

Méthodes de réduction de la variance (2)

Prof. Mohamed El Merouani

Université Abdelmalek Essaâdi
Faculté des Sciences de Tétouan
Département de Mathématiques

2018/2019

Parmi les Méthodes de réduction de la variance qui reste à voir, on trouve :

- Méthode de stratification.
- Valeur moyenne.
- Conditionnement.
- Méthode Quasi-Monte-Carlo.

Méthode de stratification :

On veut calculer une intégrale de la forme $I = E(h(X))$ avec X variable de densité f sur un ensemble D .

Cette écriture nous donne l'idée d'une première méthode de Monte-Carlo :

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^n h(X_i)$$

avec X_1, \dots, X_n i.i.d. de même loi que X .

L'erreur commise

$$I - \hat{I}_n = E(h(X)) - \frac{1}{n} \sum_{i=1}^n h(X_i)$$

est approximativement de loi $\mathcal{N}(0, \sigma/\sqrt{n})$ avec $\sigma^2 = \text{Var}(h(X))$.

Méthode de stratification :

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

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

avec pour tout i , $I_i = E(h(X)/X \in D_i)$ et $p_i = P(X \in D_i)$ (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 :

$$I_i \approx \hat{I}_i = \frac{1}{n} \sum_{j=1}^{n_i} h(X_j^{(i)})$$

avec des variables $X_1^{(i)}, X_2^{(i)}, \dots, X_{n_i}^{(i)}$ i.i.d. $\sim \mathcal{L}(X/X \in D_i)$.

Méthode de stratification :

Chacune de ces approximations a un coût numérique n_i (puisqu'on fait n_i boucles pour calculer la somme ci-dessus). Nous en déduisons une deuxième approximation :

$$I \approx \sum_{i=1}^m p_i \hat{I}_i$$

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

$$I - \left(\sum_{i=1}^m p_i \hat{I}_i \right) = \sum_{i=1}^m p_i (I_i - \hat{I}_i)$$

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

$$\begin{aligned} \sigma_i^2 &= \text{Var}(h(X)/X \in D_i) \\ &= E \left((h(X) - E(h(X)/X \in D_i))^2 / X \in D_i \right) \end{aligned}$$

$$\begin{aligned}\sigma_i^2 &= \frac{1}{p_i} E \left((h(X) - E(h(X)/X \in D_i))^2 \mathbb{I}_{\{X \in D_i\}} \right) \\ &= p_i^{-1} E(h(X)^2 \mathbb{I}_{\{X \in D_i\}}) + p_i^{-1} E(h(X)/X \in D_i)^2 E(\mathbb{I}_{\{X \in D_i\}}) \\ &\quad - 2p_i^{-1} E(h(X)/X \in D_i) E(h(X) \mathbb{I}_{\{X \in D_i\}}) \\ &= p_i^{-1} E(h(X)^2 \mathbb{I}_{\{X \in D_i\}}) - p_i^{-2} E(h(X) \mathbb{I}_{\{X \in D_i\}})^2\end{aligned}$$

Donc l'erreur est (approximativement) de loi $\mathcal{N}(0, \sqrt{\sum_{i=1}^m p_i^2 \sigma_i^2 / n_i})$.
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 :

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 $n \left(\frac{p_1 \sigma_1}{\sum_{i=1}^m p_i \sigma_i}, \dots, \frac{p_m \sigma_m}{\sum_{i=1}^m p_i \sigma_i} \right)$

Preuve :

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 un 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 (nabla) ∇ pour le gradient).

Méthode de stratification :

Suite de la preuve :

Nous avons

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_m}(x) \right) = \left(\frac{-p_1^2 \sigma_1^2}{x_1^2}, \dots, \frac{-p_m^2 \sigma_m^2}{x_m^2} \right)$$

$$\nabla g(x) = (1, 1, \dots, 1)$$

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 $x_i = n \frac{p_i \sigma_i}{\sum_{i=1}^m p_i \sigma_i}$, 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. ■

Méthode de stratification :

Comme nous voulons des n_i entiers, nous prenons $n_i = \lfloor n \frac{p_i \sigma_i}{\sum_{i=1}^m p_i \sigma_i} \rfloor$, 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 :

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

Lemme :

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

(La variance est donc bien réduite par la stratification)

Preuve : Nous avons

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

Méthode de stratification :

Suite de la preuve du lemme :

$$\begin{aligned} &= \sum_{i=1}^m \{p_i E(h(X)^2 / X \in D_i) - p_i E(h(X) / X \in D_i)^2\} \\ &+ \sum_{i=1}^m \{p_i E(h(X) / X \in D_i)^2\} - E \left(\sum_{i=1}^m p_i E(h(X) / X \in D_i) \right)^2 \end{aligned}$$

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

$$\sum_{i=1}^m \lambda_i x_i^2 \geq (\lambda_i x_i)^2 \tag{1}$$

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

Méthode de stratification :

Suite de la preuve du lemme :

Donc

$$\begin{aligned} \text{Var}(h(X)) &\geq \sum_{i=1}^m \{p_i E(h(X)^2/X \in D_i) - p_i E(h(X)/X \in D_i)^2\} \\ \text{(encore(1))} &\geq \left(\sum_{i=1}^m p_i (E(h(X))^2/X \in D_i) - E(h(X)/X \in D_i)^2 \right)^{1/2} \\ &\geq \left(\sum_{i=1}^m p_i \sigma_i \right)^2 \end{aligned}$$

Valeur moyenne :

On cherche à calculer une intégrale de la forme

$$I = E(h(X, Y)) = \int h(x, y) f(x, y) dx dy$$

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

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

avec des (X_i, Y_i) i.i.d. de même loi que (X, Y) .

Soit

$$g(x) = \frac{1}{m(x)} \int h(x, y) f(x, y) dy$$

avec $m(x) = \int f(x, y) dy$.

Nous avons

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

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

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

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

Lemme :

Nous avons

$$\text{Var}(g(X)) \leq \text{Var}(h(X, Y))$$

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

Preuve :

Nous avons que $E(g(X)) = E(h(X, Y))$ donc pour comparer $\text{Var}(g(X))$ et $\text{Var}(h(X, Y))$, il suffit de comparer $E(g(X)^2)$ et $E(h(X, Y)^2)$. Nous avons : $E(g(X)^2) = \int_{x,y} g(x)^2 f(x, y) dx dy$

$$= \int_{x,y} \left(\frac{\int_v h(x, v) f(x, v) dv}{\int_u f(x, u) du} \right)^2 f(x, y) dx dy$$

$$\text{(Cauchy-Shwarz)} \leq \int_{x,y} \frac{\int_v h(x, v)^2 f(x, v) dv}{\int_u f(x, u) du} f(x, y) dx dy$$

Valeur moyenne :

Suite de la preuve :

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

Donc $Var(g(X)) \leq Var(h(X,Y))$.

On cherche toujours à estimer $I = E(h(X))$ en supposant $E(h^2(X)) < \infty$. L'idée force dans cette section a été vue en cours de probabilités, à savoir : le conditionnement ne change pas la moyenne, mais réduit l'incertitude. Appliquée à notre contexte, ceci signifie que pour toute autre variable aléatoire Y , on a d'une part

$$E(E(h(X)/Y)) = E(h(X)) = I$$

et d'autre part

$$\begin{aligned}\sigma^2 = \text{Var}(h(X)) &= \text{Var}(E(h(X)/Y)) + E(\text{Var}(h(X)/Y)) \\ &\geq \text{Var}(E(h(X)/Y))\end{aligned}$$

Conditionnement :

Soit, donc, Y une variable auxiliaire de densité g dont on sait simuler des réalisations et telle que, pour tout y , on peut calculer $k(y) = E(h(X)/Y = y)$. On considère l'estimateur

$$\tilde{I}_n = \frac{1}{n} \sum_{i=1}^n k(Y_i) = \frac{1}{n} \sum_{i=1}^n E(h(X)/Y_i)$$

Ses propriétés découlent de ce qui vient d'être dit.

Proposition : (Estimation par conditionnement)

Si $E|h(X)| < \infty$, alors l'estimateur \tilde{I}_n est sans biais et convergent, c'est-à-dire $E(\tilde{I}_n) = I$ et

$$\tilde{I}_n = \frac{1}{n} \sum_{i=1}^n k(Y_i) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} E(k(Y)) = E(E(h(X)/Y)) = I$$

Proposition : (Estimation par conditionnement)

Si de plus $E(h^2(X)) < \infty$, alors

$$\sqrt{n}(\tilde{I}_n - I) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, s)$$

avec $s^2 = \text{Var}(E(h(X)/Y)) = \text{Var}(k(Y)) = E(k^2(Y)) - I^2$

A nouveau, la variance s^2 est estimée de façon immédiate par

$$\tilde{s}_n^2 = \frac{1}{n} \sum_{i=1}^n k^2(Y_i) - \tilde{I}_n^2$$

et les intervalles de confiance asymptotiques en découlent.

Exemple : "Surface du quart de disque unité"

On revient à l'exemple déjà vu de l'estimation de π . Supposons que (X, Y) suivre la loi uniforme sur le carré $C = [0, 1] \times [0, 1]$ et que $h(x, y) = \mathbb{I}_{\{x^2+y^2 \leq 1\}}$. En notant $D = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 \leq 1\}$ le quart de disque unité, on a donc

$$I = \iint_C \mathbb{I}_D(x, y) dx dy = \lambda(D) = \frac{\pi}{4}$$

Simuler des points (X_i, Y_i) uniformément dans C est très facile et la propriété précédente assure donc

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_D(X_i, Y_i) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \frac{\pi}{4} \Leftrightarrow 4 \times \hat{I}_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \pi$$

Conditionnement :

Exemple : "Surface du quart de disque unité"

Les variables X et Y étant uniformes et indépendantes, on a

$$\begin{aligned} E(\mathbb{I}_D(X, Y) / X = x) &= P(x^2 + Y^2 \leq 1) = P(Y \leq \sqrt{1 - x^2}) \\ &= \sqrt{1 - x^2} = k(x) \end{aligned}$$

ce qui conduit à l'estimateur

$$\tilde{I}_n = \frac{1}{n} \sum_{i=1}^n \sqrt{1 - X_i^2}$$

Sa variance se calcule facilement

$$s^2 = E(k^2(X)) - I^2 = \int_0^1 (1 - x^2) dx - \left(\frac{\pi}{4}\right)^2 = \frac{2}{3} - \left(\frac{\pi}{4}\right)^2 \approx 0,05$$

La variance de \hat{I}_n étant de $\sigma^2 = \frac{\pi}{4} \left(1 - \frac{\pi}{4}\right) \approx 0,17$ on gagne un facteur de 2 en écart-type, c'est-à-dire que pour le même n , on a un estimateur deux fois plus précis avec deux fois moins de variables uniformes (puisque l'on ne simule plus les X_i).

Méthode de quasi-Monte-Carlo :

La méthode de **quasi-Monte-Carlo** n'utilise pas des suites de nombres aléatoires, c'est l'analogue déterministe de la méthode de Monte-Carlo. Mais, elle se base sur le même problème que la méthode de Monte-Carlo, c'est-à-dire, l'approximation de l'intégrale d'une fonction f :

$$I = \int_{[0,1]^d} f(x)dx$$

Si $d = 1$, pour une méthode de Monte-Carlo classique, i.e. lorsque $(X_n)_{n \geq 1}$ est une suite de variables i.i.d. uniformes sur $[0, 1]$, on fait l'approximation

$$I \approx \hat{I}_n = \frac{1}{n} \sum_{i=1}^n f(X_i)$$

qui a une erreur moyenne en $O(1/\sqrt{n})$, celle-ci étant mesurée par l'écart-type. Donc, la méthode de Monte-Carlo classique a une vitesse de convergence lente.

Méthode de quasi-Monte-Carlo :

Notion de discrèpance :

Dans l'objectif de chercher des suites de nombres (autres que les nombres aléatoires ou pseudo-aléatoires) telles que la convergence soit plus rapide, les recherches ont conduit à s'intéresser aux méthodes dites de quasi-Monte-Carlo. Ces méthodes approchent I par $\frac{1}{n} \sum_{i=1}^n f(Y_i)$ où $(Y_i)_{i \in \mathbb{N}}$ est une suite déterministe. Comme nous voulons qu'il y ait asymptotiquement convergence quand $n \rightarrow +\infty$, la suite $(Y_i)_{i \in \mathbb{N}}$ doit être équirépartie, d'où la notion de discrèpance :

Soit la suite $Y = (Y_i)_{i \in \mathbb{N}}$, soit $A_n(B, Y)$ le nombre d'éléments appartenant à $B \subset [0, 1]^d$ parmi les n premiers termes de Y , c'est-à-dire $A_n(B, Y) = \sum_{i=1}^n \mathbb{I}_{(Y_i \in B)}$, et soit \mathcal{F} une famille de sous-ensembles de $[0, 1]^d$. La discrèpance des n premiers termes de Y est alors définie par :

$$D_n(\mathcal{F}, Y) = \sup_{B \in \mathcal{F}} \left| \frac{A_n(B, Y)}{n} - \lambda_d(B) \right|, \text{ où } \lambda_d \text{ est la mesure de Lebesgue}$$

(ou la mesure uniforme) sur $[0, 1]^d$.

Méthode de quasi-Monte-Carlo :

Notion de discrèpance :

Le type de discrèpance la plus utilisée est la discrèpance étoile :

$$D_n^*(Y) = D_n(\mathcal{F}^*, Y)$$

où $\mathcal{F}^* = \{\prod_{i=1}^d [0, u_i) : u_i \in [0, 1]\}$

En dimension $d = 1$, on a ainsi

$$D_n^*(Y) = \sup_{0 \leq u \leq 1} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{[0, u]}(Y_i) - u \right|$$

Autrement dit, la discrèpance mesure la distance en norme sup entre la fonction de répartition empirique et celle de la loi uniforme.

Variation de Hardy-Krause :

La variation au sens de Hardy-Krause d'une fonction $f : [0, 1]^d \rightarrow \mathbb{R}$ de classe C^d est

$$V(f) = \sum_{j=1}^d \sum_{i_1 < \dots < i_j} \int_{[0,1]^j} \left| \frac{\partial^j f}{\partial x_{i_1} \dots \partial x_{i_j}}(x(i_1, \dots, i_j)) \right| dx_{i_1} \dots dx_{i_j},$$

avec $x(i_1, \dots, i_j)$ le point de \mathbb{R}^d dont toutes les coordonnées valent 1 sauf celles de rangs (i_1, \dots, i_j) qui valent respectivement $(x_{i_1}, \dots, x_{i_j})$. Ainsi, lorsque $d = 1$, on a tout simplement

$$V(f) = \int_0^1 |f'(x)| dx = \|f'\|_1$$

Pour $d = 2$, ça se complique un peu :

$$\begin{aligned} V(f) &= \int_0^1 \left| \frac{\partial f}{\partial x_1}(x_1, 1) \right| dx_1 + \int_0^1 \left| \frac{\partial f}{\partial x_2}(1, x_2) \right| dx_2 + \\ &\quad + \iint_{[0,1]^2} \left| \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1, x_2) \right| dx_1 dx_2 \end{aligned}$$

Variation de Hardy-Krause :

Il est clair qu'en dimension supérieure, estimer la variation de Hardy-Krause devient vite inextricable : somme de $(2^d - 1)$ termes, dérivées partielles... Néanmoins, cette notion intervient de façon cruciale pour majorer la qualité d'un estimateur basé sur une séquence $(Y_n)_{n \geq 1}$.

Estimation d'erreur d'approximation de la méthode de quasi-Monte-Carlo :

L'erreur d'approximation de la méthode de quasi-Monte-Carlo est bornée par un terme proportionnel à la discrépance de la suite $(Y_n)_{n \geq 1}$. Plus précisément, l'inégalité de Koksma-Hlawka borne l'erreur par

$$\left| \frac{1}{n} \sum_{i=1}^n f(Y_i) - \int_{[0,1]^d} f(x) dx \right| \leq V(f) \times D_n^*(Y),$$

où $V(f)$ est la variation de f au sens de Hardy-Krause et $D_n^*(Y)$ la discrépance de Y à l'ordre n .

Méthode de quasi-Monte-Carlo :

Suite à discrèpanance faible :

Une suite à discrèpanance faible est une suite numérique ayant la propriété que pour tout entier N , la sous-suite Y_1, \dots, Y_N a une discrèpanance basse.

Grosso modo, la discrèpanance d'une suite est faible si la proportion des points de la suite sur un ensemble B est proche de la valeur de la mesure de B , ce qui est le cas en moyenne (mais pas pour des échantillons particuliers) pour une suite équirépartie.

Les suites à discrèpanance faible sont appelées quasi-aléatoires ou sous-aléatoires, en raison de leur utilisation pour remplacer les tirages de la loi uniforme continue. Le préfixe "quasi" précise ainsi que les valeurs d'une suite à discrèpanance faible ne sont pas aléatoires ou pseudo-aléatoires, mais ont des propriétés proches de tels tirages, permettant ainsi leur usage intéressant dans la méthode de quasi-Monte-Carlo.

Méthode de quasi-Monte-Carlo :

Suite à discrèpance faible :

Définition :

Une suite $(Y_n)_{n \geq 1}$ de $[0, 1]^d$ vérifiant $D_n^*(Y) = O((\log n)^d/n)$ est dite à discrèpance faible et une méthode d'approximation basée sur ce type de suite est dite de Quasi-Monte-Carlo.

Exemples :

Parmi les suites à discrèpance faible, on trouve les suites de Halton, les suites de Faure, de Sobol et de Niederreiter.

Pour une suite de Halton en dimension d , on peut montrer que $D_n^*(Y) = O((\log n)^d/n)$.

En dimension $d = 1$, le cas particulier de la suite de Halton est la suite de van der Corput. Dans son cas, $D_n^*(Y) = O(\log n/n)$.

Si l'on revient au problème d'intégration

$$I = \int_{[0,1]^d} f(x)dx,$$

et que l'on se donne une suite $(Y_n)_{n \geq 1}$ à discrédance faible, par exemple une suite de Halton basée sur les d premiers nombres premiers, alors l'estimateur

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^n f(Y_i)$$

est un estimateur Quasi-Monte-Carlo de l'intégrale I , dont l'erreur (déterministe) est donc en $O((\log n)^d/n)$. Rappelons qu'une méthode de Monte-Carlo classique présente une erreur (moyenne) en $O(1/\sqrt{n})$. Dès lors, les méthodes Quasi-Monte-Carlo sont typiquement compétitives et s'appliquent surtout en mathématiques financières.

Exercice :

On veut calculer $I = E(\mathbb{I}_{\{X>0\}}e^{\beta X})$ 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 par la méthode initiale (quand on tire des X_1, X_2, \dots, X_n i.i.d. de loi $\mathcal{N}(0, 1)$ et que l'on approche I par $\frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i>0\}}e^{\beta X_i}$).
- 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 :

(1) Si on cherche à calculer $E(e^{\beta X})$ avec $X \sim \mathcal{N}(0, 1)$. Nous savons que

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

Si nous tirons X_1, \dots, X_n i.i.d. de même loi que X , nous aurons

$$\frac{1}{n} \sum_{i=1}^n e^{\beta X_i} \approx E(e^{\beta X}) + \frac{\sigma}{\sqrt{n}} Y$$

avec $Y \sim \mathcal{N}(0, 1)$, $\sigma^2 = \text{Var}(e^{\beta X})$ (d'après le théorème de la limite-centrale).

Exercice :

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

$$\frac{1}{E(e^{\beta X})} \left| \frac{1}{n} \sum_{i=1}^n e^{\beta X_i} - E(e^{\beta X}) \right| \approx \frac{\sigma}{E(e^{\beta X})\sqrt{n}} E(|Y|) = \frac{\sigma}{E(e^{\beta X})\sqrt{n}} \sqrt{\frac{2}{\pi}}$$

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

Exercice :

(2) Méthode d'échantillonnage préférentiel :

Posons $h(x) = \mathbb{I}_{\{x>0\}} e^{\beta x}$, $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

Nous remarquons que

$$h(x)f(x) = \frac{\mathbb{I}_{\{X>0\}}}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-\beta)^2 + \frac{\beta^2}{2}\right)$$

Nous utilisons les mêmes notations que dans le cours et prenons

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

Nous avons

$$\frac{\tilde{f}(x)}{\int_{\mathbb{R}} \tilde{f}(y) dy} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-\beta)^2\right)$$

(c'est la densité de $\mathcal{N}(\beta, 1)$, suivant laquelle nous savons simuler).

Exercice :

(3) Méthode de variable de contrôle :

Nous avons

$$I = E(f(X))$$

$$\text{avec } f(x) = \mathbb{I}_{\{x>0\}} e^{\beta x}$$

$$= E(f(X) - h(X)) + E(h(X))$$

$$\text{avec } h(x) = e^{\beta x} \text{ et } E(h(X)) = \int_{\mathbb{R}} \frac{\exp\left(-\frac{x^2}{2} + \beta x\right)}{\sqrt{2\pi}} dx$$

$$= \int_{\mathbb{R}} \frac{\exp\left(-\frac{1}{2}(x - \beta)^2 + \frac{\beta^2}{2}\right)}{\sqrt{2\pi}} dx$$

$$= \exp\left(\frac{\beta^2}{2}\right)$$

(4) Technique de variable antithétique :

Si $X \sim \mathcal{N}(0, 1)$, alors $-X \sim \mathcal{N}(0, 1)$. Donc

$$E(h(X)) = E(h(-X)) = E\left(\frac{h(X) + h(-X)}{2}\right)$$

On en déduit une nouvelle méthode de Monte-Carlo pour approcher l'espérance.