

Planche HEC : estimation

Exercice avec préparation 1

Soit a un réel de $]0, 2[$ et f la fonction définie par :
$$f(x) = \begin{cases} \frac{x}{a} e^{-x^2/2a} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

1. **Question de cours** : Donner l'expression intégrale ainsi que la valeur du moment d'ordre 2 d'une variable aléatoire Z suivant la loi normale de paramètres 0 et a .

2. **a)** Montrer que la fonction f est une densité de probabilité.

Dans la suite de l'exercice, on considère une variable aléatoire X de densité f définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

b) Déterminer la fonction de répartition F_X de X .

c) Donner le tableau des variations des courbes de f et de F_X pour $a \in]0, 2[$ quelconque et tracez-les pour $a = 1$.

d) Montrer que X admet une espérance et la calculer.

3. On considère la variable aléatoire Y définie par : $Y = \frac{X^2}{2a}$.

Déterminer la loi de Y .

4. On suppose désormais que le paramètre a est inconnu et on souhaite l'estimer. Soit n un entier naturel supérieur ou égal à 1. On considère un échantillon $(X_1, \dots, X_n)$ définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ composé de variables aléatoires indépendantes ayant toutes la même loi que X .

On note S_n la variable aléatoire définie par $S_n = \frac{1}{2n} \sum_{k=1}^n X_k^2$.

a) Montrer que S_n est un estimateur sans biais et convergent de a .

(i.e. $\mathbb{E}(S_n) = a$ et, pour tout $\varepsilon > 0$, $\lim_{n \rightarrow +\infty} \mathbb{P}(|S_n - a| > \varepsilon) = 0$)

b) Proposer un intervalle de confiance de a au niveau de risque $\alpha \in]0, 1[$ à l'aide de S_n .

5. Soit la variable aléatoire Z définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ par :

Pour $\omega \in \Omega$, $Z(\omega) = \frac{1}{X(\omega) - 1}$ si $X(\omega) \neq 1$ et $Z(\omega) = 0$ si $X(\omega) = 1$.

Déterminer la fonction de répartition de Z .

Réponses de l'exercice avec préparation 1 :

1. Si $Z \hookrightarrow \mathcal{N}(0, a)$, alors $\mathbb{E}(Z) = 0$ et $\mathbb{V}(Z) = a$, d'où

$$\mathbb{E}(Z^2) = \mathbb{V}(Z) = a = \frac{1}{\sqrt{2\pi a}} \int_{-\infty}^{+\infty} t^2 e^{-\frac{t^2}{2a}} dt = \frac{2}{\sqrt{2\pi a}} \int_0^{+\infty} t^2 e^{-\frac{t^2}{2a}} dt$$

2. a) On vérifie que :

- f est positive sur $\mathbb{R}$
- f est continue sur $\mathbb{R}$ sauf éventuellement en 0
- $\int_{-\infty}^{+\infty} f(t) dt = 1$

b) $F_X(x) = \begin{cases} 1 - e^{-\frac{x^2}{2a}} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$

c) • Etude de f . La fonction f est dérivable sur $\mathbb{R}_+^*$ et pour tout $x > 0$, on a $f'(x) = \frac{a-x^2}{a^2} e^{-x^2/2a}$.
Tableau de variations de f sur $[0, +\infty[$:

x	0	$\sqrt{a}$	$+\infty$
Signe de $f'(x)$		+	-
Variations de f	0	$\frac{e^{-\frac{1}{2}}}{\sqrt{a}}$	0

De plus, $\lim_{x \rightarrow 0^+} \frac{f(x)-f(0)}{x-0} = \frac{1}{a} > 0$ donc on obtient la pente de la demi-tangente à droite en 0.
La fonction f est continue en 0 mais n'est pas dérivable en 0.

- Etude de F_X . La fonction F_X est dérivable deux fois sur $\mathbb{R}_+^*$ et pour tout $x > 0$, on a $F_X'(x) = f(x)$ et $F_X''(x) = f'(x)$.

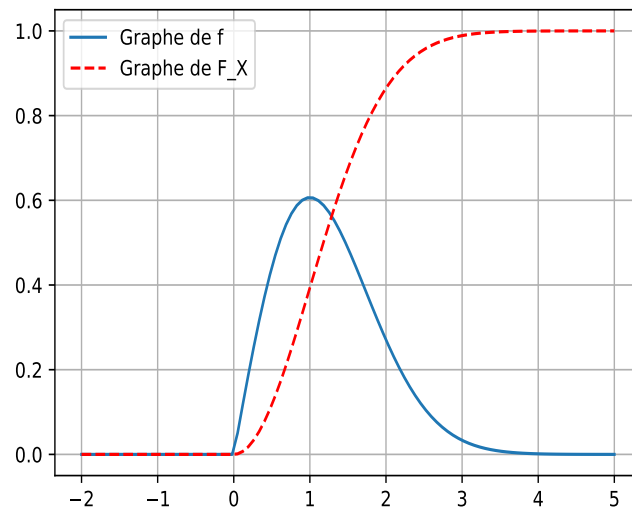
De plus, pour $x > 0$, $\frac{F_X(x)-F_X(0)}{x-0} = \frac{1-e^{-\frac{x^2}{2a}}}{x} \underset{x \rightarrow 0}{\sim} \frac{x}{2a} \underset{x \rightarrow 0}{\rightarrow} 0$ donc F_X est dérivable en 0 (on le savait par un résultat du cours, car f est continue en 0) et $F_X'(0) = 0$.

Tableau de variations de F_X sur $[0, +\infty[$:

x	0	$\sqrt{a}$	$+\infty$
Signe de $F_X''(x)$		+	-
Signe de $F_X'(x)$		+	
Variations de F_X	0	$1 - e^{-\frac{1}{2}}$	1

Donc F_X est convexe sur $[0, \sqrt{a}]$ et concave sur $[\sqrt{a}, +\infty[$.

On trace les graphes de f et F_X à l'aide des informations précédentes, pour $a = 1$:



- d) L'espérance de X existe d'après la question de cours (on fait apparaître un moment d'ordre 2 usuel) :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} t f(t) dt = \frac{1}{a} \int_0^{+\infty} t^2 e^{-\frac{t^2}{2a}} dt = \frac{\sqrt{2\pi a}}{2a} \mathbb{E}(Z^2) = \sqrt{\frac{\pi a}{2}}$$

3. $Y \hookrightarrow \mathcal{E}(1)$.

4. a) On pose $Y_k = \frac{X_k^2}{2a}$. D'après la question 3, $Y_k \hookrightarrow \mathcal{E}(1)$ donc $X_k^2 \hookrightarrow \mathcal{E}(\frac{1}{2a})$. D'où $\mathbb{E}(X_k^2) = 2a$ et $\mathbb{V}(X_k^2) = 4a^2$. Il suit :

$$\mathbb{E}(S_n) = \frac{1}{2n} \times n \times 2a = a$$

On écrit

$$S_n = \frac{1}{n} \sum_{k=1}^n \frac{X_k^2}{2}$$

avec $\mathbb{E}(\frac{X_k^2}{2}) = a$.

Tout d'abord, S_n est bien un estimateur de a , puisque S_n s'écrit comme une fonction d'un n -échantillon de la loi de X , dont l'expression ne dépend pas de a .

D'après la loi faible des grands nombres (S_n admet une variance comme combinaison linéaire de v.a.r. qui admettent une variance), S_n est un estimateur convergent de a .

- b) Soit $\varepsilon > 0$. On applique Bienaymé-Tchebychev :

$$\mathbb{P}(|S_n - a| > \varepsilon) \leq \frac{\mathbb{V}(S_n)}{\varepsilon^2} = \frac{a^2}{n\varepsilon^2} \leq \frac{4}{n\varepsilon^2} \quad \text{car } 0 < a < 2a$$

On a $\alpha = \frac{4}{n\varepsilon^2}$ ssi $\varepsilon = \frac{2}{\sqrt{\alpha n}}$. D'où $\mathbb{P}\left(|S_n - a| \leq \frac{2}{\sqrt{\alpha n}}\right) \geq 1 - \alpha$. On peut proposer l'intervalle de confiance

$$\left[S_n - \frac{2}{\sqrt{\alpha n}}, S_n + \frac{2}{\sqrt{\alpha n}}\right]$$

ou même

$$\left[\max\left(0, S_n - \frac{2}{\sqrt{\alpha n}}\right), \min\left(2, S_n + \frac{2}{\sqrt{\alpha n}}\right)\right]$$

puisque l'on sait que $a \in]0, 2[$.

5. On peut considérer que $X(\Omega) = [0, +\infty[$.

On pose $h : x \mapsto \frac{1}{x-1}$. On étudie les variations de h sur $[0, +\infty[\setminus\{1\}$, on obtient le tableau de variations :

x	0	1	$+\infty$
Signe de $h'(x)$	—		—
Variations de h	-1 	$+\infty$ 	0

On pose ensuite $\tilde{h}$ la fonction définie sur $X(\Omega) = [0, +\infty[$ par

$$\forall x \geq 0, \tilde{h}(x) = \begin{cases} h(x) & \text{si } x \neq 1 \\ 0 & \text{si } x = 1 \end{cases}$$

de sorte que $Z = \tilde{h}(X)$.

Donc $Z(\Omega) = \tilde{h}(X(\Omega)) = \tilde{h}([0, +\infty[) =]-\infty, -1] \cup [0, +\infty[$.

Soit $z \in Z(\Omega)$.

- Premier cas : $z \leq -1$. Alors

$$\begin{aligned} F_Z(z) &= \mathbb{P}\left(\left[\frac{1}{X-1} \leq z\right]\right) \\ &= \mathbb{P}\left(\left[\frac{z+1}{z} \leq X < 1\right]\right) \\ &= F_X(1) - F_X\left(\frac{z+1}{z}\right) \\ &= e^{-\frac{(z+1)^2}{2az^2}} - e^{-\frac{1}{2a}} \end{aligned}$$

- Deuxième cas : $-1 < z < 0$. Alors

$$F_Z(z) = F_Z(-1) = 1 - e^{-\frac{1}{2a}}$$

- Troisième cas : $z \geq 0$. Alors

$$\begin{aligned} F_Z(z) &= \mathbb{P}\left(\left[\frac{1}{X-1} \leq z\right]\right) \\ &= \mathbb{P}\left(\left[\frac{1}{X-1} < 0\right] \cup \left[0 \leq \frac{1}{X-1} \leq z\right]\right) \\ &= \mathbb{P}\left(\left[\frac{1}{X-1} < 0\right]\right) + \mathbb{P}\left(\left[0 \leq \frac{1}{X-1} \leq z\right]\right) && (\text{par incompatibilité}) \\ &= \mathbb{P}([X-1 < 0]) + \mathbb{P}\left(\left[X-1 \geq \frac{1}{z}\right]\right) \\ &= F_X(1) + 1 - F_X\left(\frac{z+1}{z}\right) && (\text{car } X \text{ est à densité}) \\ &= 1 - e^{-\frac{1}{2a}} + e^{-\frac{(z+1)^2}{2az^2}} \end{aligned}$$