



Département GMM - 4ème année

# Méthodes de Monte Carlo

**Aldéric Joulin**

A. Joulin  
Bureau 115 - GMM  
ajoulin@insa-toulouse.fr

**Année universitaire 2019-2020**



# Table des matières

<b>1</b>	<b>Introduction aux méthodes de Monte Carlo</b>	<b>5</b>
<b>2</b>	<b>Simulation de variables aléatoires</b>	<b>9</b>
2.1	Génération de nombres pseudo-aléatoires . . . . .	9
2.2	Simulation de variables aléatoires discrètes . . . . .	11
2.3	Méthode par inversion . . . . .	11
2.4	Simulation de quelques lois particulières . . . . .	13
2.4.1	Loi binomiale . . . . .	13
2.4.2	Loi géométrique . . . . .	14
2.4.3	Loi de Poisson . . . . .	14
2.5	Simulation de lois gaussiennes . . . . .	16
2.6	Le cas des vecteurs aléatoires . . . . .	17
2.7	Méthode du rejet . . . . .	19
<b>3</b>	<b>Techniques de réduction de variance</b>	<b>23</b>
3.0.1	Variable de contrôle . . . . .	24
3.0.2	Echantillonnage préférentiel . . . . .	25
3.0.3	Variables antithétiques . . . . .	27
3.0.4	Méthode de stratification . . . . .	29
3.0.5	Conditionnement . . . . .	31
	<b>Bibliographie</b>	<b>33</b>



# Chapitre 1

## Introduction aux méthodes de Monte Carlo

De manière générale, la simulation numérique permet d'étudier un système donné dont on connaît les interactions complexes, de mesurer les effets de certaines variations des paramètres sur le comportement global du système, voire d'expérimenter de nouvelles situations. Lorsque dans la simulation intervient un élément aléatoire, on parle alors de simulation aléatoire (ou stochastique). Parmi les nombreux exemples d'applications aussi diverses que variées, citons la simulation des flux de clients dans des réseaux de files d'attente, des portefeuilles d'actifs en mathématiques financières, la comparaison d'estimateurs en statistique ou encore la recherche d'équilibres en physique ou en économie. De surcroît, si l'on cherche une représentation fidèle des phénomènes observés, les difficultés inhérentes aux calculs souvent peu explicites deviennent insurmontables : les techniques de simulation permettent alors de les approcher numériquement.

Les méthodes dites de Monte Carlo, dont le principe repose essentiellement sur la Loi des Grands Nombres, c'est-à-dire sur une répétition d'expériences aléatoires, permettent d'évaluer une quantité donnée mais inconnue, voire de résoudre un système déterministe. Ces méthodes peuvent servir pour :

- le calcul d'intégrales (ce que nous ferons principalement dans ce cours).
- la résolution d'équations aux dérivées partielles.
- la résolution de systèmes linéaires.
- la résolution de problèmes d'optimisation (algorithme du recuit simulé, voir le cours d'Optimisation stochastique).

Bien que le système considéré soit totalement déterministe, ce type de modélisation aléatoire n'a réellement de sens que pour des situations dans lesquelles les méthodes déterministes classiques sont moins pertinentes, et il s'avère que de tels cas se multiplient dans les sciences fondamentales et en ingénierie.

Pour illustrer notre propos, considérons le problème suivant. Il s'agit d'approcher numériquement l'intégrale

$$I = \int_0^1 g(x)dx,$$

où  $g : \mathbb{R} \rightarrow \mathbb{R}$  est une fonction continue par morceaux sur  $[0, 1]$ . Diverses méthodes classiques de type déterministe existent : rectangles, trapèzes, Simpson, l'intégrale  $I$  étant approchée par une somme :

$$\sum_{i=0}^n w_i g(x_i), \quad \text{avec} \quad \sum_{i=0}^n w_i = 1,$$

et la subdivision  $(x_i)_{0 \leq i \leq n}$  bien choisie dans  $[0, 1]$ . La méthode de Monte Carlo, quant à elle, consiste à représenter cette intégrale sous la forme d'une espérance :

$$I = \mathbb{E}[g(U)],$$

où  $U$  est une variable aléatoire réelle (en abrégé v.a.r.) suivant la loi  $\mathcal{U}_{[0,1]}$  uniforme sur l'intervalle  $[0, 1]$ , et ensuite à utiliser la Loi (forte) des Grands Nombres (LGN) : si  $(U_n)_{n \in \mathbb{N}^*}$  est une suite de v.a.r. i.i.d. de loi uniforme  $\mathcal{U}_{[0,1]}$ , alors on a la convergence presque sûre (p.s.) suivante de la moyenne empirique :

$$\frac{1}{n} \sum_{i=1}^n g(U_i) \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}[g(U)].$$

En d'autres termes, si  $u_1, u_2, \dots, u_n$  sont des nombres tirés au hasard dans l'intervalle  $[0, 1]$ , alors  $n^{-1} \sum_{i=1}^n g(u_i)$  est une approximation de l'intégrale  $I$ .

À présent, la question que l'on se pose est la suivante : quelle est la vitesse de convergence dans la méthode de Monte Carlo ? C'est le Théorème Central Limite (TCL) qui nous la donne : si  $\sigma^2 = \text{Var}(g(U_1))$  désigne la variance (supposée finie) de la v.a.r.  $g(U_1)$ , alors la v.a.r.

$$\frac{\sqrt{n}}{\sigma} \left( \frac{1}{n} \sum_{i=1}^n g(U_i) - \mathbb{E}[g(U)] \right),$$

converge en loi vers une v.a.r. gaussienne centrée et réduite. Ainsi, on peut construire un intervalle de confiance asymptotique permettant d'estimer la moyenne  $\mathbb{E}[g(U)]$ , qui n'est autre que l'intégrale  $I$ . En particulier, la marge d'erreur est d'autant plus petite que l'écart-type  $\sigma$  l'est. Étant donné qu'il n'est pas toujours calculable, on l'approchera lui aussi (ou plutôt la variance  $\sigma^2$ ) par la méthode de Monte Carlo puisque

$$\frac{1}{n} \sum_{i=1}^n g(U_i)^2 - \left( \frac{1}{n} \sum_{i=1}^n g(U_i) \right)^2 \xrightarrow[n \rightarrow +\infty]{p.s.} \sigma^2,$$

ce calcul étant mené en même temps que celui de l'espérance (il s'agit ici de la variance empirique biaisée, d'espérance  $(n-1)\sigma^2/n$ , mais ceci n'a que peu d'importance). En effet, grâce au lemme de Slutsky, la convergence en loi précédente vers une v.a.r. gaussienne est préservée lorsque l'on remplace  $\sigma^2$  par son estimateur.

Au final, on en déduit donc que la vitesse de convergence est de l'ordre de  $1/\sqrt{n}$ . Cette vitesse peut paraître faible *a priori* mais, contrairement aux méthodes d'intégration déterministes classiques, elle présente l'avantage :

- d'être insensible à la dimension.
- de ne pas dépendre de la régularité de la fonction  $g$  à intégrer, pourvu que  $g(U_1)$  soit de carré intégrable.

En effet, qu'obtient-on dans le cas où l'on utilise une méthode d'intégration déterministe ? En dimension supérieure, pour une fonction  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  donnée (disons continue), il s'agit d'approcher l'intégrale

$$\int_{[0,1]^d} g(x_1, \dots, x_d) dx_1 \dots dx_d,$$

et comme mentionné précédemment, ceci peut se faire en considérant les sommes de Riemann suivantes :

$$\frac{1}{n^d} \sum_{1 \leq i_1, \dots, i_d \leq n} g\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right).$$

Supposons la fonction  $g$  ayant une certaine propriété de régularité, à savoir qu'elle est lipschitzienne sur  $\mathbb{R}^d$ , c'est-à-dire qu'il existe une constant  $C > 0$  telle que pour tous  $x, y \in \mathbb{R}^d$ ,

$$|g(x) - g(y)| \leq C \|x - y\|,$$

la norme  $\|\cdot\|$  désignant la norme euclidienne. On a alors

$$\begin{aligned} & \left| \int_{[0,1]^d} g(x_1, \dots, x_d) dx_1 \dots dx_d - \frac{1}{n^d} \sum_{1 \leq i_1, \dots, i_d \leq n} g\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right) \right| \\ &= \left| \sum_{1 \leq i_1, \dots, i_d \leq n} \int_{x_1 \in [\frac{i_1-1}{n}, \frac{i_1}{n}]} \dots \int_{x_d \in [\frac{i_d-1}{n}, \frac{i_d}{n}]} \left( g(x_1, \dots, x_d) - g\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right) \right) dx_1 \dots dx_d \right| \\ &\leq \sum_{1 \leq i_1, \dots, i_d \leq n} \int_{x_1 \in [\frac{i_1-1}{n}, \frac{i_1}{n}]} \dots \int_{x_d \in [\frac{i_d-1}{n}, \frac{i_d}{n}]} \left| g(x_1, \dots, x_d) - g\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right) \right| dx_1 \dots dx_d \\ &\leq \sum_{1 \leq i_1, \dots, i_d \leq n} \int_{x_1 \in [\frac{i_1-1}{n}, \frac{i_1}{n}]} \dots \int_{x_d \in [\frac{i_d-1}{n}, \frac{i_d}{n}]} C \sqrt{\sum_{j=1}^d \left| x_j - \frac{i_j}{n} \right|^2} dx_1 \dots dx_d \end{aligned}$$

$$\begin{aligned}
&\leq C \sum_{1 \leq i_1, \dots, i_d \leq n} \int_{x_1 \in [\frac{i_1-1}{n}, \frac{i_1}{n}]} \cdots \int_{x_d \in [\frac{i_d-1}{n}, \frac{i_d}{n}]} \frac{\sqrt{d}}{n} dx_1 \dots dx_d \\
&= \frac{C\sqrt{d}}{n}.
\end{aligned}$$

Ainsi, la précision de cette approximation est d'ordre  $1/n$ . Pour calculer la somme multiple ci-dessus, nous exécutons  $d$  boucles emboîtées, chacune de longueur  $n$ . Nous sommes donc  $n^d$  termes : on dit alors que le coût de calcul  $N$  est d'ordre  $n^d$  et la précision d'ordre  $N^{-1/d}$ . Concernant la méthode de Monte Carlo et sa généralisation à la dimension supérieure, où l'on approche alors l'intégrale par la somme

$$\frac{1}{n} \sum_{i=1}^n g(U_i^1, \dots, U_i^d),$$

les  $(U_n^1, \dots, U_n^d)_{n \in \mathbb{N}^*}$  formant une suite de vecteurs i.i.d. de taille  $d$  et suivant la loi  $\mathcal{U}_{[0,1]^d}$  uniforme sur le pavé  $[0, 1]^d$  (autrement dit, toutes les coordonnées de chacun de ces vecteurs sont i.i.d. de loi  $\mathcal{U}_{[0,1]}$ ), la LGN et le TCL s'appliquent de la même manière que dans le cas uni-dimensionnel et l'on en déduit que le coût de calcul  $N$  est d'ordre  $n$  et la précision d'ordre  $N^{-1/2}$ , qui est donc meilleure que celle obtenue par l'approche déterministe dès que la dimension est plus grande que 3 (et moins contraignante en terme de régularité de la fonction  $g$  considérée).

Pour conclure cette introduction aux méthodes de Monte Carlo, les questions que l'on se pose à présent sont les suivantes :

- comment simuler des v.a.r. i.i.d. de loi uniforme, ou plus généralement de loi donnée ?

- comment peut-on réduire la variance dans la méthode de Monte Carlo pour accélérer la convergence ?

Ces questions font respectivement l'objet des chapitres à venir.



# Chapitre 2

## Simulation de variables aléatoires

L'objectif de ce chapitre est de présenter les principales méthodes permettant de simuler une v.a. (ou, par abus de langage, une loi de probabilité) sur un ordinateur. On commence par rappeler comment “créer de l'aléa raisonnable” sur un ordinateur, ou plus précisément comment (pseudo-)simuler la loi uniforme sur  $[0, 1]$  puis dans un second temps, comment, à partir de ce point de départ, construire des méthodes pour simuler d'autres lois. Il existe deux méthodes génériques pour la simulation de v.a. : la méthode par inversion, qui est en général la première à tester mais nécessite de connaître explicitement l'inverse de la fonction de répartition, et celle par rejet, cette dernière étant plus élaborée mais permettant de traiter plus de cas grâce à des hypothèses moins restrictives. Par ailleurs, d'autres méthodes peuvent être appliquées lorsque la v.a. à simuler suit une loi spécifique.

### 2.1 Génération de nombres pseudo-aléatoires

Comment générer de l'aléa raisonnable sur un ordinateur ? D'abord, par aléa raisonnable, on entend l'idée de savoir simuler une loi (*i.e.* tirer un nombre suivant cette loi à l'aide d'un ordinateur) suffisamment riche pour qu'elle soit ensuite utilisable comme point de départ pour en simuler d'autres. Par exemple, simuler une variable de Bernoulli consiste bien à “créer” du hasard mais celui-ci n'est pas raisonnable, le hasard n'étant pas assez riche (seulement deux possibilités).

La variable suffisamment riche la plus simple est la loi  $\mathcal{U}_{[0,1]}$  uniforme sur l'intervalle  $[0, 1]$ . On a par exemple la propriété suivante.

**Proposition 2.1.1.** *Soit  $X$  un vecteur aléatoire à valeurs dans  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  (ou même dans un espace plus général) de loi notée  $\mathbb{P}_X$ . Alors, il existe une application mesurable*

$\varphi : ([0, 1], \mathcal{B}([0, 1])) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  telle que

$$\mathbb{P}_X = \mathcal{U}_{[0,1]} \circ \varphi^{-1},$$

i.e. pour tout  $A \in \mathcal{B}(\mathbb{R}^d)$ ,

$$\mathbb{P}_X(A) = \mathbb{P}(\varphi(U) \in A) \quad \text{où} \quad U \sim \mathcal{U}_{[0,1]}.$$

Ce théorème, dont nous ne ferons pas la démonstration dans un cadre général (seulement dans le cas réel), signifie qu'une loi uniforme contient suffisamment d'aléa pour simuler n'importe quelle loi. Notons cependant qu'il s'agit évidemment d'un résultat théorique et que déterminer  $\varphi$  en pratique est un problème compliqué.

On prend donc la loi uniforme comme point de départ et l'on se donne comme objectif de simuler une suite de réalisations i.i.d de cette loi.

**Définition 2.1.1.** Soit  $(x_n)_{n \in \mathbb{N}^*}$  une suite de nombres à valeurs dans  $[0, 1]$ . Alors, on dit que  $(x_n)_{n \in \mathbb{N}^*}$  est une suite de nombres aléatoires s'ils sont des réalisations d'une suite de v.a.r. i.i.d de loi  $\mathcal{U}_{[0,1]}$ .

Dans la réalité, on n'a pas la possibilité de générer une telle suite. Néanmoins en pratique, les ordinateurs utilisent une approximation d'une telle suite de nombres aléatoires, qu'on appelle suite de nombres pseudo-aléatoires. L'idée est la suivante : on construit une suite déterministe mais relativement chaotique en cherchant à minimiser le coût de calcul de cette suite. Le procédé standard consiste à considérer un nombre  $N$  "grand" et à définir la suite  $(x_n)_{n \in \mathbb{N}}$  par

$$x_n = \frac{y_n}{N}, \quad n \in \mathbb{N},$$

où  $(y_n)_{n \in \mathbb{N}}$  est définie récursivement par la relation de congruence suivante :

$$y_{n+1} = (ay_n + b) \quad \text{modulo } N, \quad n \in \mathbb{N},$$

avec  $a$  et  $b$  des entiers bien choisis compris entre 0 et  $N$ . La suite  $(y_n)_{n \in \mathbb{N}}$  est définie de manière déterministe sauf la valeur initiale  $y_0$  que l'on appelle la graine aléatoire, qui est souvent construite à partir de l'heure à laquelle est lancé le programme. Nous ne détaillerons pas plus ici la construction de la graine aléatoire. On gardera simplement en mémoire l'idée suivante : on peut créer de l'aléa en lançant le programme.

Dans ce qui suit, nous nous intéressons à plusieurs méthodes générales permettant de simuler d'autres lois à partir de la loi uniforme.

## 2.2 Simulation de variables aléatoires discrètes

Pour commencer, considérons le cas relativement simple des v.a. discrètes. On a la proposition suivante.

**Proposition 2.2.1.** *Soit  $X$  une v.a. à valeurs dans un ensemble  $\mathcal{X} = \{x_1, \dots, x_N\}$  de cardinal  $N \in \mathbb{N}^* \cup \{+\infty\}$ , dont la loi est donnée par les quantités  $\mathbb{P}(X = x_i) = p_i$ , pour tout  $i \in \{1, \dots, N\}$ . Alors, on a l'égalité en loi :*

$$X = \sum_{i=1}^N x_i 1_{[P_{i-1}, P_i[}(U),$$

où  $U \sim \mathcal{U}_{[0,1]}$  et pour tout  $i \in \{1, \dots, N\}$ ,  $P_i = \sum_{k=1}^i p_k$ , avec  $P_0 = 0$ .

*Démonstration.* La preuve est triviale d'après les égalités suivantes : pour tout  $i \in \{1, \dots, N\}$ ,

$$\mathbb{P}(X = x_i) = p_i = P_i - P_{i-1} = \mathbb{P}(P_{i-1} \leq U < P_i).$$

□

**Exemple 2.2.1.** *Dans le cas d'une v.a. de Bernoulli de paramètre  $p \in ]0, 1[$ , la représentation précédente fournit un algorithme simple pour la simuler. En effet, on simule une loi uniforme  $\mathcal{U}_{[0,1]}$  puis l'on pose*

$$X = \begin{cases} 1 & \text{si } U < p; \\ 0 & \text{si } U \geq p. \end{cases}$$

La simulation de v.a. discrètes peut être vue comme un cas particulier de la méthode par inversion que l'on va présenter ci-dessous.

## 2.3 Méthode par inversion

On suppose dans cette partie que l'on cherche à simuler une v.a.r.  $X$  de fonction de répartition  $F$  définie par

$$F(z) = \mathbb{P}(X \leq z), \quad z \in \mathbb{R}.$$

On rappelle que  $F$  est croissante, continue à droite et admettant des limites finies à gauche. On définit alors la fonction quantile de  $F$  (ou encore fonction pseudo-inverse) de la manière suivante.

**Définition 2.3.1.** La fonction quantile de  $F$  est définie par

$$F^{-1}(t) = \inf\{z \in \mathbb{R} : F(z) \geq t\}, \quad t \in ]0, 1[.$$

Notons que  $F^{-1}$  est toujours croissante car  $F$  l'est. Le cas le plus agréable est celui où  $F$  est continue et strictement croissante sur  $\mathbb{R}$ , auquel cas elle est bijective et alors la fonction quantile  $F^{-1}$  coïncide avec la fonction réciproque de  $F$ . Plus généralement, s'il existe  $a, b \in \bar{\mathbb{R}}$  tels que  $a < b$  et  $F$  réalise une bijection de  $]a, b[$  sur  $]0, 1[$ , alors  $F^{-1}$  coïncide avec la fonction réciproque de  $F$  restreinte à l'intervalle  $]a, b[$ . Ceci est notamment pratique quand la v.a.r. n'est pas à support dans  $\mathbb{R}$  tout entier, comme par exemple pour la loi exponentielle.

**Proposition 2.3.1.** Pour tout  $u \in ]0, 1[$  et tout  $x \in \mathbb{R}$ , on a

- $F(F^{-1}(u)) \geq u$  avec égalité si et seulement si  $F$  est continue au point  $F^{-1}(u)$ .
- $F^{-1}(F(x)) \leq x$  avec égalité si et seulement si  $x$  n'est pas dans un intervalle de constance de  $F$ .
- $F^{-1}(u) \leq x \Leftrightarrow u \leq F(x)$ .

*Démonstration.* Démontrons seulement le troisième point, les deux premiers en étant des conséquences directes. Définissons les ensembles

$$A = \{u \in ]0, 1[: F^{-1}(u) \leq x\}, \quad B = \{u \in ]0, 1[: F(x) \geq u\},$$

et montrons l'égalité par double inclusion.

- Supposons  $u \in A$ , i.e.  $F^{-1}(u) \leq x$ . Si  $F(x) < u$ , alors comme  $F$  est continue à droite, il existe un intervalle  $[x, x + h]$  tel que pour tout  $z \in [x, x + h]$ ,  $F(z) < u$  et dans ce cas, comme  $F$  est croissante, on en déduit que  $F^{-1}(u) \geq x + h$ . Ceci fournit une contradiction et nous permet de déduire que  $F(x) \geq u$  puis que  $A \subset B$ .

- À présent soit  $u \in B : F(x) \geq u$ . Par définition de  $F^{-1}$ , on a clairement que  $F^{-1}(u) \leq x$ . On en déduit que  $B \subset A$ . □

La méthode par inversion, qui est en général la première méthode à tester parmi les méthodes de simulation, est alors un corollaire de cette proposition.

**Corollaire 2.3.1.** Soit  $X$  une v.a.r. de fonction de répartition  $F$ . Si  $U \sim \mathcal{U}_{[0,1]}$  alors  $F^{-1}(U)$  a même loi que  $X$ .

*Démonstration.* Par la proposition précédente, on a pour tout  $x \in \mathbb{R}$ ,

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

Ainsi,  $F^{-1}(U)$  a pour fonction de répartition  $F$ , d'où la conclusion désirée. □

**Exemple 2.3.1.** Pour une loi exponentielle de paramètre  $\lambda > 0$ , notée  $\mathcal{E}(\lambda)$ , on a  $F(t) = (1 - e^{-\lambda t})1_{t>0}$ . Alors, pour tout  $u \in ]0, 1[$ ,

$$F^{-1}(u) = -\frac{1}{\lambda} \log(1 - u),$$

de sorte que si  $U \sim \mathcal{U}_{[0,1]}$ , alors  $X = -\lambda^{-1} \log(1 - U)$  suit la loi exponentielle de paramètre  $\lambda > 0$ . Comme  $U$  a même loi que  $1 - U$ , il en est de même pour  $\tilde{X} = -\lambda^{-1} \log(U)$ .

**Exemple 2.3.2.** Soit  $X$  une v.a. à valeurs dans un ensemble  $\mathcal{X} = \{x_1, \dots, x_N\}$  de cardinal  $N \in \mathbb{N}^* \cup \{+\infty\}$ , dont la loi est donnée par les quantités  $\mathbb{P}(X = x_i) = p_i$ , pour tout  $i \in \{1, \dots, N\}$ . Alors, on a

$$F(t) = \begin{cases} 0 & \text{si } t < x_1; \\ \sum_{k=1}^i p_k & \text{si } x_i \leq t < x_{i+1}; \\ 1 & \text{si } t \geq x_N. \end{cases}$$

et par conséquent, la fonction quantile vaut :

$$F^{-1}(u) = \begin{cases} x_1 & \text{si } 0 < u < p_1; \\ x_i & \text{si } \sum_{k=1}^{i-1} p_k \leq u < \sum_{k=1}^i p_k. \end{cases}$$

Ainsi, la méthode par inversion dans le cas des v.a. discrètes est bien celle énoncée précédemment.

En conclusion, on peut insister sur le fait que la méthode par inversion est a priori facile à mettre en oeuvre et très peu coûteuse lorsque  $F^{-1}$  est explicite. Malheureusement, cette propriété nécessite en général une expression explicite de  $F$ , ce qui n'est bien sûr pas toujours le cas (penser à la loi gaussienne). On doit alors recourir à d'autres méthodes.

## 2.4 Simulation de quelques lois particulières

Avant d'introduire dans la suite la seconde grande méthode de simulation, la méthode du rejet, regardons à présent comment simuler certaines lois particulières.

### 2.4.1 Loi binomiale

La loi binomiale de paramètres  $n \in \mathbb{N}^*$  et  $p \in ]0, 1[$  est donnée par

$$\mathbb{P}(X = k) = C_n^k p^k (1 - p)^{n-k}, \quad k \in \{0, \dots, n\}.$$

La fonction de répartition n'est pas explicite. Néanmoins, la v.a.  $X$  pouvant s'écrire comme une somme de  $n$  v.a. de Bernoulli indépendantes et de même paramètre  $p$ , une méthode consiste à tirer  $n$  nombres au hasard dans  $[0, 1]$  et à poser  $X = k$  si exactement  $k$  nombres sont dans l'intervalle  $[0, p]$ .

### 2.4.2 Loi géométrique

On considère la loi géométrique sur  $\mathbb{N}^*$  de paramètre  $p = 1 - q \in ]0, 1[$  :

$$\mathbb{P}(X = k) = pq^{k-1}, \quad k \in \mathbb{N}^*.$$

Cette loi peut être simulée en utilisant la méthode par inversion. Plus précisément, vu que la fonction de répartition est donnée par

$$F(t) = \begin{cases} 0 & \text{si } t < 1; \\ 1 - q^{\lfloor t \rfloor} & \text{sinon;} \end{cases}$$

où  $\lfloor t \rfloor$  désigne la partie entière de  $t$ , on a alors l'égalité en loi :

$$X = \sum_{k=1}^{\infty} k 1_{[1-q^k, 1-q^{k+1}[}(U).$$

Ainsi, on pose  $X = k$  si le nombre simulé au hasard dans  $[0, 1]$  est dans l'intervalle  $[1 - q^k, 1 - q^{k+1}[$ .

Par ailleurs, on peut la simuler d'une autre manière. La v.a.  $X$  s'écrivant comme le premier instant pour lequel on obtient pile, on simule des nombres au hasard dans  $[0, 1]$  et l'on pose  $X = k$  si le  $k$ -ème nombre tiré est le premier à tomber dans l'intervalle  $[0, p]$ . Cependant, cette méthode est moins robuste que la première au sens où elle dépend fortement du paramètre  $p$ , son coût étant justement l'espérance de  $X$  qui vaut  $1/p$  (en d'autres termes, si  $p$  est très petit, on aura tendance à simuler beaucoup de nombres au hasard).

### 2.4.3 Loi de Poisson

La loi de Poisson de paramètre  $\lambda > 0$  est définie par

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}.$$

La fonction de répartition n'étant pas explicite, on va simuler la v.a.  $X$  en se basant sur le résultat suivant.

**Proposition 2.4.1.** *Si  $(\tau_n)_{n \in \mathbb{N}^*}$  est une suite de v.a.r. i.i.d. de loi  $\mathcal{E}(\lambda)$ , alors la v.a.*

$$X = \sum_{n=1}^{\infty} n 1_{\{\tau_1 + \dots + \tau_n < 1 \leq \tau_1 + \dots + \tau_n + \tau_{n+1}\}},$$

*suit une loi de Poisson de paramètre  $\lambda$ .*

*Démonstration.* La preuve de ce résultat est quelque peu calculatoire. Tout d'abord, on a  $\mathbb{P}(X = 0) = \mathbb{P}(\tau_1 > 1) = e^{-\lambda}$ . À présent, on a pour tout  $n \in \mathbb{N}^*$ ,

$$\begin{aligned} \mathbb{P}(X = n) &= \mathbb{P}(\tau_1 + \dots + \tau_n < 1 \leq \tau_1 + \dots + \tau_n + \tau_{n+1}) \\ &= \int_{t_1=0}^1 \int_{t_2=0}^{1-t_1} \dots \int_{t_n=0}^{1-(t_1+\dots+t_{n-1})} \lambda^n e^{-\lambda(t_1+\dots+t_n)} \left( \int_{t_{n+1} \geq 1-(t_1+\dots+t_n)}^{+\infty} \lambda e^{-\lambda t_{n+1}} dt_{n+1} \right) dt_n \dots dt_1 \\ &= \lambda^n e^{-\lambda} \int_{t_1=0}^1 \int_{t_2=0}^{1-t_1} \dots \int_{t_n=0}^{1-(t_1+\dots+t_{n-1})} dt_n \dots dt_1, \end{aligned}$$

et l'on montre par récurrence sur  $n \in \mathbb{N}^*$  que l'intégrale multiple ci-dessus vaut exactement  $1/n!$ . Plus généralement, montrons que pour tout  $x > 0$ , on a

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

En effet si  $n = 1$ , on a  $\int_0^x dt_1 = \frac{x^1}{1!}$ , puis pour  $n = 2$ ,

$$\int_{t_1=0}^x \int_{t_2=0}^{x-t_1} dt_2 dt_1 = \int_{t_1=0}^x (x - t_1) dt_1 = \frac{x^2}{2!}.$$

Supposons alors la propriété vraie au rang  $n$  et démontrons-la au rang  $n + 1$  :

$$\begin{aligned} &\int_{t_1=0}^x \int_{t_2=0}^{x-t_1} \dots \int_{t_n=0}^{x-(t_1+\dots+t_{n-1})} \int_{t_{n+1}=0}^{x-(t_1+\dots+t_n)} dt_{n+1} \dots dt_1 \\ &= \int_{t_1=0}^x \left( \int_{t_2=0}^{x-t_1} \dots \int_{t_n=0}^{x-(t_1+\dots+t_{n-1})} \int_{t_{n+1}=0}^{x-(t_1+\dots+t_n)} dt_{n+1} \dots dt_2 \right) dt_1 \\ &= \int_{t_1=0}^x \frac{(x - t_1)^n}{n!} dt_1 \\ &= \frac{x^{n+1}}{(n+1)!}, \end{aligned}$$

ce qui achève la démonstration.  $\square$

Sachant que les  $\tau_n$  ont même loi que les  $-\ln(U_n)/\lambda$ , avec les  $U_n$  des v.a.r. i.i.d. de loi  $\mathcal{U}_{[0,1]}$ , on obtient l'égalité en loi :

$$X = \sum_{n=1}^{\infty} n 1_{\{U_1 U_2 \dots U_{n+1} < e^{-\lambda} \leq U_1 U_2 \dots U_n\}}.$$

Ainsi, pour simuler la loi de Poisson de paramètre  $\lambda$ , on simule une suite de nombres au hasard dans  $[0, 1]$  et l'on pose  $X = k$  si  $k$  est le dernier instant pour lequel  $u_1 u_2 \dots u_k \geq e^{-\lambda}$ .

## 2.5 Simulation de lois gaussiennes

Nous consacrons cette courte partie à la simulation des lois gaussiennes multidimensionnelles. Encore une fois, la fonction de répartition n'étant pas explicite, nous ne pouvons pas utiliser la méthode par inversion. Dans un premier temps, pour simuler une v.a.r. gaussienne centrée et réduite (le cas général est ensuite immédiat par translation/dilatation respectivement des moyenne et écart-type), on simule un vecteur gaussien centré et de matrice de covariance l'identité en dimension 2 et l'on obtient le cas unidimensionnel par projection, i.e. on ne garde qu'une seule coordonnée. Rappelons que la densité sur  $\mathbb{R}^2$  de la loi  $\mathcal{N}(0, I_2)$  est donnée par

$$g(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right).$$

Établissons tout d'abord le résultat général suivant.

**Proposition 2.5.1.** *Soit  $R \sim \mathcal{E}(1/2)$  et  $\Theta \sim \mathcal{U}_{[0, 2\pi]}$  deux v.a.r. indépendantes. Alors le vecteur bidimensionnel  $(\sqrt{R} \cos(\Theta), \sqrt{R} \sin(\Theta))$  suit la loi  $\mathcal{N}(0, I_2)$ .*

*Démonstration.* Soit  $(X_1, X_2) \sim \mathcal{N}(0, I_2)$ . On a alors pour toute fonction  $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$  mesurable positive ou bornée,

$$\mathbb{E}[\varphi(X_1, X_2)] = \int_{\mathbb{R}^2} \varphi(x_1, x_2) g(x_1, x_2) dx_1 dx_2.$$

En passant en coordonnées polaires, on obtient :

$$\mathbb{E}[\varphi(X_1, X_2)] = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{+\infty} \varphi(r \cos(\theta), r \sin(\theta)) r \exp\left(-\frac{r^2}{2}\right) dr d\theta,$$

et en posant  $\rho = r^2$ , on a

$$\mathbb{E}[\varphi(X_1, X_2)] = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{+\infty} \varphi(\sqrt{\rho} \cos(\theta), \sqrt{\rho} \sin(\theta)) \exp\left(-\frac{\rho}{2}\right) d\rho d\theta.$$

On en déduit l'égalité désirée, à savoir :

$$\mathbb{E}[\varphi(X_1, X_2)] = \mathbb{E}\left[\varphi(\sqrt{R} \cos(\Theta), \sqrt{R} \sin(\Theta))\right].$$

□

Comme corollaire immédiat du résultat précédent, on peut simuler la loi  $\mathcal{N}(0, I_2)$  par l'algorithme dit de Box-Muller, qui est directement basé sur le résultat suivant.



**Corollaire 2.5.1.** *Soit  $(U_1, U_2)$  un couple de v.a.r. i.i.d. de loi  $\mathcal{U}_{[0,1]}$ . Alors le vecteur*

$$X = \left( \sqrt{-2 \log(U_1)} \cos(2\pi U_2), \sqrt{-2 \log(U_1)} \sin(2\pi U_2) \right),$$

*suit la loi  $\mathcal{N}(0, I_2)$ . En particulier, les deux marginales suivent la loi  $\mathcal{N}(0, 1)$ .*

Ainsi, dès que l'on peut simuler une loi gaussienne uni-dimensionnelle, on peut simuler la loi  $\mathcal{N}(0, \Sigma)$  d'un vecteur gaussien sur  $\mathbb{R}^d$ , où  $\Sigma$  est une matrice de covariance  $d \times d$  diagonale et inversible. En effet, les coordonnées de ce vecteur étant indépendantes par le "miracle gaussien", on se ramène au cas uni-dimensionnel en simulant successivement les coordonnées de manière indépendante.

Considérons à présent le cas des vecteurs gaussiens centrés multi-dimensionnels dont la matrice de covariance  $\Sigma$  est une matrice symétrique définie positive générale. Les coordonnées ne sont alors plus indépendantes. Soit  $X$  un vecteur de  $\mathbb{R}^d$  de loi  $\mathcal{N}(0, \Sigma)$ . Alors on sait qu'il existe une unique matrice symétrique définie positive  $A$  telle que  $A^2 = \Sigma$ . On la note  $\sqrt{\Sigma}$ . Ainsi, si  $Z$  a pour loi  $\mathcal{N}(0, I_d)$ , le vecteur  $\sqrt{\Sigma}Z$  suit la loi  $\mathcal{N}(0, \Sigma)$ . Cependant, il se peut que le calcul de la racine carrée  $\sqrt{\Sigma}$  soit coûteux, auquel cas on préfère utiliser la factorisation de Cholesky : étant donnée  $\Sigma$  une matrice symétrique définie positive, il existe une matrice carrée réelle triangulaire inférieure  $L$  telle que  $\Sigma = LL^T$  (la factorisation est unique dès que les éléments diagonaux de la matrice  $L$  sont tous positifs). Ainsi, le vecteur  $LZ$  suit lui aussi la loi  $\mathcal{N}(0, \Sigma)$ .

## 2.6 Le cas des vecteurs aléatoires

Comme on vient de le voir pour les vecteurs gaussiens, la notion de dépendance entre les coordonnées d'un vecteur aléatoire a une importance cruciale, en particulier lorsque l'on cherche à le simuler. Dans le cas indépendant, on simule successivement les coordonnées de manière indépendante. En revanche le cas dépendant est bien entendu plus délicat et il faut alors effectuer des conditionnements successifs. Examinons en détail le cas de la dimension 2, le cas de la dimension supérieure étant obtenu par itérations successives du procédé. Soit  $(X, Y)$  un couple de v.a.r. de densité jointe  $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ . On peut écrire que

$$f(x, y) = f_X(x) f_Y^{X=x}(y), \quad x, y \in \mathbb{R},$$

où  $f_X$  est la densité de  $X$  et  $f_Y^{X=x}$  la densité conditionnelle de  $Y$  sachant que  $X = x$ . Ainsi, pour générer le couple  $(X, Y)$ , on simule  $X$  selon la densité  $f_X$  et si  $X = x$  alors on simule  $Y$  indépendamment de  $X$  et selon la loi de densité  $f_Y^{X=x}$ . On peut aussi faire l'inverse et simuler d'abord  $Y$  selon sa densité  $f_Y$  puis si  $Y = y$ , simuler la loi conditionnelle de densité  $f_X^{Y=y}$ . Bien évidemment, cette méthode n'a de sens que si l'on sait simuler facilement les

lois de densité  $f_X$  et  $f_Y^{X=x}$ , ou  $f_Y$  et  $f_X^{Y=y}$ . Enfin, mentionnons que cette méthode est aussi valable dans le cas des vecteurs aléatoires à valeurs discrètes.

De surcroît, on a vu (toujours dans le cas gaussien) que pour simuler une v.a.r., il pouvait être quelquefois plus pertinent de simuler un vecteur aléatoire en dimension plus grande puis de projeter sur une coordonnée, que de simuler directement la v.a.r. elle-même. Ceci est notamment le cas lorsque la densité de la v.a.r.  $X$  à simuler est définie par un mélange de lois discret ou continu, c'est-à-dire par une combinaison convexe du type

$$f_X(x) = \sum_{i=1}^N p_i f_i(x), \quad x \in \mathbb{R},$$

avec  $N \in \mathbb{N}^* \cup \{+\infty\}$ , la somme sur les  $p_i$  positifs ou nuls valant 1 et les  $f_i$  étant des densités sur  $\mathbb{R}$ , ou par une intégrale n'admettant pas d'expression explicite, i.e.

$$f_X(x) = \int_{\mathbb{R}} f(x, y) dy, \quad x \in \mathbb{R},$$

où  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^+$  est une densité. La méthode est la même pour ces 2 cas : on simule comme ci-dessus le couple  $(X, Y)$  puis on ne conserve que la première coordonnée (dans le cas du mélange continu la densité jointe du couple est  $f$  tandis que dans le mélange discret la "densité jointe" est donnée sur  $\mathbb{R} \times \{1, \dots, N\}$  par  $f(x, i) = p_i f_i(x)$ , la v.a.  $Y$  est définie sur  $\{1, \dots, N\}$  et de loi  $\mathbb{P}(Y = i) = p_i$  et chaque  $f_i$  est la densité conditionnelle de  $X$  sachant  $Y = i$ ).

**Exemple 2.6.1.** *Considérons la densité sur  $\mathbb{R}^2$  définie par*

$$f(x, y) = 1_{[0, +\infty[}(x) 1_{[1, +\infty[}(y) n y^{-n} e^{-xy}.$$

*On a alors que*

$$f_X(x) = 1_{[0, +\infty[}(x) \int_1^{+\infty} n y^{-n} e^{-xy}, \quad f_Y(y) = 1_{[1, +\infty[}(y) n y^{-n-1},$$

*mais aussi que pour tout  $x \geq 0$  et tout  $y \geq 1$ ,*

$$f_X^{Y=y}(x) = \frac{f(x, y)}{f_Y(y)} = y e^{-xy},$$

*la densité  $f_Y^{X=x}$ , quant à elle, n'étant pas explicite car  $f_X$  ne l'est pas. Ainsi, si l'on veut simuler la loi jointe du couple  $(X, Y)$  ayant pour densité  $f$ , on simule d'abord la v.a.r.  $Y$ , qui suit une loi de Pareto sur l'intervalle  $[1, +\infty[$  (par la méthode par inversion, il s'agit de simuler une v.a.r.  $U^{-1/n}$ , avec  $U \sim \mathcal{U}_{[0,1]}$ ) puis ensuite on simule indépendamment de  $Y$  la v.a.r.  $X \sim \mathcal{E}(y)$  (i.e. si  $Y = y$  on génère  $X = -\log(V)/y$  où  $V \sim \mathcal{U}_{[0,1]}$  est indépendante de  $U$ ).*

## 2.7 Méthode du rejet

L'idée de la méthode du rejet est de dominer la loi que l'on cherche à simuler par une autre que l'on sait simuler simplement (typiquement, par inversion). Bien que la méthode fonctionne aussi dans le cas discret, concentrons-nous sur le cas des vecteurs aléatoires à densité. On se donne  $f$  et  $g$  deux fonctions mesurables positives (et non identiquement nulles) sur  $\mathbb{R}^d$  et l'on fait les hypothèses suivantes :

- la fonction  $g$  est une densité de probabilité :  $\int_{\mathbb{R}^d} g(x)dx = 1$ .
- Il existe  $c > 0$  tel que pour tout  $x \in \mathbb{R}^d$ , on a  $f(x) \leq cg(x)$ .

L'objectif est alors de simuler la loi de densité (proportionnelle à) la fonction  $f$ . À noter que cette méthode a l'avantage d'être valable dans le cas multi-dimensionnel, même si au final on simule la loi de densité  $g$  en utilisant la méthode des conditionnements successifs. Par ailleurs, mentionnons que l'on n'a pas besoin de connaître l'intégrale de la fonction  $f$  sur  $\mathbb{R}^d$  pour utiliser la méthode du rejet (d'ailleurs, quitte à remplacer  $c$  par  $c$  fois cette intégrale, on supposera dans la suite pour simplifier que  $f$  est elle aussi une densité). En pratique, cela peut avoir un intérêt lorsque l'on ne connaît la densité de la loi à simuler qu'à une constante près et que le calcul de l'intégrale ci-dessus est coûteux.

Lorsque  $g(x) > 0$ ,  $x \in \mathbb{R}^d$ , on pose  $\alpha(x) = f(x)/cg(x)$ , qui est à valeurs dans l'intervalle  $[0, 1]$  et que l'on suppose explicite en pratique. Pour un vecteur aléatoire  $X$  de densité  $g$ , on a évidemment que  $\mathbb{P}(X \in g^{-1}(\{0\})) = 0$ , d'où le fait que  $\alpha(X)$  soit p.s. bien défini. On a alors le résultat suivant, dont la première remarque venant après la démonstration justifie la dénomination de méthode du rejet.

**Théorème 2.7.1.** *Considérons une suite  $(X_n, U_n)_{n \geq 1}$  de vecteurs aléatoires indépendants (avec de surcroît le vecteur  $X_n$  indépendant de la v.a.r.  $U_n$ ) tels que les  $X_n$  suivent la loi de densité  $g$  et les  $U_n$  la loi uniforme sur  $[0, 1]$ . Posons*

$$T = \inf\{k \geq 1 : U_k \leq \alpha(X_k)\},$$

le premier instant  $k$  pour lequel  $U_k \leq \alpha(X_k)$ . Alors la v.a. entière  $T$  suit une loi géométrique sur  $\mathbb{N}^*$  de paramètre

$$p = \mathbb{P}(U_1 \leq \alpha(X_1)),$$

appelé la probabilité d'acceptation. Le vecteur aléatoire  $X_T$  (qui est donc bien défini car  $T$  est p.s. finie) a pour densité la fonction  $f$ .

*Démonstration.* Tout d'abord, notons que la probabilité d'acceptation vaut, par indépendance de  $X_1$  et de  $U_1$ ,

$$p = \mathbb{P}(U_1 \leq \alpha(X_1)) = \int_{\mathbb{R}^d} \left( \int_0^1 1_{\{u \leq \alpha(x)\}} du \right) g(x) dx = \frac{\int_{\mathbb{R}^d} f(x) dx}{c} = \frac{1}{c}.$$

Montrons à présent que  $T$  suit une loi géométrique sur  $\mathbb{N}^*$  de paramètre  $p$ . La suite  $(X_n, U_n)_{n \in \mathbb{N}^*}$  étant formée de couples i.i.d., on a pour tout  $n \in \mathbb{N}^*$ ,

$$\begin{aligned} \mathbb{P}(T > n) &= \mathbb{P}\left(\bigcap_{k=1}^n \{U_k > \alpha(X_k)\}\right) \\ &= \prod_{k=1}^n \mathbb{P}(U_k > \alpha(X_k)) \\ &= \mathbb{P}(U_1 > \alpha(X_1))^n \\ &= (1-p)^n. \end{aligned}$$

On en déduit que la fonction de répartition de la v.a.  $T$  est bien celle d'une loi géométrique, d'où le résultat. En particulier,  $T$  est *p.s.* finie.

Il reste à montrer que le vecteur aléatoire  $X_T$  a bien la fonction  $f$  pour densité. Soit  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$  une fonction mesurable positive ou bornée. Alors, pour tout  $n \in \mathbb{N}^*$ , on a

$$\begin{aligned} \mathbb{E}[\varphi(X_n)1_{\{U_n \leq \alpha(X_n)\}}] &= \int_{\mathbb{R}^d} \left( \int_0^1 1_{\{u \leq \alpha(x)\}} du \right) \varphi(x)g(x)dx \\ &= \int_{\mathbb{R}^d} \alpha(x)\varphi(x)g(x)dx \\ &= \frac{1}{c} \int_{\mathbb{R}^d} \varphi(x)f(x)dx, \end{aligned}$$

d'où l'identité suivante :

$$\mathbb{E}[\varphi(X_n) \mid U_n \leq \alpha(X_n)] = \mathbb{E}[\varphi(Y)],$$

où  $Y$  est un vecteur aléatoire ayant pour densité la fonction  $f$  : on en déduit que la loi conditionnelle de  $X_n$  sachant l'événement  $\{U_n \leq \alpha(X_n)\}$  a pour densité  $f$ , et ce pour tout  $n \in \mathbb{N}^*$ . Enfin, on calcule  $\mathbb{E}[\varphi(X_T)]$  en utilisant cette propriété pour aboutir à la dernière ligne ci-dessous :

$$\begin{aligned} \mathbb{E}[\varphi(X_T)] &= \sum_{n=1}^{+\infty} \mathbb{E}[\varphi(X_n)1_{T=n}] \\ &= \sum_{n=1}^{+\infty} \mathbb{E}\left[\varphi(X_n)1_{\{U_n \leq \alpha(X_n)\}}1_{\bigcap_{k=1}^{n-1} \{U_k > \alpha(X_k)\}}\right] \\ &= \sum_{n=1}^{+\infty} \mathbb{E}\left[\varphi(X_n)1_{\{U_n \leq \alpha(X_n)\}}\right] \mathbb{P}\left(\bigcap_{k=1}^{n-1} \{U_k > \alpha(X_k)\}\right) \end{aligned}$$

$$\begin{aligned}
&= \sum_{n=1}^{+\infty} \mathbb{E}[\varphi(X_n) \mid U_n \leq \alpha(X_n)] p(1-p)^{n-1} \\
&= \sum_{n=1}^{+\infty} \mathbb{E}[\varphi(Y)] p(1-p)^{n-1} \\
&= \mathbb{E}[\varphi(Y)],
\end{aligned}$$

ce qui donne la conclusion désirée. À noter que pour obtenir les troisième et quatrième lignes, nous avons utilisé le fait que les couples de la suite  $(X_n, U_n)_{n \in \mathbb{N}^*}$  étaient i.i.d. La démonstration est enfin achevée.  $\square$

**Remarque 2.7.1.** *Ainsi, pour simuler la loi de densité  $f$ , l'algorithme de rejet est le suivant. Générons tout d'abord un couple  $(X_1, U_1)$  indépendant suivant respectivement la loi de densité  $g$  et la loi uniforme sur  $[0, 1]$ .*

◦ *Si  $U_1 \leq \alpha(X_1)$ , posons  $X = X_1$ .*

◦ *Sinon, rejetons  $X_1$  et recommençons en générant une suite  $(X_n, U_n)_{n \geq 2}$  i.i.d. de même loi que  $(X_1, U_1)$  jusqu'au premier instant  $k$  où  $U_k \leq \alpha(X_k)$ . Si l'on pose  $X = X_k$ , alors la densité du vecteur aléatoire  $X$  ainsi simulé est la fonction  $f$ .*

**Remarque 2.7.2.** *Si l'on ne veut pas trop de rejets lors de la simulation de  $X$ , il faut que la probabilité d'acceptation  $p = \mathbb{P}(U_1 \leq \alpha(X_1))$  soit la plus grande possible, ou encore que la constante  $c$  soit la plus petite possible (comme mentionné précédemment, on a nécessairement que  $c \geq \int_{\mathbb{R}^d} f(x) dx$ , i.e.  $c \geq 1$  dans le cas d'une densité). Par ailleurs, la constante  $c$  est reliée au temps de calcul moyen de la manière suivante : si  $\kappa$  désigne le nombre de calculs par itération de la procédure, le temps de calcul moyen est égal à  $\kappa \mathbb{E}[T]$ , avec*

$$\mathbb{E}[T] = \frac{1}{p} = \frac{c}{\int_{\mathbb{R}^d} f(x) dx}.$$

**Exemple 2.7.1.** *Considérons la fonction  $f$  définie sur  $\mathbb{R}$  par  $f(x) = e^{-x^2/2}$ . On peut vérifier que  $f(x) \leq cg(x)$  avec  $c = 2\sqrt{e}$  et  $g(x) = e^{-|x|}/2$ . La fonction  $g$  est une densité sur  $\mathbb{R}$  dont la loi (dite double-exponentielle, ou exponentielle symétrique) peut être simulée facilement. En effet, la densité  $g$  est celle du produit de deux v.a.r. indépendantes  $S$  et  $Z$  suivant respectivement la loi de Rademacher de paramètre  $1/2$  et la loi exponentielle de paramètre  $1$  (donc simulables facilement). Ainsi, on peut utiliser la méthode du rejet pour simuler la loi de densité (proportionnelle à)  $f$ , qui n'est autre que la loi normale centrée réduite. Cependant, on privilégiera en pratique la méthode par projection dans l'algorithme de Box-Muller vu précédemment.*

**Exemple 2.7.2.** *Si l'on cherche à simuler une loi uniforme sur un ensemble  $D$  inclus dans un ensemble  $P$  de  $\mathbb{R}^d$  et que l'on sait simuler simplement la loi uniforme sur  $P$*

(c'est le cas par exemple si  $P$  est un pavé de  $\mathbb{R}^d$ ), alors on peut utiliser la méthode du rejet pour simuler la loi uniforme sur  $D$ . Notons  $f(x) = 1_D(x)$  et  $g(x) = 1_P(x)/\text{vol}(P)$ , où  $\text{vol}(P)$  est le volume de l'ensemble  $P$ , i.e. sa mesure de Lebesgue. Alors,  $f(x) \leq cg(x)$  avec  $c = \text{vol}(P)$ . Dans ce cas particulier, on remarque alors que  $T$  peut se réécrire comme

$$T = \inf\{k \geq 1 : X_k \in D\},$$

avec la suite  $(X_n)_{n \in \mathbb{N}^*}$  i.i.d. suivant la loi uniforme sur  $P$ .

De la même manière, il est possible de simuler la loi d'un vecteur aléatoire conditionné à appartenir à un ensemble donné. Plus précisément, considérons  $Z$  un vecteur aléatoire admettant une densité  $g$  sur  $\mathbb{R}^d$  et soit  $A$  un borélien de  $\mathbb{R}^d$  tel que  $\mathbb{P}(Z \in A) > 0$ . Alors la densité de  $Z$  conditionné à appartenir à  $A$  est

$$f(x) = \frac{g(x)1_A(x)}{\int_A g(x)dx}.$$

En posant  $c = 1/\int_A g(x)dx$ , on a que  $f(x) \leq cg(x)$  pour tout  $x \in \mathbb{R}^d$ . Étant donné que dans le cas présent on a  $\alpha(x) = 1_A(x)$ , on obtient comme précédemment que

$$T = \inf\{k \geq 1 : X_k \in A\},$$

le premier instant  $k$  pour lequel le vecteur aléatoire  $X_k$  est dans l'ensemble  $A$ .

# Chapitre 3

## Techniques de réduction de variance

Comme nous l'avons vu en introduction, la méthode de Monte-Carlo consiste à utiliser la LGN pour approcher numériquement une intégrale (ou une somme, modulo quelques modifications mineures dans ce qui suit). Autrement dit, si l'on cherche à approcher l'intégrale

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

où  $f$  est une densité de probabilité sur  $\mathbb{R}^d$  et  $g$  est une "bonne" fonction borélienne de  $\mathbb{R}^d$  dans  $\mathbb{R}$  (i.e. l'intégrale de  $g^2f$  doit être finie), alors en interprétant cette intégrale comme l'espérance  $\mathbb{E}[g(X)]$  où  $X$  est un vecteur aléatoire à valeurs dans  $\mathbb{R}^d$ , de densité  $f$ , et en considérant une suite  $(X_n)_{n \in \mathbb{N}^*}$  de vecteurs aléatoires i.i.d. de même loi que  $X$ , la moyenne empirique des  $g(X_i)$  est le candidat idéal pour approcher cette espérance : on parle de méthode de Monte-Carlo classique. De surcroît, si l'on note  $\sigma^2$  la variance de  $g(X)$ , alors le TCL indique que la vitesse de convergence de la méthode de Monte-Carlo est de l'ordre de  $\sigma/\sqrt{n}$ , l'entier  $n$  étant la taille de l'échantillon considéré.

Ainsi, pour améliorer la vitesse de convergence, il est nécessaire d'avoir une variance  $\sigma^2$  aussi petite que possible, et c'est pourquoi l'on fait appel à des méthodes dites de réduction de variance. L'idée générale est de donner une autre représentation de la quantité à approcher,  $\mathbb{E}[g(X)]$ , sous la forme d'une espérance  $\mathbb{E}[Y]$  où  $Y$  est une autre v.a.r. dont la variance est censée être plus petite que  $\sigma^2$ . Néanmoins, il faut pouvoir comparer ce qui est comparable. Si  $\tau$  désigne le temps nécessaire pour simuler  $Y$ , alors en  $N$  unités de temps on peut simuler (la partie entière de)  $n = N/\tau$  v.a.  $Y_i$  i.i.d. de même loi que  $Y$ , auquel cas la moyenne empirique  $n^{-1} \sum_{i=1}^n Y_i$  a pour variance  $\tau \text{Var}(Y)/N$  : le critère de comparaison des estimateurs basé sur le produit temps de simulation  $\times$  variance est plus adéquat que le critère basé sur la variance seule. C'est pourquoi dans la suite, par souci de simplification, nous supposons implicitement que les v.a.r.  $g(X)$  et  $Y$  auront un coût de simulation du même ordre.

Mentionnons que toutes les méthodes de réduction de variance n'assurent pas systématiquement une diminution de la variance et c'est pourquoi en pratique le choix crucial de la v.a.r.  $Y$ , donc de l'approche à utiliser, se fait au cas par cas. Autrement dit, aucune méthode n'étant meilleure qu'une autre, le choix se fait en fonction de la situation considérée et il se peut que nous ayons à les combiner afin de diminuer sensiblement la variance.

Enfin, les différentes variances à considérer n'étant en général pas explicites, il s'agit de les approcher elles-aussi numériquement, comme mentionné dans l'introduction. Une première idée est d'utiliser la variance empirique (biaisée ou non) comme estimateur. Néanmoins, il est souvent plus commode d'utiliser l'approche suivante, dite de dédoublement des variables. Pour simuler  $\text{Var}(g(X))$  où  $X$  est un vecteur aléatoire quelconque de  $\mathbb{R}^d$  de densité  $f$ , on remarque tout d'abord que la variance satisfait l'identité

$$\text{Var}(g(X)) = \frac{1}{2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (g(x) - g(y))^2 f(x)f(y) dx dy = \frac{1}{2} \mathbb{E}[k(X, Y)],$$

où  $Y$  est une copie indépendante de  $X$  et  $k$  est la fonction positive définie sur  $\mathbb{R}^2$  par  $k(x, y) = (g(x) - g(y))^2$ . Ensuite, on approche numériquement cette espérance par la méthode de Monte-Carlo classique : au sens de la convergence p.s., on a

$$\text{Var}(g(X)) = \lim_{n \rightarrow +\infty} \frac{1}{2n} \sum_{i=1}^n k(X_i, Y_i) = \lim_{n \rightarrow +\infty} \frac{1}{2n} \sum_{i=1}^n (f(X_i) - f(Y_i))^2,$$

où la suite  $(X_n, Y_n)_{n \in \mathbb{N}^*}$  est i.i.d. de même loi que le couple  $(X, Y)$ .

### 3.0.1 Variable de contrôle

Le principe est très simple : il s'agit d'écrire l'espérance  $\mathbb{E}[g(X)]$  sous la forme

$$\mathbb{E}[g(X)] = \mathbb{E}[Y] \quad \text{avec} \quad Y = g(X) - Z + \mathbb{E}[Z],$$

où  $Z$  est une v.a.r., appelée variable de contrôle, pour laquelle on sait calculer  $\mathbb{E}[Z]$  et telle que  $\text{Var}(Y)$  soit plus petite que  $\sigma^2$ . En pratique, on essaie de trouver  $Z$  sous la forme  $Z = h(X)$  où la fonction  $h$  (appelée elle aussi variable de contrôle par abus de langage) doit être suffisamment proche de  $g$  pour assurer que  $\text{Var}(Y)$  soit petite. À noter que  $\text{Var}(Y)$ , même si elle n'est pas explicitement calculable, peut quelquefois être contrôlée (au sens "majorée") par une quantité que l'on sait calculer. Mentionnons cependant qu'il n'existe pas de résultat général permettant une réduction systématique de la variance par cette méthode.



**Exemple 3.0.3.** *Voici un exemple sans intérêt pratique mais permettant de donner une idée du principe. On souhaite approcher numériquement l'intégrale*

$$I = \int_0^1 e^u du \quad (= e - 1).$$

*Si l'on note  $U \sim \mathcal{U}_{[0,1]}$  alors  $I = \mathbb{E}[g(U)]$  où  $g(u) = e^u$ . Lorsque l'on approche cette espérance par la moyenne empirique, la variance associée se calcule explicitement et vaut*

$$\sigma^2 = \mathbb{E}[e^{2U}] - \mathbb{E}[e^U]^2 = \frac{e^2 - 1}{2} - (e - 1)^2 \approx 0,24.$$

*En utilisant le développement de Taylor de la fonction  $g$  au voisinage de 0, on peut envisager de considérer pour variable de contrôle la fonction  $h(u) = 1 + u$ , de sorte que  $\mathbb{E}[h(U)] = \int_0^1 (1 + u) du = 3/2$ . On a alors*

$$\text{Var}(g(U) - h(U)) = \text{Var}(e^U - U - 1) = \text{Var}(e^U) + \frac{1}{12} + e - 3 \approx 0,04,$$

*d'où une variance inférieure. Ainsi, l'utilisation de la variable de contrôle réduit sensiblement la variance dans cet exemple.*

**Exemple 3.0.4.** *Dans un second exemple, moins trivial, il s'agit d'approcher numériquement l'intégrale*

$$I = \int_0^1 e^{u^2} du,$$

*qui elle n'est pas explicite. En considérant pour variable de contrôle la fonction  $h(u) = 1 + u^2$ , qui est encore une fois le début du développement de Taylor de la fonction  $g(u) = e^{u^2}$  au voisinage de 0, on a alors que  $\mathbb{E}[h(U)] = 4/3$  et il nous reste à estimer numériquement les 2 variances, à savoir*

$$\text{Var}(e^{U^2}) \quad \text{et} \quad \text{Var}(e^{U^2} - 1 - U^2).$$

*Après implémentation, on voit que la variance peut être réduite d'un facteur 10 grâce à cette méthode.*

### 3.0.2 Echantillonnage préférentiel

L'idée de l'échantillonnage préférentiel (*importance sampling* en anglais) est de modifier la loi du vecteur aléatoire  $X$  sous laquelle est définie l'espérance  $\mathbb{E}[g(X)]$  pour tirer des points mieux choisis par rapport à la fonction  $g$ . Par exemple, supposons que

$X \sim \mathcal{N}(0, 1)$  et que le support de la fonction  $g$  soit contenu dans l'intervalle  $[2, 3]$ . Dans ce cas, les  $g(X_i)$  intervenant dans la mise en oeuvre de la méthode de Monte-Carlo classique ne contiendront que très peu de valeurs différentes de 0, la probabilité que les  $X_i$  soit comprise entre 2 et 3 étant infime. Ainsi, il serait donc plus approprié de modifier la loi pour tirer plus de points dans l'intervalle  $[2, 3]$ . Comment faire ?

Supposons que le vecteur aléatoire  $X$  à valeurs dans  $\mathbb{R}^d$  ait une densité  $f$ . Pour modifier l'échantillonnage, on introduit une densité de probabilité  $h$ , appelée fonction d'importance, que l'on suppose strictement positive sur le support de  $f$ . On a

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}^d} g(x)f(x)dx = \int_{\mathbb{R}^d} \frac{g(x)f(x)}{h(x)}h(x)dx = \mathbb{E}[Y],$$

où  $Y$  est une v.a.r. définie par le ratio

$$Y = \frac{g(Z)f(Z)}{h(Z)},$$

le vecteur aléatoire  $Z$  à valeurs dans  $\mathbb{R}^d$  admettant  $h$  pour densité. On a donc une autre représentation de l'espérance  $\mathbb{E}[g(X)]$ . Pour que cette nouvelle représentation génère une méthode de Monte-Carlo plus efficace, il nous reste à espérer que la variance soit diminuée, ce qui n'est pas toujours le cas et dépend en pratique du choix de la fonction  $h$ .

Calculons donc la variance de  $Y$ , que l'on suppose a priori finie :

$$\begin{aligned} \text{Var}(Y) &= \mathbb{E} \left[ \left( \frac{g(Z)f(Z)}{h(Z)} \right)^2 \right] - \mathbb{E}[g(X)]^2 \\ &= \int_{\mathbb{R}^d} \left( \frac{g(z)f(z)}{h(z)} \right)^2 h(z)dz - \mathbb{E}[g(X)]^2 \\ &= \int_{\mathbb{R}^d} \left( \frac{g(z)^2 f(z)}{h(z)} \right) f(z)dz - \mathbb{E}[g(X)]^2 \\ &= \mathbb{E} \left[ \frac{g(X)^2 f(X)}{h(X)} \right] - \mathbb{E}[g(X)]^2. \end{aligned}$$

On a donc une réduction de la variance si et seulement si

$$\mathbb{E} \left[ \frac{g(X)^2 f(X)}{h(X)} \right] \leq \mathbb{E} [g(X)^2].$$

Comment construire une densité  $h$  ayant la propriété ci-dessus ? Sans donner une réponse très précise à cette question, remarquons que la variance étant nulle si et seulement si la v.a.r. est p.s. constante (une conséquence de l'inégalité de Cauchy-Schwarz), on a que

$$\text{Var}(Y) = 0 \quad \text{si et seulement si} \quad \frac{g(z)f(z)}{h(z)} = c \quad \text{pour presque tout } z.$$

Cependant, comme  $h$  doit être une densité de probabilité, cela impose en multipliant par  $h(z)$  puis en intégrant que

$$c = \int_{\mathbb{R}^d} \frac{g(z)f(z)}{h(z)} h(z) dz = \int_{\mathbb{R}^d} g(x)f(x) dx = \mathbb{E}[g(X)].$$

Ainsi, construire une telle fonction  $h$  n'est pertinent que si l'on connaît explicitement  $\mathbb{E}[g(X)]$ , ce qui n'est pas le cas. En pratique, la méthode selon laquelle simuler le vecteur aléatoire  $Z$  ayant une telle densité  $h$  n'est donc pas implémentable. En revanche, ceci permet de comprendre comment doit être choisie  $h$  : aussi proche que possible de la fonction  $gf$ , tout en assurant qu'elle est bien une densité sur  $\mathbb{R}^d$  dont la loi est facile à simuler.

**Exemple 3.0.5.** *On souhaite approcher numériquement l'intégrale*

$$I_\alpha = \int_\alpha^{\alpha+1} e^{-x} dx \quad (= e^{-\alpha}(1 - e^{-1})),$$

où  $\alpha$  est un nombre positif ou nul quelconque. Il s'agit donc ici d'approcher la probabilité  $\mathbb{P}(X \in [\alpha, \alpha + 1])$  où  $X \sim \mathcal{E}(1)$  (donc de densité  $f(x) = e^{-x}1_{[0, +\infty[}(x)$ ), qui se réécrit comme  $\mathbb{E}[g(X)]$  où  $g = 1_{[\alpha, \alpha+1]}$ . La variance de  $g(X)$ , qui est une v.a. de Bernoulli de paramètre  $I_\alpha$ , vaut donc  $I_\alpha(1 - I_\alpha)$ . Étant donné que le produit  $gf$  est proche de la fonction  $e^{-\alpha}1_{[\alpha, \alpha+1]}$ , on choisit alors pour fonction d'importance la densité  $h = 1_{[\alpha, \alpha+1]}$ , qui est la densité d'une v.a.r.  $U + \alpha$  où  $U \sim \mathcal{U}_{[0,1]}$ . Finalement, la nouvelle variance à calculer est  $\text{Var}(e^{-U-\alpha})$  qui se révèle être plus petite que  $I_\alpha(1 - I_\alpha)$  grâce à un calcul direct : on a donc bien réduit la variance.

### 3.0.3 Variables antithétiques

Contrairement aux méthodes par variable de contrôle ou par échantillonnage préférentiel, la méthode des variables antithétiques nous assure systématiquement une réduction de variance d'un facteur au moins 2, sous réserve que l'on arrive à tirer parti de certaines symétries d'une distribution et de la corrélation négative entre deux v.a.r.

**Définition 3.0.1.** *Deux v.a.r. de carré intégrable sont dites négativement corrélées si leur covariance est négative. Si de surcroît leurs espérances et variances sont égales, elles sont dites antithétiques.*

L'idée directrice de la méthode est d'essayer de construire des v.a.r. de même loi et négativement corrélées, en se basant sur le lemme suivant, dont la démonstration est évidente.

**Lemme 3.0.1.** *Soient  $Z$  et  $Z'$  deux v.a.r. ayant même variance. Alors on a l'équivalence suivante :*

$$\text{Var}\left(\frac{Z + Z'}{2}\right) \leq \frac{\text{Var}(Z)}{2} \iff \text{Cov}(Z, Z') \leq 0.$$

**Remarque 3.0.3.** *Le  $1/2$  dans la variance de gauche n'est pas anodin : si  $Z$  et  $Z'$  ne sont pas colinéaires, alors c'est la valeur pour laquelle la variance est minimale parmi l'ensemble des combinaisons convexes de telles v.a.r., i.e., si l'on note*

$$Z^\theta = \theta Z + (1 - \theta)Z',$$

pour tout  $\theta \in [0, 1]$ , alors

$$\begin{aligned} \text{Var}(\theta Z + (1 - \theta)Z') &= \theta^2 \text{Var}(Z) + (1 - \theta)^2 \text{Var}(Z') + 2\theta(1 - \theta)\text{Cov}(Z, Z') \\ &= (\theta^2 + (1 - \theta)^2) \text{Var}(Z) + 2\theta(1 - \theta)\text{Cov}(Z, Z') \\ &= 2\theta^2 (\text{Var}(Z) - \text{Cov}(Z, Z')) - 2\theta (\text{Var}(Z) - \text{Cov}(Z, Z')) + \text{Var}(Z), \end{aligned}$$

est une parabole en  $\theta$  dont le coefficient dominant,  $\text{Var}(Z) - \text{Cov}(Z, Z')$ , est strictement positif par l'inégalité de Cauchy-Schwarz (la covariance est plus petite que le produit des racines carrées des variances, avec égalité si et seulement si les v.a.r. sont colinéaires, ce qui est impossible grâce à nos hypothèses) : elle admet donc un unique minimum et le  $\theta$  minimisant la variance vaut  $1/2$ .

On suppose à présent que la v.a.r.  $Z$  est de la forme  $g(X)$ , avec  $X$  une v.a.r. et  $g$  une bonne fonction borélienne de  $\mathbb{R}$  dans  $\mathbb{R}$ . La question est alors la suivante : comment construire la v.a.r.  $Z'$  de telle sorte que  $g(X)$  et  $Z'$  soient antithétiques ? On s'appuie sur le résultat suivant.

**Proposition 3.0.1.** *Supposons que la loi de la v.a.r.  $X$  soit invariante par une transformation borélienne  $T : \mathbb{R} \rightarrow \mathbb{R}$  décroissante. Alors les v.a.r.  $g(X)$  et  $g(T(X))$  sont antithétiques dès que la fonction  $g$  est monotone.*

*Démonstration.* Rappelons que la propriété d'invariance énoncée ci-dessus signifie que  $X$  et  $T(X)$  ont même loi. Ainsi, les v.a.r.  $g(X)$  et  $g(T(X))$  ont même espérance et variance. Il suffit maintenant de montrer que leur covariance est négative. Soit  $X'$  une copie indépendante de  $X$ . La fonction  $g$  étant supposée monotone et la transformation  $T$  décroissante, on a l'inégalité au sens p.s. :

$$(g(X) - g(X'))(g(T(X)) - g(T(X'))) \leq 0,$$

puis en passant à l'espérance,

$$\mathbb{E}[(g(X) - g(X'))(g(T(X)) - g(T(X')))] \leq 0.$$

Enfin, en développant et en utilisant l'indépendance puis l'égalité en loi, il vient

$$2\mathbb{E}[g(X)g(T(X))] - 2\mathbb{E}[g(X)]\mathbb{E}[g(T(X))] \leq 0,$$

d'où le résultat.  $\square$

Ainsi, en pratique, on applique la méthode de Monte-Carlo classique à la v.a.r.

$$Y = \frac{g(X) + g(T(X))}{2},$$

plutôt qu'à  $g(X)$  car elles ont la même espérance mais  $Y$  a une variance au moins deux fois plus petite.

**Remarque 3.0.4.** *Mentionnons que cette méthode peut être immédiatement généralisée au cas multi-dimensionnel, en adaptant les hypothèses :*

- le vecteur aléatoire  $X$  est à valeurs dans  $\mathbb{R}^d$  et ses coordonnées sont indépendantes.
- la loi de  $X$  est invariante par une transformation borélienne  $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ .
- les fonctions  $g$  et  $g \circ T$  de  $\mathbb{R}^d$  dans  $\mathbb{R}$  sont respectivement croissante et décroissante (ou décroissante et croissante) en toutes les coordonnées (les autres étant gelées).

**Remarque 3.0.5.** *Les deux exemples typiques pour lesquels la méthode des variables antithétiques s'applique (quasi-)systématiquement sont les cas uniforme et gaussien.*

◦ Lorsque  $X \sim \mathcal{N}(0, I_d)$ , alors  $T$  peut être toute transformation orthogonale, auquel cas les hypothèses de monotonie vont porter sur les fonctions  $g$  et  $g \circ T$ . La transformation  $T(x) = -x$  convient aussi, auquel cas  $g$  doit être supposée croissante ou décroissante en toutes les coordonnées.

◦ Lorsque  $U \sim \mathcal{U}_{[0,1]^d}$ , la transformation  $T(x) = \mathbf{1} - x$  convient, où  $\mathbf{1}$  est le vecteur  $d$ -dimensionnel dont les coordonnées valent toutes 1. De même, la fonction  $g$  doit être supposée croissante ou décroissante en toutes les coordonnées.

### 3.0.4 Méthode de stratification

C'est une méthode bien connue des statisticiens et souvent utilisée dans les sondages. Pour approcher numériquement l'espérance  $I = \mathbb{E}[g(X)]$  où  $X$  est un vecteur aléatoire à valeurs dans  $\mathbb{R}^d$ , on se donne une partition  $(D_i)_{i=1,\dots,m}$  de  $\mathbb{R}^d$  (appelée strates) telle que les  $p_i = \mathbb{P}(X \in D_i)$  soient connues et toutes strictement positives et l'on décompose l'espérance de la manière suivante :

$$\mathbb{E}[g(X)] = \sum_{i=1}^m \mathbb{E}[g(X)1_{D_i}(X)] = \sum_{i=1}^m p_i I_i,$$

où l'on a noté  $I_i = \mathbb{E}[g(X) \mid X \in D_i]$ . Alors on a au sens de l'approximation p.s. :

$$I = \sum_{i=1}^m p_i I_i \approx \sum_{i=1}^m p_i \tilde{I}_i,$$

où l'intégrale  $I_i$  est approchée numériquement par une moyenne empirique  $\tilde{I}_i$  à l'aide de  $n_i$  tirages i.i.d. de loi (conditionnelle) celle de  $X$  sachant  $X \in D_i$  (notons que l'on peut en général simuler cette dernière par la méthode du rejet, comme on l'a vu dans le chapitre précédent). Remarquons que ces entiers  $n_i$  doivent être supposés suffisamment grands de manière à rendre l'approximation précédente licite. Les échantillons servant à obtenir les estimateurs  $\tilde{I}_i$  étant supposés indépendants, la variance de l'estimateur de  $I$  vaut la somme des variances, à savoir

$$\sum_{i=1}^m p_i^2 \frac{\sigma_i^2}{n_i},$$

où  $\sigma_i^2 = \text{Var}(g(X) \mid X \in D_i)$ . Il est alors naturel de minimiser cette variance en les tailles  $n_i$  des différents échantillons, lorsque le nombre total de tirages est fixé égal à  $n$ , i.e.  $\sum_{i=1}^m n_i = n$ . C'est un problème de minimisation sous contrainte du type

$$\min \{ \phi(x) : x \in (\mathbb{R}^+)^m, \psi(x) = 0 \}, \quad \text{avec} \quad \phi(x) = \sum_{i=1}^m p_i^2 \frac{\sigma_i^2}{x_i} \quad \text{et} \quad \psi(x) = \sum_{i=1}^m x_i - n.$$

La méthode du lagrangien nous dit que les points critiques du problème d'optimisation sont les points  $x \in (\mathbb{R}^+)^m$  dont les vecteurs  $\nabla \phi(x)$  et  $\nabla \psi(x)$  sont colinéaires. Autrement dit, nous cherchons  $x \in (\mathbb{R}^+)^m$  et  $\lambda \in \mathbb{R}$  tels que pour tout  $i \in \{1, \dots, m\}$ , on ait  $-p_i^2 \sigma_i^2 / x_i^2 = \lambda$ . L'unique solution est donc

$$x_i = \frac{np_i \sigma_i}{\sum_{j=1}^m p_j \sigma_j},$$

et la fonction  $\phi$  étant convexe sur  $(\mathbb{R}^+)^m$ , l'unique point critique réalise le minimum global. Oubliant qu'il nous faut des entiers (ceci revient à négliger les erreurs d'arrondi), la variance minimale est alors

$$\sum_{i=1}^m \frac{p_i^2 \sigma_i^2}{n_i} = \sum_{i=1}^m p_i^2 \sigma_i^2 \frac{\sum_{j=1}^m p_j \sigma_j}{np_i \sigma_i} = \frac{1}{n} \left( \sum_{i=1}^m p_i \sigma_i \right)^2.$$

Il reste à comparer cette variance avec la variance initiale : on a

$$\text{Var}(g(X)) = \mathbb{E}[g(X)^2] - \mathbb{E}[g(X)]^2$$

$$\begin{aligned}
&= \sum_{i=1}^m p_i \mathbb{E} [g(X)^2 | X \in D_i] - \left( \sum_{i=1}^m p_i \mathbb{E} [g(X) | X \in D_i] \right)^2 \\
&= \sum_{i=1}^m p_i \text{Var} (g(X) | X \in D_i) + \sum_{i=1}^m p_i \mathbb{E} [g(X) | X \in D_i]^2 - \left( \sum_{i=1}^m p_i \mathbb{E} [g(X) | X \in D_i] \right)^2 \\
&\geq \sum_{i=1}^m p_i \text{Var} (g(X) | X \in D_i) \\
&= \sum_{i=1}^m p_i \sigma_i^2 \\
&\geq \left( \sum_{i=1}^m p_i \sigma_i \right)^2,
\end{aligned}$$

où nous avons utilisé à deux reprises l'inégalité de Cauchy-Schwarz pour obtenir les deux inégalités. Ainsi, on a bien diminué la variance initiale par la méthode de stratification. Néanmoins, le point négatif dans ce qui précède est que les  $n_i$ , choisis de manière optimale, dépendent des  $\sigma_i$  qui sont rarement explicites, ce qui limite l'utilisation de la méthode. En pratique, on approche les  $n_i$  en remplaçant la variance  $\sigma_i^2$  par un estimateur consistant calculé à partir d'un premier jeu de simulations.

Une autre façon de choisir les entiers  $n_i$  est de les prendre proportionnels à la probabilité  $p_i$ , i.e. la partie entière de  $np_i$ . Toujours en négligeant la partie entière, on a alors un estimateur de  $I$  dont la variance est égale à

$$\frac{1}{n} \sum_{i=1}^m \frac{p_i^2 \sigma_i^2}{n_i} = \frac{1}{n} \sum_{i=1}^m p_i \sigma_i^2,$$

dont on a vu précédemment qu'elle était inférieure à la variance initiale  $\text{Var}(g(X))/n$ . Ainsi, on peut toujours construire un estimateur stratifié plus efficace que l'estimateur de Monte Carlo classique, au sens où il est de variance plus faible.

### 3.0.5 Conditionnement

Terminons par une méthode qui peut se révéler intéressante dans le cas multi-dimensionnel. On cherche à approcher numériquement l'espérance  $\mathbb{E}[g(X)]$  où  $X$  est un vecteur aléatoire de  $\mathbb{R}^d$  avec  $d \geq 2$ , en exploitant une éventuelle représentation explicite de la densité de l'une des coordonnées de  $X$ . Regardons ce qu'il en est dans le cas de la dimension 2. Si  $f$  désigne la densité d'un couple de v.a.r.  $X = (X_1, X_2)$ , on souhaite

approcher l'intégrale

$$\int_{\mathbb{R}} \int_{\mathbb{R}} g(x, y) f(x, y) dx dy.$$

En posant la fonction

$$h(x) = \frac{1}{\int_{\mathbb{R}} f(x, y) dy} \int_{\mathbb{R}} g(x, y) f(x, y) dy, \quad x \in \mathbb{R},$$

on voit que l'on a

$$h(x) = \mathbb{E}[g(X_1, X_2) \mid X_1 = x],$$

et comme la densité de  $X_1$  est  $x \mapsto \int_{\mathbb{R}} f(x, y) dy$ , on obtient en intégrant l'égalité précédente que

$$\mathbb{E}[g(X_1, X_2)] = \mathbb{E}[h(X_1)].$$

L'intérêt de "désintégrer" la loi du couple pour ne considérer que l'une des deux v.a.r. réside dans le fait que la variance est forcément réduite :

$$\begin{aligned} \text{Var}(h(X_1)) &= \mathbb{E}[h(X_1)^2] - \mathbb{E}[h(X_1)]^2 \\ &= \mathbb{E}[\mathbb{E}[g(X_1, X_2) \mid X_1]^2] - \mathbb{E}[g(X_1, X_2)]^2 \\ &\leq \mathbb{E}[\mathbb{E}[g(X_1, X_2)^2 \mid X_1]] - \mathbb{E}[g(X_1, X_2)]^2 \\ &= \text{Var}(g(X_1, X_2)), \end{aligned}$$

où pour obtenir l'inégalité ci-dessus nous avons utilisé l'inégalité de Cauchy-Schwarz pour l'espérance conditionnelle. Notons que cette méthode n'est efficace que lorsque la fonction  $h$  et la densité de  $X_1$  sont explicites (si c'est plutôt celle de  $X_2$  qui l'est, on prend pour  $h$  la fonction pour laquelle on a inversé les rôles de  $x$  et de  $y$ ).



# Bibliographie

- [1] Nicolas Bouleau. Probabilités de l'ingénieur. Hermann, 1986.
- [2] Carl Graham et Denis Talay. Simulation stochastique et méthodes de Monte Carlo. Éditions de l'École Polytechnique, 2011.
- [3] Christian Robert. Méthodes de Monte Carlo par chaînes de Markov. Éditions Economica, 1996.