TP 5 – Méthode de Monte Carlo

1 Principe de la méthode

La méthode de Monte-Carlo permet d'approcher le calcul d'une intégrale, ou d'une somme difficile à calculer, en ayant recours à la loi des grands nombres.

Soient g est une fonction réelle g continue sauf éventuellement en un nombre fini de points et X une variable aléatoire. On suppose que $g(X)$ est une variable aléatoire admettant une espérance m et une variance σ^2.

La loi des grands nombres assure que si $(X_n)_{n\geqslant 1}$ est une suite de variables aléatoires indépendantes et de même loi que X, alors la suite de terme général

$$\overline{Y}_n = \frac{1}{n}\sum_{k=1}^{n} g\left(X_k\right)$$

converge en probabilité vers $m = \mathbb{E}(g(X))$.

Ainsi, pour approcher la quantité $\mathbb{E}(g(X))$, on peut utiliser une réalisation de la variable aléatoire $\overline{Y}_n$ obtenue par simulation pour n assez grand.

Lorsqu'on souhaite approcher une intégrale ou une somme, il s'agit alors de commencer par chercher à l'écrire sous la forme $\mathbb{E}(g(X))$, où X est une variable aléatoire de loi connue, facile à simuler.

Exercice 1.

1. Utiliser la méthode de Monte-Carlo pour déterminer une valeur approchée de l'intégrale

$$\int_0^1 \mathrm{e}^{-t^2}\,\mathrm{d}t.$$

2. Faire de même pour l'intégrale $\displaystyle\int_0^{+\infty} \frac{\mathrm{e}^{-t}}{1+t^4}\,\mathrm{d}t$ et pour la somme $\displaystyle\sum_{n=1}^{\infty} \frac{\sqrt{n}}{2^n}$.

2 Approximation d'une probabilité

La méthode de Monte-Carlo est la méthode utilisée pour approcher la valeur de la probabilité d'un événement par simulation. C'est d'ailleurs, sans le dire, la méthode choisie jusqu'ici à cette fin.

On commence par simuler l'expérience un grand nombre de fois et on considère les variables aléatoires mutuellement indépendantes

$$X_k = \begin{cases} 1 & \text{si } A \text{ est réalisé lors de la } k\text{-ième expérience,} \\ 0 & \text{sinon.} \end{cases}$$

On remarque que chacune de ces variables aléatoires suit la loi $\mathcal{B}(p)$, avec $p = \mathbb{P}(A)$. Alors, toujours par la loi faible des grands nombres,

$$\left(\frac{1}{n}\sum_{i=1}^{n} X_k\right) \text{ converge en probabilité vers } \mathbb{P}(A).$$

Exercice 2. Soit $n \geqslant 2$ un entier. On considère une urne contenant n boules numérotées de 1 à n. On effectue des tirages successifs d'une boule avec remise. On suppose que tous les tirages sont équiprobables. On s'arrête dès que l'on obtient une boule déjà obtenue lors d'un précédent tirage. On note T_n la variable aléatoire égale au nombre de tirages effectués.

On souhaite comparer graphiquement à l'aide de PYTHON la loi de la variable aléatoire $Y_n = \frac{\mathrm{T}_n}{\sqrt{n}}$ avec la loi qui a pour densité

$$f : x \longmapsto \begin{cases} 0 & \text{si } x \leqslant 0 \\ x\,\mathrm{e}^{-x^2/2} & \text{si } x > 0 \end{cases}.$$

1. Écrire une function `simuleY(n)` qui simule la variable aléatoire Y_n.
2. Comment approcher la fonction de répartition de Y_n par la méthode de Monte-Carlo ? La représenter graphiquement sur $[0,4]$ pour $n = 50$.
3. Représenter sur $[0,4]$ la fonction de répartition de Y. Illustrer la convergence en superposant les représentations précédentes pour différentes valeurs de n.

Exercice 3. Que retourne le code suivant ? Justifier.

```
n=10**4
X=rd.rand(n)
print(4*np.mean(np.sqrt(1-X**2)))
```

3 Estimation d'erreur

Il est utile de travailler en amont de l'utilisation de la méthode de Monte Carlo pour donner une estimation de la confiance à accorder à l'estimation. Pour ce faire, on cherche à assurer que la probabilité que l'écart entre l'estimation et la valeur recherchée soit supérieure à une précision fixée soit assez petite (on choisit souvent : inférieure à 5%).

Pour y parvenir, on peut utiliser l'inégalité de Bienaymé-Tchebychev : si (X_n) est une suite de v.a.r. indépendantes et de même loi, ayant pour espérance m et variance σ^2, alors

$$\mathbb{P}\left(\left|\overline{Y}_n - m\right| \geqslant \varepsilon\right) \leqslant \frac{\sigma^2}{n\varepsilon^2},$$

où $\overline{Y}_n = \sum\limits_{i=1}^{n} g(X_i)$.

N.B. Nous verrons par la suite qu'on peut aussi utiliser le théorème central limite à cette fin, avec de meilleurs résultats.

Exercice 4. On considère une suite de v.a.r. $(X_n)_{n\in\mathbb{N}}$ indépendantes suivantes toutes la loi $\mathcal{U}([0,1]$, et on pose pour tout $n \in \mathbb{N}$

$$\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n}\sqrt{1-X_i^2}.$$

Nous avons vu dans l'exercice 3 que pour n grand, Y_n est une valeur approchée de $\frac{\pi}{4}$.

1. Quelle valeur de n choisir pour que $\overline{Y}_n$ soit une valeur approchée de $\frac{\pi}{4}$ à la précision $\varepsilon = 10^{-2}$ avec un risque $\alpha = 5\%$, c'est-à-dire $\mathbb{P}(|\overline{Y}_n - \frac{\pi}{4}| \geqslant \varepsilon) \leqslant 0{,}05$.
2. Comment vérifier expérimentalement le résultat précédent ? Mettre cette méthode en pratique.

On peut utiliser le théorème central limite, qui donne

$$\mathbb{P}\left(\left|\overline{Y}_n - m\right| \geqslant z\frac{\sigma}{\sqrt{n}}\right) \;=\; \mathbb{P}\left(\left|\overline{Y}_n^*\right| \geqslant z\right) \;\underset{n\to\infty}{\longrightarrow}\; \mathbb{P}(|Z| \geqslant z),$$

où Z suit la loi $\mathcal{N}(0,1)$, et $\overline{Y}_n^*$ désigne la version centrée réduite de $\overline{Y}_n$. Pour $z \simeq 1{,}96$, on obtient $\mathbb{P}(|Z| \geqslant z) \simeq 5\%$.

En pratique, le théorème central limite donne de meilleurs résultats.