Cours de Probabilités, 1e année ENPC
Devoir maison sur la méthode de Monte-Carlo (obligatoire)

Principe de la méthode de Monte-Carlo. On considère une variable aléatoire $X : \Omega \to \mathbb{R}^d$ et une fonction mesurable $f : \mathbb{R}^d \to \mathbb{R}$ telle que

$$\mathbb{E}[|f(X)|] < \infty.$$

Le but de la méthode de Monte-Carlo est de calculer de façon approchée $\mathbb{E}[f(X)]$. Pour cela, on suppose que l'on sait simuler la variable aléatoire X, c'est-à-dire générer une suite i.i.d. $(X_i)_{i\geq 1}$ telle que $X_i \overset{\mathcal{L}}{=} X$. L'idée est d'approcher $\mathbb{E}[f(X)]$ par

$$M_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i), \tag{1}$$

avec n assez grand. En effet, la loi forte des grands nombres assure que $M_n \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[f(X)]$, p.s. Cela nous dit que pour n assez grand, M_n sera proche de $\mathbb{E}[f(X)]$. En revanche, nous ne savons pas à ce stade quel n choisir pour obtenir une précision ε donnée. Pour cela, nous faisons l'hypothèse supplémentaire que $\mathbb{E}[f(X)^2] < \infty$, et on pose

$$\sigma^2 = \mathbb{E}[f(X)^2] - \mathbb{E}[f(X)]^2.$$

Alors, le théorème de la limite centrale assure que

$$\frac{\sqrt{n}}{\sigma}(M_n - \mathbb{E}[f(X)]) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1). \tag{2}$$

On fait alors l'approximation que pour n suffisament grand,

$$\frac{\sqrt{n}}{\sigma}(M_n - \mathbb{E}[f(X)]) \overset{\mathcal{L}}{\approx} \mathcal{N}(0,1).$$

Comme pour $G \sim \mathcal{N}(0,1)$, $\mathbb{P}(-1,96 \leq G \leq 1,96) \approx 0,95$, on a qu'avec probabilité 95%,

$$\mathbb{E}[f(X)] \in \left[M_n - \frac{1,96\sigma}{\sqrt{n}}, M_n + \frac{1,96\sigma}{\sqrt{n}}\right].$$

L'intervalle $\left[M_n - \frac{1{,}96\sigma}{\sqrt{n}}, M_n + \frac{1{,}96\sigma}{\sqrt{n}}\right]$ est appelé intervalle de confiance à 95% pour $\mathbb{E}[f(X)]$. De même, comme $\mathbb{P}(-2,58 \leq G \leq 2,58) \approx 0,99$, l'intervalle $\left[M_n - \frac{2{,}58\sigma}{\sqrt{n}}, M_n + \frac{2{,}58\sigma}{\sqrt{n}}\right]$ est un intervalle de confiance à 99% pour $\mathbb{E}[f(X)]$.

En pratique, lorsqu'on cherche à calculer $\mathbb{E}[f(X)]$ de façon approchée, il est rare de connaître la valeur exacte de σ. Parfois, on connaît a priori un majorant $\bar{\sigma}$ de σ et dans ce cas, on sait qu'avec une probabilité au moins égale à 95%, $\mathbb{E}[f(X)]$ appartient à $\left[M_n - \frac{1{,}96\bar{\sigma}}{\sqrt{n}}, M_n + \frac{1{,}96\bar{\sigma}}{\sqrt{n}}\right]$, et cet intervalle est donc un intervalle de confiance supérieure à 95%. Néanmoins, dans la grande majorité des cas, on ne sait rien sur σ et on cherche à l'estimer également par Monte-Carlo. Ainsi,

$$V_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i)^2 - M_n^2$$

converge presque sûrement vers σ^2 par la loi forte des grands nombres. Grâce à (2) et au théorème de Slutsky, on en déduit que

$$\frac{\sqrt{n}}{\sqrt{V_n}}(M_n - \mathbb{E}[f(X)]) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

En répétant le même raisonnement que précédemment, on en déduit que

$$\left[M_n - \frac{1,96\sqrt{V_n}}{\sqrt{n}}, M_n + \frac{1,96\sqrt{V_n}}{\sqrt{n}}\right]$$

est un intervalle de confiance à 95% pour $\mathbb{E}[f(X)]$.

1 Calcul d'intégrale : sommes de Riemann ou méthode de Monte-Carlo ?

Dans cette première partie, nous étudions l'approximation numérique de l'intégrale $I = \int_{[0,1]^d} f(x)dx$, où $f : [0,1]^d \to \mathbb{R}$ est une fonction mesurable intégrable. Le but est de comparer la méthode usuelle des sommes de Riemann avec la méthode de Monte-Carlo. Nous commençons par le cas de la dimension $d = 1$.

1.1 En dimension $d = 1$

En dimension 1, les sommes de Riemann s'écrivent pour $n \in \mathbb{N}^*$,

$$S_n = \frac{1}{n}\sum_{i=0}^{n-1} f\left(\frac{i}{n}\right).$$

1. On suppose $f : [0,1] \to \mathbb{R}$ continue. On rappelle le théorème de Heine : toute application continue sur un compact est uniformément continue. Ainsi, f satisfait

   $$\forall \varepsilon > 0,\ \exists \eta > 0,\ \forall x, y \in [0,1], \quad |x - y| \leq \eta \implies |f(x) - f(y)| \leq \varepsilon.$$

   En déduire que $S_n \underset{n\to+\infty}{\longrightarrow} I$.

En fait, on peut montrer que $S_n \underset{n\to+\infty}{\longrightarrow} I$ pour une classe un peu plus grande de fonctions, les fonctions réglées qui incluent notamment les fonctions f continues par morceaux. En revanche, on ne dispose pas de vitesse de convergence à ce stade. Pour cela, on a besoin de faire des hypothèses de régularité sur f. La plus usuelle est de supposer f lipschitzienne, i.e.,

$$\exists K > 0,\ \forall x, y \in [0,1], \quad |f(x) - f(y)| \leq K|x - y|.$$

2. On suppose f Lipschitzienne. Montrer que

   $$|S_n - I| \leq \frac{K}{2n}.$$

   En déduire que pour calculer I avec une précision de $\varepsilon > 0$, il faut un nombre de calculs (et donc un temps de calcul) proportionnel à $\frac{K}{\varepsilon}$.

3. On suppose désormais f mesurable telle que $\int_0^1 f(u)^2 du < \infty$, et on note $\sigma^2 = \int_0^1 f(u)^2 du - \left(\int_0^1 f(u)du\right)^2$. Ecrire une méthode de Monte-Carlo qui permet d'approcher I. Donner l'intervalle de confiance à 95% en supposant σ connu. En déduire que pour obtenir une précision de $\varepsilon > 0$, il faut un temps de calcul proportionnel à $\frac{\sigma^2}{\varepsilon^2}$.

$\triangleright$ **À retenir :** En dimension 1, sauf éventuellement dans le cas où f n'est pas régulière, il faut préférer les sommes de Riemann (temps de calcul en $O(\varepsilon^{-1})$) à la méthode de Monte-Carlo (temps de calcul en $O(\varepsilon^{-2})$). Si f est plus régulière, on préfèrera utiliser la méthode des trapèzes ($f \in \mathcal{C}^2$) ou de Simpson ($f \in \mathcal{C}^4$) pour obtenir des temps de calcul respectivement en $O(\varepsilon^{-1/2})$ et $O(\varepsilon^{-1/4})$.

1.2 En dimension d quelconque

Pour $d \geq 1$, les sommes de Riemann s'écrivent

$$S_n = \frac{1}{n^d} \sum_{i_1=0}^{n-1} \cdots \sum_{i_d=0}^{n-1} f\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right).$$

4. Montrer que

$$I - S_n = \sum_{i_1=0}^{n-1} \cdots \sum_{i_d=0}^{n-1} \int_{\frac{i_1}{n}}^{\frac{i_1+1}{n}} \cdots \int_{\frac{i_d}{n}}^{\frac{i_d+1}{n}} f(x_1, \dots, x_d) - f\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right) dx_1 \dots dx_d.$$

5. On suppose $f : [0,1]^d \to \mathbb{R}$ lipschitzienne :

$$\forall x, y \in [0,1]^d, \quad |f(x) - f(y)| \leq K\|x - y\|_1.$$

   Montrer que $|S_n - I| \leq \frac{Kd}{2n}$. En déduire que pour obtenir une précision de $\varepsilon > 0$, il faut un temps de calcul proportionnel à $\left(\frac{Kd}{2}\right)^d \frac{1}{\varepsilon^d}$.

6. On suppose désormais f telle que $\int_{[0,1]^d} f(x)^2 dx < \infty$, et on pose $\sigma^2 = \int_{[0,1]^d} f(x)^2 dx - \left(\int_{[0,1]^d} f(x) dx\right)^2$. Ecrire une méthode de Monte-Carlo qui permet d'approcher I et donner l'intervalle de confiance à 95% avec σ connu. En déduire qu'il faut un temps de calcul proportionnel à $\frac{d\sigma^2}{\varepsilon^2}$ pour une précision de ε.

$\triangleright$ **À retenir :** En pratique, la dimension d et la fonction f sont imposées par le problème que l'on considère. Pour une précision de ε, le temps demandé par le calcul des sommes de Riemann est en $O(\varepsilon^{-d})$ lorsque $\varepsilon \to 0$. On parle de "malédiction de la dimension" : plus la dimension est grande, plus le calcul est gourmand en temps : il croît exponentiellement avec la dimension. En revanche, pour la méthode de Monte-Carlo, le temps de calcul est en $O(\varepsilon^{-2})$, quelle que soit la dimension. Bien sûr, lorsque f est régulière, on peut améliorer les sommes de Riemann avec des méthodes de type trapèzes mais qui sont frappées aussi par la malédiction de la dimension. Pour fixer les idées (car il faudrait nuancer selon les cas), la méthode de Monte-Carlo s'avère être la plus efficace à partir de $d \approx 4$.

2 Améliorations de la méthode de Monte-Carlo : méthodes de réduction de variance

La méthode de Monte-Carlo permet de calculer des intégrales avec une vitesse de convergence qui ne dépend pas de la dimension. Cette vitesse de convergence est en $\sigma/\sqrt{n}$. La décroissance en $1/\sqrt{n}$ est celle du théorème de la limite centrale, et on ne peut pas la changer sauf si on utilise une autre méthode que la méthode de Monte-Carlo. Par conséquent, la seule chose que l'on peut espérer améliorer est la variance σ^2 (ou, de façon équivalente, diminuer la valeur de l'estimateur V_n de la variance) : par exemple une diminution d'un facteur 10 de σ^2 permet d'obtenir la même précision avec 10 fois moins de simulations. Nous allons donner ici un aperçu de méthodes génériques pour réduire la variance de l'estimateur de Monte-Carlo. Par souci de simplicité, nous présentons ces méthodes en dimension $d = 1$. Elles se généralisent sans encombre à une dimension quelconque, et c'est là qu'elles sont pertinentes pour des applications concrètes.

2.1 Méthode de fonction d'importance

Pour fixer les idées, on considère que X est une variable aléatoire réelle de densité p. Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction mesurable telle que $\mathbb{E}[|f(X)|] < \infty$. On souhaite calculer $\mathbb{E}[f(X)]$ par la méthode de Monte-Carlo. Soit Y une variable aléatoire réelle de densité q strictement positive.

1. Montrer que $f(Y)\frac{p(Y)}{q(Y)}$ est une variable aléatoire intégrable et que

$$\mathbb{E}\left[f(Y)\frac{p(Y)}{q(Y)}\right] = \mathbb{E}[f(X)].$$

On suppose que :

— on sait simuler les variables aléatoires X et Y,

— les densités p et q sont connues et calculables.

Soit $(Y_i)_{i\geq 1}$ une suite i.i.d telle que $Y_i \overset{\mathcal{L}}{=} Y$. Alors, on peut approcher $\mathbb{E}[f(X)]$ par M_n défini en (1) ou bien par

$$\tilde{M}_n = \frac{1}{n}\sum_{i=1}^{n} f(Y_i)\frac{p(Y_i)}{q(Y_i)}.$$

On préfèrera $\tilde{M}_n$ à M_n si $\tilde{\sigma}^2 = \mathbb{E}\left[\left(f(Y)\frac{p(Y)}{q(Y)}\right)^2\right] - \mathbb{E}\left[f(X)\right]^2 < \sigma^2$. Le choix de la loi de Y est donc crucial pour mettre en oeuvre cette méthode.

2. **Exemple.** Soit $g : [0,1] \to \mathbb{R}$ mesurable de carré intégrable. On cherche à calculer $I = \int_0^1 g(x)e^{-x}dx$ par la méthode de Monte-Carlo. On se donne une suite i.i.d. $(U_i)_{i\geq 1}$ de variables aléatoires uniformes sur $[0,1]$.
   (a) Rappeler la loi de $-\log(U_1)$.
   (b) Montrer que $M_n = \frac{1}{n}\sum_{i=1}^{n} g(U_i)e^{-U_i}$ et $\tilde{M}_n = \frac{1}{n}\sum_{i=1}^{n} g(-\log(U_i))\mathbf{1}_{\{-\log(U_i)\leq 1\}}$ convergent presque sûrement vers I. Lequel de ces deux estimateurs préférer ?

2.2 Méthode des variables antithétiques

Cette méthode utilise des symétries de la loi que l'on simule pour avoir "gratuitement" de nouveaux tirages. Par exemple, on sait que pour $U \sim \mathcal{U}([0,1])$, $1-U$ suit également la loi $\mathcal{U}([0,1])$. Prenons l'exemple où on souhaite calculer $I = \int_0^1 f(x)dx$, avec f mesurable de carré intégrable. La méthode de Monte-Carlo usuelle consiste à utiliser

$$M_n = \frac{1}{n}\sum_{i=1}^{n} f(U_i),$$

où $(U_i)_{i\geq 1}$ est une suite i.i.d. de variables aléatoires uniformes sur $[0,1]$. La méthode des variables antithétiques consiste à utiliser :

$$\tilde{M}_n = \frac{1}{n}\sum_{i=1}^{n} \frac{f(U_i) + f(1-U_i)}{2}.$$

1. Montrer que $\tilde{M}_n$ converge presque sûrement vers I.
2. Montrer que $\mathrm{Var}(\frac{f(U)+f(1-U)}{2}) \leq \mathrm{Var}(f(U))$.

$\tilde{M}_n$ converge donc toujours mieux que M_n. Cependant, si dans le calcul numérique de M_n et $\tilde{M}_n$, le temps de calcul pour évaluer f est bien plus grand que celui nécessaire à générer un nombre aléatoire, il est plus judicieux de comparer $\tilde{M}_n$ avec M_{2n}. En effet ces deux estimateurs demandent tous les deux $2n$ calculs de la fonction f. On a

$$\begin{aligned}
\mathrm{Var}(M_{2n}) &= \frac{1}{2n}\mathrm{Var}(f(U)),\\
\mathrm{Var}(\tilde{M}_n) &= \frac{1}{2n}[\mathrm{Var}(f(U)) + \mathrm{Cov}(f(U), f(1-U))].
\end{aligned}$$

Par conséquent, on préfèrera $\tilde{M}_n$ à M_{2n} si

$$\mathrm{Cov}(f(U), f(1-U)) \leq 0.$$

C'est généralement ce critère qui est utilisé pour décider si on utilise ou pas la méthode des variables antithétiques.

3. **Exemple : cas où f est monotone.** On suppose $g, h : [0,1] \to \mathbb{R}$ mesurables et de carré intégrable. On suppose g croissante et h décroissante.
   (a) Montrer que $\mathrm{Cov}(g(U), h(U)) \leq 0$. (On pourra calculer $\mathbb{E}[(g(U_1) - g(U_2))(h(U_1) - h(U_2))]$ avec U_1 indépendante de U_2.)
   (b) En déduire que si f est monotone,

$$\mathrm{Cov}(f(U), f(1-U)) \leq 0.$$

2.3 Méthode des variables de contrôle

On considère le cadre donné en introduction avec une variable aléatoire réelle X et une fonction $f : \mathbb{R} \to \mathbb{R}$ telle que $\mathbb{E}[f(X)^2] < \infty$. On cherche à calculer $\mathbb{E}[f(X)]$ par méthode de Monte-Carlo. On suppose que l'on connaît une fonction $g : \mathbb{R} \to \mathbb{R}$ de carré intégrable telle que $\mathbb{E}[g(X)]$ est connu explicitement et telle que $\mathrm{Var}(g(X)) > 0$. On considère alors, pour $\lambda \in \mathbb{R}$,

$$M_n^\lambda = \frac{1}{n}\sum_{i=1}^n (f(X_i) + \lambda g(X_i)) - \lambda \mathbb{E}[g(X)].$$

Par la loi forte des grands nombres, M_n^λ converge presque sûrement vers $\mathbb{E}[f(X)]$. Le théorème de la limite centrale donne la convergence en loi de $\frac{\sqrt{n}}{\sigma_\lambda}(M_n^\lambda - \mathbb{E}[f(X)])$ vers $\mathcal{N}(0,1)$, où

$$\begin{aligned}
\sigma_\lambda^2 &= \mathbb{E}[(f(X) + \lambda g(X))^2] - \mathbb{E}[f(X) + \lambda g(X)]^2\\
&= \mathrm{Var}(f(X)) + 2\lambda \mathrm{Cov}(f(X), g(X)) + \lambda^2 \mathrm{Var}(g(X)).
\end{aligned}$$

Le meilleur choix de λ est $\lambda^* = -\frac{\mathrm{Cov}(f(X),g(X))}{\mathrm{Var}(g(X))}$, pour lequel σ_λ^2 atteint son minimum

$$\underline{\sigma}^2 = \frac{\mathrm{Var}(f(X))\mathrm{Var}(g(X)) - \mathrm{Cov}(f(X), g(X))^2}{\mathrm{Var}(g(X))}.$$

On a toujours $\underline{\sigma}^2 \leq \sigma^2$ par construction. Le problème pour appliquer cette méthode est qu'il faut connaître λ^*, ce qui n'est a priori pas le cas. La solution est de l'estimer aussi par Monte-Carlo. Pour $h : \mathbb{R} \to \mathbb{R}$, on introduit la notation

$$M_n(h) := \frac{1}{n}\sum_{i=1}^n h(X_i).$$

On a alors par la loi forte des grands nombres :

$$\frac{M_n(fg) - M_n(f)M_n(g)}{M_n(g^2) - M_n(g)^2} \underset{n\to+\infty}{\longrightarrow} \frac{\mathrm{Cov}(f(X), g(X))}{\mathrm{Var}(g(X))}, \quad \text{p.s.}$$

On définit alors

$$\tilde{M}_n = M_n(f) - \frac{M_n(fg) - M_n(f)M_n(g)}{M_n(g^2) - M_n(g)^2}(M_n(g) - \mathbb{E}[g(X)]).$$

1. En utilisant le théorème de Slutsky, montrer que $\frac{\sqrt{n}}{\underline{\sigma}}(\tilde{M}_n - \mathbb{E}[f(X)]) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$, puis que

$$\frac{\sqrt{n}}{\sqrt{\tilde{V}_n}}(\tilde{M}_n - \mathbb{E}[f(X)]) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

avec $\tilde{V}_n = M_n(f^2) - M_n(f)^2 - \frac{(M_n(fg)-M_n(f)M_n(g))^2}{M_n(g^2)-M_n(g)^2}$.

En particulier,

$$\left[\tilde{M}_n - \frac{1,96\sqrt{\tilde{V}_n}}{\sqrt{n}}, \tilde{M}_n + \frac{1,96\sqrt{\tilde{V}_n}}{\sqrt{n}}\right]$$

est un intervalle de confiance à 95% pour $\mathbb{E}[f(X)]$.

La variable aléatoire $g(X)$ est appelée variable de contrôle. On peut généraliser cette méthode en considérant plusieurs variables de contrôle $g_1(X), \ldots, g_m(X)$ dont on connaît explicitement l'espérance.

2. **Exemple.** On considère un ensemble $A \subset [0,1]^2$ tel que $A' = \{(u,v) \in [0,1]^2, u + v \leq 1\} \subset A$. Par exemple, A peut être le quart de disque $\{(u,v) \in [0,1]^2, u^2 + v^2 \leq 1\}$ ou plus généralement l'ensemble $\{(u,v) \in [0,1]^2, u^\alpha + v^\beta \leq 1\}$ avec $\alpha, \beta \geq 1$. On cherche à calculer l'aire de A par méthode de Monte-Carlo, que l'on note $|A|$.
   (a) On considère $(U_i, V_i)_{i\geq 1}$ une suite i.i.d., avec $U_1, V_1 \sim \mathcal{U}([0,1])$ et indépendantes. Montrer que $M_n = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_A(U_i, V_i)$ converge p.s. vers $|A|$, puis que $\frac{\sqrt{n}}{\sqrt{M_n(1-M_n)}}(M_n - |A|) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$. Construire un intervalle de confiance à 95% pour $|A|$.
   (b) Calculer $\mathbb{E}[\mathbf{1}_{A'}(U_1, V_1)]$ et $\mathrm{Var}(\mathbf{1}_{A'}(U_1, V_1))$. Donner en fonction de $|A|$ la valeur de $\lambda \in \mathbb{R}$ qui minimise $\sigma_\lambda^2 = \mathbb{E}[(\mathbf{1}_A(U_1,V_1) + \lambda\mathbf{1}_{A'}(U_1,V_1))^2] - \mathbb{E}[\mathbf{1}_A(U_1,V_1) + \lambda\mathbf{1}_{A'}(U_1,V_1)]^2$ ainsi que la valeur minimale $\underline{\sigma}^2 = \min_{\lambda\in\mathbb{R}} \sigma_\lambda^2$.
   Vérifier que $\underline{\sigma}^2 = 0$ lorsque $|A| = 1/2$. Expliquer pourquoi ce résultat est attendu.
   (c) On note

$$\tilde{M}_n = M_n + 2(M_n - 1)\left[\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{A'}(U_i, V_i) - 1/2\right] = 1 + 2(M_n - 1)\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{A'}(U_i, V_i).$$

   Construire un intervalle de confiance à 95% à partir de $\tilde{M}_n$.