

Chapter 2

Méthodes de Monte Carlo

Références:

- [F] Fishman, A first course in Monte Carlo, chap 2.
- [R] Rubinstein, Simulation and the Monte Carlo method,
- [B] Bouleau, Probabilités de l'ingénieur, chap 4.

Cadre: Soit $g : \mathbb{R}^d \rightarrow \mathbb{R}$ intégrable. On veut calculer une valeur approchée de

$$I = \int_{\mathbb{R}^d} g(x) dx.$$

Cette intégrale peut par exemple provenir d'un problème concret: en fiabilité, calculer la durée moyenne de vie (Mean Time To Failure MTTF) est souvent impossible analytiquement.

Hypothèses:

1. $\int_{\mathbb{R}^d} g^2(x) dx < +\infty$.
2. Il existe une densité f sur \mathbb{R}^d telle que $\int_{\mathbb{R}^d} \frac{g^2(x)}{f(x)} dx < +\infty$.
3. On sait simuler une variable aléatoire de densité f et on a à notre disposition une suite $(X_i)_{i \in \mathbb{N}}$ de vaiaid de densité f .

Buts:

1. Donner une valeur approchée \hat{I}_n de I en fonction de X_1, \dots, X_n .
2. Ecrire l'algorithme.
3. Etudier sa convergence et estimer l'erreur.
4. Améliorer la vitesse et comparer à d'autres méthodes de calcul d'intégrales.

2.1 Loi des grands nombres et estimateur

Pour calculer $I = \int_{\mathbb{R}^d} g(x) dx$, l'idée est de l'écrire comme l'espérance d'une fonction de la variable aléatoire X que l'on sait simuler. On pose pour tout $i \in \mathbb{N}$, $Y_i = \frac{g(X_i)}{f(X_i)}$. Les variables

aléatoires $(Y_i)_{i \in \mathbb{N}}$ sont indépendantes et de même loi; leur espérance commune est

$$\mathbb{E}(Y_1) = \int_{\mathbb{R}^d} \frac{g(x)}{f(x)} f(x) dx = I.$$

On pense alors au résultat suivant:

Théorème 2.1.1 (Loi forte des grands nombres) Soit $(Y_i)_{i \in \mathbb{N}}$ des variables aléatoires indépendantes identiquement distribuées et intégrables. Alors, presque sûrement et dans L^1 ,

$$\frac{1}{n} \sum_{i=1}^n Y_i \xrightarrow{n \rightarrow +\infty} \mathbb{E}Y_1.$$

On introduit naturellement l'estimateur suivant:

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

Comme $\mathbb{E}(\hat{I}_n) = I$, l'estimateur est *sans biais*.

Soit $(X_i)_{i \in \mathbb{N}}$ une suite de v.a. de densité f . Alors

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^n \frac{g(X_i)}{f(X_i)} \xrightarrow{n \rightarrow +\infty} I = \int_{\mathbb{R}^d} g(x) dx \quad p.s. \text{ et dans } L^1.$$

2.2 Variance, erreurs et intervalle de confiance

2.2.1 Variance et erreurs

♣ **Exercice 1:** Soit $(Y_i)_{i \in \mathbb{N}}$ des variables aléatoires indépendantes identiquement distribuées admettant un moment d'ordre 2. On pose $\hat{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i$. Calculer l'espérance et la variance de \hat{Y}_n en fonction de l'espérance et la variance de Y_1 .

Calculons la variance de Y_1 :

$$\mathbb{E}(Y_1^2) = \int_{\mathbb{R}^d} \frac{g^2(x)}{f^2(x)} f(x) dx = \int_{\mathbb{R}^d} \frac{g^2(x)}{f(x)} dx < +\infty \text{ et } \text{var}(Y_1) = \sigma^2 = \int_{\mathbb{R}^d} \frac{g^2(x)}{f(x)} dx - I^2.$$

On a immédiatement $\text{var}(\hat{I}_n) = \sigma^2/n$.

Théorème 2.2.1 (Théorème Central Limite) Soit $(Y_i)_{i \in \mathbb{N}}$ des variables aléatoires indépendantes identiquement distribuées et admettant un moment d'ordre 2. Alors, pour tout $-\infty \leq a < b \leq +\infty$,

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(a \leq \sqrt{\frac{n}{\text{var}(Y_1)}} \left(\frac{1}{n} \sum_{i=1}^n Y_i - \mathbb{E}Y_1 \right) \leq b \right) = \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a),$$

où Z est une variable aléatoire gaussienne centrée réduite et Φ sa fonction de répartition.

Ceci implique que $\sqrt{\frac{n}{\sigma^2}}(\hat{I}_n - I)$ converge en loi vers $\mathcal{N}(0, 1)$, et signifie "moralement" que $\hat{I}_n - I \simeq \sqrt{\frac{\sigma^2}{n}}Z$, où Z est une gaussienne centrée réduite. Il est donc naturel d'introduire l'erreur standard $\sqrt{\frac{\sigma^2}{n}}$ et l'erreur relative $\frac{1}{I}\sqrt{\frac{\sigma^2}{n}}$.

Cependant, la variance σ^2 est en général inconnue (en particulier, elle dépend de I). On va donc la remplacer par l'estimateur classique de la variance:

Estimateur de la variance: $s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \hat{I}_n)^2 = \frac{1}{n-1} \left(\sum_{i=1}^n Y_i^2 - n\hat{I}_n^2 \right).$
--

On sait (voir cours de statistique) que s_n^2 converge p.s. vers σ^2 . Pour estimer l'erreur standard et l'erreur relative, on remplace donc σ^2 par s_n^2 :

erreur standard: $\sqrt{\frac{\sigma^2}{n}} \simeq \sqrt{\frac{s_n^2}{n}}$ erreur relative: $\frac{1}{I}\sqrt{\frac{\sigma^2}{n}} \simeq \frac{1}{\hat{I}_n}\sqrt{\frac{s_n^2}{n}}.$
--

♦ **En pratique:** Dans l'algorithme de Monte Carlo pour calculer \hat{I}_n , il sera judicieux d'ajouter le calcul de s_n^2 pour avoir en même temps une estimation de l'erreur.

2.2.2 Intervalle de confiance par TCL

On peut maintenant donner des intervalles de confiance pour notre estimation: dans l'utilisation du théorème central limite, on remplace la variance σ^2 par son estimation s_n^2 :

$$\mathbb{P}(|\hat{I}_n - I| \leq \varepsilon) = \mathbb{P}\left(\sqrt{\frac{n}{s_n^2}}|\hat{I}_n - I| \leq \sqrt{\frac{n}{s_n^2}}\varepsilon\right) \simeq 2\Phi\left(\sqrt{\frac{n}{s_n^2}}\varepsilon\right) - 1$$

Pour un intervalle de confiance au niveau $\alpha = 95\%$, on prend

$$2\Phi\left(\sqrt{\frac{n}{s_n^2}}\varepsilon\right) - 1 = \alpha = 0,95 \quad \Leftrightarrow \quad \Phi\left(\sqrt{\frac{n}{s_n^2}}\varepsilon\right) = \frac{\alpha + 1}{2} = 0,975$$

et donc $\sqrt{\frac{n}{s_n^2}}\varepsilon = 1,96$, ou encore $\varepsilon = 1,96\sqrt{\frac{s_n^2}{n}}$.

L'intervalle de confiance pour I au niveau 95% est $\left[\hat{I}_n - 1,96\sqrt{\frac{s_n^2}{n}}, \hat{I}_n + 1,96\sqrt{\frac{s_n^2}{n}} \right].$ L'intervalle de confiance au niveau α est $\left[\hat{I}_n - \Phi^{-1}\left(\frac{\alpha + 1}{2}\right)\sqrt{\frac{s_n^2}{n}}, \hat{I}_n + \Phi^{-1}\left(\frac{\alpha + 1}{2}\right)\sqrt{\frac{s_n^2}{n}} \right].$
--

Pour un niveau de confiance α fixé, la largeur de l'intervalle de confiance décroît en $\sqrt{\sigma^2/n}$. On dit que la méthode de Monte Carlo a une vitesse de convergence en $1/\sqrt{n}$.

♦ **En pratique:** On peut aussi parfois majorer la variance, et utiliser l'inégalité de Tchebitcheff pour obtenir des intervalles de confiance.

2.3 Algorithme de Monte Carlo

On suppose qu'on a à notre disposition une procédure `simulf` dont des différents appels simulent une suite de vaidd de densité f . La fonction `cdfnor("X",0,1,p,1-p)` donne l'inverse de la fonction de répartition de la loi normale centrée réduite en p :

$$x = \text{cdfnor}("X", 0, 1, p, 1 - p) \Leftrightarrow \mathbb{P}(Z \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) dt = p,$$

où Z suit la loi normale centrée réduite.

Dans l'algorithme qui suit, on décide d'un niveau de confiance α (proche de 1), et de la largeur de l'intervalle de confiance 2δ au niveau de confiance α . Le nombre n de simulations nécessaires est décidé par un test, et est par conséquent aléatoire:

```
entrées
function [n,I,e] = montecarlo(alpha,delta)
// estimation de I par montecarlo
// alpha: niveau de confiance alpha, 2delta: largeur intervalle de confiance
// n: nombre de simulations effectuees, I: valeur estimee, e:erreur standart
p=(alpha+1)/2; z= cdfnor("X",0,1,p,1-p)};
x=simulf; y=g(x)/f(x);
n=1; S1=y; S2=y^2; V=(S2-S1^2/n)/(n-1);
while (z*sqrt(V/n)>delta);
    x=simulf; y=g(x)/f(x);
    n=n+1; S1=S1+y; S2=S2+y^2; V=(S2-S1^2/n)/(n-1);
end;
I=S1/n; e=sqrt(V);
```

Commentaires

on stocke $\sum_{i=1}^n Y_i$ dans $S1$, $\sum_{i=1}^n Y_i^2$ dans $S2$, l'estimateur de la variance s_n^2 dans V . Le test regarde si la demi-largeur de l'intervalle de confiance après n simulations, $z\sqrt{\frac{s_n^2}{n}}$ est supérieure à celle souhaitée, δ , auquel cas on refait une simulation.

Conclusions

1. L'algorithme est extrêmement facile à mettre en place.
2. Il ne suppose pas de régularité sur g , autre que l'intégrabilité.
3. L'erreur pour n simulations est de l'ordre de $\sqrt{\sigma^2/n}$, ce qui justifie l'appellation "méthode de vitesse $1/\sqrt{n}$ ":
 - Cette vitesse est indépendante de la dimension: c'est donc un avantage quand la dimension est grande, par contre en dimension petite, cette méthode est peu compétitive par rapport aux méthodes d'analyse numérique, en particulier quand la fonction à intégrer est plus régulière.

- On voit l'importance de la variance σ^2 . Une voie pour améliorer l'erreur sera d'essayer de réduire la variance.

4. C'est une méthode aléatoire: le critère de sortie de l'algorithme est aléatoire.

2.4 Choix d'une méthode de Monte Carlo

Supposons qu'on ait le choix entre deux méthodes de Monte Carlo MC1 et MC2,

	densité	var	coût par simulation	nombre de simulations	coût total	erreur std
MC1	f_1	σ_1^2	c_1	n_1	$n_1 c_1$	$\sqrt{\sigma_1^2/n_1}$
MC2	f_2	σ_2^2	c_2	n_2	$n_2 c_2$	$\sqrt{\sigma_2^2/n_2}$

laquelle doit-on choisir? On rappelle que $\sigma^2 = \text{var}_f \left(\frac{g(X)}{f(X)} \right) = \int_{R^d} \frac{g^2(x)}{f(x)} dx - \left(\int_{R^d} g(x) dx \right)^2$.

Avec erreurs standards égales: $\sqrt{\frac{\sigma_1^2}{n_1}} = \sqrt{\frac{\sigma_2^2}{n_2}} \Leftrightarrow \frac{\sigma_1^2}{n_1} = \frac{\sigma_2^2}{n_2}$.

Pour une même erreur, la meilleure méthode est alors celle qui coûte le moins cher en calcul: on compare donc les coûts totaux de simulation $n_i c_i$:

$$\frac{n_1 c_1}{n_2 c_2} = \frac{\sigma_1^2 c_1}{\sigma_2^2 c_2}.$$

Avec coûts de simulation égaux: $n_1 c_1 = n_2 c_2$.

Pour un même coût de simulation, la meilleure méthode est alors celle qui offre la plus petite erreur standard: on compare donc les erreurs standard, ou plutôt les erreurs standard au carré σ_i^2/n_i :

$$\frac{\sigma_1^2 n_2}{\sigma_2^2 n_1} = \frac{\sigma_1^2 c_1}{\sigma_2^2 c_2}.$$

Dans les deux cas, la meilleure méthode est celle qui minimise $c_i \sigma_i^2$.

On a vu que dans l'erreur $\sqrt{\sigma^2/n}$, le facteur $1/\sqrt{n}$ est intrinsèque, et que pour réduire l'erreur, on peut essayer de diminuer la variance σ^2 . Il faudra cependant veiller à ce que cette diminution de variance ne soit pas compensée par une augmentation exagérée des coûts de calcul.

2.5 Exemples

Nous allons essayer plusieurs versions de la méthode de Monte Carlo pour la calcul approché de π . Afin de pour les comparer, nous allons écrire à chaque fois la variance de l'estimation après n simulations sous la forme C/n et comparer les différents C .

(I) Calcul approché de π : Technique du hit or miss

On regarde le carré $[0, 1]^2$, et on trace le quart de disque D centré en 0 et de rayon 1. On remarque que l'aire du quart de disque vaut $\pi/4$. Soit $(X_i)_{i \in \mathbb{N}}$ des v.a. de loi uniforme sur $[0, 1]^2$, on pose

$$Y_i = \mathbf{1}_{X_i \in D}.$$

On a $\mathbb{E}Y_1 = \mathbb{P}(X_1 \in D) = \pi/4$, et donc $\frac{1}{n} \sum_{i=1}^n Y_i$ est un estimateur de $\pi/4$. La variance est ici

$$C_{(I)} = \text{var}Y_1 = \frac{\pi}{4}(1 - \frac{\pi}{4}) \simeq 0,168.$$

(II) Calcul approché de π : Moyenne empirique sur $[0, 1]$

L'équation du quart de cercle est donnée par $g(x) = \sqrt{1 - x^2}$. Soit $(X_i)_{i \in \mathbb{N}}$ des v.a. de loi uniforme sur $[0, 1]$, on pose

$$Y_i = g(X_i).$$

On a $\mathbb{E}Y_1 = \mathbb{P}(X_1 \in D) = \pi/4$, et $\frac{1}{n} \sum_{i=1}^n Y_i$ est un autre estimateur de $\pi/4$. La variance est ici

$$C_{(II)} = \text{var}Y_1 = \int_0^1 g^2(x)dx - (\pi/4)^2 = \int_0^1 (1 - x^2)dx - (\pi/4)^2 = \frac{2}{3} - (\frac{\pi}{4})^2 \simeq 0,050.$$

Comparaison

La variance est plus faible dans le second cas, la seconde méthode est donc meilleure. C'est encore vrai dans le cas général d'une fonction $g : [0, 1] \rightarrow [0, 1]$ générale: la technique de la moyenne empirique a une plus petite variance que la technique hit or miss (Cf Rubinstein paragraphe 4.2.3)

♣ **Exercice 2:** Implémenter ces deux méthodes. Donner des intervalles de confiance à l'aide du TCL.

♣ **Exercice 3:** On cherche une valeur approchée de $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$, pour $\alpha > 0$ fixé.

1) Donner un algorithme associé à la méthode de la moyenne. On pourra utiliser des v.a. exponentielles dont on optimisera éventuellement le paramètre.

2) On suppose $\alpha = \pi$.

a) Effectuer une majoration (éventuellement grossière) de la variance associée à la méthode de MC utilisée. (Dans 1), on a utilisé une moyenne empirique de v.a. i.i.d. La variance associée est la variance d'une de ces v.a.)

b) En déduire des intervalles de confiance à l'aide de l'inégalité de Bienayme-Tchebichev. Soit $\epsilon > 0$ fixé. Quel n faut-il choisir pour que la probabilité que l'erreur soit plus petite que 0.02 soit plus grande que 0,99 ?

c) Refaire le b) en utilisant cette fois le TCL.

2.6 Techniques de réduction de variance

But: On a vu que l'erreur standard est

$$\sqrt{\frac{\sigma^2}{n}} \quad \text{avec} \quad \sigma^2 = \text{var}_f \frac{g(X)}{f(X)} = \int_{\mathbb{R}^d} \frac{g^2(x)}{f(x)} dx - I^2.$$

Le facteur en $1/\sqrt{n}$ étant intrinsèque à la méthode de Monte Carlo, on cherche à réduire la variance σ^2 , pour diminuer l'erreur commise à nombre de simulation fixé.

2.6.1 Choix de la densité f : importance sampling ou échantillonnage préférentiel

Dans l'estimation précédente, on a la choix de la densité f . Si le support de g est un ensemble E de mesure de Lebesgue finie de \mathbb{R}^d , la première idée pour choisir f est de simplement prendre la loi uniforme sur E : c'est ce qu'on appelle la Méthode de Monte Carlo Standard. Ce n'est cependant pas nécessairement le meilleur choix possible.

On a vu que l'erreur était contrôlée par la variance

$$\text{var}(Y_1) = \sigma^2 = \int_{\mathbb{R}^d} \frac{g^2(x)}{f(x)} dx - I^2.$$

Théorème 2.6.1 *La variance minimale est égale à $(\int_{\mathbb{R}^d} |g(x)| dx)^2 - I^2$, et est atteinte pour*

$$f_0(x) = \frac{|g(x)|}{\int_{\mathbb{R}^d} |g(t)| dt}.$$

Démonstration: Pour montrer que f_0 réalise bien le minimum de la variance, il suffit de vérifier que pour toute densité f :

$$\int_{\mathbb{R}^d} \frac{g^2(x)}{f(x)} dx \geq \int_{\mathbb{R}^d} \frac{g^2(x)}{|g(x)|} dx \int_{\mathbb{R}^d} |g(t)| dt = \left(\int_{\mathbb{R}^d} |g(t)| dt \right)^2.$$

Mais, par l'inégalité de Cauchy-Schwarz,

$$\left(\int_{\mathbb{R}^d} |g(t)| dt \right)^2 = \left(\int_{\mathbb{R}^d} \frac{|g(t)|}{f(t)^{1/2}} f(t)^{1/2} dt \right)^2 \leq \int_{\mathbb{R}^d} \frac{|g(t)|^2}{f(t)} dt \int_{\mathbb{R}^d} f(t) dt = \int_{\mathbb{R}^d} \frac{g(t)^2}{f(t)} dt.$$

La valeur du minimum se vérifie facilement. ■

Remarquons que, dans le cas particulier où $g > 0$, la densité optimale est $f = g/I$, auquel cas la variance serait nulle. Ce choix est bien sûr d'intérêt purement théorique, puisque I est inconnu. Cependant...

Echantillonnage préférentiel: Afin de réduire la variance, on essaie d'adapter la densité f à la fonction à intégrer g : on a intérêt à prendre f grande dans les régions où $|g|$ est grande et petite dans les régions où $|g|$ est petite.

(III) Calcul approché de π : Echantillonnage préférentiel

On reprend la méthode de la moyenne empirique, en essayant de choisir une densité f meilleure que la densité de la loi uniforme sur $[0, 1]$: on peut prendre par exemple une fonction affine décroissante sur $[0, 1]$, comme $f(x) = \frac{3}{2} - x$. Alors

$$\int_0^1 \frac{g^2(x)}{f(x)} dx = \int_0^1 \frac{1-x^2}{\frac{3}{2}-x} dx = \int_0^1 \left(x + \frac{3}{2} - \frac{5}{4} \frac{1}{\frac{3}{2}-x} \right) dx = \frac{1}{2} + \frac{3}{2} - \frac{5}{4} \ln(3) \simeq 0,6267.$$

et donc

$$C_{(III)} = \text{var}Y_1 = \int_0^1 \frac{g^2(x)}{f(x)} dx - I^2 = 0,0099.$$

2.6.2 Stratified sampling ou échantillonnage stratifié

L'idée est de découper le domaine d'intégration en un nombre fini m de parties D_1, D_2, \dots, D_m deux-à-deux disjointes:

$$I = \int_{\mathbb{R}^d} g(x) dx = \sum_{i=1}^m I_{D_i}, \text{ avec } I_{D_i} = \int_{D_i} g(x) dx,$$

et d'estimer alors séparément chacun des I_{D_i} .

Soit f une densité sur \mathbb{R}^d fixée. On va comparer les deux estimateurs suivants

- Estimateur de la moyenne empirique. Soient $(X_j)_{j \in \mathbb{N}}$ des vaids de densité f , l'estimateur de la moyenne empirique avec n estimations est comme précédemment

$$\hat{I}_n = \frac{1}{n} \sum_{j=1}^n \frac{g(X_j)}{f(X_j)}.$$

- Estimateur avec stratification. On note $p_i = \int_{D_i} f(x) dx$, et, sur D_i , on renormalise f pour obtenir une densité $f_i = \frac{f}{p_i} \mathbf{1}_{D_i}$. Pour chaque morceau D_i , on estime I_i avec n_i simulations: soient $(X_j^i)_{j \in \mathbb{N}}$ des vaids de densité f_i , l'estimateur correspondant est $\hat{I}_{n_i}^i = \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{g(X_j^i)}{f_i(X_j^i)}$. L'estimateur de I par échantillonnage stratifié avec $\sum_{i=1}^m n_i$ simulations s'obtient en sommant ces m estimations:

$$\hat{I}_n^{strat} = \sum_{i=1}^m \hat{I}_{n_i}^i = \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{g(X_j^i)}{f_i(X_j^i)}.$$

Afin de comparer ces deux estimations, on fait le même nombre de simulations pour les deux: $n = \sum_{i=1}^m n_i$, et on compare les variances.

Dans le premier cas, la variance pour n simulations est:

$$\begin{aligned}
V_1 &= \frac{1}{n} \text{var}_f \left(\frac{g(X)}{f(X)} \right) = \int_D \left(\frac{g(x)}{f(x)} - I \right)^2 f(x) dx \\
&= \frac{1}{n} \sum_{i=1}^m p_i \int_{D_i} \left(\frac{g(x)}{f(x)} - I \right)^2 f_i(x) dx = \frac{1}{n} \sum_{i=1}^m p_i \int_{D_i} \left(\frac{g(x)}{f(x)} - \frac{I_i}{p_i} + \frac{I_i}{p_i} - I \right)^2 f_i(x) dx \\
&= \frac{1}{n} \sum_{i=1}^m \frac{1}{p_i} \int_{D_i} \left(\frac{g(x)}{f_i(x)} - I_i \right)^2 f_i(x) dx + \frac{1}{n} \sum_{i=1}^m p_i \int_{D_i} \left(\frac{I_i}{p_i} - I \right)^2 f_i(x) dx \\
&= \frac{1}{n} \sum_{i=1}^m \frac{1}{p_i} \text{var}_{f_i} \left(\frac{g(X)}{f_i(X)} \right) + \frac{1}{n} \sum_{i=1}^m \frac{I_i^2}{p_i} - \frac{1}{n} I^2.
\end{aligned}$$

Pour l'estimateur de la moyenne empirique de I_i , la variance pour n_i simulations est $\frac{1}{n_i} \text{var}_{f_i} \left(\frac{g(X)}{f_i(X)} \right) = \frac{1}{n_i} \sigma_i^2$. Pour l'estimation stratifiée, avec n_i simulations pour chaque morceau I_i , la variance est, vu que les estimations des différents morceaux sont indépendantes,

$$V_2 = \sum_{i=1}^m \frac{1}{n_i} \text{var}_{f_i} \left(\frac{g(X)}{f_i(X)} \right) = \sum_{i=1}^m \frac{1}{n_i} \sigma_i^2,$$

à comparer à la précédente, quand $n = \sum_{i=1}^m n_i$. On minimise donc cette variance en les n_i , sous la contrainte $n = \sum_{i=1}^m n_i$. On trouve (extrema liés)

$$n_i = n \left(\frac{\sigma_i^2}{\sum_{j=1}^m \sigma_j^2} \right),$$

ce qui donne une variance minimale pour l'estimation stratifiée avec n simulations de

$$V_2 = \frac{m}{n} \sum_{j=1}^m \sigma_j^2.$$

Si on choisit maintenant les D_i de sorte que les $p_i = \int_{D_i} f$ soient tous égaux à $1/m$, il vient

$$V_1 = \frac{1}{n} \text{var}_f \left(\frac{g(X)}{f(X)} \right) = \frac{m}{n} \sum_{i=1}^m \sigma_j^2 + \frac{1}{n} \left(m \sum_{i=1}^m I_i^2 - I^2 \right) = V_2 + \frac{1}{n} \left(m \sum_{i=1}^m I_i^2 - I^2 \right).$$

Mais, sur \mathbb{R}^m , on a $\|x\|_1 \leq \sqrt{m} \|x\|_2$, et donc $m \sum_{i=1}^m I_i^2 - I^2 \geq 0$. La variance de l'estimateur avec échantillonnage stratifié est plus petite que celle de l'estimateur de départ.

Échantillonnage stratifié: On essaie de diminuer la variance pour une densité fixée f : on partitionne le support de g en domaines $(D_i)_{1 \leq i \leq m}$ de même masse pour f , puis on estime séparément l'intégrale de g sur chacun des D_i . Dans l'idéal, le nombre n_i de simulations pour le morceau D_i doit être proportionnel à $\sigma_i^2 = \text{var}_{f_i} \left(\frac{g(X)}{f_i(X)} \right)$: on pourra estimer grossièrement ces variances afin d'avoir une idée du choix des n_i .

L'idée est un peu la même pour l'échantillonnage préférentiel, mais pour une densité fixée: pour diminuer la variance, il faut faire davantage de simulations dans les domaines D_i où la variance $\sigma_i^2 = \text{var}_{f_i} \left(\frac{g(X)}{f_i(X)} \right)$ est grande.

(IV) Calcul approché de π : échantillonnage stratifié

On reprend pour f la densité de la loi uniforme sur $[0, 1]$, et $g(x) = \sqrt{1-x^2}$. On fait une stratification en $m = 2$ morceaux:

$$\begin{aligned} D_1 &= [0, 1/2] & D_2 &= [1/2, 1] \\ f_1 &= 2\mathbf{1}_{D_1} & f_2 &= 2\mathbf{1}_{D_2} \\ I_1 &= \int_{D_1} g = \frac{\pi}{12} + \frac{\sqrt{3}}{8} \simeq 0,4783 & I_2 &= \int_{D_2} g = \frac{\pi}{6} - \frac{\sqrt{3}}{8} \simeq 0,3071 \\ \int_{D_1} g^2 f_1 &= \frac{5}{6} & \int_{D_2} g^2 f_2 &= \frac{1}{2} \\ \sigma_1^2 &= \text{var}_{f_1}(g/f_1) = \int_{D_1} g^2 f_1 - I_1^2 \simeq 0,6046 & \sigma_2^2 &= \text{var}_{f_2}(g/f_2) = \int_{D_2} g^2 f_2 - I_2^2 \simeq 0,4057 \end{aligned}$$

Pour optimiser la stratification, on doit donc répartir les n simulations en

$$n_1 = n \left(\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \right) \simeq 0,6n \quad \text{et} \quad n_2 = n \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \right) \simeq 0,4n$$

et on obtient une réduction de variance (par rapport à (II)) de l'ordre de

$$V_1 - V_2 = \frac{1}{n} (2(I_1^2 + I_2^2) - I^2) \simeq 0,03 \frac{1}{n}, \text{ d'où } C_{(IV)} = 0,03.$$

2.6.3 Variable de contrôle

L'idée est la suivante: on veut estimer une intégrale $I = \mathbb{E}(Y)$, et on connaît une va C de moyenne μ et corrélée à Y , appelée variable de contrôle. Pour $\beta > 0$, on pose $Y(\beta) = Y - \beta(C - \mu)$, de sorte que $\mathbb{E}(Y(\beta)) = \mathbb{E}(Y) = I$. On estime alors I par la moyenne empirique de $Y(\beta)$, de même loi que Y , et la variance pour n simulations est alors $\frac{1}{n} \text{var}(Y(\beta))$, avec

$$\text{var}(Y(\beta)) = \text{var}(Y - \beta(C - \mu)) = \text{var}Y + \beta^2 \text{var}C - 2\beta \text{Cov}(Y, C),$$

qui peut être inférieure à $\text{var}Y$ si β est bien choisi.

On optimise en β , afin de réduire la variance au maximum:

$$\beta^* = \frac{\text{Cov}(Y, C)}{\text{var}C} \text{ et } \text{var}(Y(\beta^*)) = (1 - \rho_{Y,C}^2) \text{var}Y.$$

Donc plus C est corrélée à Y , plus la réduction de variance est importante.

Variable de contrôle: C , de moyenne μ connue, et corrélée à Y

$$Y(\beta) = Y - \beta(C - \mu), \quad \beta^* = \frac{\text{Cov}(Y, C)}{\text{var}C}, \quad \text{var}(Y(\beta^*)) = (1 - \rho_{Y,C}^2) \text{var}Y.$$

♦ **En pratique:** La méthode présente deux difficultés. Premièrement, on ne connaît pas forcément une variable de contrôle C , c'est-à-dire une va de moyenne connue et corrélée à Y . Deuxièmement, même si on connaît une variable de contrôle C , la valeur optimale β^* dépend de la covariance entre C et Y qui est souvent inconnue et par conséquent doit être estimée.

Remarque: Cette méthode se généralise au cas d'un vecteur de contrôle $C = (C_1, C_2, \dots, C_m)^t$ de moyenne $\mu \in \mathbb{R}^m$. Pour $\beta \in \mathbb{R}^m$, on pose $Y(\beta) = Y - \beta^t(C - \mu)$, et

$$\text{var}(Y(\beta)) = \text{var}Y + \beta^t \Gamma_c \beta - 2\text{Cov}(Y, C)^t \beta,$$

où Γ_c est la matrice de covariance de C , et $Cov(Y, C)$ est le vecteur dont les composantes sont les $Cov(Y, C_i)$. Le choix optimal pour β est alors:

$$\beta^* = \Gamma_c^{-1} Cov(Y, C) \text{ et } \text{var}(Y(\beta^*)) (1 - R_{Y,C}^2) \text{var} Y.$$

où $R_{Y,C} = \frac{Cov(Y, C)^t \Gamma_c^{-1} Cov(Y, C)}{\text{var} Y}$.

(V) Calcul approché de π : variable de contrôle

Reprendre la densité uniforme, $Y = \sqrt{1 - U^2}$ et comme variable de contrôle $C = (1 - U)^2$.

$$\mathbb{E}(Y) = \frac{\pi}{4}, \text{ var}(Y) = \frac{2}{3} - \left(\frac{\pi}{4}\right)^2, \mathbb{E}(C) = \frac{1}{3}, \text{ var}(C) = \frac{4}{45}, \rho_{Y,C} = 0,8.$$

Si on prend $\beta^* = \frac{Cov(C, Y)}{\text{var}(C)} \simeq 9$, on trouve $\text{var}(Y(\beta^*)) \simeq 0,2 \text{var}(Y)$, d'où

$$C_{(V)} \simeq 0,01.$$

2.6.4 Variables antithétiques

On suppose que $I = \mathbb{E}Y = \mathbb{E}Z$ est l'intégrale à estimer.

$$I = \mathbb{E}\left(\frac{1}{2}(Y + Z)\right) \quad \text{et} \quad \text{var}\left(\frac{1}{2}(Y + Z)\right) = \frac{1}{4}\text{var}Y + \frac{1}{4}\text{var}Z + \frac{1}{2}Cov(Y, Z).$$

Si Y et Z sont négativement corrélées, alors on peut espérer faire diminuer la variance.

Supposons par exemple que l'on doive estimer $I = \int_0^1 g(x)dx$. Soit $(U_i)_{i \in \mathbb{N}}$ une suite de vaids de loi uniforme sur $[0, 1]$. L'estimateur de la moyenne empirique $\frac{1}{n} \sum_{i=1}^n g(U_i)$ pour n simulations a pour variance $\frac{1}{n} V_1$ avec

$$V_1 = \text{var}(g(U)).$$

Comme $x \rightarrow 1 - x$ laisse la mesure dx invariante sur $[0, 1]$, on peut prendre comme variables antithétiques $Y = g(U)$ et $Z = g(1 - U)$ avec U de loi uniforme sur $[0, 1]$. L'estimateur pour n simulations s'écrit:

$$\hat{I}_n = \frac{1}{2n} \sum_{i=1}^n (g(U_i) + g(1 - U_i)),$$

et a une variance $\frac{1}{n} V_2$, avec

$$V_2 = \frac{1}{2}(\text{var}(g(U)) + Cov(g(U), g(1 - U))).$$

Mais attention, si le nombre de simulations est le même, le temps de calcul est deux fois plus important dans le second cas, donc pour que la méthode des variables antithétiques soit intéressante, il faut que $V_2 < \frac{1}{2} V_1$.

Proposition 2.6.2 Si g est de classe C^1 monotone avec $g(0) \neq g(1)$, alors $V_2 < \frac{1}{2} V_1$.

Démonstration: On suppose g croissante, et donc $g(1) > g(0)$. On doit montrer que $Cov(g(U), g(1 - U)) < 0$, autrement dit que

$$\int_0^1 g(u)g(1 - u)du < I^2.$$

On pose $\phi(x) = \int_0^x g(1 - t)dt - xI$. Alors $\phi(0) = \phi(1) = 0$ et

$$\phi'(x) = g(1 - x) - I$$

est une fonction décroissante. Le théorème de la moyenne assure que $\phi'(0) > 0$ et $\phi'(1) < 0$, et donc que $\phi(x) > 0$ pour tout $x \in]0, 1[$, donc

$$0 < \int_0^1 \phi(x)g'(x)dx = - \int_0^1 \phi'(x)g(x)dx = - \int_0^1 g(u)g(1 - u)du + I^2.$$

Le cas décroissant se traite de la même manière. ■

Variables antithétiques: pour une fonction g monotone sur $[0, 1]$, si $(U_i)_{i \in \mathbb{N}}$ est une suite de vauid de loi uniforme sur $[0, 1]$, l'estimateur avec variables antithétiques

$$\hat{I}_n = \frac{1}{2n} \sum_{i=1}^n (g(U_i) + g(1 - U_i))$$

est plus efficace que l'estimateur $\frac{1}{n} \sum_{i=1}^n g(U_i)$.

Calcul de π , (VI)

Appliquer cette méthode au calcul de π .