Méthodes de Monte Carlo et processus stochastiques

Aurélien Alfonsi – Stefano De Marco – Emmanuel Gobet

MAP 556 - PC 3 (TP Python) - 3 et 5 octobre 2018

On se donne un modèle défini par la relation $Y = f(X)$. L'objectif est d'estimer la valeur $\mathbb{E}[g(Y)]$, pour une certaine fonction g sur l'ensemble des sorties Y. On suppose que $g(Y)$ est de carré intégrable.

EXERCICE 1 - Variables de contrôle

Les modèles "boîte noire" représentant des systèmes entrée-sortie de codes numériques sont souvent couteux en temps de calcul. Supposons d'avoir un modèle entrée-sortie réduit $Y_r = f_r(X)$ plus facile à simuler, pour différentes valeurs des entrées X, que le modèle $Y = f(X)$, et tel que la quantité $m_r = \mathbb{E}[g(Y_r)]$ est connue. On supposera de plus $\mathbb{E}[g(Y_r)^2] < \infty$.

On note $(X_i)_{1\leq i\leq n}$ une suite de copies indépendantes de la variable d'entrée X, et l'on pose

$$I_n = \frac{1}{n}\sum_{i=1}^{n} g(f(X_i)), \qquad I_n^c = m_r + \frac{1}{n}\sum_{i=1}^{n}\big(g(f(X_i)) - g(f_r(X_i))\big).$$

1. Vérifier que I_n et I_n^c sont des estimateurs non biasées de $\mathbb{E}[g(Y)]$, et calculer leur variances.

2. *(compléter le fichier* `Exo1_Q2.py`*)* On suppose que les variables d'entrée X suivent une loi uniforme sur $[0,1]$, $f(x) = e^x$, $f_r(x) = 1 + x$, et $g(y) = y$. Simuler les deux estimateurs et leurs intervalles de confiance asymptotiques à 95%. Tracer les courbes des moyennes empiriques I_n, I_n^c. Quel est le gain en termes de nombre de simulations de l'estimateur I_n^c, pour la même précision asymptotique que la méthode de Monte–Carlo naive?

3. *(compléter le fichier* `Exo1_Q3.py`*)* On note $(I_n^j)_{1\leq i\leq M}$ et $(I_n^{c,j})_{1\leq i\leq M}$ les estimations empiriques associées à M tirages indépendants des deux estimateurs. Evaluer explicitement $m = \mathbb{E}[g(Y)]$ et tracer les histogrammes des erreurs $(I_n^j - m)_{1\leq j\leq M}$ et $(I_n^{c,j} - m)_{1\leq j\leq M}$ pour $M = 1000$ et différentes valeurs de n.

4. *Variable de contrôle optimale*: on considère maintenant l'estimateur

$$I_n^\lambda = \lambda\, m_r + \frac{1}{n}\sum_{i=1}^{n}\big(g(f(X_i)) - \lambda\, g(f_r(X_i))\big), \qquad \lambda \in \mathbb{R}.$$

Proposer un choix du paramètre λ. Tracer les trajectoires des moyennes empiriques I_n^λ ainsi obtenues, puis l'histogramme des erreurs de cet estimateur.

EXERCICE 2 - Échantillonnage antithétique & Méthodes de stratification

On suppose que les variables d'entrée X sont de loi uniforme entre -1 et $+1$. On se place dans la situation $f(x) = e^x$, et $g(y) = y$, où l'on veut estimer $\mathbb{E}[g(Y)] = \mathbb{E}[e^X]$.

1. **Méthode de Monte-Carlo***(compléter le fichier* `Exo2_Q1.py`*)*: On note I_n la moyenne empirique de n copies indépendantes de X. Vérifier que l'on a $\mathbb{E}[e^X] = \sinh(1) \simeq 1.18$, et $\mathrm{Var}(e^X) = \frac{1}{2}(1 - e^{-2}) \simeq 0.43$. Tracer les courbes des moyennes empiriques obtenues sur plusieurs réalisations, et comparer avec la quantité recherchée.

2. **Échantillonnage antithétique** *(compléter le fichier* `Exo2_Q2.py`*)*: Vérifier que $(f(u) - f(v))(f(-u) - f(-v)) \leq 0$, pour tout $(u, v) \in [-1, 1]$, et en déduire une technique d'échantillonnage antithétique fondée sur la simulation de n variables uniformes sur $[-1, 1]$. On note I'_n l'estimateur empirique correspondant. Tracer les courbes des moyennes empiriques obtenues. Quel est le gain en termes de nombre de simulations de cet estimateur, pour la même précision asymptotique que la méthode de Monte-Carlo naive?

3. **Stratification proportionnelle :** Proposer une technique de stratification proportionnelle fondée sur la simulation de variables uniformes sur $[-1, 0]$, et uniformes sur $[0, 1]$. On note J_n l'estimateur correspondant. Tracer les trajectoires de l'estimateur en fonction de n. Cet estimateur satisfait-il un TCL? Quel est le gain en termes de nombre de simulations de cet estimateur, pour la même précision asymptotique que la méthode Monte-Carlo naive? Peut-on calculer explicitement $\mathrm{Var}(J_n)$? (voir question 5)

4. **Stratification non proportionnelle :** Proposer une technique de stratification fondée sur la simulation d'une proportion $n_1 = rn$ de variables uniformes sur $[-1, 0]$, et $n_2 = (1 - r)n$ de variables uniformes sur $[0, 1]$, avec $r \in (0, 1)$ tel que n_1, n_2 soient entiers. Tracer les trajectoires de ces estimateurs en fonction de n, pour différentes valeurs de r.

5. **Stratification optimale sur les deux strates** $(S_1, S_2) = ([-1, 0], [0, 1])$**:** Calculer les valeurs des paramètres

$$\begin{aligned} \sigma_1^2 &:= \mathbb{E}[e^{2X} \mid X \in [-1, 0]] - \mathbb{E}[e^X \mid X \in [-1, 0]]^2 \\ \sigma_2^2 &:= \mathbb{E}[e^{2X} \mid X \in [0, 1]] - \mathbb{E}[e^X \mid X \in [0, 1]]^2 \end{aligned}$$

et proposer une technique de stratification optimale. Simuler l'estimateur empirique associé, tracer ses trajectoires.

EXERCICE 3 - Échantillonnage d'importance

On suppose que les variables d'entrée X sont de loi gaussienne centrée réduite. On se place dans la situation $f(x) = x$, et $g(y) = (y - 2)^+$.

1. *(compléter le fichier* `Exo3_Q1.py`*)*: Vérifier par intégration numérique (quadrature), puis par simulation que $\mathbb{E}[g(Y)] = m \simeq 8.5 \times 10^{-3}$ et $\mathrm{Var}(g(Y)) = \sigma^2 \simeq 5.8 \times 10^{-3}$. On pourra utiliser la fonction `scipy.integrate.quad`: vérifier dans la documentation de la fonction ses variables d'entrée et de sortie. Tracer les trajectoires des moyennes empiriques $I_n = \frac{1}{n} \sum_{i=1}^n (Y_i - 2)^+$ obtenues à partir de n copies indépendantes $(Y_i)_{1 \leq i \leq n}$ de Y, puis l'histogramme des erreurs.

2. **Échantillonnage d'importance :**

   (a) Montrer que si Z est une variable aléatoire gaussienne réduite et centrée en θ, on a

   $$\mathbb{E}[g(Y)] = \mathbb{E}\Big[g(Z) e^{-\theta Z + \frac{\theta^2}{2}}\Big]$$

   Quel est l'intérêt d'une telle formule?

   (b) Proposer un estimateur de $\mathbb{E}[g(Y)]$ basé sur la simulation de la loi gaussienne réduite et centrée en 2. On note J_n cet estimateur. Vérifier par intégration numérique, puis par simulation que $\mathrm{Var}\Big(g(Z) e^{-\theta Z + \frac{\theta^2}{2}}\Big) = \sigma_J^2 \simeq 9.7 \times 10^{-5}$. Tracer les trajectoires de l'estimateur J_n, comparer avec la quantité recherchée. Quel est le gain en termes de nombre de simulations de cet estimateur, pour la même précision asymptotique que la méthode de Monte Carlo naive?

3. Comparer graphiquement les histogrammes des erreurs des estimateurs I_n et J_n, pour $n = 1000$.