

MACS - Quantification des incertitudes pour la simulation

DM4 - Année 2022-2023

1 Approches Monte-Carlo multiniveaux

On s'intéresse au calcul de l'intégrale :

$$\mathcal{I} = \int_{-1.5}^{1.5} z^2 \sin(z^2) dz.$$

On dispose pour cela de plusieurs codes :

- un code "parfait" : $x \mapsto y_1(x) = 3(3x - 1.5)^2 \sin((3x - 1.5)^2)$;
- un code "précis" : $x \mapsto y_2(x) = -3.22(x - 0.5)^2 + 345.8(x - 0.5)^4 - 993.8(x - 0.5)^6$;
- un code "approché" : $x \mapsto y_3(x) = 2.5(3x - 1.5)^2$.

On suppose que chaque appel à y_1, y_2, y_3 coûte C_1, C_2, C_3 respectivement, avec $C_1 > C_2 > C_3$. L'objectif de l'exercice est d'optimiser les appels à y_1, y_2, y_3 pour permettre la meilleure estimation de \mathcal{I} à coût total fixe.

1. Soit X une variable aléatoire uniformément distribuée sur $[0, 1]$. On note X_1, \dots, X_N N copies indépendantes et de même loi que X . En déduire un estimateur Monte Carlo de \mathcal{I} , que l'on note \hat{I}_N .

2. Calculer la moyenne et la variance de \hat{I}_N .

3. On donne $\text{Var}(3(3X - 1.5)^2 \sin((3X - 1.5)^2)) \approx 3.99$. En déduire le nombre d'appels à y_1 nécessaires pour une estimation de \mathcal{I} à 10^{-2} près, avec une confiance de 95%.

4. Afin d'obtenir une estimation plus précise de \mathcal{I} à un coût équivalent, on introduit les estimateurs \hat{I}_1, \hat{I}_2 et \hat{I}_3 , tels que :

$$\hat{I}_1 = \frac{1}{N_1} \sum_{n=1}^{N_1} y_1(X_n) - y_2(X_n), \quad \hat{I}_2 = \frac{1}{N_2} \sum_{n=N_1+1}^{N_1+N_2} y_2(X_n) - y_3(X_n) \quad \hat{I}_3 = \frac{1}{N_3} \sum_{n=N_1+N_2+1}^{N_1+N_2+N_3} y_3(X_n).$$

Proposer un nouvel estimateur Monte Carlo (en le justifiant) pour \mathcal{I} , que l'on notera $\tilde{I}_{N_1, N_2, N_3}$.

5. Calculer le coût total, noté \mathcal{C} , associé à l'évaluation de $\tilde{I}_{N_1, N_2, N_3}$ en fonction de N_1, N_2, N_3 et C_1, C_2, C_3 .

6. Montrer que la variance de $\tilde{I}_{N_1, N_2, N_3}$ est égale à $\frac{\text{Var}(y_1(X) - y_2(X))}{N_1} + \frac{\text{Var}(y_2(X) - y_3(X))}{N_2} + \frac{\text{Var}(y_3(X))}{N_3}$.

7*. En notant $V_1 = \text{Var}(y_1(X) - y_2(X))$, $V_2 = \text{Var}(y_2(X) - y_3(X))$ et $\text{Var}(y_3(X))$, calculer les valeurs de N_1 , N_2 et N_3 permettant de minimiser la variance de $\tilde{I}_{N_1, N_2, N_3}$ sous la contrainte que le coût total est égal à $\mathcal{C} = N \times C_1$. Les valeurs $C_1, C_2, C_3, V_1, V_2, V_3, N$ sont supposées être des données connues, et on résoudra le problème d'optimisation en considérant N_1, N_2, N_3 comme des quantités continues.

8*. En déduire la précision de l'estimation à 95% de \mathcal{I} obtenue pour ces valeurs de N_1, N_2, N_3 , pour $V_1 \approx 0.000345$, $V_2 \approx 0.180$, $V_3 \approx 2.83$, $C_1/C_2 = 3$, $C_1/C_3 = 100$, en supposant $N = 6.10^5$. Commenter l'intérêt d'une telle approche.