Echantillonnage préférentiel

Pierre Gloaguen

02/04/2020

Annonces:

- Premier rendu pour le 13 Avril (exo5 du TD1)
- À faire en binome (à compléter sur le lien envoyé par mail)
- Corrections des exos 1 à 4 du TD1 mis en ligne
- On regardera ces corrections et vos questions aujourd'hui

Rappel de l'épisode précédent

- Principes des méthodes de Monte Carlo;
- Approximation d'intégrales (espérances) par simulation;
- Fonctionne grâce à la loi des grands nombres, IC grâce au TCL.

Objectif du cours

- Présentation de l'échantillonnage préférentiel
  - Extension naturelle des méthodes de MC simples;
  - Améliore l'efficacité dans certains cas;
  - Utile quand on ne sait pas simuler selon une loi donnée;

Objectif du cours

- Présentation de l'échantillonnage préférentiel
  - Extension naturelle des méthodes de MC simples;
  - Améliore l'efficacité dans certains cas;
  - Utile quand on ne sait pas simuler selon une loi donnée;
- Motivation: cas de Monte Carlo problématiques;
- Définition;
- Propriétés: Analyse de la variance de ce nouvel estimateur;
- Illustration.

Echantillonnage préférentiel

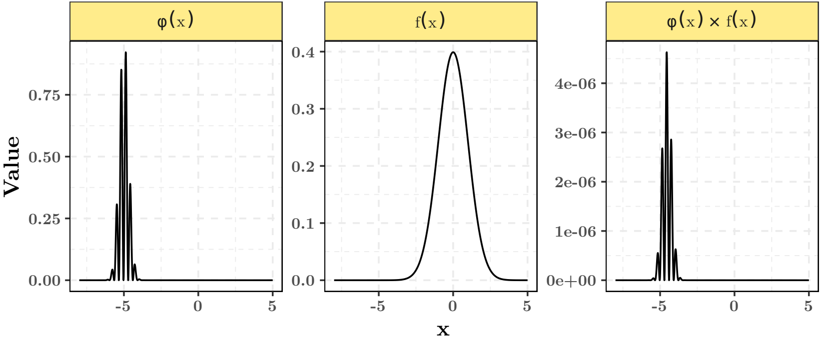
Exemple problématique:

On veut calculer

$$I = \mathbb{E}\left[\varphi(X)\right] = \int_{\mathbb{R}} \varphi(x) f(x) \mathrm{d}x$$

où $X \sim \mathcal{N}(0, 1)$ (de densité $f(x)$) et

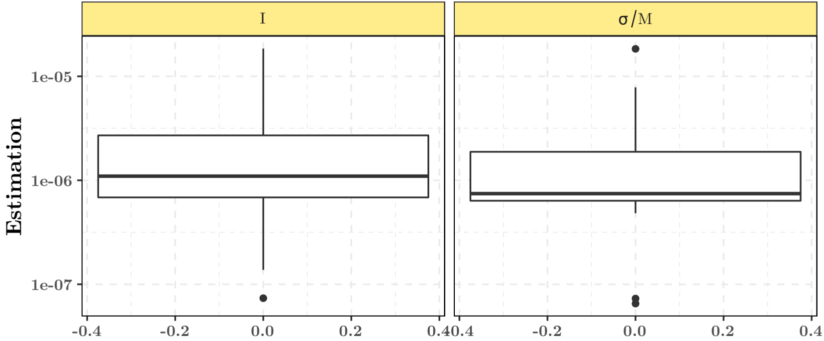
$$\varphi(x) = \sin^2(x) e^{-5(x-5)^2}$$



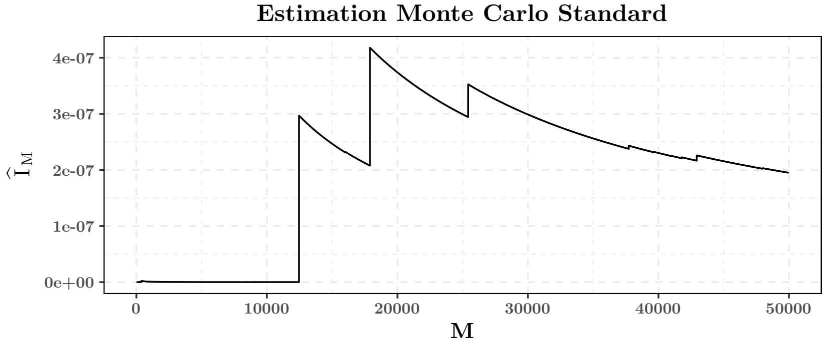
Estimateur Monte Carlo naturel

On tire $M = 50000$ points dans une loi $\mathcal{N}(0,1)$, et on calcule la moyenne empirique.

Résultats obtenus:



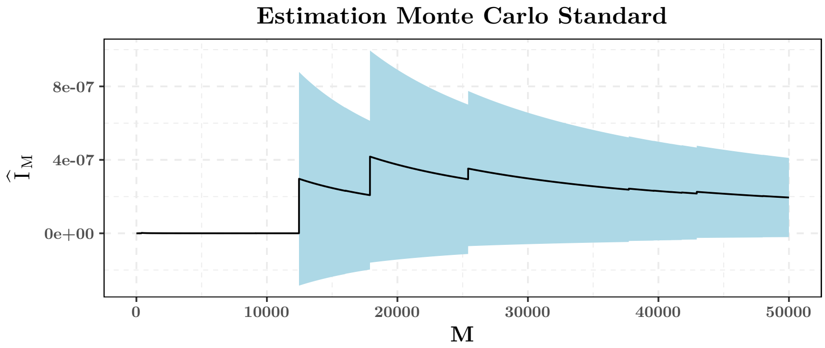
Une trajectoire d'estimation



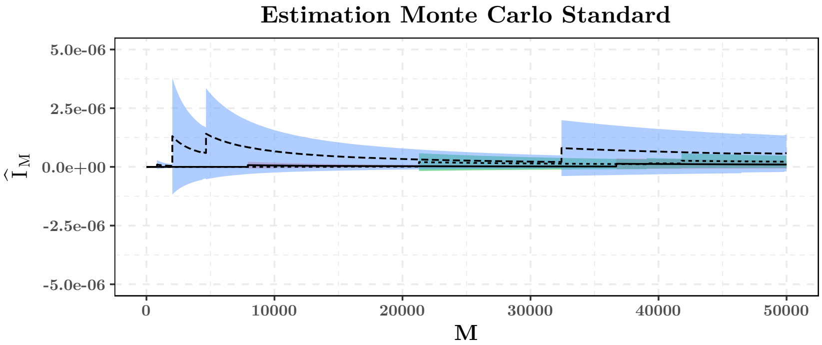
L'estimation est très instable.

Estimateur Monte Carlo naturel

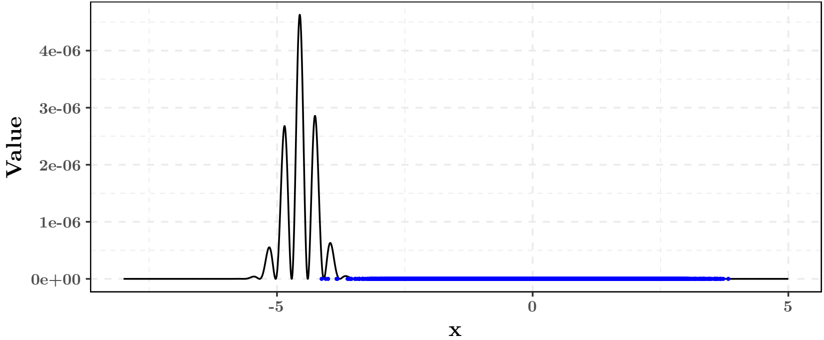
Estimation de la variance et intervalle de confiance asymptotique:



Manque de chance?



Origine du problème



On échantillonne loin des régions importantes!

Echantillonnage préférentiel

On cherche à estimer une intégrale du type:

$$I = \mathbb{E}_f[\varphi(X)] = \int_{\mathcal{D}_f} \varphi(x) f(x) \mathrm{d}x$$

où $\mathcal{D}_f \subset \mathbb{R}^d$, et f est une densité de probabilité sur $\mathcal{D}_f$ (donc $f(x) = 0$ pour $x \notin \mathcal{D}_f$) et X une v.a. de loi f.

Echantillonnage préférentiel

On cherche à estimer une intégrale du type:

$$I = \mathbb{E}_f[\varphi(X)] = \int_{\mathcal{D}_f} \varphi(x) f(x) \mathrm{d}x$$

où $\mathcal{D}_f \subset \mathbb{R}^d$, et f est une densité de probabilité sur $\mathcal{D}_f$ (donc $f(x) = 0$ pour $x \notin \mathcal{D}_f$) et X une v.a. de loi f.

Soit g une densité de probabilité sur $\mathcal{D}_g \supseteq \mathcal{D}_f$ telle que $x \in \mathcal{D}_f \Rightarrow g(x) > 0$ et Y une variable aléatoire de loi g, alors:

$$I = \int_{\mathcal{D}_f} \varphi(x) \frac{f(x)}{g(x)} g(x) \mathrm{d}x$$

Echantillonnage préférentiel

On cherche à estimer une intégrale du type:

$$I = \mathbb{E}_f[\varphi(X)] = \int_{\mathcal{D}_f} \varphi(x) f(x) \mathrm{d}x$$

où $\mathcal{D}_f \subset \mathbb{R}^d$, et f est une densité de probabilité sur $\mathcal{D}_f$ (donc $f(x) = 0$ pour $x \notin \mathcal{D}_f$) et X une v.a. de loi f.

Soit g une densité de probabilité sur $\mathcal{D}_g \supseteq \mathcal{D}_f$ telle que $x \in \mathcal{D}_f \Rightarrow g(x) > 0$ et Y une variable aléatoire de loi g, alors:

$$\begin{aligned} I &= \int_{\mathcal{D}_f} \varphi(x) \frac{f(x)}{g(x)} g(x) \mathrm{d}x \\ &= \int_{\mathcal{D}_g} \varphi(x) \frac{f(x)}{g(x)} g(x) \mathrm{d}x \qquad \text{as } x \notin \mathcal{D}_f \Rightarrow f(x) = 0 \end{aligned}$$

Echantillonnage préférentiel

On cherche à estimer une intégrale du type:

$$I = \mathbb{E}_f[\varphi(X)] = \int_{\mathcal{D}_f} \varphi(x) f(x) \mathrm{d}x$$

où $\mathcal{D}_f \subset \mathbb{R}^d$, et f est une densité de probabilité sur $\mathcal{D}_f$ (donc $f(x) = 0$ pour $x \notin \mathcal{D}_f$) et X une v.a. de loi f.

Soit g une densité de probabilité sur $\mathcal{D}_g \supseteq \mathcal{D}_f$ telle que $x \in \mathcal{D}_f \Rightarrow g(x) > 0$ et Y une variable aléatoire de loi g, alors:

$$\begin{aligned}
I &= \int_{\mathcal{D}_f} \varphi(x) \frac{f(x)}{g(x)} g(x) \mathrm{d}x \\
&= \int_{\mathcal{D}_g} \varphi(x) \frac{f(x)}{g(x)} g(x) \mathrm{d}x \qquad \text{as } x \notin \mathcal{D}_f \Rightarrow f(x) = 0 \\
&= \mathbb{E}_g\left[\varphi(Y) \frac{f(Y)}{g(Y)}\right]
\end{aligned}$$

Echantillonnage preferentiel

Comme estimateur de I, on peut ainsi proposer l'estimateur:

$$\hat{I}_M^{IS} = \frac{1}{M}\sum_{k=1}^{M}\frac{f(Y_k)}{g(Y_k)}\varphi(Y_k) = \frac{1}{M}\sum_{k=1}^{M} W(Y_k)\varphi(Y_k)$$

où $Y_1, \ldots, Y_M$ est un échantillon i.i.d. de variables aléatoires sur $\mathbb{R}^d$ de densité g.

Echantillonnage preferentiel

Comme estimateur de I, on peut ainsi proposer l'estimateur:

$$\hat{I}_M^{IS} = \frac{1}{M}\sum_{k=1}^{M}\frac{f(Y_k)}{g(Y_k)}\varphi(Y_k) = \frac{1}{M}\sum_{k=1}^{M} W(Y_k)\varphi(Y_k)$$

où $Y_1, \ldots, Y_M$ est un échantillon i.i.d. de variables aléatoires sur $\mathbb{R}^d$ de densité g.

Remarques:

- Comme $Y_k \sim g$, $g(Y_k)$ est p.s. $\neq 0$

Echantillonnage preferentiel

Comme estimateur de I, on peut ainsi proposer l'estimateur:

$$\hat{I}_M^{IS} = \frac{1}{M}\sum_{k=1}^{M} \frac{f(Y_k)}{g(Y_k)}\varphi(Y_k) = \frac{1}{M}\sum_{k=1}^{M} W(Y_k)\varphi(Y_k)$$

où $Y_1, \ldots, Y_M$ est un échantillon i.i.d. de variables aléatoires sur $\mathbb{R}^d$ de densité g.

Remarques:

- Comme $Y_k \sim g$, $g(Y_k)$ est p.s. $\neq 0$
- La variable aléatoire $W(Y_k) = \frac{f(Y_k)}{g(Y_k)}$ est appelée **poids d'importance** de Y_k.

Echantillonnage preferentiel

Comme estimateur de I, on peut ainsi proposer l'estimateur:

$$\hat{I}_M^{IS} = \frac{1}{M}\sum_{k=1}^{M}\frac{f(Y_k)}{g(Y_k)}\varphi(Y_k) = \frac{1}{M}\sum_{k=1}^{M} W(Y_k)\varphi(Y_k)$$

où $Y_1, \ldots, Y_M$ est un échantillon i.i.d. de variables aléatoires sur $\mathbb{R}^d$ de densité g.

Remarques:

- Comme $Y_k \sim g$, $g(Y_k)$ est p.s. $\neq 0$
- La variable aléatoire $W(Y_k) = \frac{f(Y_k)}{g(Y_k)}$ est appelée **poids d'importance** de Y_k.
- Quand $f = g$, on a l'estimateur MC standard (et chaque poids vaut 1).

Quel intérêt?

On peut choisir g afin d'échantillonner dans les zones d'importance!

Biais:

$$\begin{aligned}
\mathbb{E}_g[\hat{I}_M^{IS}] &= \frac{1}{M}\sum_{k=1}^{M}\int_{\mathcal{D}_g}\frac{f(x)}{g(x)}\varphi(x)g(x)\mathrm{d}x \\
&= \int_{\mathcal{D}_f} f(x)\varphi(x)\mathrm{d}x = I
\end{aligned}$$

Donc, cet estimateur reste sans biais

Quel intérêt?

On peut choisir g afin d'échantillonner dans les zones d'importance!

Biais:

$$\mathbb{E}_g[\hat{I}_M^{IS}] = \frac{1}{M}\sum_{k=1}^{M}\int_{\mathcal{D}_g}\frac{f(x)}{g(x)}\varphi(x)g(x)\mathrm{d}x$$
$$= \int_{\mathcal{D}_f} f(x)\varphi(x)\mathrm{d}x = I$$

Donc, cet estimateur reste sans biais

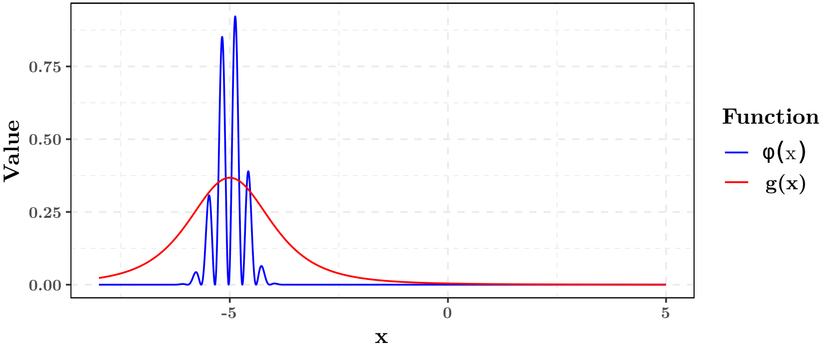
Variance:

$$\mathbb{V}_g[\hat{I}_1^{IS}] = \left[\mathbb{E}_g[(W(Y)\varphi(Y))^2] - I^2\right] = \int_{\mathcal{D}_g}\frac{(\varphi(y)f(y) - Ig(y))^2}{g(y)}\mathrm{d}y$$

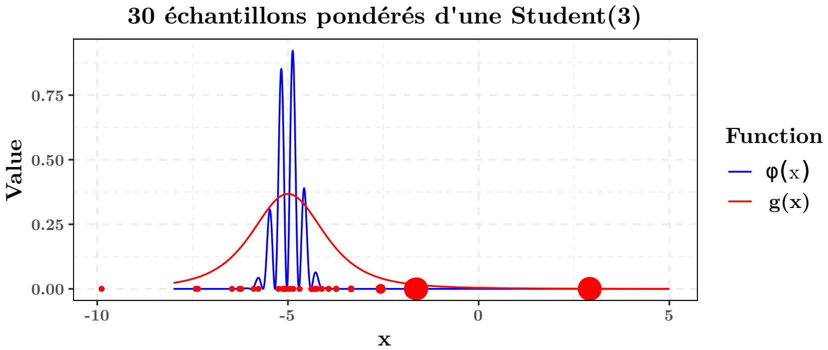
- La variance peut être très réduite si $g(y) \overset{\approx}{\propto} \varphi(y)f(y)$!
- Ceci peut guider le choix de g!

Exemple

$g(x)$ est la densité d'une loi de Student $\mathcal{T}(3)$.

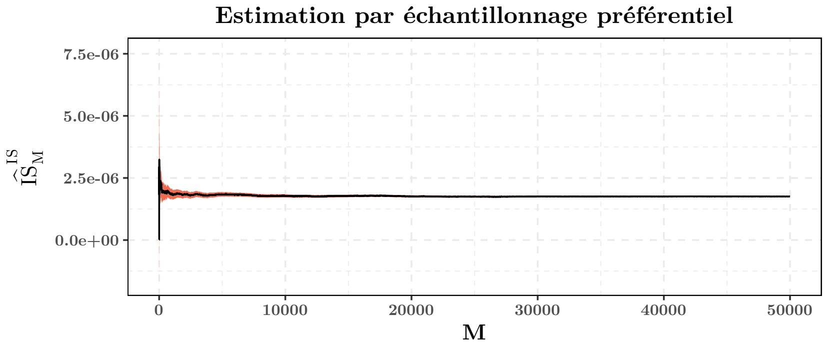


Exemple

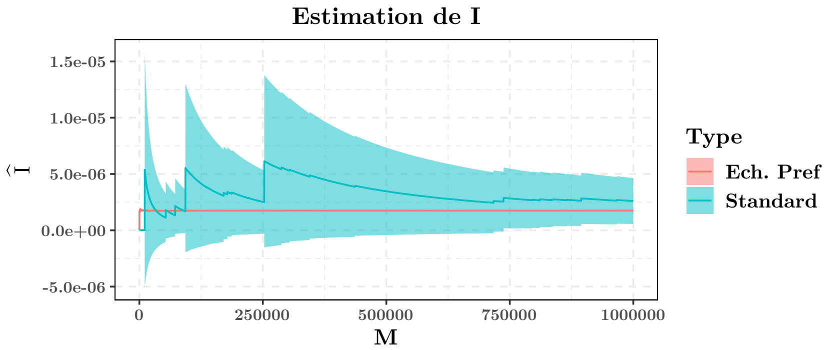


30 échantillons pondérés d'une Student(3)
Value
0.75
0.50
0.25
0.00
-10
-5
0
5
x
Function
φ(x)
g(x)

Estimation de I (et IC asymptotique)



Comparaison sur cet exemple



Echantillonnage préférentiel

Avantages

- Très utile pour l'estimation de quantités petites (probabilités d'évènements rares).

Echantillonnage préférentiel

Avantages

- Très utile pour l'estimation de quantités petites (probabilités d'évènements rares).
- Peut amener à une forte réduction de variance

Echantillonnage préférentiel

Avantages

- Très utile pour l'estimation de quantités petites (probabilités d'évènements rares).
- Peut amener à une forte réduction de variance
- Peut aussi être utiliser quand on ne sait pas simuler selon f !

Echantillonnage préférentiel

Avantages

- Très utile pour l'estimation de quantités petites (probabilités d'évènements rares).
- Peut amener à une forte réduction de variance
- Peut aussi être utiliser quand on ne sait pas simuler selon f!

Attention!

- Nécessite le choix de g! Pas toujours évident (notamment en grande dimension)!

Echantillonnage préférentiel

Avantages

- Très utile pour l'estimation de quantités petites (probabilités d'évènements rares).
- Peut amener à une forte réduction de variance
- Peut aussi être utiliser quand on ne sait pas simuler selon f !

Attention!

- Nécessite le choix de g! Pas toujours évident (notamment en grande dimension)!
- Un mauvais g peut amener à un estimateur de variance infinie! (voir TD).
- Un bon choix de g est souvent "problème dépendant", (ne conviendra que pour $\mathbb{E}[\varphi(X)]$ pour un φ spécifique).

Echantillonnage préférentiel normalisé

Problématique

Objectif, calculer:

$$I = \mathbb{E}[\varphi(X)] = \int_{\mathcal{D}_f} \varphi(x) f(x) d(x)$$

Supposons que f ne soit connue qu'à une constante près:

$$f(x) = \frac{\overbrace{f^{(u)}(x)}^{\text{Connu}}}{\underbrace{\int_{\mathcal{D}_f} f^{(u)}(z)\mathrm{d}z}_{\text{Inconnu}}}.$$

- Pour une densité de proposition g, on ne peut plus calculer le poids d'importance!

Problématique

Objectif, calculer:

$$I = \mathbb{E}[\varphi(X)] = \int_{\mathcal{D}_f} \varphi(x) f(x) d(x)$$

Supposons que f ne soit connue qu'à une constante près:

$$f(x) = \underbrace{\frac{\overbrace{f^{(u)}(x)}^{\text{Connu}}}{\int_{\mathcal{D}_f} f^{(u)}(z)\mathrm{d}z}}_{\text{Inconnu}}.$$

- Pour une densité de proposition g, on ne peut plus calculer le poids d'importance!
- Ce cas est en pratique très fréquent en inférence bayésienne!

Echantillonnage préférentiel normalisé

Si on dispose d'une loi de proposition g et de $Y_1, \ldots, Y_M$ tirés indépendemment selon g, alors l'estimateur:

$$\hat{I}_M^{IS,u} = \sum_{k=1}^{M} \frac{f^{(u)}(Y_k)/g(Y_k)}{\sum_{\ell=1}^{M} f^{(u)}(Y_\ell)/g(Y_\ell)} \varphi(Y_k)$$

est un estimateur consistant (convergence en proba.) de I.

Echantillonnage préférentiel normalisé

Si on dispose d'une loi de proposition g et de $Y_1, \ldots, Y_M$ tirés indépendemment selon g, alors l'estimateur:

$$\hat{I}_M^{IS,u} = \sum_{k=1}^{M} \frac{f^{(u)}(Y_k)/g(Y_k)}{\sum_{\ell=1}^{M} f^{(u)}(Y_\ell)/g(Y_\ell)} \varphi(Y_k)$$

est un estimateur consistant (convergence en proba.) de I.

- Estimateur biaisé pour M petit;
- Peut amener à un estimateur de variance plus faible que l'échantillonnage préférentiel classique.

Autres méthodes de réduction de variance

- Conditionnement
- Variables de contrôles
- Quasi Monte Carlo

Voir références dans le poly.