ESSEC I 2023

On s'intéresse dans ce sujet à la méthode de Stein, introduite par Charles Stein (1920/2016) en 1972, dont les développements et applications sont nombreux.

Les parties 1 et 2 concernent la justification de la méthode, elles sont indépendantes.

Dans la partie 3, on s'intéresse à l'estimation en un point d'une densité d'une loi de probabilité. Cette partie peut être traitée indépendamment des deux premières parties.

Dans la partie 4, on met en œuvre la méthode de Stein, vue dans les parties 1 et 2, pour établir des convergences « uniformes » en loi et on démontre le résultat admis dans la partie 3. Cette partie est indépendante de la partie 3 à l'exception de sa dernière question.

Dans tout le problème :

- les variables aléatoires considérées sont définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.
- si X est une variable aléatoire, $\mathbb{E}(X)$ et $\mathbb{V}(X)$ désignent respectivement, lorsqu'elles existent, l'espérance et la variance de X.
- W désigne l'ensemble des fonctions h de classe $\mathcal{C}^1$ sur $\mathbb{R}$ telles que :

$$\forall x \in \mathbb{R},\ \left|h'(x)\right| \leqslant 1$$

- N est une variable aléatoire qui suit la loi normale $(0,1)$.
- on admet que si X est une variable aléatoire possédant une espérance et $h \in W$, $\mathbb{E}(h(X))$ existe. On note en particulier c_h l'espérance de $h(N)$.
- On note Φ la fonction de répartition de la loi normale $(0,1)$ définie par, pour tout $x \in \mathbb{R}$, $\Phi(x) = \displaystyle\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}}\, dt$. On rappelle que c'est la primitive sur $\mathbb{R}$, qui vaut $\frac{1}{2}$ en 0, de la fonction $\varphi : t \mapsto \dfrac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}}$.

Partie 1 - Transformation de Stein

Soit $h \in W$. On définit sur $\mathbb{R}$, la fonction $\theta : x \mapsto \dfrac{\Phi(x)}{\varphi(x)}$ et la fonction f_h par,

$$f_h : x \mapsto \theta(-x) \int_{-\infty}^{x} h'(t)\Phi(t)\, dt + \theta(x) \int_{x}^{+\infty} h'(t)(1-\Phi(t))\, dt$$

lorsque ces intégrales convergent.

L'objectif principal de cette partie est d'obtenir, pour X une variable aléatoire admettant une espérance, une expression de $\mathbb{E}(h(X)) - \mathbb{E}(h(N))$ qui ne fait pas intervenir N directement.

1. a) Montrer que pour tout $x \geqslant 0$ et $t \in [x, +\infty[$, $0 \leqslant x\varphi(t) \leqslant t\varphi(t)$. En déduire que :

$$\forall x \geqslant 0,\ 0 \leqslant x(1-\Phi(x)) \leqslant \varphi(x)$$

(on remarquera que pour tout $t \in \mathbb{R}$, $\varphi'(t) = -t\varphi(t)$)

Commentaire

La remarque faite par l'énoncé peut sembler anodine, mais la formule $\varphi'(t) = -t\varphi(t)$ sera utilisée à de nombreuses reprises dans cette partie pour ne pas avoir à revenir aux formules explicites de $\varphi(t)$ et $\varphi'(t)$. La difficulté n'est donc pas de démontrer cette formule, mais bien d'y penser à chaque fois qu'elle sera utile.

Démonstration. Soit $x \geqslant 0$ et soit $t \in [x, +\infty[$.

$$\begin{aligned} &0 \leqslant x \leqslant t \\ \text{donc } &0 \leqslant x\varphi(t) \leqslant t\varphi(t) \qquad \textit{(car } \varphi(t) \geqslant 0, \ \varphi \textit{ étant une densité)} \end{aligned}$$

Soit $B \geqslant x$. Par croissance de l'intégrale, les bornes étant rangées dans l'ordre croissant ($x \leqslant B$), on a

$$\begin{aligned} &0 \leqslant \int_x^B x\varphi(t)\ dt \leqslant \int_x^B t\varphi(t)\ dt \\ \text{donc } &0 \leqslant x\int_x^B \varphi(t)\ dt \leqslant -\int_x^B \varphi'(t)\ dt \\ \text{donc } &0 \leqslant x\int_x^B \varphi(t)\ dt \leqslant -\left[\ \varphi(t)\ \right]_x^B \\ \text{donc } &0 \leqslant x\int_x^B \varphi(t)\ dt \leqslant \varphi(x) - \varphi(B) \end{aligned}$$

Or,

$$\lim_{B\to+\infty} \int_x^B \varphi(t)\ dt = \int_x^{+\infty} \varphi(t)\ dt = \int_{-\infty}^{+\infty} \varphi(t)\ dt - \int_{-\infty}^{x} \varphi(t)\ dt = 1 - \Phi(x)$$

et

$$\lim_{B\to+\infty} \varphi(B) = 0$$

donc, par passage à la limite lorsque B tend ves $+\infty$, on obtient bien

$$\boxed{0 \leqslant x(1-\Phi(x)) \leqslant \varphi(x).}$$

□

b) Procéder de façon analogue pour montrer que : $\forall x \leqslant 0,\ -\varphi(x) \leqslant x\Phi(x) \leqslant 0$.

Démonstration. Soit $x \leqslant 0$ et soit $t \in]-\infty, x]$.

$$\begin{aligned} &t \leqslant x \leqslant 0 \\ \text{donc } &t\varphi(t) \leqslant x\varphi(t) \leqslant 0 \end{aligned}$$

Soit $B \leqslant x$. Par croissance de l'intégrale, les bornes étant rangées dans l'ordre croissant ($B \leqslant x$), on a

$$\begin{aligned}
\int_B^x t\varphi(t)\ dt &\leqslant \int_B^x x\varphi(t)\ dt \leqslant 0 \\
\text{donc } -\int_B^x \varphi'(t)\ dt &\leqslant x\int_B^x \varphi(t)\ dt \leqslant 0 \\
\text{donc } -\left[\ \varphi(t)\ \right]_B^x &\leqslant x\int_B^x \varphi(t)\ dt \leqslant 0 \\
\text{donc } \varphi(B) - \varphi(x) &\leqslant x\int_B^x \varphi(t)\ dt \leqslant 0
\end{aligned}$$

Or,

$$\lim_{B\to-\infty} \int_B^x \varphi(t)\ dt = \int_{-\infty}^x \varphi(t)\ dt = \Phi(x)$$

et

$$\lim_{B\to-\infty} \varphi(B) = 0$$

donc, par passage à la limite lorsque B tend ves $-\infty$, on obtient bien

$$\boxed{-\varphi(x) \leqslant x\Phi(x) \leqslant 0.}$$

□

c) En déduire à l'aide d'une intégration par parties, pour tout x réel, la convergence des intégrales qui suivent et montrer que :

$$\int_{-\infty}^x \Phi(t)\ dt = x\Phi(x) + \varphi(x) \quad \text{et} \quad \int_x^{+\infty} (1-\Phi(t))\ dt = -x(1-\Phi(x)) + \varphi(x) \qquad (R_1)$$

Démonstration. Soit $x \in \mathbb{R}$.

- Remarquons tout d'abord que les fonctions Φ et $1-\Phi$ sont continues sur $\mathbb{R}$ donc $\displaystyle\int_{-\infty}^x \Phi(t)\ dt$ est impropre en $-\infty$ et $\displaystyle\int_x^{+\infty} (1-\Phi(t))\ dt$ est impropre en $+\infty$.
- Soit $B \leqslant 0$. On procède par intégration par parties en posant

$$\left|\begin{array}{ll} u'(t) = 1 & u(t) = t \\ v(t) = \Phi(t) & v'(t) = \varphi(t) \end{array}\right.$$

Cette intégration par parties est valide car les fonctions u et v sont $\mathcal{C}^1$ sur le segment $[B, x]$.

$$\begin{aligned}
\int_B^x \Phi(t)\ dt &= \left[\ t\Phi(t)\ \right]_B^x - \int_B^x t\varphi(t)\ dt \\
&= x\Phi(x) - B\Phi(B) + \int_B^x \varphi'(t)\ dt \\
&= x\Phi(x) - B\Phi(B) + \varphi(x) - \varphi(B)
\end{aligned}$$

Premièrement,

$$\lim_{B\to-\infty} \varphi(B) = 0$$

Ensuite, d'après la question ***1.b)***, $B \leqslant 0$ donc $-\varphi(B) \leqslant B\Phi(B) \leqslant 0$.
Par théorème d'encadrement, on en déduit que

$$\lim_{B \to -\infty} B\Phi(B) = 0$$

$$\boxed{\int_{-\infty}^{x} \Phi(t)\ dt \text{ converge et } \int_{-\infty}^{x} \Phi(t)\ dt = x\Phi(x) + \varphi(x)}$$

- Soit $B \geqslant 0$. On procède par intégration par parties en posant

$$\left| \begin{array}{ll} u'(t) = 1 & u(t) = t \\ v(t) = 1 - \Phi(t) & v'(t) = -\varphi(t) \end{array} \right.$$

Cette intégration par parties est valide car les fonctions u et v sont $\mathcal{C}^1$ sur le segment $[x, B]$.

$$\begin{aligned} \int_x^B (1 - \Phi(t))\ dt &= [\ t(1 - \Phi(t))\ ]_x^B - \int_x^B -t\varphi(t)\ dt \\ &= B(1 - \Phi(B)) - x(1 - \Phi(x)) - \int_x^B \varphi'(t)\ dt \\ &= B(1 - \Phi(B)) - x(1 - \Phi(x)) + \varphi(x) - \varphi(B) \end{aligned}$$

Premièrement,

$$\lim_{B \to +\infty} \varphi(B) = 0$$

Ensuite, d'après la question ***1.a)***, $B \geqslant 0$ donc $0 \leqslant B(1 - \Phi(B)) \leqslant \varphi(B)$.
Par théorème d'encadrement, on en déduit que

$$\lim_{B \to +\infty} B(1 - \Phi(B)) = 0$$

$$\boxed{\int_x^{+\infty} (1 - \Phi(t))\ dt \text{ converge et } \int_x^{+\infty} (1 - \Phi(t))\ dt = -x(1 - \Phi(x)) + \varphi(x)}$$

□

2. a) Montrer que pour tous réels x et y,

$$|h(x) - h(y)| \leqslant |x - y|, \quad \text{puis que } |h(x)| \leqslant |x| + |h(0)|$$

Démonstration. La fonction h est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ et vérifie : pour tout $x \in \mathbb{R}$, $|h'(x)| \leqslant 1$.
Par inégalité des accroissements finis, on obtient :

$$\forall (x, y) \in \mathbb{R}^2,\ |h(x) - h(y)| \leqslant |x - y|$$

Soit $x \in \mathbb{R}$. D'après l'inégalité précédente :

$$|h(x) - h(0)| \leqslant |x|$$

d'où, par inégalité triangulaire,

$$|h(x)| = |(h(x) - h(0)) + h(0)| \leqslant |h(x) - h(0)| + |h(0)| \leqslant |x| + |h(0)|$$

□

b) Pour tout x réel, justifier la convergence de $\displaystyle\int_{-\infty}^{x} h'(t)\Phi(t)\ dt$ et montrer que :

$$\int_{-\infty}^{x} h'(t)\Phi(t)\ dt = h(x)\Phi(x) - \int_{-\infty}^{x} h(t)\varphi(t)\ dt$$

On admet de même que, $\displaystyle\int_{x}^{+\infty} h'(t)(1-\Phi(t))\ dt$ *converge et que,*

$$\int_{x}^{+\infty} h'(t)(1-\Phi(t))\ dt = -h(x)(1-\Phi(x)) + \int_{x}^{+\infty} h(t)\varphi(t)\ dt$$

Démonstration. Soit $x \in \mathbb{R}$.

- pour tout $t \in \,]-\infty, x]$, $|h'(t)\Phi(t)| = |h'(t)|\,\Phi(t) \leqslant \Phi(t)$ car $h \in W$
- les fonctions $t \mapsto |h'(t)\Phi(t)|$ et $t \mapsto \Phi(t)$ sont continues sur $\mathbb{R}$
  ($h \in W$ donc h est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ donc h' est continue sur $\mathbb{R}$)
- l'intégrale $\displaystyle\int_{-\infty}^{x} \Phi(t)\ dt$ converge (cf question ***1.c)***)

Par théorème de comparaison pour les intégrales généralisées de fonctions continues et positives, il vient que

$$\boxed{\int_{-\infty}^{x} h'(t)\Phi(t)\ dt \text{ converge absolument donc converge}}$$

Soit $B \leqslant 0$. On procède ensuite par intégration par parties en posant

$$\left|\begin{array}{ll} u'(t) = h'(t) & u(t) = h(t) \\ v(t) = \Phi(t) & v'(t) = \varphi(t) \end{array}\right.$$

Cette intégration par parties est valide car les fonctions u et v sont $\mathcal{C}^1$ sur le segment $[B, x]$.

$$\begin{aligned} \int_{B}^{x} h'(t)\Phi(t)\ dt &= [\ h(t)\Phi(t)\ ]_{B}^{x} - \int_{B}^{x} h(t)\varphi(t)\ dt \\ &= h(x)\Phi(x) - h(B)\Phi(B) - \int_{B}^{x} h(t)\varphi(t)\ dt \end{aligned}$$

Montrons que $\displaystyle\lim_{B\to-\infty} h(B)\Phi(B) = 0$.

$$\begin{aligned} |h(B)\Phi(B)| &= |h(B)|\,\Phi(B) \\ &\leqslant (|B| + |h(0)|)\Phi(B) && \textit{(cf question } \textbf{\textit{2.a)}}\textit{)} \\ &\leqslant -B\Phi(B) + |h(0)|\,\Phi(B) && \textit{(car } B \leqslant 0\textit{)} \end{aligned}$$

D'après la question ***1.b)***, $\displaystyle\lim_{B\to-\infty} B\Phi(B) = 0$ et, a fortiori, $\displaystyle\lim_{B\to-\infty} \Phi(B) = 0$.

On en déduit que $\displaystyle\lim_{B\to-\infty} -B\Phi(B) + |h(0)|\,\Phi(B) = 0$ et, par théorème d'encadrement,

$$\boxed{\lim_{B\to-\infty} h(B)\Phi(B) = 0.}$$

Or,

$$\int_{B}^{x} h(t)\varphi(t)\ dt = h(x)\Phi(x) - h(B)\Phi(B) - \int_{B}^{x} h'(t)\Phi(t)\ dt$$

et tous les termes de droite admettent une limite finie lorsque B tend vers $-\infty$.

Ainsi, $\displaystyle\int_{-\infty}^{x} h(t)\varphi(t)\ dt$ converge et on peut écrire, par passage à la limite lorsque B tend vers $-\infty$:

$$\boxed{\int_{-\infty}^{x} h'(t)\Phi(t)\ dt = h(x)\Phi(x) - \int_{-\infty}^{x} h(t)\varphi(t)\ dt.}$$

□

c) En déduire que, pour tout x réel :

$$-\int_{-\infty}^{x} h'(t)\Phi(t)\ dt + \int_{x}^{+\infty} h'(t)(1-\Phi(t))\ dt = c_h - h(x)$$

Démonstration. D'après la question précédente :

$$\begin{aligned}
&-\int_{-\infty}^{x} h'(t)\Phi(t)\ dt + \int_{x}^{+\infty} h'(t)(1-\Phi(t))\ dt \\
&= -\left(h(x)\Phi(x) - \int_{-\infty}^{x} h(t)\varphi(t)\ dt\right) - h(x)(1-\Phi(x)) + \int_{x}^{+\infty} h(t)\varphi(t)\ dt \\
&= \cancel{-h(x)\Phi(x)} + \int_{-\infty}^{x} h(t)\varphi(t)\ dt - h(x) + \cancel{h(x)\Phi(x)} + \int_{x}^{+\infty} h(t)\varphi(t)\ dt \\
&= \int_{-\infty}^{+\infty} h(t)\varphi(t)\ dt - h(x) \qquad \textit{(par Chasles)}
\end{aligned}$$

Remarquons maintenant que φ est une densité de N. La fonction h est continue sur $\mathbb{R}$ et, d'après l'énoncé, $h(N)$ admet une espérance. On en déduit par théorème de transfert que l'intégrale $\displaystyle\int_{-\infty}^{+\infty} h(t)\varphi(t)\ dt$ converge absolument (ce que l'on savait déjà) et que

$$\int_{-\infty}^{+\infty} h(t)\varphi(t)\ dt = \mathbb{E}(h(N)) = c_h$$

D'où

$$\boxed{-\int_{-\infty}^{x} h'(t)\Phi(t)\ dt + \int_{x}^{+\infty} h'(t)(1-\Phi(t))\ dt = c_h - h(x).}$$

□

3. a) Établir que pour tout $x \in \mathbb{R}$,

$$\begin{aligned}
\theta'(x) &= 1 + x\theta(x) \\
\theta''(x) &= x + (1+x^2)\theta(x) \\
\theta(-x)\Phi(x) &= \theta(x)(1-\Phi(x))
\end{aligned}$$

Démonstration. La fonction θ est de classe $\mathcal{C}^2$ sur $\mathbb{R}$ comme quotient de deux fonctions de classe $\mathcal{C}^2$ sur $\mathbb{R}$ dont le dénominateur ne s'annule pas.
Soit $x \in \mathbb{R}$.

$$\begin{aligned}
\theta'(x) &= \frac{\varphi(x)^2 - \Phi(x)\varphi'(x)}{\varphi(x)^2} \\
&= \frac{\varphi(x)^2 + \Phi(x)x\varphi(x)}{\varphi(x)^2} \\
&= 1 + x\frac{\Phi(x)}{\varphi(x)} \\
&= 1 + x\theta(x)
\end{aligned}$$

donc

$$\begin{aligned}\theta''(x) &= \theta(x) + x\theta'(x) \\ &= \theta(x) + x(1 + x\theta(x)) \\ &= \theta(x) + x + x^2\theta(x) \\ &= x + (1 + x^2)\theta(x)\end{aligned}$$

D'autre part,

$$\begin{aligned}\theta(-x)\Phi(x) &= \frac{\Phi(-x)}{\varphi(-x)}\Phi(x) \\ &= \frac{\Phi(-x)}{\varphi(x)}\Phi(x) && \textit{(car } \varphi \textit{ est paire)} \\ &= \Phi(-x)\frac{\Phi(x)}{\varphi(x)} \\ &= (1 - \Phi(x))\theta(x) && \textit{(par propriété de } \Phi\textit{)}\end{aligned}$$

□

b) En déduire que f_h est une fonction de classe $\mathcal{C}^1$ sur $\mathbb{R}$ qui vérifie, pour tout x réel :

$$f_h'(x) - xf_h(x) = c_h - h(x)$$

Pourquoi peut-on alors affirmer que f_h est de classe $\mathcal{C}^2$ sur $\mathbb{R}$?

Démonstration.

- Notons $g_1 : x \mapsto \displaystyle\int_0^x h'(t)\Phi(t)\ dt$.
  La fonction $t \mapsto h'(t)\Phi(t)$ est continue sur $\mathbb{R}$ donc g_1 est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ par théorème fondamental de l'analyse et, pour tout $x \in \mathbb{R}$,
  $$g_1'(x) = h'(x)\Phi(x)$$
- De manière analogue, la fonction $g_2 : x \mapsto \displaystyle\int_x^0 h'(t)(1 - \Phi(t))\ dt$ est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ et, pour tout $x \in \mathbb{R}$,
  $$g_2'(x) = -h'(x)(1 - \Phi(x))$$
- Avec ces notations et par relation de Chasles, on a, pour tout $x \in \mathbb{R}$,
  $$f_h(x) = \theta(-x)\left(\int_{-\infty}^0 h'(t)\Phi(t)\ dt + g_1(x)\right) + \theta(x)\left(\int_0^{+\infty} h'(t)(1 - \Phi(t))\ dt + g_2(x)\right)$$
  On en déduit que, par somme et produit de fonctions de classe $\mathcal{C}^1$ sur $\mathbb{R}$
  $$\boxed{f_h \text{ est de classe } \mathcal{C}^1 \text{ sur } \mathbb{R}}$$
  Soit $x \in \mathbb{R}$. En utilisant les questions ***3.a)*** et ***2.c)***, on obtient

$$\begin{aligned}
f_h'(x) &= -\theta'(-x)\int_{-\infty}^{x} h'(t)\Phi(t)\ dt + \theta(-x)h'(x)\Phi(x)\\
&\quad + \theta'(x)\int_{x}^{+\infty} h'(t)(1-\Phi(t))\ dt - \theta(x)h'(x)(1-\Phi(x))\\
&= h'(x)\left(\cancel{\theta(-x)\Phi(x) - \theta(x)(1-\Phi(x))}\right)\\
&\quad - (1 - x\theta(-x))\int_{-\infty}^{x} h'(t)\Phi(t)\ dt + (1 + x\theta(x))\int_{x}^{+\infty} h'(t)(1-\Phi(t))\ dt\\
&= -\int_{-\infty}^{x} h'(t)\Phi(t)\ dt + \int_{x}^{+\infty} h'(t)(1-\Phi(t))\ dt\\
&\quad + x\theta(-x)\int_{-\infty}^{x} h'(t)\Phi(t)\ dt + x\theta(x)\int_{x}^{+\infty} h'(t)(1-\Phi(t))\ dt\\
&= c_h - h(x) + xf_h(x)
\end{aligned}$$

D'où

$$\boxed{f_h'(x) - xf_h(x) = c_h - h(x).}$$

Pour tout $x \in \mathbb{R}$, $f_h'(x) = c_h - h(x) + xf_h(x)$.
Or,

- f_h est de classe $\mathcal{C}^1$ sur $\mathbb{R}$.
- h est de classe $\mathcal{C}^1$ sur $\mathbb{R}$.

Par somme et produit, il vient que f_h' est de classe $\mathcal{C}^1$ sur $\mathbb{R}$, *i.e.*

$$\boxed{f_h \text{ est de classe } \mathcal{C}^2 \text{ sur } \mathbb{R}.}$$

□

c) En conclure que, si X est une variable aléatoire admettant une espérance,

$$|\mathbb{E}(h(X)) - \mathbb{E}(h(N))| = \left|\mathbb{E}(f_h'(X) - Xf_h(X))\right|$$

Démonstration. Montrons que $\mathbb{E}(f_h'(X) - Xf_h(X))$ existe.
D'après la question précédente, pour tout $x \in \mathbb{R}$,

$$f_h'(x) - xf_h(x) = c_h - h(x)$$

On en déduit que

$$f_h'(X) - Xf_h(X) = c_h - h(X)$$

Or, c_h est une constante et $h(X)$ admet une espérance (car $h \in W$ et X admet une espérance) donc $c_h - h(X)$ admet une espérance par transformation affine.

$$\boxed{\mathbb{E}(f_h'(X) - Xf_h(X)) \text{ existe.}}$$

On peut alors écrire

$$\begin{aligned}
\left|\mathbb{E}(f_h'(X) - Xf_h(X))\right| &= |\mathbb{E}(c_h - h(X))|\\
&= |\mathbb{E}(c_h) - \mathbb{E}(h(X))| && \textit{(par linéarité de l'espérance)}\\
&= |c_h - \mathbb{E}(h(X))| && \textit{(car } c_h \textit{ est une constante)}\\
&= |\mathbb{E}(h(N)) - \mathbb{E}(h(X))| && \textit{(par définition de } c_h\textit{)}
\end{aligned}$$

D'où

$$\boxed{|\mathbb{E}(h(X)) - \mathbb{E}(h(N))| = |\mathbb{E}(f_h'(X) - X f_h(X))|.}$$

□

4. Majoration de $|f_h|$.

a) Montrer, en utilisant les égalités (R_1), que pour tout x réel :

$$\theta(-x) \int_{-\infty}^{x} \Phi(t)\ dt + \theta(x) \int_{x}^{+\infty} (1 - \Phi(t))\ dt = 1$$

Démonstration. Soit $x \in \mathbb{R}$.

$$\begin{aligned}
&\theta(-x) \int_{-\infty}^{x} \Phi(t)\ dt + \theta(x) \int_{x}^{+\infty} (1 - \Phi(t))\ dt \\
=&\theta(-x)\left(x\Phi(x) + \varphi(x)\right) + \theta(x)\left(-x(1-\Phi(x)) + \varphi(x)\right) \\
=&x\theta(-x)\Phi(x) + \theta(-x)\varphi(x) - x\theta(x)(1-\Phi(x)) + \theta(x)\varphi(x) \\
=&x\left(\cancel{\theta(-x)\Phi(x) - \theta(x)(1-\Phi(x))}\right) + (\theta(-x) + \theta(x))\varphi(x)
\end{aligned}$$

Or,

$$\begin{aligned}
\theta(-x) + \theta(x) &= \frac{\Phi(-x)}{\varphi(-x)} + \frac{\Phi(x)}{\varphi(x)} \\
&= \frac{1-\Phi(x)}{\varphi(x)} + \frac{\Phi(x)}{\varphi(x)} && \textit{(car } \varphi \textit{ est paire et par propriété de } \Phi\textit{)} \\
&= \frac{1}{\varphi(x)}
\end{aligned}$$

donc $(\theta(-x) + \theta(x))\varphi(x) = 1$ et finalement :

$$\boxed{\theta(-x) \int_{-\infty}^{x} \Phi(t)\ dt + \theta(x) \int_{x}^{+\infty} (1 - \Phi(t))\ dt = 1.}$$

□

b) En déduire que pour tout x réel : $|f_h(x)| \leqslant 1$.

Démonstration. Soit $x \in \mathbb{R}$.

$$\begin{aligned}
|f_h(x)| &= \left|\theta(-x) \int_{-\infty}^{x} h'(t)\Phi(t)\ dt + \theta(x) \int_{x}^{+\infty} h'(t)(1-\Phi(t))\ dt\right| \\
&\leqslant |\theta(-x)| \left|\int_{-\infty}^{x} h'(t)\Phi(t)\ dt\right| + |\theta(x)| \left|\int_{x}^{+\infty} h'(t)(1-\Phi(t))\ dt\right| && \textit{(inégalité triangulaire)} \\
&\leqslant \theta(-x) \left|\int_{-\infty}^{x} h'(t)\Phi(t)\ dt\right| + \theta(x) \left|\int_{x}^{+\infty} h'(t)(1-\Phi(t))\ dt\right| && (\forall y \in \mathbb{R},\ \theta(y) \geqslant 0) \\
&\leqslant \theta(-x) \int_{-\infty}^{x} \left|h'(t)\Phi(t)\right|\ dt + \theta(x) \int_{x}^{+\infty} \left|h'(t)(1-\Phi(t))\right|\ dt && \textit{(inégalité triangulaire)} \\
&\leqslant \theta(-x) \int_{-\infty}^{x} \left|h'(t)\right|\Phi(t)\ dt + \theta(x) \int_{x}^{+\infty} \left|h'(t)\right|(1-\Phi(t))\ dt && (\forall t \in \mathbb{R},\ 0 \leqslant \Phi(t) \leqslant 1) \\
&\leqslant \theta(-x) \int_{-\infty}^{x} \Phi(t)\ dt + \theta(x) \int_{x}^{+\infty} (1-\Phi(t))\ dt && \textit{(}h \in W \textit{ et croissance de l'intégrale)} \\
&= 1 && \textit{(cf question 4.a))}
\end{aligned}$$

□

5. *Majoration de* $|f_h''|$.

a) Montrer que pour tout x réel :

$$\theta''(-x)\int_{-\infty}^{x} \Phi(t)\ dt + \theta''(x)\int_{x}^{+\infty} (1-\Phi(t))\ dt = 1$$

Démonstration. Soit $x \in \mathbb{R}$.

$$\begin{aligned}
&\theta''(-x)\int_{-\infty}^{x} \Phi(t)\ dt + \theta''(x)\int_{x}^{+\infty} (1-\Phi(t))\ dt \\
&= (-x+(1+(-x)^2)\theta(-x))\int_{-\infty}^{x} \Phi(t)\ dt \\
&\quad + (x+(1+x^2)\theta(x))\int_{x}^{+\infty} (1-\Phi(t))\ dt && \textit{(cf question 3.a))} \\
&= x\left(-\int_{-\infty}^{x} \Phi(t)\ dt + \int_{x}^{+\infty} (1-\Phi(t))\ dt\right) \\
&\quad + (1+x^2)\left(\theta(-x)\int_{-\infty}^{x} \Phi(t)\ dt + \theta(x)\int_{x}^{+\infty} (1-\Phi(t))\ dt\right) \\
&= x(-x\Phi(x) - \varphi(x) - x(1-\Phi(x)) + \varphi(x)) && \textit{(cf } (R_1)\textit{)} \\
&\quad + (1+x^2) \times 1 && \textit{(cf question 4.a))} \\
&= x(-\cancel{x\Phi(x)} - \cancel{\varphi(x)} - x + \cancel{x\Phi(x)} + \cancel{\varphi(x)}) + 1 + x^2 \\
&= -x^2 + 1 + x^2 \\
&= 1
\end{aligned}$$

On a bien

$$\boxed{\theta''(-x)\int_{-\infty}^{x} \Phi(t)\ dt + \theta''(x)\int_{x}^{+\infty} (1-\Phi(t))\ dt = 1.}$$

□

b) Établir pour tout x réel l'égalité :

$$f_h''(x) = -h'(x) + \theta''(-x)\int_{-\infty}^{x} h'(t)\Phi(t)\ dt + \theta''(x)\int_{x}^{+\infty} h'(t)(1-\Phi(t))\ dt$$

Démonstration. Soit $x \in \mathbb{R}$.

- D'une part, $f_h'(x) = xf_h(x) + c_h - h(x)$ (cf question ***3.b)***) donc, en dérivant (ce qui est possible car f est de classe $\mathcal{C}^2$ sur $\mathbb{R}$) :

$$\begin{aligned}
f_h''(x) &= f_h(x) + xf_h'(x) - h'(x) \\
&= f_h(x) + x(xf_h(x) + c_h - h(x)) - h'(x) \\
&= -h'(x) + (1+x^2)f_h(x) + x(c_h - h(x))
\end{aligned}$$

- D'autre part, en reprenant la logique du calcul de la question précédente :

$$\begin{aligned}
&\theta''(-x)\int_{-\infty}^{x} h'(t)\Phi(t)\ dt + \theta''(x)\int_{x}^{+\infty} h'(t)(1-\Phi(t))\ dt \\
&= (-x+(1+(-x)^2)\theta(-x))\int_{-\infty}^{x} h'(t)\Phi(t)\ dt \\
&\quad + (x+(1+x^2)\theta(x))\int_{x}^{+\infty} h'(t)(1-\Phi(t))\ dt \qquad \textit{(cf question \textbf{3.a)})} \\
&= x\left(-\int_{-\infty}^{x} h'(t)\Phi(t)\ dt + \int_{x}^{+\infty} h'(t)(1-\Phi(t))\ dt\right) \\
&\quad + (1+x^2)\left(\theta(-x)\int_{-\infty}^{x} h'(t)\Phi(t)\ dt + \theta(x)\int_{x}^{+\infty} h'(t)(1-\Phi(t))\ dt\right) \\
&= x(c_h - h(x)) + (1+x^2)f_h(x) \qquad \textit{(cf question \textbf{2.c)})}
\end{aligned}$$

On a bien

$$\boxed{f_h''(x) = -h'(x) + \theta''(-x)\int_{-\infty}^{x} h'(t)\Phi(t)\ dt + \theta''(x)\int_{x}^{+\infty} h'(t)(1-\Phi(t))\ dt.}$$

□

c) Étudier les variations sur $\mathbb{R}$ de la fonction $x \mapsto \Phi(x) + \dfrac{x}{1+x^2}\varphi(x)$. En déduire son signe et le signe de θ''.
En conclure que, pour tout x réel : $|f_h''(x)| \leqslant 2$.

Démonstration. On note $g : x \mapsto \Phi(x) + \dfrac{x}{1+x^2}\varphi(x)$. La fonction g est dérivable comme somme et produit de fonctions dérivables sur $\mathbb{R}$. Soit $x \in \mathbb{R}$.

$$\begin{aligned}
g'(x) &= \varphi(x) + \frac{1+x^2-2x^2}{(1+x^2)^2}\varphi(x) + \frac{x}{1+x^2}\varphi'(x) \\
&= \varphi(x) + \frac{1-x^2}{(1+x^2)^2}\varphi(x) - \frac{x}{1+x^2}x\varphi(x) \\
&= \frac{\varphi(x)}{(1+x^2)^2}\left((1+x^2)^2 + 1 - x^2 - x^2(1+x^2)\right) \\
&= \frac{\varphi(x)}{(1+x^2)^2}\left(1+2x^2+x^4+1-x^2-x^2-x^4\right) \\
&= \frac{2\varphi(x)}{(1+x^2)^2} > 0
\end{aligned}$$

Donc

$$\boxed{g \text{ est strictement croissante sur } \mathbb{R}.}$$

De plus,

- $\lim\limits_{x\to-\infty} \Phi(x) = 0$ car Φ est une fonction de répartition
- $\lim\limits_{x\to-\infty} \dfrac{x}{1+x^2} = 0$
- $\lim\limits_{x\to-\infty} \varphi(x) = 0$

donc

$$\lim_{x\to-\infty} g(x) = 0 + 0 \times 0 = 0$$

On en déduit que

$$\boxed{\text{pour tout } x \in \mathbb{R},\ g(x) > 0.}$$

Commentaire

Le tableau de variations de g est :

x	$-\infty$ $\quad$ $+\infty$
Signe de $g'(x)$	$+$
Variations de g	$0 \nearrow 1$

D'autre part,

$$\begin{aligned}
\theta''(x) &= x + (1+x^2)\theta(x) && \textit{(cf question } \textbf{\textit{3.a)}}\textit{)} \\
&= x + (1+x^2)\frac{\Phi(x)}{\varphi(x)} \\
&= \frac{1+x^2}{\varphi(x)}\left(\frac{x}{1+x^2}\varphi(x) + \Phi(x)\right) \\
&= \frac{1+x^2}{\varphi(x)} g(x)
\end{aligned}$$

On en déduit que

$$\boxed{\text{pour tout } x \in \mathbb{R},\ \theta''(x) > 0.}$$

Ensuite,

$$\begin{aligned}
&\left|f_h''(x)\right| \\
&= \left|-h'(x) + \theta''(-x)\int_{-\infty}^{x} h'(t)\Phi(t)\ dt + \theta''(x)\int_{x}^{+\infty} h'(t)(1-\Phi(t))\ dt\right| \\
&\leqslant \left|h'(x)\right| + \left|\theta''(-x)\right|\int_{-\infty}^{x} \left|h'(t)\Phi(t)\right|\ dt + \left|\theta''(x)\right|\int_{x}^{+\infty} \left|h'(t)(1-\Phi(t))\right|\ dt \quad \textit{(inégalité triangulaire)} \\
&\leqslant 1 + \theta''(-x)\int_{-\infty}^{x} \left|h'(t)\right|\Phi(t)\ dt + \theta''(x)\int_{x}^{+\infty} \left|h'(t)\right|(1-\Phi(t))\ dt \\
&\leqslant 1 + \theta''(-x)\int_{-\infty}^{x} \Phi(t)\ dt + \theta''(x)\int_{x}^{+\infty} (1-\Phi(t))\ dt \quad \textit{(} h \in W \textit{ et croissance de l'intégrale)} \\
&\leqslant 1 + 1 = 2 \quad \textit{(cf question } \textbf{\textit{5.a)}}\textit{)}
\end{aligned}$$

On a bien

$$\boxed{|f_h''(x)| \leqslant 2.}$$

□

Partie 2 - Majoration uniforme de la distance de Kolmogorov

Dans la suite du problème, si X est une variable aléatoire de fonction de répartition F_X, on définit, pour tout x réel, $d_X(x)$ la distance de Kolmogorov au point x entre la loi de X et la loi normale centrée réduite par :

$$d_X(x) = |F_X(x) - \Phi(x)|$$

On définit, pour tout x réel, la fonction h_x sur $\mathbb{R}$ par $h_x(t) = \begin{cases} 1 & \text{si } t \leqslant x \\ 0 & \text{si } t > x \end{cases}$.

Commentaire

Contrairement à ce qui a pu se voir les années précédentes, le concepteur du sujet a préféré ici ne pas introduire la notation des fonctions indicatrices d'intervalles. Cependant, la fonction h_x n'est rien d'autre que l'indicatrice de l'intervalle $]-\infty, x]$:

$$h_x = \mathbb{1}_{]-\infty,x]}$$

La question ***6.*** est alors une question classique de reconnaissance d'une loi de Bernoulli construite à l'aide d'une indicatrice.

On définit aussi la fonction γ sur $\mathbb{R}$ par :

$$\gamma(t) = \begin{cases} 1 & \text{si } t < 0 \\ 0 & \text{si } t > 1 \\ 1 - 3t^2 + 2t^3 & \text{si } t \in [0,1] \end{cases}$$

Soit X une variable aléatoire.

6. Pour tout x réel, quelle est la loi de la variable aléatoire $h_x(X)$? En déduire que $\mathbb{E}(h_x(X))$ existe et vaut $F_X(x)$.

Démonstration. Soit $x \in \mathbb{R}$. La fonction h_x est à valeurs dans $\{0,1\}$ donc $h_x(X)(\Omega) \subset \{0,1\}$. On en déduit que $h_x(X)$ suit une loi de Bernoulli. Notons p son paramètre. Par définition(s) :

$$\begin{aligned} p &= \mathbb{P}([h_x(X) = 1]) \\ &= \mathbb{P}([X \leqslant x]) \\ &= F_X(x) \end{aligned}$$

Finalement,

$$\boxed{h_x(X) \hookrightarrow \mathcal{B}(F_X(x)), \text{ donc } h_x(X) \text{ admet une espérance et } \mathbb{E}(h_x(X)) = F_X(x).}$$

□

7. a) Écrire une fonction **Python** `gamma(t)` qui calcule et renvoie la valeur de $\gamma(t)$, t étant donné.

Démonstration.

```python
def gamma(t):
    if t < 0:
        return 1
    elif t > 1:
        return 0
    else:
        return 1 - 3 * t**2 + 2 * t**3
```

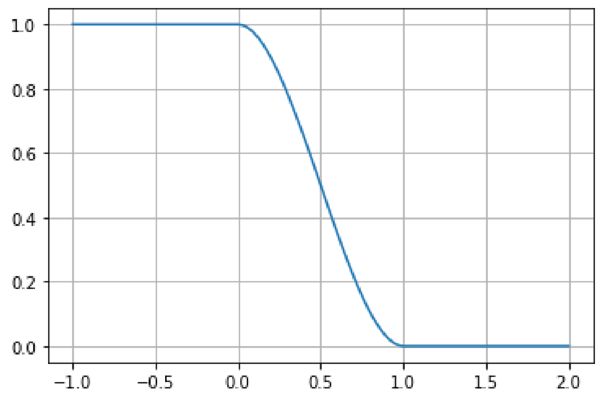
□

b) Utiliser la fonction précédente pour écrire un script qui affiche le graphe de γ sur le segment $[-1, 2]$ dans un repère.

Démonstration.

```python
import numpy as np
import matplotlib.pyplot as plt
abscisse = np.linspace(-1,2,100)
ordonnees = [gamma(t) for t in abscisse]
plt.plot(abscisse, ordonnees)
plt.grid()
```

En compilant ce script, on obtient le graphe : □



8. a) Montrer que γ est continue sur $\mathbb{R}$, de classe $\mathcal{C}^1$ sur $\mathbb{R}$ privé de 0 et 1.

Démonstration. La fonction γ est de classe $\mathcal{C}^1$

- sur l'intervalle **ouvert** $]-\infty, 0[$ car constante sur cet intervalle
- sur l'intervalle **ouvert** $]1, +\infty[$ car constante sur cet intervalle
- sur l'intervalle **ouvert** $]0, 1[$ car polynomiale sur cet intervalle

Ainsi,

$$\boxed{\gamma \text{ est de classe } \mathcal{C}^1 \text{ sur } \mathbb{R} \text{ privé de 0 et 1.}}$$

Il reste à montrer que γ est continue en 0 et en 1.

- Continuité en 0 :

× $\gamma(0) = 1$

× $\displaystyle\lim_{t \to 0^-} \gamma(t) = 1$

× $\displaystyle\lim_{t \to 0^+} \gamma(t) = 1$

donc

$$\boxed{\gamma \text{ est continue en } 0.}$$

- Continuité en 1 :

× $\gamma(1) = 1 - 3 + 2 = 0$

× $\displaystyle\lim_{t \to 1^-} \gamma(t) = 1 - 3 + 2 = 0$

× $\displaystyle\lim_{t \to 1^+} \gamma(t) = 0$

donc

$$\boxed{\gamma \text{ est continue en } 1.}$$

□

b) Étudier les variations de γ sur $[0,1]$ et montrer que pour tout $t \in \mathbb{R}$, $\gamma(t) \in [0,1]$.

Démonstration. D'après la question précédente, γ est continue sur $[0,1]$ et dérivable sur $]0,1[$.
Soit $t \in]0,1[$.

$$\begin{aligned} \gamma'(t) &= -6t + 6t^2 \\ &= 6t(t-1) \end{aligned}$$

$6t > 0$ et $t - 1 < 0$ donc $\gamma'(t) < 0$.
On en déduit que

$$\boxed{\gamma \text{ est strictement décroissante sur } [0,1].}$$

Le tableau de variations de γ est :

t	0		1
Signe de $\gamma'(t)$	‖	$-$	‖
Variations de γ	1	$\searrow$	0

D'après ce tableau de variations, pour tout $t \in [0,1]$, $\gamma(t) \in [0,1]$.
De plus, pour tout $t < 0$, $\gamma(t) = 1 \in [0,1]$ et pour tout $t > 1$, $\gamma(t) = 0 \in [0,1]$.
D'où

$$\boxed{\text{pour tout } t \in \mathbb{R},\ \gamma(t) \in [0,1].}$$

□

c) Établir que γ est dérivable en 1 et que $\gamma'(1) = 0$.
On montrerait de même que γ est dérivable en 0 et que $\gamma'(0) = 0$. On l'admet.

Démonstration. Soit $t \neq 1$. Puisque $\gamma(1) = 0$, on a $\dfrac{\gamma(t) - \gamma(1)}{t-1} = \dfrac{\gamma(t)}{t-1}$.
On sépare ensuite deux cas.

- Premier cas : $t > 1$

$$\begin{aligned}\frac{\gamma(t)-\gamma(1)}{t-1} &= \frac{0}{t-1}\\ &= 0\end{aligned}$$

donc $\lim\limits_{t\to 1^+} \dfrac{\gamma(t)-\gamma(1)}{t-1} = 0.$

- Deuxième cas : $0 < t < 1$

$$\begin{aligned}\frac{\gamma(t)-\gamma(1)}{t-1} &= \frac{1-3t^2+2t^3}{t-1}\\ &= \frac{(t-1)(2t^2-t-1)}{t-1}\\ &= 2t^2-t-1\end{aligned}$$

donc $\lim\limits_{t\to 1^-} \dfrac{\gamma(t)-\gamma(1)}{t-1} = 2-1-1 = 0.$

Finalement, $\lim\limits_{t\to 1} \dfrac{\gamma(t)-\gamma(1)}{t-1} = 0$ donc

$$\boxed{\gamma \text{ est dérivable en } 1 \text{ et } \gamma'(1) = 0.}$$

□

d) Justifier que γ est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ et que pour tout t réel $|\gamma'(t)| \leqslant \dfrac{3}{2}$.

Démonstration. On a vu dans les questions précédentes que

- γ est dérivable sur $\mathbb{R}$
- γ est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ privé de 0 et 1.

Il reste à montrer que γ' est continue en 0 et en 1.

- En 0 :
  - × pour tout $t < 0$, $\gamma'(t) = 0$ donc $\lim\limits_{t\to 0^-} \gamma'(t) = 0 = \gamma'(0)$
  - × pour tout $0 < t < 1$, $\gamma'(t) = 6t(t-1)$ donc $\lim\limits_{t\to 0^+} \gamma'(t) = 0 = \gamma'(0)$
- En 1 :
  - × pour tout $0 < t < 1$, $\gamma'(t) = 6t(t-1)$ donc $\lim\limits_{t\to 1^-} \gamma'(t) = 0 = \gamma'(1)$
  - × pour tout $t > 1$, $\gamma'(t) = 0$ donc $\lim\limits_{t\to 1^+} \gamma'(t) = 0 = \gamma'(1)$

Donc

$$\boxed{\gamma \text{ est de classe } \mathcal{C}^1 \text{ sur } \mathbb{R}.}$$

Soit $t \in \mathbb{R}$.

- Si $t \leqslant 0$, $\gamma'(t) = 0$ donc $|\gamma'(t)| \leqslant \dfrac{3}{2}$.
- Si $t \geqslant 1$, $\gamma'(t) = 0$ donc $|\gamma'(t)| \leqslant \dfrac{3}{2}$.

- Si $0 < t < 1$, $\gamma'(t) = 6t(t-1)$.
  Or, l'inégalité $p(1-p) \leqslant \dfrac{1}{4}$ valable pour tout $p \in [0,1]$ est classique et on en déduit que

$$\begin{aligned} \left|\gamma'(t)\right| &= |6t(t-1)| \\ &= 6t(1-t) \\ &\leqslant 6 \times \frac{1}{4} \\ &= \frac{3}{2} \end{aligned}$$

Finalement, on a bien

$$\boxed{\text{pour tout } t \text{ réel } |\gamma'(t)| \leqslant \frac{3}{2}.}$$

□

On suppose dans la suite de cette partie que X admet une espérance et on considère un réel M_X qui vérifie, pour tout $h \in W$, $|\mathbb{E}(h(X)) - \mathbb{E}(h(N))| \leqslant M_X$.

Commentaire

L'énoncé n'est pas très clair ici. Est-ce que l'on suppose seulement que X admet une espérance ou est-ce que l'on suppose aussi l'existence d'une telle constante M_X ? Si l'on suppose seulement que X admet une espérance, il faudrait alors que l'énoncé admette l'existence de M_X, ce qui n'est pas écrit explicitement, ou bien demander de fournir une preuve, ce qui n'est pas non plus fait. Est-ce que le concepteur du sujet considère que cette existence est évidente ? Cela paraît peu crédible.
Détaillons l'argument. On suppose seulement que X admet une espérance. Soit $h \in W$. D'après la question ***2.a)***,

$$|h(X) - h(N)| \leqslant |X - N| \leqslant |X| + |N|$$

On remarque ensuite que,
- si Y est une variable aléatoire, Y admet une espérance si et seulement si $|Y|$ admet une espérance *(on peut faire la preuve dans le cas discret ou le cas à densité, mais pas dans le cas général)*
- en cas d'existence, par inégalité triangulaire :

$$|\mathbb{E}(Y)| \leqslant \mathbb{E}(|Y|)$$

  (ici on peut faire une preuve générale en utilisant la croissance de l'espérance)

Puisque X et N admettent chacune une espérance, toutes les variables aléatoires en jeu admettent une espérance et on peut écrire, par linéarité et par croissance de l'espérance :

$$|\mathbb{E}(h(X)) - \mathbb{E}(h(N))| = |\mathbb{E}(h(X) - h(N))| \leqslant \mathbb{E}(|h(X) - h(N)|) \leqslant \mathbb{E}(|X|) + \mathbb{E}(|N|)$$

On pose alors $M_X = \mathbb{E}(|X|) + \mathbb{E}(|N|)$, qui ne dépend pas de h et qui vérifie bien la propriété voulue.
Le concepteur a sans doute fait le choix de ne pas tester les candidat·es sur ce raisonnement abstrait, préférant poser d'autres questions. Il aurait alors été préférable d'écrire :
On suppose dans la suite de cette partie que X admet une espérance. On admet alors qu'il existe un réel M_X qui vérifie, pour tout $h \in W$, $|\mathbb{E}(h(X)) - \mathbb{E}(h(N))| \leqslant M_X$.

9. Soit $t > 0$ et x un réel. Pour tout $y \in \mathbb{R}$, on pose $k_x(y) = \gamma\left(\dfrac{y-x}{t}\right)$.

a) Montrer que pour tout y réel, $h_x(y) \leqslant k_x(y)$.

Démonstration. Soit $y \in \mathbb{R}$.

- Premier cas : $y \leqslant x$. Alors
  - × $h_x(y) = 1$.
  - × $y - x \leqslant 0$ et $t > 0$ donc $\dfrac{y-x}{t} \leqslant 0$ donc $k_x(y) = \gamma\left(\dfrac{y-x}{t}\right) = 1$ par définition de γ.

  On a bien $h_x(y) \leqslant k_x(y)$.
- Deuxième cas : $y > x$. Alors
  - × $h_x(y) = 0$.
  - × pour tout $u \in \mathbb{R}$, $\gamma(u) \in [0, 1]$ (cf question ***8.b)***) donc $k_x(y) = \gamma\left(\dfrac{y-x}{t}\right) \geqslant 0$

  On a bien $h_x(y) \leqslant k_x(y)$.

$$\boxed{\text{Pour tout } y \text{ réel, } h_x(y) \leqslant k_x(y).}$$

□

b) *On admet l'existence de* $\mathbb{E}(k_x(X))$ *et de* $\mathbb{E}(k_x(N))$. Justifier l'inégalité suivante :

$$\mathbb{E}(h_x(X)) - \mathbb{E}(h_x(N)) \leqslant \mathbb{E}(k_x(X)) - \mathbb{E}(k_x(N)) + \mathbb{E}(k_x(N)) - \mathbb{E}(h_x(N))$$

Commentaire

On peut démontrer l'existence de $\mathbb{E}(k_x(X))$ et de $\mathbb{E}(k_x(N))$ en utilisant le résultat admis dans le préambule du sujet. On démontre un peu plus loin que $g : u \mapsto \dfrac{2t}{3} k_x(u)$ est une fonction dans W. Ce résultat pourrait être démontré dès maintenant. On en déduit alors que $g(X)$ et $g(N)$ existent. Or, $k_x(X)$ (resp. $k_x(N)$) est une transformée affine de $g(X)$ (resp. $g(N)$) donc $k_x(X)$ et $k_x(N)$ admettent également une espérance.

Démonstration. D'après la question précédente :

$$[h_x(X) \leqslant k_x(X)] = \Omega$$

donc

$$\mathbb{P}([h_x(X) \leqslant k_x(X)]) = 1$$

Par croissance de l'espérance :

$$\mathbb{E}(h_x(X)) \leqslant \mathbb{E}(k_x(X))$$

On en déduit que

$$\begin{aligned}\mathbb{E}(h_x(X)) - \mathbb{E}(h_x(N)) &\leqslant \mathbb{E}(k_x(X)) - \mathbb{E}(h_x(N)) \\ &\leqslant \mathbb{E}(k_x(X)) - \mathbb{E}(k_x(N)) + \mathbb{E}(k_x(N)) - \mathbb{E}(h_x(N))\end{aligned}$$

Commentaire

Il peut paraître assez surprenant de rajouter un terme « qui ne sert à rien » de manière artificielle à la toute fin, alors qu'on aurait pu s'arrêter à la majoration

$$\mathbb{E}(h_x(X)) - \mathbb{E}(h_x(N)) \leqslant \mathbb{E}(k_x(X)) - \mathbb{E}(h_x(N))$$

Cette opération est en fait assez courante lorsque l'on souhaite majorer une différence de deux termes qui diffèrent de deux paramètres (premier paramètre : la fonction k_x ou la fonction h_x, deuxième paramètre : la variable aléatoire X ou la variable aléatoire N). On diminue la difficulté en ne faisant varier qu'un seul paramètre à la fois, mais on a alors deux termes à majorer. C'est ce que fait le sujet en nous faisant majorer le terme $\mathbb{E}(k_x(N)) - \mathbb{E}(h_x(N))$ dès la question suivante.
De manière plus abstraite, on écrit souvent

$$f(x, y) - f(a, b) = (f(x, y) - f(x, b)) + (f(x, b) - f(a, b))$$

et on majore alors le terme $f(x, y) - f(x, b)$ puis le terme $f(x, b) - f(a, b)$ afin d'obtenir une majoration de $f(x, y) - f(a, b)$.

$\square$

c) Montrer que $\mathbb{E}(k_x(N)) - \mathbb{E}(h_x(N)) = \displaystyle\int_x^{x+t} k_x(u)\varphi(u)du$.

Démonstration. Calculons tout d'abord les deux termes séparément :

- La fonction k_x est continue sur $\mathbb{R}$ comme composée de γ et d'une fonction affine toutes les deux continues sur $\mathbb{R}$ (cf question ***8.a)***). De plus, il est admis que $k_x(N)$ admet une espérance. Par théorème de transfert :

$$\mathbb{E}(k_x(N)) = \int_{-\infty}^{+\infty} k_x(u)\varphi(u)du$$

Soit $u \in \mathbb{R}$, on a :

$$\begin{aligned} k_x(u) = 0 &\iff \gamma\left(\frac{u-x}{t}\right) = 0 \\ &\iff \frac{u-x}{t} \geqslant 1 && \textit{(cf question } \textbf{8.}\textit{)} \\ &\iff u - x \geqslant t && \textit{(car } t > 0\textit{)} \\ &\iff u \geqslant x + t \end{aligned}$$

donc

$$\mathbb{E}(k_x(N)) = \int_{-\infty}^{x+t} k_x(u)\varphi(u)du$$

-

$$\begin{aligned} \mathbb{E}(h_x(N)) &= F_N(x) && \textit{(cf question } \textbf{6.}\textit{)} \\ &= \Phi(x) \\ &= \int_{-\infty}^{x} \varphi(u)du \end{aligned}$$

Soit $u \in \mathbb{R}$, on a :

$$\begin{aligned} k_x(u) = 1 &\iff \gamma\left(\frac{u-x}{t}\right) = 1 \\ &\iff \frac{u-x}{t} \leqslant 0 && \textit{(cf question } \textbf{\textit{8.}}\textit{)} \\ &\iff u - x \leqslant 0 && \textit{(car } t > 0\textit{)} \\ &\iff u \leqslant x \end{aligned}$$

donc

$$\mathbb{E}(h_x(N)) = \int_{-\infty}^{x} k_x(u)\varphi(u)du$$

Finalement,

$$\boxed{\mathbb{E}(k_x(N)) - \mathbb{E}(h_x(N)) = \int_{-\infty}^{x+t} k_x(u)\varphi(u)du - \int_{-\infty}^{x} k_x(u)\varphi(u)du = \int_{x}^{x+t} k_x(u)\varphi(u)du}$$

Commentaire

Cette question est difficile car elle demande une prise d'initiative : faire apparaître $k_x(u)$ là où il n'est pas naturellement.
La première idée doit être de penser au théorème de transfert lorsqu'on doit montrer une égalité entre une espérance et une intégrale. A partir de là, il faut s'efforcer de faire apparaître les termes de la formule de fin. Une bonne approche est de conjecturer des égalités au brouillon qui permettraient de faire le lien entre la formule que l'on a démontré et celle que l'on cherche à obtenir, puis à vérifier que ces conjectures sont vraies.

□

d) Établir que la fonction g, définie sur $\mathbb{R}$ par $g : u \mapsto \dfrac{2t}{3}k_x(u)$, appartient à W. En déduire que :

$$\mathbb{E}(h_x(X)) - \mathbb{E}(h_x(N)) \leqslant \frac{3}{2t}M_X + \frac{t}{\sqrt{2\pi}} \leqslant \frac{3}{2t}M_X + \frac{t}{2}$$

Démonstration. La fonction k_x est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ comme composée de γ et d'une fonction affine toutes les deux de classe $\mathcal{C}^1$ sur $\mathbb{R}$ (cf question ***8.d)***). Ainsi, g est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ comme transformée affine de k_x. Soit $u \in \mathbb{R}$.

$$\begin{aligned} |g'(u)| &= \left|\frac{2t}{3}k_x'(u)\right| \\ &= \left|\frac{2t}{3}\frac{1}{t}\gamma'\left(\frac{u-x}{t}\right)\right| \\ &= \frac{2}{3}\left|\gamma'\left(\frac{u-x}{t}\right)\right| \\ &\leqslant \frac{2}{3} \times \frac{3}{2} && \textit{(cf question 8.d)} \\ &= 1 \end{aligned}$$

donc

$$\boxed{g \in W.}$$

D'après l'énoncé, pour tout $h \in W$, $|\mathbb{E}(h(X)) - \mathbb{E}(h(N))| \leqslant M_X$. On applique ce résultat à la fonction g :

$$\begin{aligned}
& |\mathbb{E}(g(X)) - \mathbb{E}(g(N))| \leqslant M_X \\
\text{donc } & \left|\mathbb{E}\left(\frac{2t}{3}k_x(X)\right) - \mathbb{E}\left(\frac{2t}{3}k_x(N)\right)\right| \leqslant M_X \\
\text{donc } & \frac{2t}{3}\left|\mathbb{E}\left(k_x(X)\right) - \mathbb{E}\left(k_x(N)\right)\right| \leqslant M_X \qquad \textit{(par linéarité de l'espérance et car } t>0\textit{)} \\
\text{donc } & \left|\mathbb{E}\left(k_x(X)\right) - \mathbb{E}\left(k_x(N)\right)\right| \leqslant \frac{3}{2t}M_X
\end{aligned}$$

D'après les questions ***9.b)***, ***9.c)*** et ce qui précède :

$$\begin{aligned}
\mathbb{E}(h_x(X)) - \mathbb{E}(h_x(N)) &\leqslant \mathbb{E}(k_x(X)) - \mathbb{E}(k_x(N)) + \mathbb{E}(k_x(N)) - \mathbb{E}(h_x(N)) \\
&\leqslant |\mathbb{E}(k_x(X)) - \mathbb{E}(k_x(N))| + \mathbb{E}(k_x(N)) - \mathbb{E}(h_x(N)) \\
&\leqslant \frac{3}{2t}M_X + \int_x^{x+t} k_x(u)\varphi(u)du
\end{aligned}$$

Il reste à montrer que $\displaystyle\int_x^{x+t} k_x(u)\varphi(u)du \leqslant \frac{t}{\sqrt{2\pi}}$.

On le fait en appliquant à deux reprises la croissance de l'intégrale, les bornes étant rangées dans l'ordre croissant ($t > 0$ donc $x \leqslant x+t$) :

$$\begin{aligned}
\int_x^{x+t} k_x(u)\varphi(u)du &\leqslant \left|\int_x^{x+t} k_x(u)\varphi(u)du\right| \\
&\leqslant \int_x^{x+t} |k_x(u)\varphi(u)|\,du \qquad \textit{(par inégalité triangulaire et car } x \leqslant x+t\textit{)} \\
&\leqslant \int_x^{x+t} |k_x(u)|\,\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{u^2}{2}}du \\
&\leqslant \int_x^{x+t} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{u^2}{2}}du \qquad \textit{(} |k_x(u)| \leqslant 1 \textit{ cf question } \textbf{\textit{8.b)}}\textit{)} \\
&\leqslant \int_x^{x+t} \frac{1}{\sqrt{2\pi}}du \qquad \textit{(} -\frac{u^2}{2} \leqslant 0 \textit{ donc } e^{-\frac{u^2}{2}} \leqslant 1\textit{)} \\
&= \frac{t}{\sqrt{2\pi}}
\end{aligned}$$

D'où

$$\boxed{\mathbb{E}(h_x(X)) - \mathbb{E}(h_x(N)) \leqslant \frac{3}{2t}M_X + \frac{t}{\sqrt{2\pi}}.}$$

Ensuite, $\pi \geqslant 3$ donc $2\pi \geqslant 6 \geqslant 4$ donc $\sqrt{2\pi} \geqslant \sqrt{4} = 2$ donc $\dfrac{t}{\sqrt{2\pi}} \leqslant \dfrac{t}{2}$ (car $t > 0$).

On a bien

$$\boxed{\frac{3}{2t}M_X + \frac{t}{\sqrt{2\pi}} \leqslant \frac{3}{2t}M_X + \frac{t}{2}.}$$

□

On admet de même, qu'en utilisant la fonction k_{x-t}, on a :

$$\mathbb{E}(h_x(N)) - \mathbb{E}(h_x(X)) \leqslant \frac{3}{2t} M_X + \frac{t}{2}$$

10. En étudiant la fonction $t \mapsto \dfrac{3}{2t} M_X + \dfrac{t}{2}$ sur $]0, +\infty[$, en déduire que, pour tout x réel,

$$|\mathbb{E}(h_x(X)) - \mathbb{E}(h_x(N))| \leqslant \sqrt{3M_X}, \text{ puis que } d_X(x) \leqslant \sqrt{3M_X} \qquad (R_2)$$

Démonstration. Soit $\psi : t \mapsto \dfrac{3}{2t} M_X + \dfrac{t}{2}$. La fonction ψ est dérivable sur $]0, +\infty[$. Soit $t \in]0, +\infty[$.

$$\begin{aligned} \psi'(t) &= -\frac{3}{2t^2} M_X + \frac{1}{2} \\ &= \frac{t^2 - 3M_X}{2t^2} \\ &= \frac{(t - \sqrt{3M_X})(t + \sqrt{3M_X})}{2t^2} && \textit{(car } M_X \geqslant 0\textit{)} \end{aligned}$$

Ainsi,

$$\begin{aligned} \psi'(t) \geqslant 0 &\iff \frac{(t - \sqrt{3M_X})(t + \sqrt{3M_X})}{2t^2} \geqslant 0 \\ &\iff t - \sqrt{3M_X} \geqslant 0 && \textit{(car } 2t^2 > 0 \textit{ et } t + \sqrt{3M_X} \geqslant 0\textit{)} \\ &\iff t \geqslant \sqrt{3M_X} \end{aligned}$$

et, de plus,

$$\psi(\sqrt{3M_X}) = \frac{3M_X}{2\sqrt{3M_X}} + \frac{\sqrt{3M_X}}{2} = \frac{\sqrt{3M_X}}{2} + \frac{\sqrt{3M_X}}{2} = \sqrt{3M_X}$$

Le tableau de variations de ψ est donc :

t	0		$\sqrt{3M_X}$		$+\infty$
Signe de $\psi'(t)$	‖	$-$	0	$+$	
Variations de ψ	‖ $+\infty$	↘	$\sqrt{3M_X}$	↗	$+\infty$

Soit $x \in \mathbb{R}$. D'après la question ***9.d)***, l'encadrement

$$-\psi(t) \leqslant \mathbb{E}(h_x(X)) - \mathbb{E}(h_x(N)) \leqslant \psi(t)$$

i.e. , l'inégalité

$$|\mathbb{E}(h_x(X)) - \mathbb{E}(h_x(N))| \leqslant \psi(t)$$

est valable pour tout réel $t > 0$.
On peut donc choisir $t = \sqrt{3M_X}$ et obtenir :

$$\boxed{|\mathbb{E}(h_x(X)) - \mathbb{E}(h_x(N))| \leqslant \psi(\sqrt{3M_X}) = \sqrt{3M_X}.}$$

Il reste à interpréter le terme de gauche. D'après la question ***6.***,

$$\mathbb{E}(h_x(X)) = F_X(x) \qquad \text{et} \qquad \mathbb{E}(h_x(N)) = F_N(x) = \Phi(x)$$

d'où

$$\boxed{d_X(x) = |F_X(x) - \Phi(x)| = |\mathbb{E}(h_x(X)) - \mathbb{E}(h_x(N))| \leqslant \sqrt{3M_X}.}$$

□

Partie 3 - Estimation d'une densité

On considère X une variable aléatoire à densité de fonction de répartition F et de densité de probabilité f qui dépendent d'un paramètre inconnu θ, où $\theta \in \Theta$, Θ un intervalle de $\mathbb{R}$.

Soit a un point de continuité de f, fixé. On souhaite estimer $f(a)$.

Par exemple, si X suit la loi exponentielle de paramètre θ et $a > 0$, on souhaite estimer $\theta e^{-\theta a}$.

On dispose pour tout $\theta \in \Theta$, d'une suite de variables aléatoires $(X_i)_{i\geqslant 1}$ indépendantes de même loi que X.

On choisit une suite $(h_n)_{n\geqslant 1}$ de réels strictement positifs tels que :

$$\lim_{n\to+\infty} h_n = 0 \text{ et } \lim_{n\to+\infty} nh_n = +\infty$$

Pour tout $n \in \mathbb{N}^*$, et $\omega \in \Omega$, on définit :

$$C_n(\omega) \text{ comme le nombre d'indices } i \in [\![1, n]\!] \text{ tels que } X_i(\omega) \in \left]a - h_n, a + h_n\right[$$

et $f_n(\omega) = \dfrac{1}{2nh_n} C_n(\omega)$.

Commentaire

La notation f_n peut faire penser à une fonction d'une variable réelle, mais c'est bien une variable aléatoire qui est définie ici. D'ailleurs, si doute il y avait, la question ***13.*** parle de son espérance.

11. On suppose que l'on dispose d'un fichier `stats.csv` qui comporte une colonne nommée `salaire`. On considère que les valeurs de cette colonne constituent la réalisation d'un échantillon de la loi de X dont la taille dépasse 10000.

a) Après avoir exécuté `import pandas as pd`, quelle(s) instruction(s) permet(tent) de lire dans le fichier `stats.csv` les valeurs de la colonne `salaire` et d'affecter cette série `pandas` obtenue à une variable `échantillon` ?
On supposera que le fichier `stats.csv` se trouve dans le répertoire de travail.

Démonstration.

```
donnees = pd.read_csv(stats)
échantillon = donnees['salaire']
```

□

b) On souhaite calculer et afficher $f_n(\omega)$ pour a donné, lorsque la réalisation d'un échantillon $(X_1(\omega), \ldots, X_n(\omega))$ de la loi de X est représentée en **Python** par `échantillon` et, pour tout $n \in \mathbb{N}^*$, $h_n = \dfrac{1}{\sqrt{n}}$.
Compléter le script suivant pour qu'il réalise cette tâche.

```
a = float(input('a='))
n = échantillon.count()
h = 1 / np.sqrt(n)
C = 0
for i in range(n):
    if ... and ...:
        ... += 1
print(C / ...)
```

Démonstration.

```
a = float(input('a='))
n = échantillon.count()
h = 1 / np.sqrt(n)
C = 0
for i in range(n):
    if échantillon[i] > a - h and échantillon[i] < a + h:
        C += 1
print(C / (2 * n * h))
```

□

12. Montrer que C_n suit une loi binomiale de paramètres (n, p_n) en précisant l'expression de p_n en fonction de a et h_n.
En déduire que $\mathbb{E}(f_n)$ existe et vaut : $\dfrac{F(a+h_n)-F(a-h_n)}{2h_n}$.

Démonstration. L'expérience aléatoire consiste en la répétition de n épreuves de Bernoulli indépendantes et identiques (les variables aléatoires X_i étant indépendantes de même loi), le succès étant : « $X_i(\omega) \in \,]a-h_n, a+h_n[$ ». On note p_n la probabilité de succès.
La variable aléatoire C_n compte le nombre de succès.
On en déduit que

$$\boxed{C_n \hookrightarrow \mathcal{B}(n, p_n).}$$

De plus,

$$\begin{aligned} p_n &= \mathbb{P}([a-h_n < X < a+h_n]) \\ &= F(a+h_n) - F(a-h_n) && \textit{(car } X \textit{ est à densité)} \end{aligned}$$

$$\boxed{p_n = F(a+h_n) - F(a-h_n).}$$

La variable aléatoire C_n admet une espérance. On en déduit que f_n admet une espérance en tant que transformée affine de C_n et

$$\begin{aligned} \mathbb{E}(f_n) &= \mathbb{E}\left(\frac{1}{2nh_n} C_n\right) \\ &= \frac{1}{2nh_n}\mathbb{E}(C_n) && \textit{(par linéarité)} \\ &= \frac{1}{2\cancel{n}h_n}\cancel{n}p_n \\ &= \frac{F(a+h_n)-F(a-h_n)}{2h_n} \end{aligned}$$

Commentaire

On peut faire dans cette question un peu de rétro-ingénierie et s'aider du résultat donné dans l'énoncé pour trouver p_n si on bloque sur le calcul.
L'énoncé nous donne $\mathbb{E}(f_n) = \dfrac{F(a+h_n) - F(a-h_n)}{2h_n}$.
On cherche p_n et on sait que $\mathbb{E}(C_n) = np_n$. Or, $f_n = \dfrac{1}{2nh_n}C_n$ donc $C_n = 2nh_n f_n$. D'où

$$np_n = 2nh_n \frac{F(a+h_n) - F(a-h_n)}{2h_n} = n(F(a+h_n) - F(a-h_n))$$

donc

$$p_n = F(a+h_n) - F(a-h_n)$$

□

13. a) En utilisant la dérivabilité de F en a, montrer que $\lim\limits_{n\to+\infty} \mathbb{E}(f_n) = f(a)$.

Démonstration. a est un point de continuité de f donc F est dérivable en a. La fonction F admet alors un développement limité à l'ordre 1 au point a qui s'écrit :

$$F(a+h) = F(a) + hF'(a) + \underset{h\to0}{o}(h) = F(a) + hf(a) + \underset{h\to0}{o}(h)$$

Puisque $h_n \underset{n\to+\infty}{\longrightarrow} 0$, on peut écrire que :

$$F(a+h_n) = F(a) + h_n f(a) + \underset{n\to+\infty}{o}(h_n) \qquad \text{et} \qquad F(a-h_n) = F(a) - h_n f(a) + \underset{n\to+\infty}{o}(h_n)$$

D'où

$$\begin{aligned}\mathbb{E}(f_n) &= \frac{F(a+h_n) - F(a-h_n)}{2h_n} \\ &= \frac{(F(a) + h_n f(a) + \underset{n\to+\infty}{o}(h_n)) - (F(a) - h_n f(a) + \underset{n\to+\infty}{o}(h_n))}{2h_n} \\ &= \frac{2h_n f(a) + \underset{n\to+\infty}{o}(h_n)}{2h_n} \\ &= f(a) + \underset{n\to+\infty}{o}(1)\end{aligned}$$

donc on a bien

$$\boxed{\mathbb{E}(f_n) \underset{n\to+\infty}{\longrightarrow} f(a).}$$

□

b) Montrer que $\mathbb{V}(f_n)$ existe et que $\lim\limits_{n\to+\infty} \mathbb{V}(f_n) = 0$.

Démonstration. La variable aléatoire C_n admet une variance. On en déduit que f_n admet une variance en tant que transformée affine de C_n et

$$\begin{aligned}
\mathbb{V}(f_n) &= \mathbb{V}\left(\frac{1}{2nh_n}C_n\right) \\
&= \frac{1}{(2nh_n)^2}\mathbb{V}(C_n) && \textit{(par propriété de la variance)} \\
&= \frac{1}{(2nh_n)^2}np_n(1-p_n) \\
&= \frac{1-p_n}{2nh_n}\frac{p_n}{2h_n} \\
&= \frac{1-p_n}{2nh_n}\mathbb{E}(f_n)
\end{aligned}$$

Or, on a vu à la question ***13.a)*** que $p_n \underset{n\to+\infty}{\longrightarrow} 0$ et $\mathbb{E}(f_n) \underset{n\to+\infty}{\longrightarrow} f(a)$.
De plus, par hypothèse, $nh_n \underset{n\to+\infty}{\longrightarrow} +\infty$. Il vient alors

$$\boxed{\mathbb{V}(f_n) \underset{n\to+\infty}{\longrightarrow} 0 \times f(a) = 0.}$$

□

On suppose désormais, que $f(a) > 0$, que pour tout $n \in \mathbb{N}^*$, $p_n \in]0,1[$, que F est de classe $\mathcal{C}^2$ au voisinage de a et que $\lim\limits_{n\to+\infty} nh_n^3 = 0$.

On note pour tout $n \geqslant 1$, $\sigma_n = \sqrt{np_n(1-p_n)}$ et $\theta_n = \sqrt{2h_n f(a)}$.

On définit les variables aléatoires : $D_n = \dfrac{C_n - np_n}{\sigma_n}$ et $\hat{f}_n = \dfrac{\theta_n\sqrt{n}}{f(a)}(f_n - f(a))$.

14. a) En utilisant le développement limité de F à l'ordre 2 au point a, montrer que :

$$p_n \underset{n\to+\infty}{=} 2h_n f(a) + o(h_n^2) \underset{n\to+\infty}{=} \theta_n^2 + o(h_n^2)$$

Démonstration. F est de classe $\mathcal{C}^2$ au voisinage de a donc admet un développement limité d'ordre 2 au point a qui s'écrit :

$$F(a+h) = F(a) + hF'(a) + \frac{1}{2}h^2F''(a) + \underset{h\to0}{o}(h^2) = F(a) + hf(a) + \frac{1}{2}h^2F''(a) + \underset{h\to0}{o}(h^2)$$

Puisque $h_n \underset{n\to+\infty}{\longrightarrow} 0$, on peut écrire que :

$$F(a+h_n) = F(a) + h_nf(a) + \frac{1}{2}h_n^2F''(a) + \underset{n\to+\infty}{o}(h_n^2)$$

et

$$F(a-h_n) = F(a) - h_nf(a) + \frac{1}{2}h_n^2F''(a) + \underset{n\to+\infty}{o}(h_n^2)$$

D'où

$$\begin{aligned}
p_n &= F(a+h_n) - F(a-h_n) \\
&= F(a) + h_nf(a) + \frac{1}{2}h_n^2F''(a) + \underset{n\to+\infty}{o}(h_n^2) - \left(F(a) - h_nf(a) + \frac{1}{2}h_n^2F''(a) + \underset{n\to+\infty}{o}(h_n^2)\right) \\
&= 2h_nf(a) + \underset{n\to+\infty}{o}(h_n^2) \\
&= \theta_n^2 + \underset{n\to+\infty}{o}(h_n^2)
\end{aligned}$$

par définition de θ_n. □

b) En déduire que : $p_n \underset{n\to+\infty}{\sim} 2h_n f(a)$, puis que $\lim\limits_{n\to+\infty} np_n = +\infty$.

Démonstration. On a, d'après la question précédente,

$$p_n = 2h_n f(a) + \underset{n\to+\infty}{o}(h_n^2)$$

Or, $f(a) > 0$ et pour tout $n \in \mathbb{N}^*$, $h_n > 0$, donc

$$\frac{p_n}{2h_n f(a)} = 1 + \underset{n\to+\infty}{o}(h_n)$$

$h_n \underset{n\to+\infty}{\longrightarrow} 0$ donc $1 + \underset{n\to+\infty}{o}(h_n) \underset{n\to+\infty}{\longrightarrow} 1$ donc $\dfrac{p_n}{2h_n f(a)} \underset{n\to+\infty}{\longrightarrow} 1$ donc

$$\boxed{p_n \underset{n\to+\infty}{\sim} 2h_n f(a).}$$

On en déduit que

$$np_n \underset{n\to+\infty}{\sim} 2nh_n f(a)$$

Or, $2f(a) > 0$ et $nh_n \underset{n\to+\infty}{\longrightarrow} +\infty$ donc, par produit, $2nh_n f(a) \underset{n\to+\infty}{\longrightarrow} +\infty$ et finalement

$$\boxed{np_n \underset{n\to+\infty}{\longrightarrow} +\infty.}$$

□

c) Montrer que $\hat{f}_n = \dfrac{\sigma_n}{\theta_n\sqrt{n}} D_n + \sqrt{n}\left(\dfrac{p_n}{\theta_n} - \theta_n\right)$ et que l'on a :

$$\lim_{n\to+\infty} \frac{\sigma_n}{\theta_n\sqrt{n}} = 1 \quad ; \quad \lim_{n\to+\infty} \sqrt{n}\left(\frac{p_n}{\theta_n} - \theta_n\right) = 0$$

Démonstration.

- D'une part :

$$\begin{aligned}
\hat{f}_n &= \frac{\theta_n\sqrt{n}}{f(a)}\left(f_n - f(a)\right) \\
&= \frac{\theta_n\sqrt{n}}{f(a)} f_n - \theta_n\sqrt{n} \\
&= \frac{\theta_n\sqrt{n}C_n}{2nh_n f(a)} - \theta_n\sqrt{n} \\
&= \frac{\theta_n\sqrt{n}C_n}{\theta_n^2 n} - \theta_n\sqrt{n} \\
&= \frac{1}{\theta_n\sqrt{n}} C_n - \theta_n\sqrt{n}
\end{aligned}$$

- D'autre part :

$$\begin{aligned}
\frac{\sigma_n}{\theta_n\sqrt{n}}D_n + \sqrt{n}\left(\frac{p_n}{\theta_n} - \theta_n\right) &= \frac{C_n - np_n}{\theta_n\sqrt{n}} + \sqrt{n}\frac{p_n - \theta_n^2}{\theta_n} \\
&= \frac{C_n - np_n}{\theta_n\sqrt{n}} + n\frac{p_n - \theta_n^2}{\theta_n\sqrt{n}} \\
&= \frac{C_n - \cancel{np_n} + \cancel{np_n} - n\theta_n^2}{\theta_n\sqrt{n}} \\
&= \frac{1}{\theta_n\sqrt{n}}C_n - \frac{n\theta_n^2}{\theta_n\sqrt{n}} \\
&= \frac{1}{\theta_n\sqrt{n}}C_n - \theta_n\sqrt{n}
\end{aligned}$$

D'où

$$\boxed{\hat{f}_n = \frac{\sigma_n}{\theta_n\sqrt{n}}D_n + \sqrt{n}\left(\frac{p_n}{\theta_n} - \theta_n\right).}$$

De plus,

$$\begin{aligned}
\frac{\sigma_n}{\theta_n\sqrt{n}} &= \frac{\sqrt{np_n(1-p_n)}}{\theta_n\sqrt{n}} \\
&= \frac{\sqrt{p_n(1-p_n)}}{\theta_n} \\
&= \frac{\sqrt{p_n(1-p_n)}}{\sqrt{2h_n f(a)}} \\
&\underset{n\to+\infty}{\sim} \frac{\sqrt{p_n(1-p_n)}}{\sqrt{p_n}} \\
&= \sqrt{1-p_n}
\end{aligned}$$

et puisque $p_n \underset{n\to+\infty}{\longrightarrow} 0$, il vient

$$\boxed{\frac{\sigma_n}{\theta_n\sqrt{n}} \underset{n\to+\infty}{\longrightarrow} 1.}$$

Enfin,

$$\begin{aligned}
\sqrt{n}\left(\frac{p_n}{\theta_n} - \theta_n\right) &= \sqrt{n}\frac{p_n - \theta_n^2}{\theta_n} \\
&= \sqrt{n}\frac{\underset{n\to+\infty}{o}(h_n^2)}{\sqrt{2h_n f(a)}} && \textit{(cf question } \textbf{\textit{14.a)}}\textit{)} \\
&= \underset{n\to+\infty}{o}\left(\sqrt{n}h_n^{\frac{3}{2}}\right) \\
&= \underset{n\to+\infty}{o}\left(\sqrt{nh_n^3}\right)
\end{aligned}$$

Or, par hypothèse, $nh_n^3 \underset{n\to+\infty}{\longrightarrow} 0$ donc $\sqrt{nh_n^3} \underset{n\to+\infty}{\longrightarrow} 0$ et finalement,

$$\boxed{\sqrt{n}\left(\frac{p_n}{\theta_n} - \theta_n\right) \underset{n\to+\infty}{\longrightarrow} 0.}$$

Commentaire

Le nombre de notations intervenant dans la formule $\hat{f}_n = \dfrac{\sigma_n}{\theta_n\sqrt{n}}D_n + \sqrt{n}\left(\dfrac{p_n}{\theta_n} - \theta_n\right)$ et sa preuve est assez impressionnant. Il est difficile dans ces conditions de trouver du premier coup comment faire un calcul direct en partant d'un terme et en arrivant à l'autre.
La meilleure méthode dans ces cas là est souvent de calculer et de simplifier les deux termes en parallèle au brouillon, en essayant de tout exprimer en fonction des mêmes notations.
Il est normal pour un calcul difficile comme celui-ci de devoir tester plusieurs pistes. Le résultat étant donné dans l'énoncé, il ne faut pas hésiter à l'admettre si l'intuition ne vient pas après quelques échecs.

$\square$

On admet, dans la suite de cette partie, que $(\hat{f}_n)_{n\geqslant 1}$ converge en loi vers N ce qui implique que pour tout $(x,y)\in\mathbb{R}^2$, avec $x\leqslant y$, on a :

$$\lim_{n\to+\infty} \mathbb{P}(x\leqslant \hat{f}_n \leqslant y) = \Phi(y)-\Phi(x)$$

15. Soit $\alpha\in\,]0,1[$. On pose $\eta_\alpha = t_\alpha^2$ où t_α est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale $(0,1)$.

a) Montrer que $\lim\limits_{n\to+\infty} \mathbb{P}\left((f(a))^2 - \left(2f_n + \dfrac{\eta_\alpha}{2nh_n}\right)f(a) + f_n^2 \leqslant 0\right) = 1-\alpha$.

Démonstration. Par définition $\Phi(t_\alpha) = 1-\frac{\alpha}{2}$.
De plus,

$$\begin{aligned}
\left[(f(a))^2 - \left(2f_n + \frac{\eta_\alpha}{2nh_n}\right)f(a) + f_n^2 \leqslant 0\right] &= \left[(f(a))^2 - 2f_nf(a) - \frac{\eta_\alpha}{2nh_n}f(a) + f_n^2 \leqslant 0\right] \\
&= \left[(f_n - f(a))^2 - \frac{\eta_\alpha}{2nh_n}f(a) \leqslant 0\right] \\
&= \left[(f_n - f(a))^2 \leqslant \frac{\eta_\alpha}{2nh_n}f(a)\right] \\
&= \left[(f_n - f(a))^2 \leqslant \frac{t_\alpha^2}{2nh_n}f(a)\right] \\
&= \left[\left(\frac{\theta_n\sqrt{n}}{f(a)}(f_n - f(a))\right)^2 \leqslant \frac{n\theta_n^2}{f(a)^2}\frac{t_\alpha^2}{2nh_n}f(a)\right] \\
&= \left[\hat{f}_n^2 \leqslant \frac{n2h_nf(a)}{f(a)^2}\frac{t_\alpha^2}{2nh_n}f(a)\right] \\
&= \left[\hat{f}_n^2 \leqslant t_\alpha^2\right] \\
&= \left[-t_\alpha \leqslant \hat{f}_n \leqslant t_\alpha\right]
\end{aligned}$$

Par convergence en loi de $(\hat{f}_n)_{n\geqslant 1}$ vers N, on a

$$\mathbb{P}\left(\left[-t_\alpha \leqslant \hat{f}_n \leqslant t_\alpha\right]\right) \underset{n\to+\infty}{\longrightarrow} \Phi(t_\alpha) - \Phi(-t_\alpha) = 2\Phi(t_\alpha) - 1 = 1-\alpha$$

D'où

$$\boxed{\lim_{n\to+\infty} \mathbb{P}\left((f(a))^2 - \left(2f_n + \frac{\eta_\alpha}{2nh_n}\right) f(a) + f_n^2 \leqslant 0\right) = 1-\alpha.}$$

Commentaire

On commence à construire dans cette question un intervalle de confiance asymptotique pour $f(a)$ au niveau de confiance $1-\alpha$.
Il faut donc nécessairement faire apparaître $\hat{f}_n$ et t_α pour utiliser la convergence en loi qui est admise.
On se souvient alors que

$$\hat{f}_n = \frac{\theta_n\sqrt{n}}{f(a)}(f_n - f(a))$$

donc il faut faire apparaître le terme $f_n - f(a)$.
On fait le lien avec les termes $(f(a))^2$ et f_n^2 qui font penser au développement de $(f_n - f(a))^2$.

$\square$

b) On note, pour $n \geqslant 1$, $\Delta_n = \sqrt{\left(f_n + \dfrac{\eta_\alpha}{4nh_n}\right)^2 - f_n^2}$.

Montrer que :

$$\lim_{n\to+\infty} \mathbb{P}\left(f(a) \in \left[f_n + \frac{\eta_\alpha}{4nh_n} - \Delta_n, f_n + \frac{\eta_\alpha}{4nh_n} + \Delta_n\right]\right) = 1-\alpha$$

Démonstration.

$$\begin{aligned}
\left[(f(a))^2 - \left(2f_n + \frac{\eta_\alpha}{2nh_n}\right) f(a) + f_n^2 \leqslant 0\right] &= \left[(f(a))^2 - 2f(a)\left(f_n + \frac{\eta_\alpha}{4nh_n}\right) + f_n^2 \leqslant 0\right] \\
&= \left[\left(f(a) - \left(f_n + \frac{\eta_\alpha}{4nh_n}\right)\right)^2 - \left(f_n + \frac{\eta_\alpha}{4nh_n}\right)^2 + f_n^2 \leqslant 0\right] \\
&= \left[\left(f(a) - \left(f_n + \frac{\eta_\alpha}{4nh_n}\right)\right)^2 \leqslant \left(f_n + \frac{\eta_\alpha}{4nh_n}\right)^2 - f_n^2\right] \\
&= \left[\left(f(a) - \left(f_n + \frac{\eta_\alpha}{4nh_n}\right)\right)^2 \leqslant \Delta_n^2\right] \\
&= \left[-\Delta_n \leqslant f(a) - \left(f_n + \frac{\eta_\alpha}{4nh_n}\right) \leqslant \Delta_n\right] \\
&= \left[f_n + \frac{\eta_\alpha}{4nh_n} - \Delta_n \leqslant f(a) \leqslant f_n + \frac{\eta_\alpha}{4nh_n} + \Delta_n\right] \\
&= \left[f(a) \in \left[f_n + \frac{\eta_\alpha}{4nh_n} - \Delta_n, f_n + \frac{\eta_\alpha}{4nh_n} + \Delta_n\right]\right]
\end{aligned}$$

D'après la question ***15.a)***, on a bien

$$\boxed{\lim_{n\to+\infty} \mathbb{P}\left(f(a) \in \left[f_n + \tfrac{\eta_\alpha}{4nh_n} - \Delta_n, f_n + \tfrac{\eta_\alpha}{4nh_n} + \Delta_n\right]\right) = 1-\alpha.}$$

Commentaire

Puisque l'on cherche à construire un intervalle de confiance asymptotique pour $f(a)$ au niveau de confiance $1-\alpha$, il faut réussir à isoler $f(a)$ dans l'événement dont on contrôle la probabilité (celui de la question précédente).
On se rend alors compte que l'expression

$$(f(a))^2 - \left(2f_n + \frac{\eta_\alpha}{2nh_n}\right) f(a) + f_n^2$$

n'est rien d'autre qu'un trinôme du second degré en $f(a)$. On met donc en place dans cette question le calcul classique de mise sous forme canonique d'un trinôme du second degré.
On peut aussi remarquer que

$$\begin{aligned}\left(f_n + \frac{\eta_\alpha}{4nh_n}\right)^2 - f_n^2 &= f_n^2 + 2f_n \frac{\eta_\alpha}{4nh_n} + \left(\frac{\eta_\alpha}{4nh_n}\right)^2 - f_n^2 \\ &= 2f_n \frac{\eta_\alpha}{4nh_n} + \left(\frac{\eta_\alpha}{4nh_n}\right)^2 \geqslant 0\end{aligned}$$

donc Δ_n est bien définie.

□

Partie 4 - Convergence « uniforme » en loi vers la loi normale

Soit $n \in \mathbb{N}^*$. On considère n variables aléatoires $X_1, \ldots, X_n$ indépendantes centrées qui possèdent un moment d'ordre 3. On admet alors que ces variables aléatoires possèdent une variance.

On pose, pour tout $k \in [\![1, n]\!]$, $\mathbb{E}(X_k^2) = v_k$, $S_n = \sum\limits_{k=1}^{n} X_k$, $Y_k = S_n - X_k$ et on suppose que $\sum\limits_{k=1}^{n} v_k = 1$.

Soit f une fonction de classe $\mathcal{C}^2$ sur $\mathbb{R}$ telle que pour tout x réel, $|f(x)| \leqslant 1$ et $|f''(x)| \leqslant 2$.

On admet que si X et Y sont des variables aléatoires possédant une espérance alors $\mathbb{E}(Yf(X))$ et $\mathbb{E}(f'(X))$ existent.

16. a) Montrer que $\sum\limits_{k=1}^{n} v_k \mathbb{E}(f'(S_n) - f'(Y_k)) + \sum\limits_{k=1}^{n} \mathbb{E}(X_k^2 f'(Y_k)) = \mathbb{E}(f'(S_n))$.

Démonstration.

- Les variables aléatoires X_k possèdent chacune une variance donc possèdent chacune une espérance.
- On en déduit que S_n et Y_k admettent également une espérance comme sommes de variables aléatoires admettant chacune un espérance.
- D'après le résultat admis, $f'(S_n)$ et $f'(Y_k)$ admettent donc une espérance.
  Par suite, $f'(S_n) - f'(Y_k)$ admet une espérance.
- Pour finir, $Y_k = S_n - X_k = X_1 + \cdots + X_{k-1} + X_{k+1} + \cdots + X_n$ donc par lemme des coalitions, Y_k et X_k sont indépendantes. Toujours par lemme des coalitions, X_k^2 et $f'(Y_k)$ sont aussi indépendantes. Puisque ces deux variables aléatoires admettent une espérance, on peut conclure que $X_k^2 f'(Y_k)$ admet une espérance.

- On peut maintenant faire le calcul :

$$\begin{aligned}
&\sum_{k=1}^{n} v_k \mathbb{E}(f'(S_n) - f'(Y_k)) + \sum_{k=1}^{n} \mathbb{E}(X_k^2 f'(Y_k)) \\
=&\sum_{k=1}^{n} v_k (\mathbb{E}(f'(S_n)) - \mathbb{E}(f'(Y_k))) + \sum_{k=1}^{n} \mathbb{E}(X_k^2 f'(Y_k)) && \textit{(par linéarité)} \\
=&\sum_{k=1}^{n} v_k \mathbb{E}(f'(S_n)) - \sum_{k=1}^{n} v_k \mathbb{E}(f'(Y_k)) + \sum_{k=1}^{n} \mathbb{E}(X_k^2 f'(Y_k)) \\
=&\mathbb{E}(f'(S_n)) \sum_{k=1}^{n} v_k - \cancel{\sum_{k=1}^{n} \mathbb{E}(X_k^2)\mathbb{E}(f'(Y_k))} + \cancel{\sum_{k=1}^{n} \mathbb{E}(X_k^2)\mathbb{E}(f'(Y_k))} && \textit{(par indépendance)} \\
=&\mathbb{E}(f'(S_n)) && \textit{(car } \sum_{k=1}^{n} v_k = 1\textit{)}
\end{aligned}$$

□

b) Montrer que $\displaystyle\sum_{k=1}^{n} \mathbb{E}(X_k [f(S_n) - f(Y_k)]) = \sum_{k=1}^{n} \mathbb{E}(X_k f(S_n)) = \mathbb{E}(S_n f(S_n))$.

Démonstration.

- Les variables aléatoires X_k et S_n admettent une espérance donc $X_k f(S_n)$ admet une espérance (cf résultat admis). Par somme $S_n f(S_n) = \displaystyle\sum_{k=1}^{n} X_k f(S_n)$ admet aussi une espérance.
- Les variables aléatoires X_k et Y_k admettent une espérance donc $X_k f(Y_k)$ admet une espérance (cf résultat admis).
- Par somme, $X_k f(S_n) - X_k f(Y_k) = X_k(f(S_n) - f(Y_k))$ admet une espérance.
- Les variables Y_k et X_k sont indépendantes donc par lemme des coalitions, $f(Y_k)$ et X_k sont aussi indépendantes.
- On peut maintenant faire le calcul :

$$\begin{aligned}
&\sum_{k=1}^{n} \mathbb{E}(X_k [f(S_n) - f(Y_k)]) \\
=&\sum_{k=1}^{n} \mathbb{E}(X_k f(S_n) - X_k f(Y_k)) \\
=&\sum_{k=1}^{n} (\mathbb{E}(X_k f(S_n)) - \mathbb{E}(X_k f(Y_k))) && \textit{(par linéarité)} \\
=&\sum_{k=1}^{n} \mathbb{E}(X_k f(S_n)) - \sum_{k=1}^{n} \mathbb{E}(X_k f(Y_k)) \\
=&\sum_{k=1}^{n} \mathbb{E}(X_k f(S_n)) - \sum_{k=1}^{n} \mathbb{E}(X_k)\mathbb{E}(f(Y_k)) && \textit{(par indépendance)} \\
=&\sum_{k=1}^{n} \mathbb{E}(X_k f(S_n)) && \textit{(car } \forall k \in [\![1, n]\!],\ \mathbb{E}(X_k) = 0\textit{)} \\
=&\mathbb{E}\left(\sum_{k=1}^{n} X_k f(S_n)\right) && \textit{(par linéarité)} \\
=&\mathbb{E}\left(f(S_n) \sum_{k=1}^{n} X_k\right) \\
=&\mathbb{E}\left(f(S_n) S_n\right)
\end{aligned}$$

□

c) En déduire que :

$$\mathbb{E}(f'(S_n) - S_n f(S_n)) = \sum_{k=1}^{n} v_k \mathbb{E}(f'(S_n) - f'(Y_k)) + \sum_{k=1}^{n} \mathbb{E}\left(X_k \left[X_k f'(Y_k) - (f(S_n) - f(Y_k))\right]\right)$$

Démonstration.

- Les variables aléatoires $f'(S_n)$ et $S_n f(S_n)$ admettent une espérance donc $f'(S_n) - S_n f(S_n)$ admet une espérance par somme.
- Les variables aléatoires $X_k^2 f'(Y_k)$, $X_k f(S_n)$ et $X_k f(Y_k)$ admettent une espérance donc la variable aléatoire

$$X_k \left[X_k f'(Y_k) - (f(S_n) - f(Y_k))\right] = X_k^2 f'(Y_k) - X_k f(S_n) + X_k f(Y_k)$$

  admet une espérance par somme.
- On peut maintenant faire le calcul :

$$\begin{aligned}
& \mathbb{E}(f'(S_n) - S_n f(S_n)) \\
= & \mathbb{E}(f'(S_n)) - \mathbb{E}(S_n f(S_n)) && \textit{(par linéarité)} \\
= & \sum_{k=1}^{n} v_k \mathbb{E}(f'(S_n) - f'(Y_k)) + \sum_{k=1}^{n} \mathbb{E}(X_k^2 f'(Y_k)) && \textit{(cf question } \textbf{\textit{16.a)}}\textit{)} \\
& - \sum_{k=1}^{n} \mathbb{E}(X_k \left[f(S_n) - f(Y_k)\right]) && \textit{(cf question } \textbf{\textit{16.b)}}\textit{)} \\
= & \sum_{k=1}^{n} v_k \mathbb{E}(f'(S_n) - f'(Y_k)) \\
& + \sum_{k=1}^{n} \mathbb{E}(X_k^2 f'(Y_k) - X_k \left[f(S_n) - f(Y_k)\right]) && \textit{(par linéarité)} \\
= & \sum_{k=1}^{n} v_k \mathbb{E}(f'(S_n) - f'(Y_k)) + \sum_{k=1}^{n} \mathbb{E}\left(X_k \left[X_k f'(Y_k) - (f(S_n) - f(Y_k))\right]\right)
\end{aligned}$$

□

17. Soit a et b deux réels.

a) Montrer que :

$$b f'(a) - (f(a+b) - f(a)) = \int_0^1 b(f'(a) - f'(a+tb)) \, dt$$

Démonstration. La fonction f est de classe $\mathcal{C}^2$ sur $\mathbb{R}$ donc la fonction $t \mapsto f'(a+tb)$ est continue sur $\mathbb{R}$ comme composée de deux fonctions continues sur $\mathbb{R}$.

$$\begin{aligned}
\int_0^1 b(f'(a) - f'(a+tb)) \, dt &= \int_0^1 b f'(a) \, dt - \int_0^1 b f'(a+tb) \, dt && \textit{(par linéarité)} \\
&= b f'(a) - \left[\ f(a+tb)\ \right]_0^1 \\
&= b f'(a) - (f(a+b) - f(a))
\end{aligned}$$

□

b) En déduire que :

$$\left|b f'(a) - (f(a+b) - f(a))\right| \leqslant b^2$$

Démonstration. D'après la question précédente :

$$\begin{aligned}\left|bf'(a) - (f(a+b) - f(a))\right| &= \left|\int_0^1 b(f'(a) - f'(a+tb))\ dt\right| \\ &\leqslant \int_0^1 \left|b(f'(a) - f'(a+tb))\right|\ dt \quad \textit{(par inégalité triangulaire, car } 0 \leqslant 1\textit{)}\end{aligned}$$

Soit $t \in [0,1]$.

$$\left|b(f'(a) - f'(a+tb))\right| = |b| \left|f'(a) - f'(a+tb)\right|$$

Or, pour tout $x \in \mathbb{R}$, $|f''(x)| \leqslant 2$, donc par inégalité des accroissements finis :

$$\left|f'(a) - f'(a+tb)\right| \leqslant 2\,|a - (a+tb)| = 2\,|tb| = 2\,|b|\,t \qquad \text{car } t \geqslant 0$$

Par croissance de l'intégrale, les bornes étant rangées dans l'ordre croissant ($0 \leqslant 1$), on a

$$\begin{aligned}\left|bf'(a) - (f(a+b) - f(a))\right| &\leqslant \int_0^1 \left|b(f'(a) - f'(a+tb))\right|\ dt \\ &\leqslant \int_0^1 2\,|b|^2\,t\ dt \\ &= 2b^2 \int_0^1 t\ dt \\ &= 2b^2 \times \frac{1}{2} \\ &= b^2\end{aligned}$$

On a bien

$$\boxed{|bf'(a) - (f(a+b) - f(a))| \leqslant b^2.}$$

□

c) En conclure que :

$$\left|\mathbb{E}(f'(S_n) - S_n f(S_n))\right| \leqslant 2 \sum_{k=1}^{n} v_k \mathbb{E}(|X_k|) + \sum_{k=1}^{n} \mathbb{E}(|X_k|^3)$$

puis, grâce à l'inégalité (R_2), que, pour tout x réel :

$$d_{S_n}(x) \leqslant \sqrt{3\left(2 \sum_{k=1}^{n} v_k \mathbb{E}(|X_k|) + \sum_{k=1}^{n} \mathbb{E}(|X_k|^3)\right)} \tag{R_3}$$

Démonstration. D'après la question ***16.c)*** et par inégalité triangulaire :

$$\begin{aligned}&\left|\mathbb{E}(f'(S_n) - S_n f(S_n))\right| \\ &\leqslant \sum_{k=1}^{n} |v_k| \left|\mathbb{E}(f'(S_n) - f'(Y_k))\right| + \sum_{k=1}^{n} \left|\mathbb{E}\left(X_k \left[X_k f'(Y_k) - (f(S_n) - f(Y_k))\right]\right)\right| \\ &\leqslant \sum_{k=1}^{n} v_k \left|\mathbb{E}(f'(S_n) - f'(Y_k))\right| + \sum_{k=1}^{n} \left|\mathbb{E}\left(X_k \left[X_k f'(Y_k) - (f(S_n) - f(Y_k))\right]\right)\right|\end{aligned}$$

En effet, $X_k^2 \geqslant 0$ donc $v_k \geqslant 0$ par positivité de l'espérance.
Soit $k \in [\![1, n]\!]$. Majorons maintenant chaque terme séparément :

•

$$\begin{aligned}\left|\mathbb{E}(f'(S_n)-f'(Y_k))\right| &\leqslant \mathbb{E}(\left|f'(S_n)-f'(Y_k)\right|) && \textit{(par inégalité triangulaire)}\\ &\leqslant \mathbb{E}(2\left|S_n-Y_k\right|) && \textit{(par IAF et par croissance de l'espérance)}\\ &= 2\mathbb{E}(|X_k|) && \textit{(par linéarité)}\end{aligned}$$

•

$$\begin{aligned}&\left|\mathbb{E}\left(X_k\left[X_k f'(Y_k)-(f(S_n)-f(Y_k))\right]\right)\right|\\ \leqslant&\mathbb{E}\left(\left|X_k\left[X_k f'(Y_k)-(f(S_n)-f(Y_k))\right]\right|\right) && \textit{(par inégalité triangulaire)}\\ =&\mathbb{E}\left(|X_k|\left|X_k f'(Y_k)-(f(S_n)-f(Y_k))\right|\right)\\ =&\mathbb{E}\left(|X_k|\left|X_k f'(Y_k)-(f(Y_k+X_k)-f(Y_k))\right|\right)\\ \leqslant&\mathbb{E}\left(|X_k|\,X_k^2\right) && \textit{(cf question \textbf{17.b)} et par croissance de l'espérance)}\\ =&\mathbb{E}\left(|X_k|\,|X_k|^2\right)\\ =&\mathbb{E}\left(|X_k|^3\right)\end{aligned}$$

On en déduit, en sommant toutes ces inégalités, que :

$$\boxed{|\mathbb{E}(f'(S_n)-S_n f(S_n))| \leqslant 2\sum_{k=1}^{n} v_k\mathbb{E}(|X_k|)+\sum_{k=1}^{n}\mathbb{E}(|X_k|^3).}$$

Notons maintenant

$$M_X = 2\sum_{k=1}^{n} v_k\mathbb{E}(|X_k|)+\sum_{k=1}^{n}\mathbb{E}(|X_k|^3)$$

Soit $h\in W$. D'après la partie 1, la fonction f_h vérifie :

- f_h est de classe $\mathcal{C}^2$ sur $\mathbb{R}$
- pour tout $x\in\mathbb{R}$, $|f_h(x)|\leqslant 1$
- pour tout $x\in\mathbb{R}$, $|f_h''(x)|\leqslant 2$

On peut donc appliquer la majoration précédente avec $f=f_h$:

$$\left|\mathbb{E}(f_h'(S_n)-S_n f_h(S_n))\right|\leqslant M_X$$

Or, d'après la question ***3.c)*** (la variable S_n admet bien une espérance),

$$|\mathbb{E}(h(S_n))-\mathbb{E}(h(N))| = \left|\mathbb{E}(f_h'(S_n)-S_n f_h(S_n))\right|$$

On a donc démontré que, pour tout $h\in W$,

$$|\mathbb{E}(h(S_n))-\mathbb{E}(h(N))|\leqslant M_X$$

On peut donc appliquer la question ***10.*** pour obtenir que, pour tout x réel :

$$d_{S_n}(x)\leqslant\sqrt{3M_X}$$

i.e.

$$\boxed{\text{pour tout } x \text{ réel, } d_{S_n}(x)\leqslant\sqrt{3\left(2\sum_{k=1}^{n} v_k\mathbb{E}(|X_k|)+\sum_{k=1}^{n}\mathbb{E}(|X_k|^3)\right)}.}$$

Commentaire

Il manque des justifications dans cette preuve. En effet, nous n'avons pas montré que toutes les espérances écrites existent bien. En particulier, pourquoi $\mathbb{E}(|X_k|^3)$ existe ?
Nous sommes dans l'obligation d'admettre le résultat suivant :
« Si Y est une variable aléatoire, alors Y admet une espérance si et seulement si $|Y|$ admet une espérance. »
(seul le sens direct est utile ici)
Nous avons les outils pour faire la preuve dans le cas discret ou le cas à densité, mais pas dans le cas général.
Ce commentaire est à mettre en relation avec celui fait à la partie 2 sur un résultat qui était admis sans vraiment le dire. Cela fait maintenant deux fois qu'une réelle difficulté est mise sous le tapis dans ce sujet, en raison du fait que nous n'avons pas d'outils théoriques au programme pour définir l'espérance d'une variable aléatoire quelconque.
En plus de cette difficulté technique, cette question est délicate parce qu'elle fait le bilan des deux premières parties ainsi que du début de la quatrième partie.

□

Une définition - *Dans la suite du sujet, si $(X_n)_{n\geqslant 1}$ est une suite de variables aléatoires réelles et $(\delta_n)_{n\geqslant 1}$ une suite réelle de limite nulle qui vérifient,*

$$\forall n \in \mathbb{N}^*, \forall x \in \mathbb{R}, d_{X_n}(x) \leqslant \delta_n$$

on dira alors que $(X_n)_{n\geqslant 1}$ converge uniformément en loi vers N.
On remarque, et on l'admet pour la suite, que si $(X_n)_{n\geqslant 1}$ converge uniformément en loi vers N alors $(X_n)_{n\geqslant 1}$ converge en loi vers N.

Commentaire

La définition de convergence (simple) en loi vers N, qui est au programme, s'écrit :

$$\forall x \in \mathbb{R}, F_{X_n}(x) \underset{n\to+\infty}{\longrightarrow} \Phi(x)$$

i.e.

$$\forall x \in \mathbb{R}, d_{X_n}(x) \underset{n\to+\infty}{\longrightarrow} 0$$

Pour chaque x fixé, la suite $(d_{X_n}(x))_{n\geqslant 1}$ des distances au point x converge vers 0. La *vitesse de convergence* peut alors dépendre de x. Cela peut poser des problèmes dans certaines preuves, où l'on a besoin que cette vitesse ne dépende pas du point x considéré. C'est à cet effet que l'on introduit la notion de convergence « uniforme ».
Écrire

$$\forall n \in \mathbb{N}^*, \forall x \in \mathbb{R}, d_{X_n}(x) \leqslant \delta_n$$

c'est affirmer que **toutes** les suites $(d_{X_n}(x))_{n\geqslant 1}$ convergent vers 0 au moins aussi vite que la suite $(\delta_n)_{n\geqslant 1}$.

18. *Une première application.* On suppose dans cette question que $(Z_k)_{k\geqslant 1}$ est une suite de variables aléatoires indépendantes, suivant la même loi et admettant des moments d'ordre 1 à 3.

On note pour $i \in [\![1, 3]\!]$, $s_i = \mathbb{E}(|Z_k - \mathbb{E}(Z_k)|^i)$, $\sigma = \sqrt{s_2}$ et , $X_k = \dfrac{Z_k - \mathbb{E}(Z_k)}{\sigma\sqrt{n}}$ pour tout $k \in \mathbb{N}^*$.

On suppose que $\sigma \neq 0$.

On utilise les notations de la question précédente.

a) Montrer que l'on peut appliquer l'inégalité (R_3) qui donne ici :

$$d_{S_n}(x) \leqslant \sqrt{3\frac{2\sigma^2 s_1 + s_3}{\sigma^3\sqrt{n}}}$$

Démonstration.

- Les variables aléatoires Z_k sont indépendantes donc les variables aléatoires $X_k = \dfrac{Z_k - \mathbb{E}(Z_k)}{\sigma\sqrt{n}}$ sont indépendantes par lemme des coalitions.
- Les variables aléatoires Z_k admettent chacune des moments d'ordre 1 à 3 donc les variables aléatoires X_k admettent également chacune des moments d'ordre 1 à 3 (par transformation affine).
- Soit $k \geqslant 1$. Par linéarité :
$$\mathbb{E}(X_k) = \mathbb{E}\left(\frac{Z_k - \mathbb{E}(Z_k)}{\sigma\sqrt{n}}\right) = \frac{1}{\sigma\sqrt{n}}\mathbb{E}\left(Z_k - \mathbb{E}(Z_k)\right) = \frac{1}{\sigma\sqrt{n}}(\mathbb{E}(Z_k) - \mathbb{E}(Z_k)) = 0$$
donc les variables aléatoires X_k sont centrées.
-
$$\begin{aligned}\sum_{k=1}^{n} v_k &= \sum_{k=1}^{n} \mathbb{E}(X_k^2)\\ &= \sum_{k=1}^{n} \mathbb{E}\left(\left(\frac{Z_k - \mathbb{E}(Z_k)}{\sigma\sqrt{n}}\right)^2\right)\\ &= \sum_{k=1}^{n} \frac{1}{n\sigma^2}\mathbb{E}\left((Z_k - \mathbb{E}(Z_k))^2\right)\\ &= \sum_{k=1}^{n} \frac{1}{ns_2}s_2\\ &= 1\end{aligned}$$

Nous sommes donc dans les conditions d'application de la majoration (R_3) qui s'écrit : pour tout x réel,

$$d_{S_n}(x) \leqslant \sqrt{3M_X}$$

où

$$M_X = 2\sum_{k=1}^{n} v_k\mathbb{E}(|X_k|) + \sum_{k=1}^{n} \mathbb{E}(|X_k|^3)$$

Or,

$$\begin{aligned}\mathbb{E}(|X_k|) &= \mathbb{E}\left(\left|\frac{Z_k - \mathbb{E}(Z_k)}{\sigma\sqrt{n}}\right|\right)\\ &= \frac{1}{\sigma\sqrt{n}}\mathbb{E}(|Z_k - \mathbb{E}(Z_k)|)\\ &= \frac{s_1}{\sigma\sqrt{n}}\end{aligned}$$

et

$$\begin{aligned}\mathbb{E}(|X_k|^3) &= \mathbb{E}\left(\left|\frac{Z_k - \mathbb{E}(Z_k)}{\sigma\sqrt{n}}\right|^3\right)\\ &= \frac{1}{(\sigma\sqrt{n})^3}\mathbb{E}(|Z_k - \mathbb{E}(Z_k)|^3)\\ &= \frac{s_3}{(\sigma\sqrt{n})^3}\end{aligned}$$

donc

$$\begin{aligned} M_X &= 2\sum_{k=1}^{n} v_k \frac{s_1}{\sigma\sqrt{n}} + \sum_{k=1}^{n} \frac{s_3}{(\sigma\sqrt{n})^3} \\ &= \frac{2s_1}{\sigma\sqrt{n}} \sum_{k=1}^{n} v_k + n\frac{s_3}{(\sigma\sqrt{n})^3} \\ &= \frac{2s_1}{\sigma\sqrt{n}} + n\frac{s_3}{(\sigma\sqrt{n})^3} && \textit{(car } \sum_{k=1}^{n} v_k = 1\textit{)} \\ &= \frac{2s_1}{\sigma\sqrt{n}} + \frac{s_3}{\sigma^3\sqrt{n}} \\ &= \frac{2\sigma^2 s_1 + s_3}{\sigma^3\sqrt{n}} \end{aligned}$$

d'où

$$\boxed{d_{S_n}(x) \leqslant \sqrt{3\frac{2\sigma^2 s_1 + s_3}{\sigma^3\sqrt{n}}}.}$$

□

b) En déduire que $(S_n)_{n\geqslant 1}$ converge uniformément en loi vers N, donc converge en loi vers N. Quel résultat du cours nous aurait permis d'obtenir cette dernière convergence directement ?

Démonstration. σ, s_1 et s_3 ne dépendent pas de n (ce sont des constantes) donc

$$\delta_n = \sqrt{3\frac{2\sigma^2 s_1 + s_3}{\sigma^3\sqrt{n}}} \underset{n\to+\infty}{\longrightarrow} 0$$

D'après la question précédente :

$$\forall n \in \mathbb{N}^*, \forall x \in \mathbb{R}, d_{S_n}(x) \leqslant \delta_n$$

donc

$$\boxed{(S_n)_{n\geqslant 1} \text{ converge uniformément en loi vers } N}$$

D'après la remarque, on a donc également !

$$\boxed{(S_n)_{n\geqslant 1} \text{ converge en loi vers } N}$$

On aurait pu obtenir cette dernière convergence par une application directe du théorème central limite.

En effet, notons $W_n = \sum_{k=1}^{n} Z_k$ et $m = \mathbb{E}(Z_k)$.

- les variables aléatoires Z_k sont indépendantes
- les variables aléatoires Z_k suivent la même loi
- les variables aléatoires Z_k admettent une espérance et une variance non nulle

Donc $W_n^* \underset{n\to+\infty}{\overset{\mathscr{L}}{\longrightarrow}} N$.

Il reste à montrer que $S_n = W_n^*$.

$$\begin{aligned} S_n &= \sum_{k=1}^{n} X_k \\ &= \sum_{k=1}^{n} \frac{Z_k - m}{\sigma\sqrt{n}} \\ &= \frac{W_n - nm}{\sigma\sqrt{n}} \\ &= W_n^* && \textit{(calcul classique)} \end{aligned}$$

□

19. *Une deuxième application.* On suppose dans cette question que $Z_1, \ldots, Z_n$ sont des variables aléatoires indépendantes suivant la même loi de Bernoulli de paramètre $p_n \in]0,1[$.

On pose $\sigma_n = \sqrt{np_n(1-p_n)}$ et $X_k = \dfrac{Z_k - p_n}{\sigma_n}$ pour tout $k \in [\![1,n]\!]$.

a) Montrer que $\mathbb{E}(|X_k|) = \dfrac{2\sigma_n}{n}$ et $\mathbb{E}(|X_k|^3) \leqslant \dfrac{2}{n\sigma_n}$.

Démonstration. Soit $k \in [\![1,n]\!]$. Représentons les lois des différentes variables aléatoires en jeu sous forme de tableaux :

- Loi de Z_k :

$x \in Z_k(\Omega)$	0	1
$\mathbb{P}([Z_k = x])$	$1-p_n$	p_n

- Loi de X_k :

$x \in X_k(\Omega)$	$-\dfrac{p_n}{\sigma_n}$	$\dfrac{1-p_n}{\sigma_n}$
$\mathbb{P}([X_k = x])$	$1-p_n$	p_n

- Loi de $|X_k|$:

$x \in \lvert X_k\rvert(\Omega)$	$\dfrac{p_n}{\sigma_n}$	$\dfrac{1-p_n}{\sigma_n}$
$\mathbb{P}([\lvert X_k\rvert = x])$	$1-p_n$	p_n

- Loi de $|X_k|^3$:

$x \in \lvert X_k\rvert^3(\Omega)$	$\left(\dfrac{p_n}{\sigma_n}\right)^3$	$\left(\dfrac{1-p_n}{\sigma_n}\right)^3$
$\mathbb{P}\left(\left[\lvert X_k\rvert^3 = x\right]\right)$	$1-p_n$	p_n

Remarquons que $\dfrac{p_n}{\sigma_n} = \dfrac{1-p_n}{\sigma_n}$ si et seulement si $p_n = \dfrac{1}{2}$. La formule utilisée pour calculer les espérances ci-dessous reste vraie même dans ce cas.

Les variables aléatoires $|X_k|$ et $|X_k|^3$ sont finies donc admettent chacune une espérance.

$$\begin{aligned}
\mathbb{E}(|X_k|) &= \frac{p_n}{\sigma_n}(1-p_n) + \frac{1-p_n}{\sigma_n}p_n \\
&= \frac{2p_n(1-p_n)}{\sigma_n} \\
&= \frac{2np_n(1-p_n)}{n\sigma_n} \\
&= \frac{2\sigma_n^2}{n\sigma_n} \\
&= \frac{2\sigma_n}{n}
\end{aligned}$$

et

$$\begin{aligned}
\mathbb{E}(|X_k|^3) &= \left(\frac{p_n}{\sigma_n}\right)^3 (1-p_n) + \left(\frac{1-p_n}{\sigma_n}\right)^3 p_n \\
&= \frac{p_n(1-p_n)}{\sigma_n^3}\left(p_n^2 + (1-p_n)^2\right) \\
&= \frac{2}{n\sigma_n}\frac{np_n(1-p_n)}{2\sigma_n^2}\left(p_n^2 + (1-p_n)^2\right) \\
&= \frac{2}{n\sigma_n}\frac{\sigma_n^2}{2\sigma_n^2}\left(p_n^2 + (1-p_n)^2\right) \\
&= \frac{2}{n\sigma_n}\frac{1}{2}\left(p_n^2 + (1-p_n)^2\right) \\
&\leqslant \frac{2}{n\sigma_n}\frac{1}{2}(1+1) && \textit{(car } p_n^2 \leqslant 1 \textit{ et } (1-p_n)^2 \leqslant 1\textit{)} \\
&= \frac{2}{n\sigma_n}
\end{aligned}$$

D'où

$$\boxed{\mathbb{E}(|X_k|) = \frac{2\sigma_n}{n} \text{ et } \mathbb{E}(|X_k|^3) \leqslant \frac{2}{n\sigma_n}.}$$

□

b) En déduire que, pour tout x réel :

$$d_{S_n}(x) \leqslant 2\sqrt{3\left(\frac{\sigma_n}{n} + \frac{1}{2\sigma_n}\right)}$$

Démonstration. On démontre comme à la question ***18.a)*** que l'on peut appliquer l'inégalité (R_3). Soit $x \in \mathbb{R}$.

$$d_{S_n}(x) \leqslant \sqrt{3M_X}$$

où

$$\begin{aligned}
M_X &= 2\sum_{k=1}^{n} v_k \mathbb{E}(|X_k|) + \sum_{k=1}^{n} \mathbb{E}(|X_k|^3) \\
&\leqslant 2\sum_{k=1}^{n} v_k \frac{2\sigma_n}{n} + \sum_{k=1}^{n} \frac{2}{n\sigma_n} && \textit{(cf question } \textbf{\textit{19.a)}}\textit{)} \\
&= \frac{4\sigma_n}{n}\sum_{k=1}^{n} v_k + \frac{2}{\sigma_n} \\
&= 4\left(\frac{\sigma_n}{n} + \frac{1}{2\sigma_n}\right) && \textit{(car } \sum_{k=1}^{n} v_k = 1\textit{)}
\end{aligned}$$

donc

$$\boxed{d_{S_n}(x) \leqslant \sqrt{3\times 4\left(\frac{\sigma_n}{n} + \frac{1}{2\sigma_n}\right)} = 2\sqrt{3\left(\frac{\sigma_n}{n} + \frac{1}{2\sigma_n}\right)}.}$$

□

c) Justifier le résultat suivant :
si pour tout $n \in \mathbb{N}^*$, T_n est une variable aléatoire qui suit la loi binomiale (n, p_n) avec $\lim\limits_{n\to+\infty} p_n = 0$ et $\lim\limits_{n\to+\infty} np_n = +\infty$ alors, $\left(\dfrac{T_n - np_n}{\sqrt{np_n(1-p_n)}}\right)_{n\geqslant 1}$ converge uniformément en loi vers N.

Démonstration. On note $T_n = \sum\limits_{k=1}^{n} Z_k$ où les Z_k sont indépendantes et suivent toutes la même loi de Bernoulli de paramètre p_n.
On note comme précédemment $\sigma_n = \sqrt{np_n(1-p_n)}$ et $X_k = \dfrac{Z_k - p_n}{\sigma_n}$.
Alors

$$\begin{aligned}
S_n &= \sum_{k=1}^{n} X_k \\
&= \sum_{k=1}^{n} \frac{Z_k - p_n}{\sigma_n} \\
&= \frac{T_n - np_n}{\sigma_n} \\
&= \frac{T_n - np_n}{\sqrt{np_n(1-p_n)}}
\end{aligned}$$

et donc il faut montrer que $(S_n)_{n\geqslant 1}$ converge uniformément en loi vers N.
Notons $\delta_n = 2\sqrt{3\left(\dfrac{\sigma_n}{n} + \dfrac{1}{2\sigma_n}\right)}$. D'après la question précédente, pour tout $x \in \mathbb{R}$,

$$d_{S_n}(x) \leqslant \delta_n$$

Montrons que $\delta_n \underset{n\to+\infty}{\longrightarrow} 0$.

- $p_n \underset{n\to+\infty}{\longrightarrow} 0$ par hypothèse donc $1 - p_n \underset{n\to+\infty}{\longrightarrow} 1$ donc $1 - p_n \underset{n\to+\infty}{\sim} 1$ donc

$$\sigma_n \underset{n\to+\infty}{\sim} \sqrt{np_n}$$

et donc $\sigma_n \underset{n\to+\infty}{\longrightarrow} +\infty$ car $np_n \underset{n\to+\infty}{\longrightarrow} +\infty$ par hypothèse.

- Par suite :

$$\frac{\sigma_n}{n} \underset{n\to+\infty}{\sim} \sqrt{\frac{p_n}{n}}$$

et donc $\dfrac{\sigma_n}{n} \underset{n\to+\infty}{\longrightarrow} 0$.

On en déduit que $\delta_n \underset{n\to+\infty}{\longrightarrow} 0$ et donc

$$\boxed{\left(\frac{T_n - np_n}{\sqrt{np_n(1-p_n)}}\right)_{n\geqslant 1} \text{ converge uniformément en loi vers } N.}$$

$\square$

20. *Un petit lemme.* Soit $(V_n)_{n\geqslant 1}$ une suite de variables aléatoires réelles qui converge uniformément en loi vers N. Soit $(a_n)_{n\geqslant 1}$ une suite de réels strictement positifs qui converge vers a, tel que $a > 0$, et $(b_n)_{n\geqslant 1}$ une suite de réels qui converge vers b.

a) Soit X une variable aléatoire et (α, β) un couple de réels avec $\alpha > 0$. On note $F_{\alpha X+\beta}$ et $F_{\alpha N+\beta}$ les fonctions de répartition respectives de $\alpha X + \beta$ et $\alpha N + \beta$.

Montrer que, pour tout x réel,

$$|F_{\alpha X+\beta}(x) - F_{\alpha N+\beta}(x)| = d_X\left(\frac{x-\beta}{\alpha}\right)$$

Démonstration. Soit $x \in \mathbb{R}$.

$$\begin{aligned}
|F_{\alpha X+\beta}(x) - F_{\alpha N+\beta}(x)| &= |\mathbb{P}([\alpha X + \beta \leqslant x]) - \mathbb{P}([\alpha N + \beta \leqslant x])| \\
&= |\mathbb{P}([\alpha X \leqslant x - \beta]) - \mathbb{P}([\alpha N \leqslant x - \beta])| \\
&= \left|\mathbb{P}\left(\left[X \leqslant \frac{x-\beta}{\alpha}\right]\right) - \mathbb{P}\left(\left[N \leqslant \frac{x-\beta}{\alpha}\right]\right)\right| \qquad \textit{(car } \alpha > 0\textit{)} \\
&= \left|F_X\left(\frac{x-\beta}{\alpha}\right) - \Phi\left(\frac{x-\beta}{\alpha}\right)\right| \\
&= d_X\left(\frac{x-\beta}{\alpha}\right)
\end{aligned}$$

$\square$

b) Montrer que pour tout x réel,

$$\lim_{n\to+\infty} (\mathbb{P}(a_n V_n + b_n \leqslant x) - \mathbb{P}(a_n N + b_n \leqslant x)) = 0$$

Démonstration. Soit (δ_n) une suite de limite nulle telle que

$$\forall n \in \mathbb{N}^*, \forall x \in \mathbb{R}, d_{V_n}(x) \leqslant \delta_n$$

Soit $x \in \mathbb{R}$. D'après la question précédente :

$$\begin{aligned}
|\mathbb{P}(a_n V_n + b_n \leqslant x) - \mathbb{P}(a_n N + b_n \leqslant x)| &= |F_{a_n V_n + b_n}(x) - F_{a_n N + b_n}(x)| \\
&= d_{V_n}\left(\frac{x-b_n}{a_n}\right) \qquad \textit{(}\forall n \in \mathbb{N}^*, a_n > 0\textit{)} \\
&\leqslant \delta_n
\end{aligned}$$

$\delta_n \underset{n\to+\infty}{\longrightarrow} 0$ donc, par théorème d'encadrement,

$$\boxed{\lim_{n\to+\infty} \left(\mathbb{P}(a_nV_n + b_n \leqslant x) - \mathbb{P}(a_nN + b_n \leqslant x)\right) = 0.}$$

□

c) Établir que $\lim\limits_{n\to+\infty} \mathbb{P}(a_nN + b_n \leqslant x) = \mathbb{P}(aN + b \leqslant x)$ puis en déduire que $(a_nV_n + b_n)_{n\geqslant 1}$ converge en loi vers $aN + b$. Quelle est la loi de la variable aléatoire $aN + b$?

Démonstration. Soit $x \in \mathbb{R}$.

$$\begin{aligned} \mathbb{P}(a_nN + b_n \leqslant x) - \mathbb{P}(aN + b \leqslant x) &= \mathbb{P}\left(N \leqslant \frac{x - b_n}{a_n}\right) - \mathbb{P}\left(N \leqslant \frac{x - b}{a}\right) \quad a_n > 0 \textit{ et } a > 0 \\ &= \Phi\left(\frac{x - b_n}{a_n}\right) - \Phi\left(\frac{x - b}{a}\right) \end{aligned}$$

Or, $a_n \underset{n\to+\infty}{\longrightarrow} a$ et $b_n \underset{n\to+\infty}{\longrightarrow} b$ donc $\dfrac{x - b_n}{a_n} \underset{n\to+\infty}{\longrightarrow} \dfrac{x - b}{a}$. De plus, Φ est continue en $\dfrac{x - b}{a}$. Donc

$$\Phi\left(\frac{x - b_n}{a_n}\right) \underset{n\to+\infty}{\longrightarrow} \Phi\left(\frac{x - b}{a}\right)$$

i.e.

$$\boxed{\lim_{n\to+\infty} \mathbb{P}(a_nN + b_n \leqslant x) = \mathbb{P}(aN + b \leqslant x).}$$

De plus,

$$\begin{aligned} &\mathbb{P}(a_nV_n + b_n \leqslant x) - \mathbb{P}(aN + b \leqslant x) \\ &= \left(\mathbb{P}(a_nV_n + b_n \leqslant x) - \mathbb{P}(a_nN + b_n \leqslant x)\right) + \left(\mathbb{P}(a_nN + b_n \leqslant x) - \mathbb{P}(aN + b \leqslant x)\right) \end{aligned}$$

et

- $\lim\limits_{n\to+\infty} \left(\mathbb{P}(a_nV_n + b_n \leqslant x) - \mathbb{P}(a_nN + b_n \leqslant x)\right) = 0$
- $\lim\limits_{n\to+\infty} \left(\mathbb{P}(a_nN + b_n \leqslant x) - \mathbb{P}(aN + b \leqslant x)\right) = 0$

donc

$$\mathbb{P}(a_nV_n + b_n \leqslant x) - \mathbb{P}(aN + b \leqslant x) \underset{n\to+\infty}{\longrightarrow} 0$$

Ceci étant vrai pour tout $x \in \mathbb{R}$,

$$\boxed{a_nV_n + b_n \xrightarrow[n\to+\infty]{\mathscr{L}} aN + b.}$$

$N \hookrightarrow \mathcal{N}(0, 1)$ et $a > 0$ donc, par transformation affine, $aN + b$ suit une loi normale.

- $\mathbb{E}(aN + b) = a\mathbb{E}(N) + b = b$
- $\mathbb{V}(aN + b) = a^2\mathbb{V}(N) = a^2$

donc

$$\boxed{aN + b \hookrightarrow \mathcal{N}\left(b, a^2\right)}$$

□

21. On reprend les notations de la partie 3.

a) Justifier que $(D_n)_{n\geqslant 1}$ converge uniformément en loi vers N.

Démonstration. D'après les résultats de la partie 3, on a, pour tout $n \in \mathbb{N}^*$,

$$D_n = \frac{C_n - np_n}{\sigma_n} = \frac{C_n - np_n}{\sqrt{np_n(1-p_n)}}$$

où

- $C_n \hookrightarrow \mathcal{B}(n, p_n)$
- $\lim\limits_{n\to+\infty} p_n = 0$ (car $h_n \underset{n\to+\infty}{\longrightarrow} 0$)
- $\lim\limits_{n\to+\infty} np_n = +\infty$ (car $nh_n \underset{n\to+\infty}{\longrightarrow} +\infty$)

D'après la question ***19.c)***, $\left(\dfrac{C_n - np_n}{\sqrt{np_n(1-p_n)}}\right)_{n\geqslant 1}$ converge uniformément en loi vers N, *i.e.*

$$\boxed{(D_n)_{n\geqslant 1} \text{ converge uniformément en loi vers } N.}$$

□

b) En utilisant les résultats des questions ***14.*** et ***20.***, en déduire que la suite $(\hat{f}_n)_{n\geqslant 1}$ converge en loi vers N.

Démonstration. Récapitulons les résultats de la question ***14.*** :

- pour tout $n \in \mathbb{N}^*$, $\hat{f}_n = a_n D_n + b_n$ où $a_n = \dfrac{\sigma_n}{\theta_n \sqrt{n}}$ et $b_n = \sqrt{n}\left(\dfrac{p_n}{\theta_n} - \theta_n\right)$
- pour tout $n \in \mathbb{N}^*$, $a_n > 0$ et $\lim\limits_{n\to+\infty} a_n = 1 > 0$
- $\lim\limits_{n\to+\infty} b_n = 0$

D'après la question précédente, $(D_n)_{n\geqslant 1}$ converge uniformément en loi vers N.
Ainsi, d'après la question ***20.c)***, $(a_n D_n + b_n)_{n\geqslant 1}$ converge en loi vers $1 \times N + 0 = N$, *i.e.*

$$\boxed{(\hat{f}_n)_{n\geqslant 1} \text{ converge en loi vers } N.}$$

□