

TP3 Sujet

February 10, 2023

1 Méthode de Monte Carlo

Tout d'abord, on illustre numériquement les deux résultats probabilistes sur lesquels reposent la méthode dite de Monte Carlo:

- la loi forte de grands nombres,
- le théorème central limit (TCL).

On considère ensuite un premier exemple d'estimateur de Monte Carlo et l'importance de l'intervalle de confiance (IC) dans lequel se trouve la valeur recherchée avec probabilité grande (0.95).

Enfin on applique la méthode de Monte Carlo à un exemple multidimensionnel où on illustre l'efficacité de 2 méthodes de réduction de variance:

- variables antithétiques,
- variable de contrôle.

```
[ ]: import numpy as np
from scipy import stats
import matplotlib.pyplot as plt
import seaborn as sns
sns.set_theme()
from numpy.random import default_rng
rng = default_rng()
```

1.1 Illustration de la loi des grands nombres

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires *i.i.d.* de carré intégrable. On définit les suites $(m_n)_{n \geq 1}$ et $(\sigma_n^2)_{n \geq 2}$ (non définie pour $n = 1$) de la façon suivante

$$m_n = \frac{1}{n} \sum_{k=1}^n X_k \quad \text{et} \quad \sigma_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - m_n)^2 \quad \text{pour } n \geq 2$$

et on veut illustrer la Loi Forte des Grands Nombres et le Théorème Central Limite (étendu en utilisant le lemme de Slutsky pour remplacer $\sigma^2 = \text{var}(X_1)$ par l'estimateur σ_n^2) c'est à dire les convergences

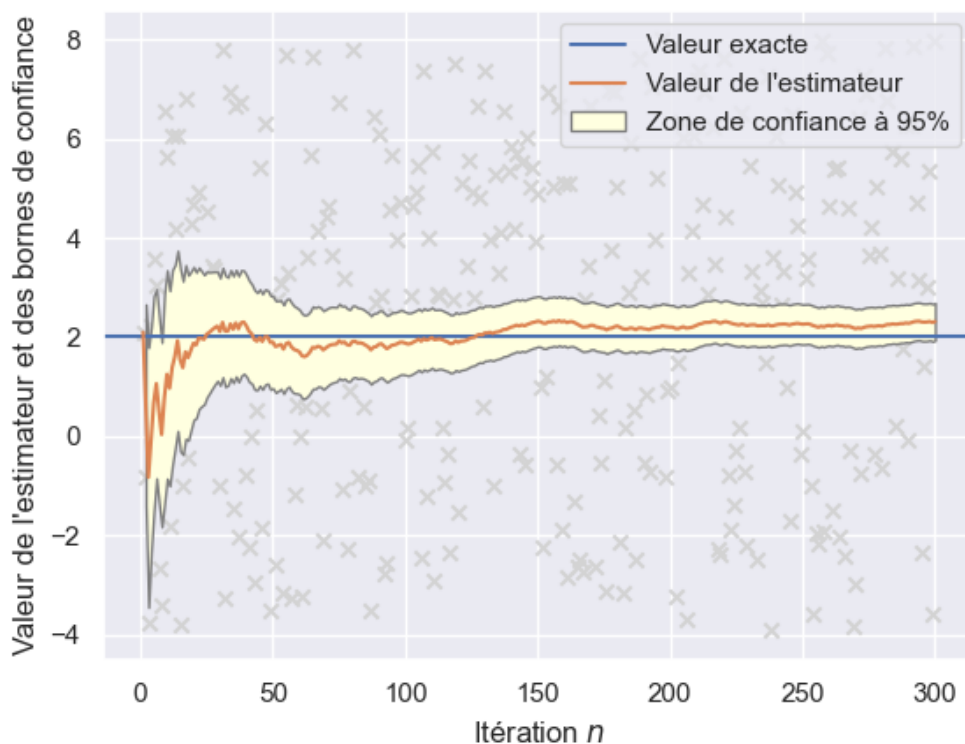
$$m_n \xrightarrow{p.s.} m \quad \text{et} \quad \sqrt{n} \left(\frac{m_n - m}{\sigma_n} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Plus précisément on construit l'intervalle de confiance (asymptotique) à 95% à partir du TCL c'est à dire

$$\text{pour } n \text{ grand } \mathbf{P}\left(m \in \left[m_n - \frac{1.96\sigma_n}{\sqrt{n}}, m_n + \frac{1.96\sigma_n}{\sqrt{n}}\right]\right) \simeq 0.95$$

1.1.1 Question: LFGN loi uniforme

Reproduire le tracé suivant où les points (les croix 'x') sont les réalisations X_n (en fonction de n) de loi uniforme sur $[-4, 8]$. La ligne bleue (couleur 'C0', première couleur de la palette utilisée) correspond à la moyenne m , la courbe orangée (couleur 'C1') correspond à la suite m_n et les lignes grises correspondent aux bornes de l'intervalle de confiance. La zone de confiance en jaune s'obtient par la méthode `fill_between` de `ax`.



[]:

1.1.2 Question: LFGN loi de Cauchy

Reprendre rapidement l'exemple précédent en remplaçant la loi uniforme par la loi de Cauchy. On obtient des réalisations de la loi de Cauchy en utilisant la méthode `standard_cauchy` de l'objet `rng`. Répliquer plusieurs fois le tracé (avec l'axe des ordonnées restreint à $[-10, 10]$) pour différentes valeurs de $n = 100\,000$. Qu'en pensez-vous?

[]:

1.2 Illustration du TCL

On veut illustrer la répartition de l'erreur renormalisée $\varepsilon_n = \sqrt{n} \left(\frac{m_n - m}{\sigma_n} \right)$ pour différentes valeurs de n . Lorsque n est grand cette erreur renormalisée est proche de la loi normale centrée réduite, c'est ce qu'on veut vérifier numériquement. Pour illustrer cette répartition, il est nécessaire de répliquer un grand nombre de fois l'erreur c'est à dire de considérer un échantillon $(\varepsilon_n^{(j)})_{j=1, \dots, M}$ de taille M et de construire l'histogramme de cet échantillon.

Attention: en pratique il n'est pas nécessaire de répliquer M fois l'estimateur m_n pour approcher m . L'estimateur de la variance v_n suffit pour donner la zone de confiance autour de m_n . C'est une information importante donnée par le TCL.

1.2.1 Question: TCL loi uniforme

Dans le cas de la loi uniforme sur $[-4, 8]$ vérifier la répartition de l'erreur renormalisée ε_n pour $n = 10$ puis $n = 1\,000$ à partir d'un échantillon de taille $M = 100\,000$.

[]:

[]:

1.3 Un premier exemple d'estimateur de Monte Carlo

On va mettre en oeuvre un estimateur de Monte Carlo pour calculer

$$I(\beta) = \mathbf{E}[\exp(\beta G)] \quad \text{où } G \sim \mathcal{N}(0, 1) \text{ et } \beta \in \mathbf{R}.$$

La valeur exacte $I(\beta) = \exp(\beta^2/2)$ est connue mais cet exemple permet d'illustrer l'importance des bornes de l'intervalle de confiance (et donc de l'estimation de la variance) dans une méthode de Monte Carlo. La seule valeur moyenne $I_n = \frac{1}{n} \sum_{k=1}^n X_k$ n'est pas suffisante pour déterminer I .

1.3.1 Question: fonction monte_carlo

Ecrire une fonction `monte_carlo(sample, proba=0.95)` qui à partir d'un échantillon `sample` de réalisations indépendantes $(X_k)_{k=1, \dots, n}$ renvoie un tuple qui contient:

- la moyenne de l'estimateur Monte Carlo de $I = \mathbf{E}[X]$,
- l'estimateur de la variance asymptotique apparaissant dans le TCL,
- les bornes inférieures et supérieures de l'intervale de confiance de niveau de probabilité `proba`.

[]:

1.3.2 Question: premier exemple

En utilisant la fonction `monte_carlo`, reproduire le tableau suivant où chaque ligne représente un résultat pour une valeur de $\beta \in \{0.2, 0.5, 1, 2, 3, 5\}$:

- la première colonne est la valeur moyenne I_n ,
- la deuxième colonne l'estimateur de la variance,
- les colonnes 3 et 4 sont les bornes inférieures et supérieures de l'IC à 95%,

- la colonne 5 contient la valeur exacte $\mathbf{E}[\exp(\beta G)] = \exp(\beta^2/2)$.

Ce tableau est obtenu pour $n = 1\,000\,000$. Comment interpréter ce tableau?

```
[1]: import pandas as pd
df = pd.read_pickle("data/first_df.pkl")
df
```

```
[1]:
```

	mean	var	low	high	exact
0.2	1.020551	4.245536e-02	1.020147	1.020955	1.020201
0.5	1.134027	3.650418e-01	1.132843	1.135212	1.133148
1.0	1.651060	4.676691e+00	1.646821	1.655298	1.648721
2.0	7.402685	2.379946e+03	7.307068	7.498301	7.389056
3.0	87.915075	8.558333e+06	82.181273	93.648877	90.017131
5.0	121963.825619	6.439313e+14	72228.169429	171699.481809	268337.286521

```
[ ]:
```

```
[ ]:
```

1.4 Option panier: un exemple multidimensionnel

On considère $d \geq 2$ actifs financiers dont la loi à l'instant $T > 0$ est modélisée par une loi log-normale c'est à dire

$$\forall i \in \{1, \dots, d\}, \quad S_T^i = S_0^i \exp\left((r - \frac{\sigma_i^2}{2})T + \sigma_i \sqrt{T} \tilde{G}_i\right)$$

où le vecteur $(\tilde{G}_1, \dots, \tilde{G}_d)$ est gaussien centré de matrice de covariance Σ et les constantes $r > 0$, $\sigma_i > 0$ sont fixées. Il s'agit d'actifs financiers $(S_t^i)_{t \in [0, T]}$, $1 \leq i \leq d$, modélisés par un processus de Black-Scholes multidimensionnel. On introduit la matrice L triangulaire inférieure obtenue par la décomposition de Cholesky de la matrice $\Sigma = LL^\top$.

A l'aide de cette matrice L , on définit la fonction $\Phi : \mathbf{R}^d \rightarrow \mathbf{R}^d$ telle que

$$(S_T^1, \dots, S_T^d) = \Phi(G_1, \dots, G_d) \quad \text{ou encore} \quad S_T^i = \Phi_i(G_1, \dots, G_d)$$

où $(G_1, \dots, G_d) \sim \mathcal{N}(0, I_d)$ (l'égalité précédente est à considérer en loi).

On s'intéresse au prix d'une option européenne (aussi appelé produit dérivé européen) sur le panier de ces d actifs financiers, c'est à dire qu'on veut calculer

$$\mathbf{E}[X] \quad \text{avec} \quad X = \left(\frac{1}{d} \sum_{i=1}^d S_T^i - K \right)_+.$$

1.4.1 Question: initialisation

Définir les paramètres globaux $d = 10$, $T = 1$, $r = 0.01$, $S_0^i = 100$ (pour tous les actifs), $\sigma_i = i/(2d)$ (on dit que certains actifs sont plus volatiles que d'autres) et la matrice de corrélation Σ définie par $\Sigma_{i,i} = 1$ et $\Sigma_{i,j} = \rho \in [0, 1]$ pour $i \neq j$, avec $\rho = 0.2$.

Initialiser la matrice L en utilisant la fonction `np.linalg.cholesky`.

[]:

1.4.2 Question: simulation d'un échantillon d'actifs

Définir la fonction python **phi** qui transforme le vecteur (G_1, \dots, G_d) en un vecteur (S_T^1, \dots, S_T^d) (tous les paramètres sont des variables globales pour simplifier l'écriture du code). L'appel suivant doit fonctionner

```
G = rng.standard_normal(size=d)
phi(G)
```

Si on veut implémenter un estimateur Monte Carlo il faut travailler avec des échantillons *i.i.d.* $(S_T^{(j)})_{j=1, \dots, n}$ où $S_T^{(j)} = (S_T^{(j),1}, \dots, S_T^{(j),d}) \in \mathbf{R}^d$. Modifier votre fonction **phi** pour création un tel échantillon à partir de l'appel suivant:

```
sample_G = rng.standard_normal(size=(d, n))
phi(sample_G)
```

(il faut utiliser la technique du broadcasting en **numpy**, c'est très important à connaître en pratique).

[]:

1.4.3 Question: estimateur Monte Carlo

Définir une fonction $\psi : \mathbf{R}^d \times \mathbf{R}_+ \rightarrow \mathbf{R}_+$ telle que

$$\psi(G_1, \dots, G_d, K) = \left(\frac{1}{d} \sum_{i=1}^d \Phi_i(G_1, \dots, G_d) - K \right)_+$$

dans une fonction python appelée **psi**. Cette fonction doit fonctionner avec un échantillon $(G_1^{(j)}, \dots, G_d^{(j)})_{j=1, \dots, n}$.

Ecrire et programmer l'estimateur de Monte Carlo pour estimer la quantité $\mathbf{E}[X] = \mathbf{E}[\psi(G_1, \dots, G_d, K)]$ où $(G_1, \dots, G_d) \sim \mathcal{N}(0, I_d)$.

Pour différentes valeur de $K \in \{80, 90, 100, 110, 120\}$ et $n = 100\,000$ vous devez obtenir le tableau suivant:

```
[2]: df = pd.read_pickle("data/basket_mc.pkl")
df
```

```
[2]:
```

	mean	var	lower	upper
80	21.394471	228.318772	21.300818	21.488123
90	12.860460	181.187947	12.777032	12.943889
100	6.655165	111.553749	6.589702	6.720627
110	2.998650	54.132985	2.953049	3.044252
120	1.204158	21.976278	1.175102	1.233213

[]:

1.4.4 Question: variables antithétiques

Sur le même modèle que précédemment, implémenter la méthode de Monte Carlo avec réduction de variance par variables antithétiques c'est à dire basée sur la représentation:

$$\mathbf{E}[X] = \mathbf{E}\left[\frac{1}{2}(\psi(G_1, \dots, G_d, K) + \psi(-G_1, \dots, -G_d, K))\right]$$

Calculer le ratio de variance (variance de la méthode naïve divisée par variance par variables antithétiques) pour les différentes valeurs de K .

Que signifie ce ratio de variance?

[]:

1.5 Option panier: une variable de contrôle

Dans le cas de la dimension 1 ($d = 1$), le prix est donnée par une formule fermée, on appelle cette formule la formule de Black-Scholes. Pour une option Basket (en dimension $d \geq 2$) on approche le prix par Monte Carlo mais on peut utiliser des approximations pour construire un problème unidimensionnel proche du produit Basket. Ces approximations servent de variables de contrôles: **on ne rajoute pas une erreur, on retire de la variance.**

On rappelle que, en posant $\mu_i = r - \frac{1}{2}\sigma_i^2$,

$$X = \left(\frac{1}{d} \sum_{i=1}^d S_0^i e^{\mu_i T + \sigma_i \sqrt{T} G_i} - K \right)_+$$

et en introduisant $a_0^i = \frac{S_0^i}{\sum_{j=1}^d S_0^j}$ (t.q. $\sum a_0^i = 1$) et $\bar{S}_0 = \frac{1}{d} \sum_{i=1}^d S_0^i$ on a

$$X = \left(\bar{S}_0 \sum_{i=1}^d a_0^i e^{\mu_i T + \sigma_i \sqrt{T} G_i} - K \right)_+.$$

La variable de contrôle proposée est obtenue en échangeant l'exponentielle et la moyenne pondérée par les poids $(a_0^i)_{i=1, \dots, d}$:

$$Y = (\bar{S}_0 e^Z - K)_+ \quad \text{avec} \quad Z = \sum_{i=1}^d a_0^i (\mu_i T + \sigma_i \sqrt{T} G_i)$$

La variable aléatoire Z suit une loi gaussienne $Z \sim \mathcal{N}(mT, s^2 T)$ avec

$$m = \sum_{i=1}^d a_0^i \mu_i \quad \text{et} \quad s^2 = \sum_{i=1}^d \left(\sum_{j=1}^d a_0^i \sigma_i L_{ij} \right)^2.$$

Ainsi l'espérance de la variable de contrôle Y est connue par la formule de Black-Scholes, car elle correspond au prix d'un call de strike K d'un actif Black-Scholes de dimension 1, de valeur initiale \bar{S}_0 , de taux $\rho = m + \frac{1}{2}s^2$ et de volatilité s (à un facteur d'actualisation près... attention à ça). On a donc

$$e^{-\rho T} \mathbf{E}[Y] = P_{BS}(\bar{S}_0, \rho, s, T, K),$$

où

$$P_{\text{BS}}(x, r, \sigma, T, K) = xF_{\mathcal{N}(0,1)}(d_1) - Ke^{-rT}F_{\mathcal{N}(0,1)}(d_2),$$

avec $F_{\mathcal{N}(0,1)}$ est la fonction de répartition de la loi normale centrée réduite et la notation

$$d_1 = \frac{1}{\sigma\sqrt{T}} \left(\log\left(\frac{x}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)T \right) \quad \text{et} \quad d_2 = d_1 - \sigma\sqrt{T}$$

1.5.1 Question: préliminaires pour la variable de contrôle

- Définir la fonction `price_call_BS` qui code la fonction $P_{\text{BS}}(x, r, \sigma, T, K)$ définie ci-dessus.
- Initialiser les paramètres $\bar{S}_0, (a_0^i)_{i=1,\dots,d}, m, s^2$ et ρ .
- Calculer $\mathbf{E}[Y]$ par la formule fermée.
- Calculer $\mathbf{E}[Y]$ par un estimateur Monte Carlo à partir de réalisations de $(G_1^{(j)}, \dots, G_d^{(j)}), j \in \{1, \dots, n\}$.
- Vérifier que tout est cohérent.

[]:

1.5.2 Question: MC avec variable de contrôle

Implémenter l'estimateur de Monte Carlo avec variable de contrôle pour le calcul de $\mathbf{E}[X]$ c'est à dire

$$\mathbf{E}[X] = \mathbf{E}[\psi(G_1, \dots, G_d, K) - (Y - \mathbf{E}[Y])],$$

où Y est la variable de contrôle introduite précédemment et $\mathbf{E}[Y]$ est calculée par la formule fermée. Comparer les ratios de variance pour les différentes valeurs de K .

[]: