^*, X_n)}{Z(X_n, U_n)Q(X_n, X_{n+1}^*)} \right).$$

- Si $V_{n+1} \leq \alpha(X_n, U_n, X_{n+1}^*, U_{n+1}^*)$ alors $X_{n+1} = X_{n+1}^*$ et $U_{n+1} = U_{n+1}^*$. Sinon $X_{n+1} = X_n$ et $U_{n+1} = U_n$.
- (1) On remarque que pour tout $(x, y) \in E \times F$, $(x', y') \mapsto Q(x, x')K(x', y')$ définit une mesure de probabilité sur $E \times F$ (on a bien $\sum_{x' \in E} \sum_{y' \in F} Q(x, x')K(x', y') = 1$). On notera $\bar{Q}((x, y), \cdot)$ cette mesure. Montrer que \bar{Q} est un noyau de Markov irréductible sur $E \times F$.
 - (2) Montrer que $((X_n, U_n))_{n \geq 0}$ est une chaîne de Metropolis. Préciser son noyau de proposition et sa loi cible. Nous noterons \bar{P} sa transition.
 - (3) Montrer que \bar{P} est irréductible.
 - (4) Montrer que \bar{P} est apériodique.
 - (5) Montrer que $(X_n)_{n \geq 0}$ converge en loi vers une loi proportionnelle à z .

EXERCICE 4.2. Soient p et q des probabilités sur E (espace fini) avec $\forall x, 0 < p(x) \leq cq(x)$ où c constante > 0 . On prend des variables Y_n i.i.d. de loi q , indépendantes de X_0 . On définit par récurrence une chaîne (X_n) :

- On tire X_0 suivant q .
- Quand on a X_n , on tire $U_n \sim \mathcal{U}([0, 1])$ indépendamment de toutes les autres variables et :

$$X_{n+1} = \begin{cases} Y_{n+1} & \text{si } U_n \leq \frac{p(Y_{n+1})}{cq(Y_{n+1})} \\ X_n & \text{sinon} \end{cases}.$$

- (1) Montrer que (X_n) est une chaîne de Markov.
- (2) Calculer la probabilité de transition $P(x, y)$ de (X_n) .
- (3) Quel rapport y-a-t-il entre cette chaîne et une méthode de rejet classique ?
- (4) Calculer μP pour une probabilité μ . En déduire que la loi de X_n converge vers une unique probabilité invariante égale à p .

EXERCICE 4.3. On se place sur $E = \mathbb{Z}$. Soit q loi sur E telle que $q(x) = q(-x)$. Soit p une loi telle que $p(x) > 0, \forall x$. On définit une chaîne de Markov (X_n) par :

- $X_0 \sim p$
- On tire $U_n \sim \mathcal{U}([0, 1])$ indépendante de toutes les autres variables et Z_n de loi q indépendante de toutes les autres variables. On pose $Y_n = X_n + Z_n$ et :

$$X_{n+1} = \begin{cases} Y_n & \text{si } U_n \leq \inf \left(1, \frac{p(Y_n)}{p(X_n)} \right) \\ X_n & \text{sinon} \end{cases}.$$

- (1) Notons Q le noyau tel que Y_n est de loi $Q(X_n, \cdot)$. Montrer que Q est symétrique.
- (2) Soit $\pi(x) = \frac{1}{Z} \exp(-\beta H(x))$ une loi sur E (où $H(x) = x^3$, $\beta = 3$ et Z est la constante de normalisation, non connue). Comment approcher $\int_E f(x)\pi(dx)$ à l'aide d'un algorithme de Métropolis ?

EXERCICE 4.4 (Couplage). On suppose que l'on a un noyau de Markov irréductible Q sur un espace E fini (ou dénombrable) et π probabilité sur E telle que $\pi Q = \pi$. On suppose qu'il existe une probabilité λ sur E et $\epsilon > 0$ tels que $\forall x, y \in E$,

$$\epsilon \lambda(y) \leq Q(x, y).$$

On construit deux chaînes de Markov (X_n) et (Y_n) par :

- On tire X_0 suivant π et Y_0 (indépendamment de X_0) suivant π' .

- Si on a tiré X_0, \dots, X_n et Y_0, \dots, Y_n , on tire $U_n \sim \mathcal{U}([0, 1])$ indépendamment de toutes les autres variables.
- Si $U_n \leq \epsilon$, on tire $X_{n+1} = Y_{n+1}$ suivant λ .
- Sinon
 - Si $X_n = Y_n$, on tire $X_{n+1} = Y_{n+1}$ suivant la loi $\frac{1}{1-\epsilon}(Q(X_n, \cdot) - \epsilon\lambda(\cdot))$ (c'est bien une loi de probabilité).
 - Sinon on tire X_{n+1} suivant $\frac{1}{1-\epsilon}(Q(X_n, \cdot) - \epsilon\lambda(\cdot))$ et Y_{n+1} suivant $\frac{1}{1-\epsilon}(Q(X_n, \cdot) - \epsilon\lambda(\cdot))$ (indépendamment).

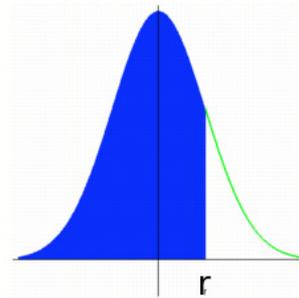
- (1) Montrer que (X_n) et (Y_n) sont des chaînes de Markov de transition Q .
- (2) Montrer que pour presque tout ω , $\exists n$ tel que $X_n(\omega) = Y_n(\omega)$.
- (3) On prend f une fonction bornée. On rappelle que $\frac{1}{n} \sum_{k=1}^n f(X_k) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \sum_{x \in E} f(x)\pi(x)$.
Montrer que

$$\frac{1}{n} \sum_{k=1}^n f(Y_k) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \sum_{x \in E} f(x)\pi(x).$$

Annexe A

Table de la loi normale

$P(X \leq r)$ avec $X \sim N(0,1)$



	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

FIGURE A.0.1. Table de la loi normale

Fonctions, intégrales et sommes usuelles

Nous rappelons $(\forall a, b)$

$$\int_a^b e^{-x} dx = [-e^{-x}]_a^b = -e^{-b} + e^{-a}.$$

Par intégration par parties, nous avons

$$\int_a^b x e^{-x} dx = [-x e^{-x}]_a^b + \int_a^b e^{-x} dx.$$

Et $\forall \alpha$,

$$\int_a^b x^{\alpha-1} e^{-x} dx = [-x^{\alpha-1} e^{-x}]_a^b + (\alpha-1) \int_a^b x^{\alpha-2} e^{-x} dx.$$

Formule du binôme (de Newton) :

$$\forall x, y \in \mathbb{R}, \forall 0 \leq k \leq n (k, n \in \mathbb{N}), (x+y)^n = \sum_{k=0}^n C_n^k x^k y^{n-k},$$

où $C_n^k = \frac{n!}{k!(n-k)!}$ (c'est nombre de parties à k éléments dans un ensemble à n éléments) ($n! = 1 \times 2 \times 3 \times \dots \times n$).

Somme géométrique :

$$\forall p \in \mathbb{R}, \forall n \in \mathbb{N}, \sum_{k=0}^n p^k = \frac{1-p^{n+1}}{1-p},$$

$$\text{si, de plus, } |p| < 1, \sum_{k=0}^{+\infty} p^k = \frac{1}{1-p}.$$

Somme arithmétique

$$1 + 2 + \dots + n = \frac{n(n+1)}{2}.$$

Somme de carrés

$$1^2 + 2^2 + \dots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

Exponentielle :

$$\forall x \in \mathbb{R}, e^x = \sum_{n=0}^{+\infty} \frac{x^n}{n!}.$$

Cosinus hyperbolique :

$$\forall x \in \mathbb{R}, \cosh(x) = \frac{e^x + e^{-x}}{2} = \sum_{n=0}^{+\infty} \frac{x^{2n}}{(2n)!}.$$

Sinus hyperbolique :

$$\forall x \in \mathbb{R}, \sinh(x) = \frac{e^x - e^{-x}}{2} = \sum_{n=0}^{+\infty} \frac{x^{2n+1}}{(2n+1)!}.$$

Bibliographie

- [DB01] Pierre Del Moral and Nathalie Bartoli. *Simulation et algorithmes stochastiques : une introduction avec applications*. Cepadue, Toulouse, 2001.
- [DRR06] Pierre Del Moral, Bruno Rémillard, and Sylvain Rubenthaler. *Une introduction aux probabilités*. Ellipses, Paris, 2006.
- [Dur96] Richard Durrett. *Probability : theory and examples*. Duxbury Press, Belmont, CA, second edition, 1996.
- [JP03] Jean Jacod and Philip Protter. *L'essentiel en théorie des probabilités*. Cassini, Paris, 2003.
- [Par08] Etienne Pardoux. *Markov processes and applications*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester; Dunod, Paris, french edition, 2008. Algorithms, networks, genome and finance.
- [Wil91] David Williams. *Probability with martingales*. Cambridge Mathematical Text books. Cambridge University Press, Cambridge, 1991.

Liste des symboles

$:=$	Définition
$\#$	Cardinal
\sqcup	Partition
\mathbb{E}	Espérance
$\lfloor \dots \rfloor$	Partie entière inférieure
\log	Logarithme népérien
$\mathcal{B}(p)$	Loi de Bernoulli de paramètre p
$\mathcal{C}_b^+(\dots)$	Fonctions continues bornées positives
$\mathcal{E}(\lambda)$	Loi exponentielle de paramètre λ
\mathcal{N}	Loi normale
\mathcal{U}	Loi uniforme
∇	Gradient
\mathbb{P}	Probabilité
$\perp\!\!\!\perp$	Indépendant
\propto	Proportionnel à
\sim	Suivre la loi, de loi, ...
\mathbb{V}	Variance
C	Nombre de parties (...)

Index

- Aiguilles de Buffon, 4, 6
- Appels successifs, 1
- Apériodique, 29

- Box-Müller, 12

- convergence en loi, 2
- Cosinus hyperbolique, 43
- Couplage, 39

- Espace de configurations, 33
- Estimateur de la variance, 3
- Exponentielle, 43

- Fonction d'importance, 21
- Formule du binôme, 43

- Générateur de nombres aléatoires, 9

- Histogramme, 16

- i.i.d., 1
- Intervalle de confiance, 2
- Invariante, 29
- Irréductible, 30

- Logarithme népérien, 12
- Loi a posteriori, 35
- Loi cible, 30
- Loi conditionnelle, 15
- loi exponentielle, 11
- Loi forte des grands nombres, 2

- Marche simple dans \mathbb{Z} ., 30
- MCMC, 29
- Monte-Carlo, 1
- Méthode de rejet, 14, 28

- Nombre de parties (...), 43
- Noyau de proposition, 30
- Noyau symétrique, 32

- p.s., 2
- Partition., 25
- Poisson, 13
- Proposition, 30
- Pseudo-inverse, 9

- Sinus hyperbolique, 43
- Somme arithmétique, 43
- Somme de carrés, 43
- Somme géométrique, 43
- statistique d'ordre, 20

- Symétrique, 29

- Temps de calcul, 16
- Théorème de la limite centrale, 2
- Théorème ergodique, 29

- v. a. r., 9
- VaR (« value at risk »), 19
- Variance, 2
- Variance de la méthode, 3

- Échantillonneur de Gibbs, 33