

# Méthodes de réduction de la variance

Prof. Mohamed El Merouani

Université Abdelmalek Essaâdi  
Faculté des Sciences de Tétouan  
Département de Mathématiques

2018/2019

- Dans la séance précédente, on voulait estimer  $I = E(h(X))$  par la méthode de Monte-Carlo.
- On aboutit à un estimateur  $\hat{I}_n$  avec une erreur en  $\sigma/\sqrt{n}$ .
- Cette erreur d'approximation se réduit au fur et à mesure que le nombre d'observations  $n$  augmente, mais cette réduction est assez lente.
- Cependant, comme cette erreur dépend aussi positivement de la variance  $\sigma^2$ , une idée est de chercher à réduire cette dernière pour améliorer l'approximation de la méthode de Monte-Carlo.

Une autre raison pour laquelle on optera pour réduire la variance est :

- On veut estimer  $I = E(h(X)) = \int h(x)f(x)dx$ , et si la fonction  $h$  prend ses plus grandes valeurs là où la densité  $f$  est très faible, i.e. là où  $X$  a très peu de chances de tomber, et réciproquement, l'estimateur Monte-Carlo Classique  $\hat{I}_n = \frac{1}{n} \sum_{i=1}^n h(X_i)$  est très mauvais puisqu'à moins de prendre  $n$  très grand, il va estimer environ 0 même si  $I$  vaut 1.

## 1- Événements rares :

La variable  $X$  suivant une loi normale centrée réduite, on veut estimer la probabilité

$$P(X > 6) = E [\mathbb{I}_{X>6}] = \int_{\mathbb{R}} \mathbb{I}_{X>6} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

La commande `1-pnorm(6)` montre qu'elle est de l'ordre de  $10^{-9}$ . Ceci signifie qu'à moins de prendre  $n$  de l'ordre d'au moins un milliard, on a toutes les chances d'obtenir  $\hat{I}_n = 0$

## Exemple 2 :

Soit  $m$  un réel,  $X \sim \mathcal{N}(m, 1)$  et  $h(x) = \exp\left(-mx + \frac{m^2}{2}\right)$ . Pour tout  $m$ , on a donc  $I = E(h(X)) = \int_{\mathbb{R}} h(x)f(x)dx = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1$  Or 95% des  $X_i$  tombent dans l'intervalle  $[m - 2, m + 2]$  tandis que  $h(x) \exp\left(-\frac{m^2}{2}\right)$  tend très vite vers 0 quand  $m$  augmente. Ainsi, dès lors que  $m$  est grand, on obtient  $\hat{I}_n$  proche de 0, ce qui n'est pas une très bonne approximation de la valeur cherchée  $I = 1$

Plusieurs Méthodes de réduction de la variance ont été proposées. En fait, il s'agit de trouver une autre approximation  $\tilde{I}_n$  de  $I$  ayant une variance plus petite que celle de  $\hat{I}_n$ . Parmi ces Méthodes on trouve :

- Méthode de l'échantillonnage préférentiel.
- Méthode de la variable de contrôle.
- Méthode de la variable antithétique.
- Méthode de stratification.
- Méthode Quasi-Monte-Carlo.
- Valeur moyenne ou conditionnement

# Méthode de l'échantillonnage préférentiel

L'idée de l'échantillonnage préférentiel, ou échantillonnage pondéré, ou "importance sampling", est de tirer des points non pas suivant la densité  $f$  de  $X$  mais selon une densité auxiliaire  $\tilde{f}$  réalisant un compromis entre les régions de l'espace où  $h$  est grande et où la densité  $f$  est élevée, quitte à rétablir le tir ensuite en tenant compte du fait que la loi de simulation  $\tilde{f}$  n'est pas la loi initiale  $f$ .

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

$$\begin{aligned} E(h(X)) &= \int_{\mathbb{R}^d} h(x) f(x) dx = \int_{\mathbb{R}^d} \frac{h(x) f(x)}{\tilde{f}(x)} \tilde{f}(x) dx \\ &= E \left( \frac{h(Y) f(Y)}{\tilde{f}(Y)} \right) \end{aligned}$$

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

$$E(h(X)) \approx \frac{1}{n} \sum_{i=1}^n h(X_i)$$

$$E(h(X)) \approx \frac{1}{n} \sum_{i=1}^n \frac{h(Y_i)f(Y_i)}{\tilde{f}(Y_i)}$$

avec des  $X_i, i = 1, \dots, n$  i.i.d. de même loi que  $X$  et des  $Y_i, i = 1, \dots, n$  de même loi que  $Y$ .

La deuxième méthode est plus intéressante que la première si

$$\text{Var} \left( \frac{h(Y)f(Y)}{\tilde{f}(Y)} \right) \leq \text{Var}(h(X))$$

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

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

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

# Méthode de l'échantillonnage préférentiel

Mais ceci conduit à l'idée d'approcher  $\tilde{f}(x) = \frac{h(x)f(x)}{E(h(x))}$  par

$$\frac{\text{Approx}(h(x)f(x))}{\int \text{Approx}(h(x)f(x))dx}$$

## Exemple :

Soit à estimer ou évaluer  $I = \int_0^1 \cos \frac{\pi x}{2} dx$

On note que  $I = E(\cos(\frac{\pi X}{2}))$  où  $X \sim \mathcal{U}(0, 1)$ .

L'approximation initiale de  $I$  par la méthode de Monte-Carlo est donc définie par :

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^n \cos \left( \frac{\pi X_i}{2} \right) \quad (1)$$

où les  $X_i$  sont des valeurs simulées d'une variable  $X \sim \mathcal{U}(0, 1)$ .

# Méthode de l'échantillonnage préférentiel

Exemple :

Afin d'améliorer cette approximation avec la méthode de l'échantillonnage préférentiel, on peut approcher  $\cos(\frac{\pi x}{2})$  par son développement limité au voisinage de 0, soit :

$$|h(x)f(x)| = \cos\left(\frac{\pi x}{2}\right) \approx 1 - \frac{\pi^2 x^2}{8} \approx 1 - x^2$$

et prendre

$$\tilde{f}(x) = \frac{h(X)f(X)}{E(h(X))} = \frac{1 - x^2}{\int_0^1 (1 - x^2) dx} = \frac{3}{2}(1 - x^2)$$

Une approximation améliorée de  $I$  est alors donnée par :

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

où les  $Y_i$  sont des valeurs simulées d'une variable  $Y$  selon la densité  $\tilde{f}$  et

$$\frac{h(x)f(x)}{\tilde{f}(x)} = \frac{\cos(\frac{\pi x}{2})}{\frac{3}{2}(1-x^2)}.$$

Nous récapitulons par cette démarche à suivre :

- 1 Approcher  $|h(x)f(x)|$  par une autre fonction  $\tilde{f}_1$  telle que l'on sache simuler suivant la densité

$$\tilde{f} = \frac{\tilde{f}_1}{\int \tilde{f}_1(x)dx}$$

- 2 Soit  $Y$  une v.a. de densité  $\tilde{f}$ .  
Comparer  $Var(h(X))$  et  $Var\left(\frac{h(Y)f(Y)}{\tilde{f}(Y)}\right)$ .

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

# Méthode de l'échantillonnage préférentiel

Exemple :

Dans l'exemple précédent on pourrait aussi faire :

$$|h(x)f(x)| \approx \tilde{f}_1 = \cos\left(\frac{\pi x}{2}\right) = 1 - x$$

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

La fonction  $\tilde{f}_1$  est (grossièrement) proche de  $h(x)f(x)$ .

Si  $U \sim \mathcal{U}([0, 1])$ , alors  $1 - \sqrt{U}$  est de loi de densité  $\tilde{f}$  (nous trouvons ce résultat par la méthode d'inversion de la fonction de répartition).

Donc, nous savons simuler la loi de densité  $\tilde{f}$ .

# Méthode de l'échantillonnage préférentiel

Exemple : "Comparaison des variances sous R"

Dans le programme suivant, nous estimons les variances des deux méthodes. Nous trouvons  $\approx 1.10^{-1}$  pour la méthode de l'équation (1) de l'approximation initiale par la méthode de Monte-Carlo, et  $7.10^{-3}$  pour la méthode d'échantillonnage préférentiel par  $\tilde{f}$ . La variance est donc réduite.

```
simu<-function(t)
{
u=runif(1,0,1)
return(1-sqrt(u))
}
s=0
for (i in 1:n)
{
u=runif(1,0,1)
```

# Méthode de l'échantillonnage préférentiel

Exemple : "Comparaison des variances sous R"

```
v=runif(1,0,1)
u1=cos(pi*u/2)
v1=cos(pi*v/2)
s=s+(u1-v1)^2
}
s=s/(2*n)
cat("Variance de la première méthode:",s)
s=0
for(i in 1:n)
{
v=simu(1)
v1=cos(pi*v/2)/(2*(1-v))
w=simu(1)
```

# Méthode de l'échantillonnage préférentiel

Exemple : "Comparaison des variances sous R"

```
w1=cos(pi*w/2)/(2*(1-w))  
s=s+(v1-w1)^2  
}  
s=s/(2*n)  
cat("Variance de la méthode d'échantillonnage  
préférentiel:",s)
```

# Méthode de la variable de contrôle

Supposons, maintenant, que nous voulons calculer  $E(f(X))$  avec  $X$  une variable aléatoire. La première méthode consiste à faire l'approximation

$$E(f(X)) \approx \frac{1}{n} \sum_{i=1}^n f(X_i)$$

avec les  $X_i, i = 1, \dots, n$  i.i.d. de même loi que  $X$  (toujours pour "n grand"). Si on sait calculer  $E(h(X))$  pour une certaine fonction  $h$ , alors on peut aussi faire l'approximation

$$E(f(X)) \approx \frac{1}{n} \sum_{i=1}^n (f(X_i) - h(X_i)) + E(h(X))$$

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

$$\begin{aligned} E(f(X)) - \left( \frac{1}{n} \sum_{i=1}^n (f(X_i) - h(X_i)) + E(h(X)) \right) &= \\ &= E(f(X) - h(X)) - \frac{1}{n} \sum_{i=1}^n (f(X_i) - h(X_i)) \end{aligned}$$

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

La méthode est donc la suivante :

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

## Exemple :

On cherche à calculer  $I = \int_0^1 \exp(x^2) dx$  par une méthode de Monte-Carlo.

Nous avons  $I = E(e^{X^2})$  avec  $X \sim \mathcal{U}([0, 1])$ . Donc nous pouvons faire l'approximation  $I \approx \frac{1}{n} \sum_{i=1}^n e^{X_i^2}$  avec  $X_1, X_2, \dots, X_n$  i.i.d.  $\sim \mathcal{U}([0, 1])$  (toujours pour "n grand"). Nous avons donc une première méthode de Monte-Carlo. Avec les notations ci-dessus, nous avons  $f(x) = \exp(x^2)$ . Remarquons que  $h \mapsto 1 + x^2$  est proche de  $f$  sur  $[0, 1]$  (c'est le début du développement limité de  $f$  en 0). Nous savons calculer

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

Nous pouvons faire l'approximation

$$I \approx \frac{1}{n} \sum_{i=1}^n \left( e^{X_i^2} - 1 - X_i^2 \right) + E(h(X))$$

# Méthode de la variable de contrôle

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

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

# Méthode de la variable de contrôle

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

## Méthode de la variable antithétique :

Supposons que l'on cherche à calculer  $I = \int_{\Delta} h(x)f(x)dx$  où  $f$  est une densité de probabilité d'une variable  $X$  ayant pour support  $\Delta \subset \mathbb{R}^d$ , donc  $I = E(h(X))$ . On sait qu'une approximation initiale de  $I$  par la méthode de Monte-Carlo est donnée par  $\hat{I} = \frac{1}{n} \sum_{i=1}^n h(X_i)$  où les  $X_i$  sont des valeurs simulées d'une variable  $X$  selon la densité  $f$ . Soit  $Y = \varphi(X)$  une transformation bijective définie sur  $\Delta$  (telle que  $|\det(\text{Jac } \varphi)(x)| = 1$ ) et continue préservant la loi de  $X$ . Autrement dit,  $X$  et  $Y$  ont la même loi de probabilité. On peut, alors, écrire :

$$I = \int_{\Delta} h(\varphi^{-1}(y))f(\varphi^{-1}(y)) \left( \frac{d\varphi^{-1}(y)}{dy} \right) dy$$

soit,

$$I = \int_{\Delta} h(\varphi^{-1}(y))g(y)dy$$

où  $g$  est la densité de probabilité de  $Y$ . Mais, comme la transformation  $\varphi$  préserve la loi de probabilité  $g = f$  sur  $\Delta$  et donc :

$$I = \int_{\Delta} h(\varphi^{-1}(y))f(y)dy$$

ou encore

$$I = \frac{1}{2} \int_{\Delta} (h(\varphi^{-1}(y)) + h(y)) f(y)dy = \frac{1}{2} E (h(X) + h(\varphi^{-1}(X)))$$

où  $X$  est une variable aléatoire ayant  $f$  pour densité de probabilité. À la place de  $I$  on peut alors proposer  $\tilde{I} = \frac{1}{2n} \sum_{i=1}^n (h(X_i) + h(\varphi^{-1}(X_i)))$  où les  $X_i$  sont des valeurs simulées d'une variable  $X$  selon  $f$ .

## Cas particulier :

Soit à calculer  $I = \int_0^1 h(x) dx$

On note que  $I = E(h(X))$  où  $X \sim \mathcal{U}([0, 1])$ .

L'approximation initiale de  $I$  est alors donnée par :

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

où les  $X_i$  sont des valeurs simulées d'une variable  $X \sim \mathcal{U}([0, 1])$ . On note qu'on peut aussi écrire :

$$I = \int_0^1 h(1-x) dx$$

ou encore :

$$I = \frac{1}{2} \int_0^1 (h(x) - h(1-x)) dx$$

# Méthode de la variable antithétique :

Cas particulier :

D'où une autre approximation de  $I$  par la méthode de Monte-Carlo :

$$\tilde{I} = \frac{1}{2n} \sum_{i=1}^n (h(X_i) + h(1 - X_i))$$

où les  $X_i$  sont des valeurs simulées d'une variable  $X \sim \mathcal{U}([0, 1])$ .

On peut alors montrer que  $Var(\tilde{I}) \leq Var(\hat{I})$ .

**En effet :**

$$Var(\tilde{I}) = \frac{1}{4n} (Var(h(X)) + Var(h(1 - X))) + 2Cov(h(X), h(1 - X))$$

Puisque  $X \sim \mathcal{U}([0, 1])$ , la variable  $(1 - X)$  suit également la loi uniforme continue sur  $[0, 1]$ . Les variables  $h(X)$  et  $h(1 - X)$  suivent alors la même loi et donc :

$$Var(h(X)) = Var(h(1 - X))$$

# Méthode de la variable antithétique :

Cas particulier :

D'autre part, on sait que :

$$\text{Cov}(h(X), h(1 - X)) \leq \sqrt{\text{Var}(h(X))\text{Var}(h(1 - X))} = \text{Var}(h(X))$$

(Inégalité de Hölder)

En conséquence,

$$\text{Var}(\tilde{I}) \leq \frac{1}{4n} (4\text{Var}(h(X))) = \text{Var}(\hat{I})$$

L'approximation  $\tilde{I}$  a une variance plus faible que celle de  $\hat{I}$ .

## Exercice 1 :

Appliquer différentes méthodes de réduction de la variance pour estimer  $P(X > 3)$ , où  $X$  est une variable aléatoire de loi  $\mathcal{N}(0, 1)$ .

## Exercice 2 :

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.