TP 1 Computational Statistics

Matthieu Boyer

21 octobre 2025

1 Exercice 1

1.1 Question 1

Par définition de l'espérance, on a :

$$
\begin{aligned}
\mathbb{E}_\theta(X_1) &= \int_{\mathbb{R}} x p_\theta(x)\,\mathrm{d}x \\
&= \int_{\mathbb{R}} x \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x)\,\mathrm{d}x \\
&= \int_0^\theta x \frac{1}{\theta}\,\mathrm{d}x \\
&= \left.\frac{x^2}{2\theta}\right|_{x=0}^{x=\theta} \\
&= \frac{\theta^2}{2\theta} = \frac{\theta}{2}
\end{aligned}
$$

On remarque que ce calcul est indépendant de la variable aléatoire d'entrée considérée.

En se donnant une seule (puisqu'on n'a qu'un seul paramètre) fonction T mesurable avec un premier moment possédant une forme fermée $e(\theta)$, on va estimer θ par la solution de l'équation suivante :

$$
e(\theta) = \frac{1}{n} \sum_{i=1}^{n} T(X_i)
$$

En considérant par exemple, $T = \mathrm{id}$, qui vérifie bien nos hypothèses, puisqu'alors $e(\theta) = \mathbb{E}_\theta(X_i) = \frac{\theta}{2}$ pour tout i est bien fini et ne dépend bien que de θ :

$$
\boxed{\hat{\theta}_1 = \frac{2}{n} \sum_{i=1}^{n} X_i}
$$

1.2 Question 2

Le risque quadratique peut s'écrire :

$$
\mathbb{E}_\theta[(\hat{\theta}_1 - \theta)^2] = \mathbb{V}_\theta(\hat{\theta}_1) + \mathcal{B}(\hat{\theta}_1)^2
$$

où $\mathbb{V}$ désigne la variance et $\mathcal{B}$ désigne le biais. Ici, les variables X_i étant indépendantes :

$$
\mathbb{V}_{\hat{\theta}_1} = \frac{2}{n} \sum \mathbb{V}[X_i] = \frac{2}{n} \frac{\theta^2}{6} = \frac{\theta^2}{3n}
$$

Comme de plus :

$$\mathcal{B}(\hat{\theta}_1) = \frac{2}{n}\sum \mathbb{E}[X_i] - \theta = 0$$

$\boxed{\text{Le risque quadratique vaut } R\left(\hat{\theta}_1\right) = \frac{\theta^2}{3n}.}$

1.3 Question 3

On veut calculer :

$$\hat{\theta}_2 \in \operatorname{argmax}_\theta \Pi_{i=1}^N p_\theta(X_i) = \operatorname{argmax}_\theta \prod_{i=1}^n \frac{1}{\theta}\mathbb{1}_{[0,\theta]}(X_i)$$

puisque les variables X_i sont supposées indépendantes. En maximisant ci-dessus, puisque $x \mapsto \frac{1}{x^n}$ est strictement décroissante, on trouve :

$$\boxed{\hat{\theta}_2 = \max_i X_i}$$

En effet, un paramètre plus petit rendrait le produit ci-dessus nul.

1.4 Question 4

Pour $\hat{\theta}_2$, la fonction cumulative de la distribution est donnée par $P(\hat{\theta}_2 \leq t) = \frac{t^n}{\theta^n}$ pour $t \in [0, \theta]$. On a donc, par la formule de König-Huygens :

$$\mathbb{V}(\hat{\theta}_2) = \int_0^\theta n\frac{t^{n-1}}{\theta^n}t^2 \,\mathrm{d}t - \left(\int_0^\theta n\frac{t^{n-1}}{\theta^n}t \,\mathrm{d}t\right)^2 = \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2} = \frac{(n^2+n-n^2)\theta^2}{(n+1)^2(n+2)}$$

En calculant le biais :

$$\mathcal{B}(\hat{\theta}_2) = \mathbb{E}(\hat{\theta}_2) - \theta = \frac{n\theta}{n+1} - \theta = -\frac{\theta}{n+1}$$

On trouve donc :

$$\boxed{R(\hat{\theta}_2) = \mathbb{V}(\hat{\theta}_2) + \mathcal{B}(\hat{\theta}_2) = \frac{2\theta^2}{(n+1)(n+2)}}$$

1.5 Question 5

On compare les risques quadratiques quand $n \to \infty$. On a $R(\hat{\theta}_1) \sim \frac{\theta^2}{n}$ et $R(\hat{\theta}_2) \sim \frac{\theta^2}{n^2}$, et donc $R(\hat{\theta}_2) = o(R(\hat{\theta}_1))$. $\boxed{\text{Pour cette mesure de risque, il vaut donc mieux utiliser l'estimateur } \hat{\theta}_2 = \max X_i}$ puisqu'il a un risque significativement plus petit (quand le nombre d'échantillon grandit).

2 Exercice 2

2.1 Question 1

Le passage $(X, Y) \leftrightarrow (R, \Theta)$ se fait par un difféomorphisme sur $\mathbb{R}^2 \setminus 0$ dont l'inverse est :

$$\begin{cases} R = \sqrt{X^2+Y^2} \\ \Theta = \arctan(Y/X) \end{cases}$$

On calcule le jacobien de ce difféomorphisme :

$$J = \begin{vmatrix} \cos(\Theta) & -R\sin(\Theta) \\ \sin(\Theta) & R\cos(\Theta) \end{vmatrix} = R\cos^2(\Theta) + R\sin^2(\Theta) = R$$

Puisque R et Θ sont indépendants, par la formule de changement de variables :

$$f_{X,Y}(x,y) = f_{R,\Theta}(r,\theta)\,|J|^{-1} \overset{R\perp\Theta}{=} f_R(r) f_\Theta(\theta) \times |J|^{-1}$$

On calcule alors :

$$f_{X,Y} = \frac{r\exp(-r^2/2)}{2\pi} \times \frac{1}{r} = \frac{\exp(-(x^2+y^2)/2)}{2\pi} = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2) \times \frac{1}{\sqrt{2\pi}}\exp(-y^2/2)$$

$\boxed{\text{Ainsi, on vérifie bien que } X \perp Y \text{ et que les deux variables suivent des lois normales de paramètres 0, 1.}}$

2.2 Question 2

On suppose qu'on est capables d'échantillonner des lois uniformes de manière indépendante. On utilise l'algorithme naïf suivant :

Algorithme 1 Algorithme Naïf

1. On échantillonne $U_1, U_2 \sim \mathcal{U}([0,1])$ de manière indépendante.
2. On pose $R = \sqrt{-2\log(U_1)}$ et $\Theta = 2\pi U_2$.
3. On renvoie $(X, Y) = (R\cos(\Theta), Y\sin(\Theta))$.

La formule de la distribution de Rayleigh montre que R suit bien une distribution de Rayleigh et Θ suit bien une loi uniforme sur $[0, 2\pi]$. Par la question 1, $\boxed{X \text{ et } Y \text{ sont bien des lois normales de paramètres 0, 1 indépendantes.}}$

2.3 Question 3

2.3.1 a)

À chaque itération, V_1 et V_2 suivent une loi uniforme sur $[-1, 1]$ (par translation et dilatation des U_i). La boucle continuant tant que $V_1^2 + V_2^2 > 1$, c'est à dire tant que le point est hors du disque unité, à la fin de la boucle, $\boxed{(V_1, V_2) \text{ est uniformément distribué sur le disque unité.}}$

2.3.2 b)

La probabilité que V_1, V_2 soit dans le disque unité est le rapport des aires du disque unité et du carré $[-1, 1]^2$, puisque la distribution est uniforme pour la mesure de Lebesgue sur $\mathbb{R}^2$, c'est-à-dire $\pi/4$.

Le nombre d'itérations suit donc une loi géométrique avec probabilité de succès $p = \frac{\pi}{4}$, $\boxed{\text{le nombre d'étapes est donc } \frac{4}{\pi}.}$

2.3.3 c)

On fait un passage en coordonnées polaires : $(V_1, V_2) = (\sqrt{V}\cos(\theta), \sqrt{V}\sin(\theta))$. La distribution de $\sqrt{V} = r, \theta$ étant donnée la distribution uniforme est :

$$f_{R,\Theta}(r,\theta) = \frac{r}{\pi}, 0 \le r \le 1, 0 \le \theta < 2\pi$$

En écrivant $V = r^2$, de jacobien $2r$:

$$f_{V,\Theta}(v,\theta) = \frac{\sqrt{v}}{\pi}\frac{1}{2\sqrt{v}} = \frac{1}{2\pi}$$

On obtient donc bien que $V \sim \mathcal{U}([0,1])$ et $\Theta \sim \mathcal{U}([0,2\pi])$ sont indépendantes. Puisque de plus :

$$T_1 = \frac{V_1}{\sqrt{V_1^2 + V_2^2}} = \frac{\sqrt{V}\cos(\Theta)}{\sqrt{V}} = \cos(\Theta)$$

$\boxed{\text{On vérifie bien que } T_1 \text{ et } V \text{ sont indépendantes}}$ (T_1 ne dépendant que d'une variable aléatoire indépendante de V) et que $\boxed{T_1 \text{ a la même distribution que } \cos(\Theta) \sim \mathcal{U}([0,2\pi]).}$

2.3.4 d)

On a : $S = \sqrt{-2\log V}$ qui a une distribution de Rayleigh de paramètre 1. Puisque qu'on définit $X = S{\cdot}T_1$ et $Y = S{\cdot}T_1$ comme dans la question 1 : $\boxed{(X,Y) \text{ sont des variables aléatoires qui suivent une loi gaussienne de paramètres } 0,\ 1.}$

3 Exercice 3

Voir notebook