Correction – TD/TP 2

Simulation de variables aléatoires et applications de Monte-Carlo

M2 Finance – Université Gustave Eiffel

Exercice 1 – Inversion d'une loi mixte : un atome et une partie continue

On a

$$\mathbb{P}(X = 0) = q, \qquad \mathbb{P}(X > x \mid X > 0) = \mathrm{e}^{-\lambda x}, \qquad x \geq 0.$$

La seconde relation signifie que, conditionnellement à l'événement $\{X > 0\}$, la variable X suit une loi exponentielle de paramètre λ.

1) **Détermination de la fonction de répartition.**
Par définition,

$$F(x) = \mathbb{P}(X \leq x).$$

Comme $X \geq 0$, si $x < 0$, l'événement $\{X \leq x\}$ est impossible. Ainsi

$$F(x) = 0, \qquad x < 0.$$

Considérons maintenant $x \geq 0$. L'événement $\{X \leq x\}$ se décompose en deux événements disjoints :

$$\{X \leq x\} = \{X = 0\} \cup \{0 < X \leq x\}.$$

Donc

$$\begin{aligned} F(x) &= \mathbb{P}(X = 0) + \mathbb{P}(0 < X \leq x) \\ &= q + \mathbb{P}(X > 0)\,\mathbb{P}(X \leq x \mid X > 0). \end{aligned}$$

Or

$$\mathbb{P}(X > 0) = 1 - q,$$

et, puisque la loi conditionnelle est exponentielle,

$$\mathbb{P}(X \leq x \mid X > 0) = 1 - \mathrm{e}^{-\lambda x}, \qquad x \geq 0.$$

On obtient donc

$$\begin{aligned} F(x) &= q + (1 - q)(1 - \mathrm{e}^{-\lambda x}) \\ &= q + 1 - q - (1 - q)\mathrm{e}^{-\lambda x} \\ &= 1 - (1 - q)\mathrm{e}^{-\lambda x}. \end{aligned}$$

Finalement,

$$\boxed{F(x) = \begin{cases} 0, & x < 0, \\ 1 - (1 - q)\mathrm{e}^{-\lambda x}, & x \geq 0. \end{cases}}$$

Le saut en 0 vaut

$$F(0) - F(0^-) = q - 0 = q.$$

C'est exactement la masse ponctuelle $\mathbb{P}(X = 0) = q$.

2) **Calcul de l'inverse généralisée.**

On rappelle

$$F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}, \qquad 0 < u < 1.$$

Premier cas : $0 < u \leq q$.

On a $F(x) = 0 < u$ pour tout $x < 0$, tandis que

$$F(0) = q \geq u.$$

Le premier point où F atteint le niveau u est donc $x = 0$. Ainsi

$$\boxed{F^{-1}(u) = 0, \qquad 0 < u \leq q.}$$

Deuxième cas : $q < u < 1$.

Cette fois, $F(0) = q < u$, donc l'inverse est strictement positif. Sur $(0, +\infty)$, la fonction F est continue et strictement croissante ; on peut donc résoudre l'équation

$$F(x) = u.$$

Cela donne successivement

$$\begin{aligned} u &= 1 - (1-q)\mathrm{e}^{-\lambda x}, \\ 1 - u &= (1-q)\mathrm{e}^{-\lambda x}, \\ \frac{1-u}{1-q} &= \mathrm{e}^{-\lambda x}, \\ \log\left(\frac{1-u}{1-q}\right) &= -\lambda x, \end{aligned}$$

d'où

$$\boxed{x = -\frac{1}{\lambda}\log\left(\frac{1-u}{1-q}\right).}$$

Ainsi

$$\boxed{F^{-1}(u) = \begin{cases} 0, & 0 < u \leq q, \\ -\dfrac{1}{\lambda}\log\left(\dfrac{1-u}{1-q}\right), & q < u < 1. \end{cases}}$$

3) **Algorithme d'inversion et vérification de sa loi.**

Soit $U \sim \mathcal{U}([0,1])$. La méthode d'inversion consiste à poser $X = F^{-1}(U)$, donc

$$\boxed{X = \begin{cases} 0, & U \leq q, \\ -\dfrac{1}{\lambda}\log\left(\dfrac{1-U}{1-q}\right), & U > q. \end{cases}}$$

Vérifions d'abord la masse en zéro :

$$\mathbb{P}(X = 0) = \mathbb{P}(U \leq q) = q,$$

car U est uniforme sur $[0,1]$.

Vérifions ensuite la loi de la partie positive. Pour $x \geq 0$, sur l'événement $\{U > q\}$,

$$\begin{aligned} X > x &\iff -\frac{1}{\lambda}\log\left(\frac{1-U}{1-q}\right) > x \\ &\iff \log\left(\frac{1-U}{1-q}\right) < -\lambda x \\ &\iff \frac{1-U}{1-q} < \mathrm{e}^{-\lambda x} \\ &\iff 1 - U < (1-q)\mathrm{e}^{-\lambda x} \\ &\iff U > 1 - (1-q)\mathrm{e}^{-\lambda x}. \end{aligned}$$

Comme $1-(1-q)\mathrm{e}^{-\lambda x} \geq q$, cette dernière condition implique déjà $U > q$. Ainsi

$$\begin{aligned}\mathbb{P}(X > x) &= \mathbb{P}\big(U > 1-(1-q)\mathrm{e}^{-\lambda x}\big)\\ &= 1-\big(1-(1-q)\mathrm{e}^{-\lambda x}\big)\\ &= (1-q)\mathrm{e}^{-\lambda x}.\end{aligned}$$

Enfin,

$$\mathbb{P}(X > x \mid X > 0) = \frac{\mathbb{P}(X > x)}{\mathbb{P}(X > 0)} = \frac{(1-q)\mathrm{e}^{-\lambda x}}{1-q} = \mathrm{e}^{-\lambda x}.$$

On retrouve bien la fonction de survie d'une exponentielle de paramètre λ.

4) **Calcul de l'espérance et de la variance.**

On peut écrire la loi de X comme un mélange : avec probabilité q, $X = 0$, et avec probabilité $1-q$, X suit une exponentielle $E \sim \mathcal{E}(\lambda)$. Par la formule des probabilités totales pour l'espérance,

$$\mathbb{E}(X) = q \times 0 + (1-q)\mathbb{E}(E).$$

Pour une exponentielle de paramètre λ,

$$\mathbb{E}(E) = \frac{1}{\lambda}.$$

Donc

$$\boxed{\mathbb{E}(X) = \frac{1-q}{\lambda}.}$$

Pour la variance, calculons d'abord le second moment. On sait que

$$\mathrm{Var}(E) = \frac{1}{\lambda^2}, \qquad \mathbb{E}(E) = \frac{1}{\lambda}.$$

Donc

$$\mathbb{E}(E^2) = \mathrm{Var}(E) + \mathbb{E}(E)^2 = \frac{1}{\lambda^2} + \frac{1}{\lambda^2} = \frac{2}{\lambda^2}.$$

Par conséquent,

$$\mathbb{E}(X^2) = q \times 0^2 + (1-q)\frac{2}{\lambda^2} = \frac{2(1-q)}{\lambda^2}.$$

Ainsi

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2\\ &= \frac{2(1-q)}{\lambda^2} - \frac{(1-q)^2}{\lambda^2}\\ &= \frac{(1-q)(2-(1-q))}{\lambda^2}\\ &= \boxed{\frac{1-q^2}{\lambda^2}}.\end{aligned}$$

Idée à retenir

Le saut de la fonction de répartition n'est pas un détail gênant : il est précisément ce qui permet à l'inverse généralisée d'envoyer tout l'intervalle $(0, q]$ sur le point 0, et donc de créer un atome de masse q.

Exercice 2 – Acceptation–rejet : comparer deux lois de proposition

La densité cible est

$$f(x) = 30x(1-x)^4 \mathbf{1}_{\{0\le x\le 1\}}.$$

Dans une méthode d'acceptation–rejet, si g est la densité utilisée pour proposer les candidats, on cherche la plus petite constante c telle que

$$f(x) \le cg(x) \quad \text{pour tout } x.$$

La probabilité d'acceptation est alors $1/c$, et le nombre moyen de propositions nécessaires pour obtenir une valeur acceptée est c.

1) **Premier choix : proposition uniforme.**
Vérifions d'abord que f est une densité. Elle est positive sur $[0,1]$, et nulle ailleurs. Il reste à vérifier que son intégrale vaut 1.
On développe

$$(1-x)^4 = 1 - 4x + 6x^2 - 4x^3 + x^4.$$

Donc

$$\begin{aligned}\int_0^1 30x(1-x)^4\,\mathrm{d}x &= 30\int_0^1 \left(x - 4x^2 + 6x^3 - 4x^4 + x^5\right)\mathrm{d}x \\ &= 30\left[\frac{x^2}{2} - \frac{4x^3}{3} + \frac{6x^4}{4} - \frac{4x^5}{5} + \frac{x^6}{6}\right]_0^1 \\ &= 30\left(\frac{1}{2} - \frac{4}{3} + \frac{3}{2} - \frac{4}{5} + \frac{1}{6}\right).\end{aligned}$$

Mettons les termes au même dénominateur 30 :

$$\frac{1}{2} - \frac{4}{3} + \frac{3}{2} - \frac{4}{5} + \frac{1}{6} = \frac{15 - 40 + 45 - 24 + 5}{30} = \frac{1}{30}.$$

Ainsi

$$\int_0^1 f(x)\,\mathrm{d}x = 30 \times \frac{1}{30} = 1.$$

Donc f est bien une densité.
Pour la proposition uniforme,

$$g_0(x) = \mathbf{1}_{[0,1]}(x).$$

Sur $[0,1]$, on a donc

$$\frac{f(x)}{g_0(x)} = 30x(1-x)^4.$$

La plus petite constante possible est

$$c_0 = \max_{0\le x\le 1} 30x(1-x)^4.$$

Posons

$$h(x) = x(1-x)^4.$$

On peut dériver directement :

$$\begin{aligned}h'(x) &= (1-x)^4 + x\,4(1-x)^3(-1) \\ &= (1-x)^3\big((1-x) - 4x\big) \\ &= (1-x)^3(1-5x).\end{aligned}$$

Sur $(0,1)$, $(1-x)^3 > 0$, donc le signe de h' est celui de $1-5x$. La fonction croît jusqu'à $x = 1/5$, puis décroît. Son maximum est donc atteint en $x = 1/5$.
On obtient

$$\begin{aligned} c_0 &= 30 \times \frac{1}{5}\left(1-\frac{1}{5}\right)^4 \\ &= 6\left(\frac{4}{5}\right)^4 \\ &= 6\frac{256}{625} \\ &= \boxed{\frac{1536}{625} \simeq 2.4576}. \end{aligned}$$

La probabilité théorique d'acceptation est donc

$$\boxed{\frac{1}{c_0} = \frac{625}{1536} \simeq 0.4069.}$$

Autrement dit, environ 40.7% des propositions sont acceptées. Le nombre moyen de propositions nécessaires pour une valeur acceptée est

$$\boxed{c_0 \simeq 2.46.}$$

2) **Deuxième choix : la densité $g_1(x) = 4(1-x)^3\mathbf{1}_{[0,1]}(x)$.**
Vérifions que g_1 est une densité. Elle est positive, et

$$\begin{aligned} \int_0^1 4(1-x)^3\,\mathrm{d}x &= 4\left[-\frac{(1-x)^4}{4}\right]_0^1 \\ &= \left[-(1-x)^4\right]_0^1 \\ &= 0-(-1) = 1. \end{aligned}$$

Sa fonction de répartition, pour $0 \le x \le 1$, est

$$\begin{aligned} G_1(x) &= \int_0^x 4(1-t)^3\,\mathrm{d}t \\ &= \left[-(1-t)^4\right]_{t=0}^{t=x} \\ &= -(1-x)^4 + 1 \\ &= \boxed{1-(1-x)^4}. \end{aligned}$$

Bien sûr, $G_1(x) = 0$ pour $x < 0$ et $G_1(x) = 1$ pour $x \ge 1$.
Pour simuler par inversion, on part de $U \sim \mathcal{U}([0,1])$ et on résout

$$u = 1-(1-x)^4.$$

Alors

$$\begin{aligned} 1-u &= (1-x)^4, \\ (1-u)^{1/4} &= 1-x, \\ x &= 1-(1-u)^{1/4}. \end{aligned}$$

Donc

$$\boxed{Y = 1-(1-U)^{1/4}}$$

a pour densité g_1. Comme $1-U$ est encore uniforme, on peut aussi écrire

$$\boxed{Y = 1-U^{1/4}}.$$

3) **Constante d'enveloppe pour le second choix.**
Pour $0 < x < 1$,

$$\begin{aligned}\frac{f(x)}{g_1(x)} &= \frac{30x(1-x)^4}{4(1-x)^3} \\ &= \frac{30}{4}x(1-x) \\ &= \frac{15}{2}x(1-x).\end{aligned}$$

La plus petite constante c_1 est donc le maximum de cette fonction sur $[0,1]$.
Posons $k(x) = x(1-x) = x - x^2$. Alors

$$k'(x) = 1 - 2x,$$

donc le maximum est atteint en $x = 1/2$. Sa valeur est

$$k\left(\frac{1}{2}\right) = \frac{1}{2}\left(1 - \frac{1}{2}\right) = \frac{1}{4}.$$

Ainsi

$$\boxed{c_1 = \frac{15}{2} \times \frac{1}{4} = \frac{15}{8} = 1.875.}$$

Le taux d'acceptation est donc

$$\boxed{\frac{1}{c_1} = \frac{8}{15} \simeq 0.5333.}$$

Environ 53.3% des valeurs proposées sont acceptées.

4) **Comparaison quantitative.**
Pour la proposition uniforme,

$$c_0 \simeq 2.4576, \qquad \frac{1}{c_0} \simeq 0.4069.$$

Pour la seconde proposition,

$$c_1 = 1.875, \qquad \frac{1}{c_1} \simeq 0.5333.$$

On peut résumer :

	uniforme g_0	$g_1(x) = 4(1-x)^3$
Constante d'enveloppe c	2.4576	1.875
Taux d'acceptation $1/c$	40.7%	53.3%
Propositions moyennes par valeur acceptée	2.4576	1.875

La seconde loi de proposition est donc plus efficace. Le nombre moyen de propositions diminue relativement de

$$\begin{aligned}\frac{c_0 - c_1}{c_0} &= 1 - \frac{c_1}{c_0} \\ &= 1 - \frac{1.875}{2.4576} \\ &\simeq 0.237.\end{aligned}$$

Il y a donc environ $\boxed{24\%}$ de propositions en moins par valeur acceptée.
Ce gain s'explique par la forme des densités : g_1 concentre davantage de masse dans la région où f est elle-même importante, ce qui permet d'utiliser une enveloppe plus serrée, donc une constante c plus petite.

Idée à retenir

Dans l'acceptation–rejet, une petite constante c signifie que l'enveloppe cg est proche de la densité cible f. C'est exactement ce que l'on veut : moins de zone "inutile" entre f et cg, donc moins de rejets.

Exercice 3 – Simulation gaussienne par la méthode polaire

On note

$$\mathcal{A} = \{0 < U^2 + V^2 < 1\}$$

l'événement d'acceptation. Comme le point $(0,0)$ a probabilité nulle, on pourrait remplacer cette définition par $\{U^2+V^2 < 1\}$ sans rien changer aux calculs probabilistes. La phrase "après acceptation" signifie que toutes les lois considérées sont des lois conditionnelles sachant $\mathcal{A}$.
On pose

$$R = \sqrt{U^2 + V^2}, \qquad \Theta = \mathrm{Arg}(U + \mathrm{i}V) \in [0, 2\pi), \qquad S = R^2.$$

Ainsi Θ est, par définition, l'angle polaire du point accepté.

1) **Loi du point accepté.**
Avant acceptation, (U,V) est uniforme sur le carré $[-1,1]^2$, donc sa densité est

$$f_{U,V}(u,v) = \frac{1}{4}\mathbf{1}_{[-1,1]^2}(u,v).$$

Le disque unité est

$$D = \{(u,v) : u^2 + v^2 < 1\}.$$

Sa probabilité d'acceptation vaut

$$\mathbb{P}(\mathcal{A}) = \frac{\mathrm{Aire}(D)}{\mathrm{Aire}([-1,1]^2)} = \frac{\pi}{4}.$$

Soit maintenant $B \subset D$ mesurable. Alors

$$\mathbb{P}((U,V) \in B \mid \mathcal{A}) = \frac{\mathbb{P}((U,V) \in B, \mathcal{A})}{\mathbb{P}(\mathcal{A})}.$$

Comme $B \subset D$, l'événement $\{(U,V) \in B\}$ implique déjà $\mathcal{A}$. Donc

$$\mathbb{P}((U,V) \in B, \mathcal{A}) = \mathbb{P}((U,V) \in B) = \frac{\mathrm{Aire}(B)}{4}.$$

Ainsi

$$\mathbb{P}((U,V) \in B \mid \mathcal{A}) = \frac{\mathrm{Aire}(B)/4}{\pi/4} = \frac{\mathrm{Aire}(B)}{\pi}.$$

C'est exactement la loi uniforme sur le disque. Sa densité conditionnelle est donc

$$\boxed{f_{U,V|\mathcal{A}}(u,v) = \frac{1}{\pi}\mathbf{1}_{\{u^2+v^2<1\}}.}$$

2) **Passage en coordonnées polaires : loi jointe de (R,Θ).**
Le changement de variables polaires est

$$T : (r,\theta) \longmapsto (u,v) = (r\cos\theta, r\sin\theta),$$

avec $0 < r < 1$ et $0 \le \theta < 2\pi$. Son jacobien est

$$J_T(r,\theta) = \det\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} = r(\cos^2\theta + \sin^2\theta) = r.$$

Par la formule de changement de variables, la densité jointe conditionnelle de (R, Θ) vaut donc

$$\begin{aligned} f_{R,\Theta|\mathcal{A}}(r,\theta) &= f_{U,V|\mathcal{A}}(r\cos\theta, r\sin\theta)\,|J_T(r,\theta)| \\ &= \frac{r}{\pi}\,\mathbf{1}_{\{0<r<1\}}\mathbf{1}_{\{0\le\theta<2\pi\}}. \end{aligned}$$

Calculons maintenant les densités marginales. Pour $0 < r < 1$,

$$\begin{aligned} f_{R|\mathcal{A}}(r) &= \int_0^{2\pi} f_{R,\Theta|\mathcal{A}}(r,\theta)\,\mathrm{d}\theta \\ &= \int_0^{2\pi} \frac{r}{\pi}\,\mathrm{d}\theta = 2r. \end{aligned}$$

Donc

$$\boxed{f_{R|\mathcal{A}}(r) = 2r\mathbf{1}_{(0,1)}(r).}$$

Pour $0 \le \theta < 2\pi$,

$$\begin{aligned} f_{\Theta|\mathcal{A}}(\theta) &= \int_0^1 f_{R,\Theta|\mathcal{A}}(r,\theta)\,\mathrm{d}r \\ &= \int_0^1 \frac{r}{\pi}\,\mathrm{d}r = \frac{1}{2\pi}. \end{aligned}$$

Ainsi

$$\boxed{\Theta \mid \mathcal{A} \sim \mathcal{U}([0,2\pi)).}$$

Enfin,

$$\begin{aligned} f_{R|\mathcal{A}}(r) f_{\Theta|\mathcal{A}}(\theta) &= (2r)\left(\frac{1}{2\pi}\right) = \frac{r}{\pi} \\ &= f_{R,\Theta|\mathcal{A}}(r,\theta) \end{aligned}$$

sur le support $(0,1)\times[0,2\pi)$. Comme la densité jointe est le produit des deux densités marginales,

$$\boxed{R \text{ et } \Theta \text{ sont indépendants conditionnellement à } \mathcal{A}.}$$

3) **Loi de $S = R^2$, puis de $\rho = \sqrt{-2\log S}$.**
Pour $0 \le s \le 1$,

$$\begin{aligned} \mathbb{P}(S \le s \mid \mathcal{A}) &= \mathbb{P}(R^2 \le s \mid \mathcal{A}) \\ &= \mathbb{P}(R \le \sqrt{s} \mid \mathcal{A}) \\ &= \int_0^{\sqrt{s}} 2r\,\mathrm{d}r \\ &= s. \end{aligned}$$

Donc

$$\boxed{S \mid \mathcal{A} \sim \mathcal{U}([0,1]).}$$

Posons maintenant

$$\rho = \sqrt{-2\log S}.$$

Pour $a \ge 0$,

$$\begin{aligned} \mathbb{P}(\rho \le a \mid \mathcal{A}) &= \mathbb{P}(\sqrt{-2\log S} \le a \mid \mathcal{A}) \\ &= \mathbb{P}(-2\log S \le a^2 \mid \mathcal{A}) \\ &= \mathbb{P}(\log S \ge -a^2/2 \mid \mathcal{A}) \\ &= \mathbb{P}(S \ge e^{-a^2/2} \mid \mathcal{A}) \\ &= 1 - e^{-a^2/2}, \end{aligned}$$

car S est uniforme sur $[0,1]$ sous la loi conditionnelle. En dérivant cette fonction de répartition,

$$\boxed{f_{\rho|\mathcal{A}}(r) = re^{-r^2/2}\mathbf{1}_{\{r\geq 0\}}.}$$

Il s'agit simplement de la densité de la variable réelle positive ρ. Nous n'avons pas encore conclu que (X,Y) est gaussien : cette conclusion viendra du calcul de la densité jointe cartésienne à la question suivante.

Comme ρ est une fonction mesurable de $S = R^2$, donc de R, et que R est indépendant de Θ conditionnellement à $\mathcal{A}$, on a aussi

$$\boxed{\rho \text{ et } \Theta \text{ sont indépendants conditionnellement à } \mathcal{A}.}$$

Par conséquent,

$$f_{\rho,\Theta|\mathcal{A}}(r,\theta) = \frac{1}{2\pi} re^{-r^2/2}\mathbf{1}_{\{r\geq 0\}}\mathbf{1}_{\{0\leq\theta<2\pi\}}.$$

4) **Densité jointe de (X,Y).**

On a

$$A = \sqrt{\frac{-2\log S}{S}} = \frac{\sqrt{-2\log S}}{R} = \frac{\rho}{R}.$$

Comme

$$U = R\cos\Theta, \qquad V = R\sin\Theta,$$

il vient

$$X = AU = \rho\cos\Theta, \qquad Y = AV = \rho\sin\Theta.$$

Le passage de (ρ,Θ) à (X,Y) est donc encore le changement de variables polaires. Pour $(x,y) \neq (0,0)$, posons

$$r = \sqrt{x^2+y^2}.$$

Le jacobien direct $(r,\theta) \mapsto (x,y)$ vaut r, donc le jacobien inverse vaut $1/r$. Ainsi

$$\begin{aligned} f_{X,Y}(x,y) &= f_{\rho,\Theta|\mathcal{A}}(r,\theta)\frac{1}{r} \\ &= \frac{1}{2\pi} re^{-r^2/2}\frac{1}{r} \\ &= \frac{1}{2\pi} e^{-(x^2+y^2)/2}. \end{aligned}$$

On factorise :

$$\frac{1}{2\pi} e^{-(x^2+y^2)/2} = \left(\frac{1}{\sqrt{2\pi}} e^{-x^2/2}\right)\left(\frac{1}{\sqrt{2\pi}} e^{-y^2/2}\right).$$

On reconnaît le produit de deux densités de loi $\mathcal{N}(0,1)$. Par conséquent,

$$\boxed{X \sim \mathcal{N}(0,1), \qquad Y \sim \mathcal{N}(0,1), \qquad X \perp Y.}$$

Autrement dit,

$$\boxed{(X,Y) \sim \mathcal{N}(0, I_2).}$$

5) **Implémentation.**

Une implémentation possible est :

```
import numpy as np
rng = np.random.default_rng(123)

values = []
while len(values) < 50_000:
    u, v = rng.uniform(-1.0, 1.0, size=2)
```

```
    s = u*u + v*v
    if 0.0 < s < 1.0:
        a = np.sqrt(-2.0*np.log(s)/s)
        values.append((a*u, a*v))

xy = np.asarray(values)
print("moyennes =", xy.mean(axis=0))
print("matrice de covariance =")
print(np.cov(xy.T))
```

On doit observer des moyennes proches de 0, des variances proches de 1 et une covariance proche de 0. Ces vérifications numériques ne constituent pas la preuve ; elles servent seulement à contrôler l'implémentation.

Idée à retenir

La preuve comporte deux changements de variables polaires distincts. Le premier décrit la loi du point uniforme accepté dans le disque. Le second transforme le couple (ρ, Θ) en coordonnées cartésiennes et permet de reconnaître directement la densité jointe d'un couple de normales indépendantes.

Exercice 4 – Choisir une loi de simulation : un exemple calculable

On considère

$$J = \int_0^{+\infty} \sqrt{x}\, e^{-x}\, \mathrm{d}x.$$

1) Si $X \sim \mathcal{E}(1)$, sa densité est

$$p_X(x) = e^{-x} \mathbf{1}_{\{x \geq 0\}}.$$

Donc

$$\begin{aligned} \mathbb{E}(\sqrt{X}) &= \int_0^{+\infty} \sqrt{x}\, p_X(x)\, \mathrm{d}x \\ &= \int_0^{+\infty} \sqrt{x}\, e^{-x}\, \mathrm{d}x \\ &= J. \end{aligned}$$

Ainsi, si $X_1, \ldots, X_n$ sont i.i.d. de loi $\mathcal{E}(1)$,

$$\boxed{\widehat{J}_n^{(1)} = \frac{1}{n} \sum_{k=1}^{n} \sqrt{X_k}}$$

est un estimateur sans biais de J.

2) La fonction Gamma est définie par

$$\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x}\, \mathrm{d}x, \qquad a > 0.$$

Ici $x^{1/2} = x^{3/2-1}$, donc

$$J = \Gamma\left(\frac{3}{2}\right).$$

En utilisant la relation $\Gamma(a+1) = a\Gamma(a)$ et $\Gamma(1/2) = \sqrt{\pi}$,

$$\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}.$$

Ainsi

$$\boxed{J = \frac{\sqrt{\pi}}{2}.}$$

3) On cherche $\mathrm{Var}(\sqrt{X})$. Par définition,

$$\mathrm{Var}(\sqrt{X}) = \mathbb{E}((\sqrt{X})^2) - \mathbb{E}(\sqrt{X})^2.$$

Mais $(\sqrt{X})^2 = X$, donc

$$\mathbb{E}((\sqrt{X})^2) = \mathbb{E}(X) = 1$$

pour une exponentielle de paramètre 1. D'autre part,

$$\mathbb{E}(\sqrt{X}) = J = \frac{\sqrt{\pi}}{2}.$$

Par conséquent,

$$\boxed{\mathrm{Var}(\sqrt{X}) = 1 - \frac{\pi}{4}.}$$

La variance de la moyenne Monte-Carlo vaut donc

$$\boxed{\mathrm{Var}(\widehat{J}_n^{(1)}) = \frac{1}{n}\left(1 - \frac{\pi}{4}\right).}$$

4) Soit maintenant $X_\lambda \sim \mathcal{E}(\lambda)$. Sa densité est

$$p_\lambda(x) = \lambda e^{-\lambda x} \mathbf{1}_{\{x \geq 0\}}.$$

On veut réécrire J comme une espérance sous cette nouvelle loi. On multiplie et divise l'intégrande par $p_\lambda(x)$:

$$\begin{aligned} J &= \int_0^{+\infty} \sqrt{x}\, e^{-x}\, \mathrm{d}x \\ &= \int_0^{+\infty} \frac{\sqrt{x}\, e^{-x}}{\lambda e^{-\lambda x}} \lambda e^{-\lambda x}\, \mathrm{d}x \\ &= \int_0^{+\infty} \frac{1}{\lambda} \sqrt{x}\, e^{-(1-\lambda)x}\, p_\lambda(x)\, \mathrm{d}x. \end{aligned}$$

Donc

$$\boxed{J = \mathbb{E}\left[\frac{1}{\lambda}\sqrt{X_\lambda}\, e^{-(1-\lambda)X_\lambda}\right].}$$

5) Posons

$$Y_\lambda = \frac{1}{\lambda}\sqrt{X_\lambda}\, e^{-(1-\lambda)X_\lambda}.$$

Son second moment est

$$\begin{aligned} \mathbb{E}(Y_\lambda^2) &= \int_0^{+\infty} \left(\frac{1}{\lambda}\sqrt{x}\, e^{-(1-\lambda)x}\right)^2 \lambda e^{-\lambda x}\, \mathrm{d}x \\ &= \frac{1}{\lambda} \int_0^{+\infty} x\, e^{-2(1-\lambda)x} e^{-\lambda x}\, \mathrm{d}x \\ &= \frac{1}{\lambda} \int_0^{+\infty} x\, e^{-(2-\lambda)x}\, \mathrm{d}x. \end{aligned}$$

Cette intégrale est finie si et seulement si

$$2 - \lambda > 0.$$

Comme $\lambda > 0$, la condition est

$$\boxed{0 < \lambda < 2.}$$

Pour $a > 0$, une intégration par parties donne

$$\begin{aligned}\int_0^{+\infty} xe^{-ax}\,\mathrm{d}x &= \left[-\frac{x}{a}e^{-ax}\right]_0^{+\infty} + \frac{1}{a}\int_0^{+\infty} e^{-ax}\,\mathrm{d}x \\ &= 0 + \frac{1}{a}\times\frac{1}{a} \\ &= \frac{1}{a^2}.\end{aligned}$$

Avec $a = 2-\lambda$,

$$\boxed{\mathbb{E}(Y_\lambda^2) = \frac{1}{\lambda(2-\lambda)^2}.}$$

Comme $\mathbb{E}(Y_\lambda) = J = \sqrt{\pi}/2$,

$$\boxed{\mathrm{Var}(Y_\lambda) = \frac{1}{\lambda(2-\lambda)^2} - \frac{\pi}{4}.}$$

6) Comme $J^2 = \pi/4$ ne dépend pas de λ, minimiser la variance revient à minimiser

$$\frac{1}{\lambda(2-\lambda)^2},$$

donc à maximiser

$$h(\lambda) = \lambda(2-\lambda)^2, \qquad 0 < \lambda < 2.$$

On dérive :

$$\begin{aligned}h'(\lambda) &= (2-\lambda)^2 - 2\lambda(2-\lambda) \\ &= (2-\lambda)\big((2-\lambda) - 2\lambda\big) \\ &= (2-\lambda)(2-3\lambda).\end{aligned}$$

Dans l'intervalle $(0,2)$, le seul point critique intérieur est

$$\lambda^\star = \frac{2}{3}.$$

Comme $h(\lambda) \to 0$ lorsque $\lambda \to 0$ ou $\lambda \to 2$, ce point critique est bien le maximum global. Ainsi

$$\boxed{\lambda^\star = \frac{2}{3}.}$$

Pour cette valeur,

$$\begin{aligned}\mathbb{E}(Y_{\lambda^\star}^2) &= \frac{1}{(2/3)(4/3)^2} \\ &= \frac{1}{(2/3)(16/9)} \\ &= \frac{27}{32}.\end{aligned}$$

La variance minimale vaut donc

$$\boxed{\frac{27}{32} - \frac{\pi}{4} \simeq 0.0584.}$$

Pour $\lambda = 1$, la variance était

$$1 - \frac{\pi}{4} \simeq 0.2146.$$

Le choix $\lambda = 2/3$ divise donc la variance par environ

$$\frac{0.2146}{0.0584} \simeq 3.7.$$

C'est un premier exemple de l'idée centrale du chapitre suivant : changer intelligemment la loi de simulation peut réduire fortement la variance sans changer la quantité estimée.

Exercice 5 – Estimation d'un paramètre exponentiel et méthode delta

Soient $X_1, \dots, X_n$ i.i.d. de loi exponentielle de paramètre $\lambda > 0$, de densité

$$f_\lambda(x) = \lambda e^{-\lambda x} \mathbf{1}_{\{x \geq 0\}}.$$

On pose

$$\overline{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k, \qquad \widehat{\lambda}_n = \frac{1}{\overline{X}_n}.$$

1) **Moments de la loi exponentielle et TCL.**
On calcule d'abord

$$\mathbb{E}(X_1) = \int_0^{+\infty} x\lambda e^{-\lambda x} \, \mathrm{d}x = \frac{1}{\lambda},$$

et

$$\mathrm{Var}(X_1) = \frac{1}{\lambda^2}.$$

Par le théorème central limite,

$$\sqrt{n} \left(\overline{X}_n - \frac{1}{\lambda} \right) \Longrightarrow \mathcal{N}\left(0, \frac{1}{\lambda^2}\right).$$

2) **Application de la méthode delta.**
On considère

$$\varphi(x) = \frac{1}{x}.$$

La valeur autour de laquelle on linéarise est

$$\theta = \mathbb{E}(X_1) = \frac{1}{\lambda}.$$

La dérivée est

$$\varphi'(x) = -\frac{1}{x^2}, \qquad \varphi'\left(\frac{1}{\lambda}\right) = -\lambda^2.$$

La méthode delta donne donc

$$\begin{aligned} \sqrt{n}\left(\widehat{\lambda}_n - \lambda\right) &= \sqrt{n}\left(\varphi(\overline{X}_n) - \varphi(1/\lambda)\right) \\ &\Longrightarrow \mathcal{N}\left(0, \left(-\lambda^2\right)^2 \frac{1}{\lambda^2}\right) \\ &= \mathcal{N}(0, \lambda^2). \end{aligned}$$

Ainsi, pour n grand,

$$\widehat{\lambda}_n \approx \mathcal{N}\left(\lambda, \frac{\lambda^2}{n}\right).$$

L'écart-type asymptotique dépend encore de λ, inconnu. On le remplace naturellement par $\widehat{\lambda}_n$. Un intervalle de confiance asymptotique à 95% est donc

$$\boxed{\left[\widehat{\lambda}_n - 1.96\frac{\widehat{\lambda}_n}{\sqrt{n}}, \widehat{\lambda}_n + 1.96\frac{\widehat{\lambda}_n}{\sqrt{n}}\right].}$$

3) **Biais exact à taille finie.**

Posons

$$T_n = \sum_{k=1}^{n} X_k.$$

Comme $\overline{X}_n = T_n/n$,

$$\widehat{\lambda}_n = \frac{n}{T_n}.$$

On admet que T_n a pour densité

$$f_{T_n}(t) = \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t} \mathbf{1}_{\{t>0\}}.$$

Pour $n \geq 2$,

$$\begin{aligned}\mathbb{E}\left(\frac{1}{T_n}\right) &= \int_0^{+\infty} \frac{1}{t} \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t} \, \mathrm{d}t \\ &= \frac{\lambda^n}{\Gamma(n)} \int_0^{+\infty} t^{n-2} e^{-\lambda t} \, \mathrm{d}t.\end{aligned}$$

Effectuons le changement de variable $u = \lambda t$, donc $t = u/\lambda$ et $\mathrm{d}t = \mathrm{d}u/\lambda$. Alors

$$\begin{aligned}\int_0^{+\infty} t^{n-2} e^{-\lambda t} \, \mathrm{d}t &= \frac{1}{\lambda^{n-1}} \int_0^{+\infty} u^{n-2} e^{-u} \, \mathrm{d}u \\ &= \frac{\Gamma(n-1)}{\lambda^{n-1}}.\end{aligned}$$

Par conséquent,

$$\mathbb{E}\left(\frac{1}{T_n}\right) = \lambda \frac{\Gamma(n-1)}{\Gamma(n)}.$$

Pour un entier n, $\Gamma(n) = (n-1)\Gamma(n-1)$, donc

$$\boxed{\mathbb{E}(1/T_n) = \frac{\lambda}{n-1}.}$$

Ainsi

$$\mathbb{E}(\widehat{\lambda}_n) = n\mathbb{E}(1/T_n) = \frac{n}{n-1}\lambda.$$

Donc $\widehat{\lambda}_n$ est biaisé vers le haut :

$$\boxed{\mathrm{Biais}(\widehat{\lambda}_n) = \frac{\lambda}{n-1}.}$$

Ce biais tend vers 0 lorsque $n \to \infty$, ce qui est compatible avec la consistance et la méthode delta.

4) **Correction exacte du biais.**

Puisque

$$\mathbb{E}(\widehat{\lambda}_n) = \frac{n}{n-1}\lambda,$$

il suffit de multiplier $\widehat{\lambda}_n$ par $(n-1)/n$. On définit

$$\boxed{\widetilde{\lambda}_n = \frac{n-1}{n}\widehat{\lambda}_n = \frac{n-1}{T_n}.}$$

Alors

$$\mathbb{E}(\widetilde{\lambda}_n) = \frac{n-1}{n}\frac{n}{n-1}\lambda = \lambda.$$

L'estimateur $\widetilde{\lambda}_n$ est donc exactement sans biais.

5) **Variance exacte de l'estimateur corrigé.**
Pour $n \geq 3$,

$$\begin{aligned}\mathbb{E}\left(\frac{1}{T_n^2}\right) &= \int_0^{+\infty} \frac{1}{t^2}\frac{\lambda^n}{\Gamma(n)} t^{n-1}e^{-\lambda t}\,\mathrm{d}t \\ &= \frac{\lambda^n}{\Gamma(n)}\int_0^{+\infty} t^{n-3}e^{-\lambda t}\,\mathrm{d}t \\ &= \frac{\lambda^n}{\Gamma(n)}\frac{\Gamma(n-2)}{\lambda^{n-2}} \\ &= \lambda^2\frac{\Gamma(n-2)}{\Gamma(n)}.\end{aligned}$$

Or

$$\Gamma(n) = (n-1)(n-2)\Gamma(n-2),$$

d'où

$$\boxed{\mathbb{E}\left(\frac{1}{T_n^2}\right) = \frac{\lambda^2}{(n-1)(n-2)}.}$$

Comme $\widetilde{\lambda}_n = (n-1)/T_n$,

$$\begin{aligned}\mathrm{Var}(\widetilde{\lambda}_n) &= (n-1)^2\,\mathrm{Var}(1/T_n) \\ &= (n-1)^2\left[\mathbb{E}(1/T_n^2) - \mathbb{E}(1/T_n)^2\right] \\ &= (n-1)^2\left[\frac{\lambda^2}{(n-1)(n-2)} - \frac{\lambda^2}{(n-1)^2}\right] \\ &= \frac{\lambda^2}{n-2}.\end{aligned}$$

Donc

$$\boxed{\mathrm{Var}(\widetilde{\lambda}_n) = \frac{\lambda^2}{n-2}.}$$

Pour n grand,

$$\frac{\lambda^2}{n-2} \sim \frac{\lambda^2}{n},$$

ce qui est exactement du même ordre que la variance asymptotique λ^2/n obtenue par la méthode delta.

Idée à retenir

La méthode delta décrit le comportement à l'échelle $n^{-1/2}$, mais elle ne dit pas que l'estimateur est sans biais pour n fini. Ici, le calcul exact montre un biais de taille $1/n$, donc plus petit que les fluctuations statistiques de taille $1/\sqrt{n}$.

Exercice 6 – TP Python : maximum d'exponentielles et simulation en grande dimension

Soient $X_1, \ldots, X_d$ indépendantes de loi exponentielle de paramètre 1, et

$$M_d = \max(X_1, \ldots, X_d).$$

Étape 1 : loi exacte de M_d

1) Pour $x < 0$, on a évidemment $\mathbb{P}(M_d \leq x) = 0$. Pour $x \geq 0$,

$$\begin{aligned}\mathbb{P}(M_d \leq x) &= \mathbb{P}(X_1 \leq x, \dots, X_d \leq x) \\ &= \prod_{j=1}^{d} \mathbb{P}(X_j \leq x) \qquad \text{(indépendance)} \\ &= \left(1 - e^{-x}\right)^d.\end{aligned}$$

Ainsi

$$\boxed{F_{M_d}(x) = \left(1 - e^{-x}\right)^d, \qquad x \geq 0.}$$

Pour inverser cette fonction, posons $u \in (0,1)$ et résolvons

$$u = \left(1 - e^{-x}\right)^d.$$

On obtient successivement

$$u^{1/d} = 1 - e^{-x}, \qquad e^{-x} = 1 - u^{1/d},$$

puis

$$\boxed{F_{M_d}^{-1}(u) = -\log\left(1 - u^{1/d}\right).}$$

Donc, si $U \sim \mathcal{U}([0,1])$,

$$\boxed{M_d = -\log\left(1 - U^{1/d}\right)}$$

a exactement la loi du maximum de d exponentielles indépendantes.

2) Pour une variable positive, on utilise l'identité

$$\mathbb{E}(M_d) = \int_0^{+\infty} \mathbb{P}(M_d > x)\,\mathrm{d}x.$$

Ici,

$$\mathbb{P}(M_d > x) = 1 - \left(1 - e^{-x}\right)^d.$$

Donc

$$\mathbb{E}(M_d) = \int_0^{+\infty} \left[1 - \left(1 - e^{-x}\right)^d\right] \mathrm{d}x.$$

Posons

$$t = e^{-x}.$$

Alors $x = -\log t$, $\mathrm{d}x = -\mathrm{d}t/t$, et lorsque x va de 0 à $+\infty$, t va de 1 à 0. Ainsi

$$\begin{aligned}\mathbb{E}(M_d) &= \int_1^0 \left[1 - (1-t)^d\right] \left(-\frac{\mathrm{d}t}{t}\right) \\ &= \int_0^1 \frac{1 - (1-t)^d}{t}\,\mathrm{d}t.\end{aligned}$$

Or, par la formule de la somme géométrique,

$$1 - (1-t)^d = t \sum_{j=0}^{d-1} (1-t)^j.$$

Donc

$$\begin{aligned}\mathbb{E}(M_d) &= \sum_{j=0}^{d-1} \int_0^1 (1-t)^j \, \mathrm{d}t \\ &= \sum_{j=0}^{d-1} \frac{1}{j+1} \\ &= \sum_{k=1}^{d} \frac{1}{k}.\end{aligned}$$

Finalement,

$$\boxed{\mathbb{E}(M_d) = H_d := 1 + \frac{1}{2} + \cdots + \frac{1}{d}.}$$

Étape 2 : deux algorithmes de simulation

Une implémentation possible est la suivante.

```
import numpy as np

rng = np.random.default_rng(321)


def sample_max_method_A(n, d):
    u = rng.random((n, d))
    x = -np.log(u)          # n*d exponentielles E(1)
    return x.max(axis=1)


def sample_max_method_B(n, d):
    u = rng.random(n)
    return -np.log(1.0 - u**(1.0/d))


def harmonic_number(d):
    return np.sum(1.0 / np.arange(1, d + 1))
```

3) Pour $d = 10$ ou $d = 100$, les deux méthodes doivent produire des moyennes empiriques proches de

$$H_d = \sum_{k=1}^{d} \frac{1}{k}.$$

Par exemple :

```
for d in [10, 100]:
    exact = harmonic_number(d)
    a = sample_max_method_A(50_000, d)
    b = sample_max_method_B(50_000, d)
    print(d, exact, a.mean(), b.mean())
```

Les deux algorithmes simulent exactement la même loi. Les différences numériques observées sont uniquement dues au hasard Monte-Carlo.

4) Pour un échantillon $Y_1, \ldots, Y_n$ simulé par l'une ou l'autre méthode, on calcule

$$\overline{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i, \qquad S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y}_n)^2.$$

L'intervalle asymptotique à 95% est

$$\boxed{\overline{Y}_n \pm 1.96 \frac{S_n}{\sqrt{n}}.}$$

Comme la valeur exacte est H_d, on peut directement vérifier si H_d appartient à l'intervalle. Une fonction possible est :

```
def ci95(values):
    n = len(values)
    mean = np.mean(values)
    s = np.std(values, ddof=1)
    half = 1.96 * s / np.sqrt(n)
    return mean, mean-half, mean+half
```

Il est normal qu'un intervalle particulier puisse manquer H_d : le niveau 95% est une propriété de répétition de la procédure, pas une garantie pour chaque expérience.

5) La différence de coût vient du nombre de variables uniformes générées.

La méthode A utilise d uniformes pour produire une réalisation de M_d. Pour n réalisations, elle génère donc nd uniformes et construit un tableau de taille $n \times d$.

La méthode B exploite la fonction de répartition connue de M_d et utilise une seule uniforme par réalisation. Pour n réalisations, elle ne génère que n uniformes.

Ainsi, du point de vue du nombre de tirages aléatoires,

$$\text{coût A} \propto nd, \qquad \text{coût B} \propto n.$$

Le gain devient important lorsque d est grand. Cela ne change pas le taux statistique $1/\sqrt{n}$ de la moyenne : cela réduit le coût nécessaire pour produire chacune des n observations.

6) Pour étudier la couverture, on peut écrire :

```
def coverage_experiment(d=20, n=1000, m=500):
    exact = harmonic_number(d)
    covered = 0

    for _ in range(m):
        values = sample_max_method_B(n, d)
        mean, lower, upper = ci95(values)
        covered += (lower <= exact <= upper)

    return covered / m

print(coverage_experiment())
```

La fréquence observée fluctue autour de 0.95. Avec seulement $m = 500$ répétitions, elle n'a aucune raison d'être exactement égale à 0.95. En augmentant m, la fréquence empirique de couverture se stabilise autour du niveau théorique si l'approximation normale est suffisamment bonne pour la taille n choisie.

Idée à retenir

Les méthodes A et B ont exactement la même loi de sortie. La différence est algorithmique : la méthode B utilise la structure analytique de la loi du maximum pour éviter de simuler explicitement les d composantes. C'est un exemple simple où une meilleure représentation probabiliste réduit fortement le coût de simulation.