TP Informatique n° 15/16
Méthode de Monte-Carlo

I. Présentation de la méthode

I.1. Rappel de la loi faible des grands nombres

Loi Faible des Grands Nombres (LfGN)

- Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes de même espérance m et de même variance.
- On note $\overline{X_n} = \dfrac{1}{n} \sum\limits_{i=1}^{n} X_i$ *(moyenne empirique)*.

On a alors : $\boxed{\forall \varepsilon > 0,\ \lim\limits_{n\to+\infty} \mathbb{P}(|\overline{X_n} - m| \geqslant \varepsilon) = 0}$

Signification dans le contexte de l'estimation

Soit X une v.a.r. qui admet un moment d'ordre 2 dont on souhaite estimer la moyenne m.
On note $(X_1, \dots, X_n)$ un échantillon de n v.a.r. toutes de même loi que X.
Afin d'estimer m, on considère l'estimateur moyenne empirique $T_n = \overline{X_n}$.

▶ Qu'énonce la LfGN sur l'estimateur T_n ?

▶ Démontrer la loi faible des grands nombres.

Rappelons tout d'abord l'inégalité de Bienaymé-Tchebychev.
Soit Y une v.a.r. admettant une variance $\mathbb{V}(Y)$.

$$\forall \lambda > 0,$$

- Soit X une v.a.r. admettant un moment d'ordre 2.
  On note $m = \mathbb{E}(X)$ et $\sigma^2 = \mathbb{V}(X)$.
- Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de v.a.r. indépendantes de même espérance m et de même variance.
  × Par linéarité de l'espérance, on a :

$$\mathbb{E}(\overline{X_n}) =$$

$$=$$

× Par indépendance, on a :

$\mathbb{V}(\overline{X_n}) \ =$

$=$

Ainsi, pour tout $\varepsilon > 0$: $\mathbb{P}\big(\qquad\qquad \big) \leqslant \qquad \underset{n\to+\infty}{\longrightarrow} 0$

I.2. Méthode de Monte-Carlo

I.2.a) Principe de la méthode

La Méthode de Monte-Carlo (MMC) est simplement l'application directe de la LfGN. Plus précisément, il s'agit d'approcher m par l'estimateur $T_n = \overline{X_n}$ pour un n suffisamment grand.

Cas pratique d'utilisation

La méthode de Monte-Carlo est souvent utilisée pour déterminer une valeur approchée d'un intégrale difficile à calculer. Pour ce faire, on agit comme suit :

× on fait apparaître l'intégrale en question sous la forme d'une espérance,
(penser au théorème de transfert)

× on approche cette espérance à l'aide de la LfGN.

Si la LfGN garantit la convergence, elle ne permet d'en mesurer la rapidité. On assortit généralement le procédé ci-dessus d'une quantification des garanties d'approximation fournie par le Théorème Central Limite (TCL).

I.2.b) Illustration sur un exemple

On souhaite déterminer une valeur approchée de $I = \displaystyle\int_0^1 (\ln(t))^2 \ dt$.

Dans la suite, on note $g : t \mapsto (\ln(t))^2$.

▶ Soit U une v.a.r. telle que $U \hookrightarrow \mathcal{U}([0,1])$. Justifier que $I = \mathbb{E}(g(U))$.

- Coder la fonction `g` en **Python**.

- Par quel appel peut-on simuler `n` v.a.r. indépendantes suivant la loi uniforme sur $[0, 1]$?

- Écrire une fonction `estimME1` qui prend en paramètre un entier `n` et qui renvoie la variable `T` simulant l'estimateur moyenne empirique associé au problème.

- Tester cette fonction 8 fois de suite avec le paramètre `n = 10**2` puis avec `n = 10**3`.

- Commenter les résultats obtenus.

- Commenter l'intérêt de l'utilisation de la MMC pour cette intégrale.

I.2.c) Quantification des garanties d'approximations par le TCL

Afin de quantifier la variabilité observée précédemment, on utilise le TCL, que l'on peut énoncer comme suit.

Intervalles de confiance asymptotique via le TCL

- Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de v.a.r. identiquement distribuées, d'espérance m et de variance σ.
- Soit $\alpha \in]0,1[$. On note $u_{1-\frac{\alpha}{2}}$ le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi $\mathcal{N}(0,1)$. Par définition, $u_{1-\frac{\alpha}{2}}$ est l'unique réel tel que : $\Phi\left(u_{1-\frac{\alpha}{2}}\right) = 1-\frac{\alpha}{2}$ *i.e.* $u_{1-\frac{\alpha}{2}} = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$.
- On note $\overline{X_n} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$ *(moyenne empirique)*.

On a alors : $$\boxed{\lim_{n\to+\infty} \mathbb{P}\left(\overline{X_n} - u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leqslant m \leqslant \overline{X_n} + u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha}$$

► Rappeler l'appel permettant le calcul de $\Phi^{-1}(1-\texttt{alpha}/2)$ où `alpha` est une variable donnée.

```
from scipy.stats import norm

x0 =
```

► Calculer l'écart-type de la v.a.r. $g(U)$.

► En déduire la précision des intervalles de confiance obtenu par le TCL dans le cas `n1 = 10**2` et `n2 = 10**3` avec `alpha = 0.05`.

- Compléter le programme suivant qui simule 100 intervalles de confiance et affiche en vert ceux qui contiennent la valeur cherchée m et en rouge les autres.
  On l'appellera une première fois avec `n1 = 10**2` puis avec `n2 = 10**3` et on affichera les graphes obtenus à côté l'un de l'autre.

```
n1 = 10**2
n2 = 10**2
m = 2
N = 100
alpha = 0.05

x0 =

subplot(1,2,1)
for i in range(N) :
    T =
    eps =
    if                  :
        plt.plot([i+1, i+1], [T-eps, T+eps], 'g')
    else :
        plt.plot([i+1, i+1], [T-eps, T+eps], 'r')
plt.plot([1, N], [m, m])
# "Astuce" pour que les ordonnées soient entre 0 et 6
plt.plot([0, 0], [0, 6], "b")

subplot(1,2,2)
for i in range(N) :
    T =
    eps =
    if                  :
        plt.plot([i+1, i+1], [T-eps, T+eps], 'g')
    else :
        plt.plot([i+1, i+1], [T-eps, T+eps], 'r')
plt.plot([1, N], [m, m])
# "Astuce" pour que les ordonnées soient entre 0 et 6
plt.plot([0, 0], [0, 6], "b")
```

- Le fait de pouvoir déterminer la valeur exacte de σ vous semble-t-il raisonnable ?

I.2.d) Quantification des garanties d'approcimations par le TCL, version 2

L'utilisation de la MMC suppose que l'on ne connaît pas, de manière précise, les valeurs de m (à estimer) et de σ. On remplace alors la valeur de σ^2 par la variance empirique. Cette méthode est validée par le théorème suivant.

Intervalles de confiance asymptotique via le TCL, version 2

- Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de v.a.r. identiquement distribuées, d'espérance m et de variance σ.
- Soit $\alpha \in]0,1[$. On note $u_{1-\frac{\alpha}{2}}$ le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi $\mathcal{N}(0,1)$. Par définition, $u_{1-\frac{\alpha}{2}}$ est l'unique réel tel que : $\Phi\left(u_{1-\frac{\alpha}{2}}\right) = 1-\frac{\alpha}{2}$ *i.e.* $u_{1-\frac{\alpha}{2}} = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$.
- On note $\overline{X_n} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$ *(moyenne empirique).*
- On note ${S_n}^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n} (X_i-\overline{X_n})^2 = \dfrac{1}{n}\left(\sum\limits_{i=1}^{n} {X_i}^2\right) - \overline{X_n}^2$ *(variance empirique).*

On a alors : $\boxed{\lim\limits_{n\to+\infty} \mathbb{P}\left(\overline{X_n} - u_{1-\frac{\alpha}{2}}\dfrac{S_n}{\sqrt{n}} \leqslant m \leqslant \overline{X_n} + u_{1-\frac{\alpha}{2}}\dfrac{S_n}{\sqrt{n}}\right) = 1-\alpha}$

► Démontrer que : $\mathbb{E}(\overline{X_n}^2) = \dfrac{1}{n}\,\mathbb{E}(X_1^2) + \dfrac{n-1}{n}\,(\mathbb{E}(X_1))^2$.

- En déduire le biais de l'estimateur ${S_n}^2$.

- Compléter le programme suivant qui simule 100 intervalles de confiance et affiche en vert ceux qui contiennent la valeur cherchée m et en rouge les autres.

```
n1 = 10**2
n2 = 10**2
m = 2
N = 100
alpha = 0.05

x0 =

subplot(1,2,1)
for i in range(N) :
    T =
    S2 =
    eps =
    if                  :
        plt.plot([i+1, i+1], [T-eps, T+eps], 'g')
    else :
        plt.plot([i+1, i+1], [T-eps, T+eps], 'r')
plt.plot([1, N], [m, m])
# "Astuce" pour que les ordonnées soient entre 0 et 6
plt.plot([0, 0], [0, 6], "b")

subplot(1,2,2)
for i in range(N) :
    T =
    S2 =
    eps =
    if                  :
        plt.plot([i+1, i+1], [T-eps, T+eps], 'g')
    else :
        plt.plot([i+1, i+1], [T-eps, T+eps], 'r')
plt.plot([1, N], [m, m])
# "Astuce" pour que les ordonnées soient entre 0 et 6
plt.plot([0, 0], [0, 6], "b")
```