examen

June 26, 2024

```
[ ]: import numpy as np
     from numpy.random import Generator, PCG64
     import math
     from heateq import Exact, Simulateur
```

```
[ ]: seed = 213731490053398181466621250222036675538
     rng = Generator(PCG64(seed))
```

1 Description des simulateurs

On considère l'équation de la chaleur unidimensionnelle vue en cours, et la même distribution de probabilité pour $\mathbf{X}$. On rappelle que la précision de $\widetilde{h}^{(K,Q)}$ est gouvernée par les entiers K et Q, et que le coût moyen d'évaluer $\widetilde{h}^{(K,Q)}(\mathbf{X})$ est proportionnel à KQ.

On définit un simulateur haute fidélité $f = \widetilde{h}^{(K=21,Q=60)}$ et un simulateur basse fidélité $g = \widetilde{h}^{(K=3,Q=15)}$.

On note $w = c_g/c_f$, où c_f (respectivement, c_g) est le coût moyen d'évaluer $f(\mathbf{X})$ (respectivement, $g(\mathbf{X})$).

Dans la suite, on note $Y = f(\mathbf{X})$ et $Z = g(\mathbf{X})$.

Question 1 : Quelle est la valeur attendue de w ? Le vérifier expérimentalement.

```
[ ]: # Permet d'obtenir un n-échantillon du vecteur aléatoire d'entrée X
     def n_echantillon_X(n):
         return np.vstack(
             (
                 rng.uniform(-math.pi, math.pi, (3, n)),
                 rng.uniform(0.001, 0.009, (1, n)),
                 rng.uniform(-1., 1., (3, n))
             )
         )
```

```
[ ]: # Simulateurs haute et basse fidélité
     f = Simulateur(21, 60)
     g = Simulateur(3, 15)
```

```
[ ]: # Espérance exacte de la solution continue
mu_exact = Exact().mu
```

2 Estimateur ACV-IS

On définit deux échantillons *indépendants* (c'est-à-dire que leurs membres sont mutuellement indépendants)

$$\mathcal{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_n\} \quad \text{et} \quad \mathcal{X}' = \{\mathbf{X}'_1, \dots, \mathbf{X}'_N\}$$

et l'estimateur par variable de contrôle approchée ACV-IS (IS pour *independent sampling*) par

$$\bar{Y}^{\text{acv-is}}_{n,N}(\alpha) = \bar{Y}_n - \alpha(\bar{Z}_n - \bar{Z}'_N)$$

où * $\bar{Y}_n = \frac{1}{n}\sum\limits_{i=1}^{n} f(\mathbf{X}_i)$, * $\bar{Z}_n = \frac{1}{n}\sum\limits_{i=1}^{n} g(\mathbf{X}_i)$, * $\bar{Z}'_N = \frac{1}{N}\sum\limits_{i=1}^{N} g(\mathbf{X}'_i)$.

Question 2 : Montrer que le coût moyen d'évaluation de $\bar{Y}^{\text{acv-is}}_{n,N}(\alpha)$ est

$$c = n(c_f + \tau c_g)$$

où $\tau = 1 + \dfrac{N}{n}$.

Question 3 : En déduire le nombre d'évaluations haute fidélité équivalentes $\tilde{n}_f = c/c_f$, exprimée en fonction de τ et w.

Question 4 : Montrer que

$$\mathbb{C}[\bar{Y}_n, \bar{Z}_n - \bar{Z}'_N] = \frac{1}{n}\mathbb{C}[Y, Z] \quad \text{et} \quad \mathbb{V}[\bar{Z}_n - \bar{Z}'_N] = \frac{n+N}{nN}\mathbb{V}[Z].$$

Question 5 : En déduire que la valeur de α qui minimise $\mathbb{V}[\bar{Y}^{\text{acv-is}}_{n,N}(\alpha)]$, notée α^*, est

$$\alpha^* = \frac{N}{n+N}\frac{\mathbb{C}[Y, Z]}{\mathbb{V}[Z]}$$

et que la variance minimale est alors

$$\mathbb{V}[\bar{Y}^{\text{acv-is}}_{n,N}(\alpha^*)] = \mathbb{V}[\bar{Y}_{\tilde{n}_f}](1 + \tau w)\left(1 - \frac{N}{n+N}\rho^2\right)$$

où $\bar{Y}_{\tilde{n}_f} = \dfrac{1}{\tilde{n}_f}\sum\limits_{i=1}^{\tilde{n}_f} f(\mathbf{X}_i)$ et ρ est le coefficient de corrélation de Pearson entre Y et Z.

Question 6 : Montrer que $\dfrac{N}{n+N} = 1 - \dfrac{1}{\tau}$ et comparer l'expression de $\mathbb{V}[\bar{Y}^{\text{acv-is}}_{n,N}(\alpha^*)]$ à celle de $\mathbb{V}[\bar{Y}^{\text{acv}}_{n,N}(\alpha^*)]$ vue en cours. En déduire que la valeur de τ qui minimise $\mathbb{V}[\bar{Y}^{\text{acv-is}}_{n,N}(\alpha^*)]$, notée τ^*, est

$$\tau^* = \sqrt{\frac{\rho^2}{w(1-\rho^2)}}$$

et que, pour un budget $\tilde{n}_f$ donné, l'estimateur ACV-IS optimal est

$$\bar{Y}^{\text{acv-is}}_{n^*,N^*}(\alpha^*) = \bar{Y}_{n^*} - \alpha^*(\bar{Z}_{n^*} - \bar{Z}'_{N^*})$$

où $\alpha^* = \left(1 - \frac{1}{\tau^*}\right)\frac{\mathbb{C}[Y,Z]}{\mathbb{V}[Z]}$, $n^* = \frac{\tilde{n}_f}{1+\tau^* w}$ et $N^* = (\tau^* - 1)n^*$.

Question 7 : Déduire du cours que

$$\beta = \frac{\mathbb{V}[\bar{Y}^{\text{acv-is}}_{n^*,N^*}(\alpha^*)]}{\mathbb{V}[\bar{Y}_{\tilde{n}_f}]} = (1+\tau^* w)^2(1-\rho^2).$$

Quelles conclusions pouvez-vous tirer sur les estimateurs ACV et ACV-IS optimaux ?

3 Expérience pilote

Question 8 : Estimer ρ^2 et τ^* à l'aide de `ns = 10000` échantillons pilotes.

Question 9 : En déduire une estimation de $\alpha^* = \left(1 - \frac{1}{\tau^*}\right)\frac{\mathbb{C}[Y,Z]}{\mathbb{V}[Z]}$ et de $\beta = (1+\tau^* w)^2(1-\rho^2)$.

[ ]:
```
ns = 10000 # échantillons pilotes
X = n_echantillon_X(ns)
```

4 Analyse de l'estimateur ACV-IS

Question 10 : Utiliser la valeur de α^* estimée précédemment (avec les échantillons pilotes) pour contruire un estimateur ACV-IS de l'espérance de Y. Faire `nr = 1000` répétitions pour des budgets (en termes de nombre d'évaluations haute fidélité équivalentes) $\tilde{n}_f = c/c_f \in \{5; 10; 20; 50; 100; 200; 500; 1000\}$.

Question 11 : Pour chaque budget, estimer le rapport de variance β à budget équivalent entre l'estimateur ACV-IS et l'estimateur Monte Carlo classique (haute fidélité). Ces estimations sont-elles conformes à la valeur de β estimée précédemment (dans l'expérience pilote) ?

Question 12 : Tracer l'espérance et l'écart-type (sous forme de barres d'erreur) des estimateurs par variable de contrôle et Monte Carlo (haute fidélité) à budget équivalent en fonction de $\tilde{n}_f$. Sur un autre graphe (en échelle log-log), tracer l'évolution de la REQM (par rapport à l'espérance exacte du problème continu, `mu_exact`) des estimateurs en fonction de $\tilde{n}_f$.

[ ]:
```
nr = 1000
budgets = [5, 10, 20, 50, 100, 200, 500, 1000]
```