

TP 5 : Estimation et méthode de Monte Carlo

On rappelle que la loi des grands nombres permet d'estimer la moyenne d'une variable aléatoire intégrable X à partir d'une suite $(X_k)_{k \in \mathbb{N}}$ de variables aléatoires indépendantes et de même loi que X . En effet, sous ces hypothèses, on a

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbb{E}X.$$

Par ailleurs, si X est de carré intégrable, de variance σ^2 , on peut obtenir une estimation de l'erreur commise grâce au théorème limite central. En effet, si l'on pose

$$\hat{I}_n(a) = \left[\frac{1}{n} \sum_{k=1}^n X_k - a \frac{\sigma}{\sqrt{n}}, \frac{1}{n} \sum_{k=1}^n X_k + a \frac{\sigma}{\sqrt{n}} \right],$$

la convergence suivante est alors vérifiée, où G désigne une variable aléatoire de loi normale centrée réduite :

$$\mathbb{P}(\mathbb{E}X \in \hat{I}_n(a)) \xrightarrow[n \rightarrow \infty]{} \mathbb{P}(|G| \leq a) = \frac{1}{\sqrt{2\pi}} \int_{-a}^a e^{-x^2/2} dx. \quad (1)$$

Autrement dit, il est possible de construire un intervalle aléatoire contenant la valeur inconnue $\mathbb{E}X$ avec une probabilité fixée a priori. Une valeur classique est $a = 1.96^1$, pour laquelle on a $\mathbb{P}(|G| \leq a) = 0.95$.

On remarquera que l'intervalle $\hat{I}_n(a)$ est connu si on observe les X_i et si on connaît leur variance, et que sa largeur est proportionnelle à $\frac{\sigma}{\sqrt{n}}$. Dans le cas où la variance est inconnue, on peut remplacer σ^2 par son approximation

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n \left(X_k - \frac{1}{n} \sum_{k=1}^n X_k \right)^2 = \frac{1}{n} \sum_{k=1}^n X_k^2 - \left(\frac{1}{n} \sum_{k=1}^n X_k \right)^2$$

et le résultat (1) est toujours valide.

1 Estimation de la moyenne d'une variable

On va mettre en œuvre la méthode de Monte Carlo pour chacun des trois exemples suivants :

- le cas d'une variable X de loi $\mathcal{N}(m, 1)$ où m est inconnu. En particulier, la variance est connue ;
- le cas d'une variable X de loi de Bernoulli de paramètre p . On remarquera que la variance $p(1-p)$ de X est majorée par $1/4$;
- le cas d'une variable X de loi $\mathcal{N}(m, \sigma^2)$ où m et σ^2 sont tous les deux inconnus.

Pour chacun des trois exemples, écrire un programme qui calcule un intervalle de confiance pour la moyenne de la variable considérée. Pour écrire le programme, on donnera une valeur arbitraire à la moyenne, et on calculera l'intervalle de confiance comme si la moyenne était inconnue.

Pour chaque exemple, on calculera 1000 intervalles de confiance, et on comptera parmi ces derniers combien contenaient effectivement la valeur théorique.

2 Influence de la variance

La méthode de Monte Carlo permet également de calculer des valeurs approchées d'intégrales, si l'on exprime ces intégrales sous forme d'espérances. On va ici considérer l'exemple de l'intégrale

$$\mathcal{I} = \int_0^1 \frac{\cos(t)}{t^{1/3}} dt.$$

Pour calculer \mathcal{I} par méthode de Monte Carlo, on peut utiliser :

1. Il s'agit en fait d'une valeur approchée, la valeur exacte étant 1.95996398454005...

— une variable U de loi uniforme sur $[0, 1]$, en remarquant que

$$\mathcal{I} = \mathbb{E} \left[\frac{\cos(U)}{U^{1/3}} \right] ;$$

— la variable $X = U^{3/2}$ où U est uniforme sur $[0, 1]$. On remarquera que X a pour densité $\frac{2}{3t^{1/3}} \mathbf{1}_{t \in [0,1]}$, de sorte que l'on a

$$\mathcal{I} = \mathbb{E}[3 \cos(X)/2].$$

Dans chacun des deux cas, calculer un intervalle de confiance, à partir de 1000 simulations indépendantes. Comment se comparent les deux résultats ?

3 Robustesse de la méthode

Un des principaux intérêts de la méthode de Monte Carlo est que son efficacité ne dépend *pas* de la dimension de l'intégrale que l'on cherche à calculer, ni de la régularité de la fonction intégrée, contrairement aux méthodes déterministes usuelles de calcul approché d'intégrales (méthode des trapèzes, des rectangles, ...). Par exemple, on peut utiliser la méthode de Monte Carlo pour estimer le volume de la boule unité B_d de dimension d , même pour d grand :

$$B_d = \left\{ (x_1, \dots, x_n) \in [-1, 1]^d, \sum_{k=1}^d x_k^2 < 1 \right\}.$$

La valeur théorique est

$$|B_d| = \frac{\pi^{d/2}}{\Gamma(d/2 + 1)}, \text{ ou } \Gamma(d/2 + 1) = \begin{cases} (d/2)! & \text{si } d \text{ est pair,} \\ \sqrt{\pi} 2^{-(d+1)/2} \prod_{i=1, i \text{ impair}}^d i & \text{si } d \text{ est impair.} \end{cases}$$

On peut calculer $|B_d|$ par Monte Carlo, en remarquant l'égalité

$$|B_d| = \mathbb{E}[2^d \mathbf{1}_{B_d}(Z)], \tag{2}$$

où $Z = (Z_1, \dots, Z_d)$, avec des Z_i indépendants de loi uniforme sur $[-1, 1]$. Autrement dit, Z suit la loi uniforme sur $[-1, 1]^d$. Le facteur 2^d qui apparaît dans (2) correspond au volume de $[-1, 1]^d$.

Mettre en œuvre la méthode de Monte Carlo pour calculer $|B_d|$ avec d grand (par exemple $d = 10$). Serait-il faisable dans ce cas d'utiliser une méthode déterministe ?