MATHÉMATIQUES II - ESSEC E 2022

Proposition de corrigé par David Meneu

Lycée Champollion - Grenoble, pour Major-Prépa

Première partie : Autour du théorème limite central

1. Soit Z une variable aléatoire de loi normale $\mathcal{N}(0,1)$, de densité $f_Z: x \mapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

On note Φ la fonction de répartition de Z définie par $\Phi(x) = \int_{-\infty}^{x} f_Z(t)\mathrm{d}t$.

a) C'est une question de cours : puisque Z est une variable à densité, alors sa fonction de répartition Φ est continue sur $\mathbb{R}$.

b) La fonction Φ est aussi de classe $\mathcal{C}^1$ sur $\mathbb{R}$, et : $\forall x \in \mathbb{R},\ \Phi'(x) = f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} > 0$, donc Φ est strictement croissante sur $\mathbb{R}$.

c) Toujours à cause du fait que Φ est une fonction de répartition : $\lim_{x\to-\infty} \Phi(x) = 0$ et $\lim_{x\to+\infty} \Phi(x) = 1$.

La fonction Φ est donc continue, strictement croissante sur $\mathbb{R}$: d'après le théorème éponyme, Φ réalise donc une bijection de $\mathbb{R} =]-\infty\,;+\infty[$ sur $]\lim_{x\to-\infty}\Phi(x)\,;\lim_{x\to+\infty}\Phi(x)[=]0\,;1[$.

d) La propriété demandée est une conséquence de la *parité* de la densité f_Z ($\forall x \in \mathbb{R},\ f_Z(-x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(-x)^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = f_Z(x)$) : pour tout réel x, les intégrales $\int_{-\infty}^{-x} f_Z(t)\mathrm{d}t$ et $\int_{x}^{+\infty} f_Z(t)\mathrm{d}t$ converge et sont égales, et d'après les règles de calcul des probabilités avec les variables à densité :

$$\begin{aligned}\forall x \in R,\quad \int_{-\infty}^{-x} f_Z(t)\mathrm{d}t = \int_{x}^{+\infty} f_Z(t)\mathrm{d}t &\iff \forall x \in \mathbb{R},\quad \mathbb{P}(Z \leqslant -x) = \mathbb{P}(Z \geqslant x)\\ &\iff \forall x \in \mathbb{R},\quad \Phi(-x) = 1 - \Phi(x).\end{aligned}$$

2. a) On récite le cours : la loi faible des grands nombres stipule que si $(X_i)_{i\in\mathbb{N}^*}$ est une suite de variables aléatoires réelles de même loi, admettant une espérance commune m ; alors la fréquence empirique $\overline{X_n}$ converge en probabilité vers la variable certaine égale à m, c'est-à-dire :

$$\forall \varepsilon > 0,\quad \lim_{n\to+\infty} \mathbb{P}\big(|\overline{X_n} - m| \geqslant \varepsilon\big) = 0.$$

b) On cite encore le cours : soit $(X_i)_{i\in\mathbb{N}^*}$ une suite de variables aléatoires réelles mutuellement indépendantes et de même loi, admettant une espérance m et une variance σ^2 communes.

Alors la variable centrée réduite associée à $\overline{X_n}$, à savoir $\frac{\overline{X_n} - m}{\sigma/\sqrt{n}} = \sqrt{n}\frac{\overline{X_n} - m}{\sigma}$ converge en loi vers une variable aléatoire suivant la loi normale centrée réduite. Comme Φ est continue sur $\mathbb{R}$, cela signifie (par définition de la convergence en loi) que :

$$\forall x \in \mathbb{R},\quad \lim_{n\to+\infty} \mathbb{P}(Z_n \leqslant x) = \Phi(x).$$

3. a) La variable aléatoire X_i est finie, donc admet une espérance qui vaut :

$$E(X_i) = 0 \cdot \mathbb{P}(X_i = 0) + 2 \cdot \mathbb{P}(X_i = 2) + 5 \cdot \mathbb{P}(X_i = 5) + 10 \cdot \mathbb{P}(X_i = 10) = 0 + 1 + 1 + 1 = 3.$$

b) De même, la variable aléatoire X_i admet un moment d'ordre 2 qui vaut :

$$E(X_i) = 0^2 \cdot \mathbb{P}(X_i = 0) + 2^2 \cdot \mathbb{P}(X_i = 2) + 5^2 \cdot \mathbb{P}(X_i = 5) + 10^2 \cdot \mathbb{P}(X_i = 10) = 0 + 2 + 5 + 10 = 17.$$

La formule de Koenig-Huygens donne la variance de X_i :

$$\mathrm{Var}(X_i) = E(X_i^2) - E(X_i)^2 = 17 - 9 = 8.$$

c) Soit $f : [0\,;1] \to \mathbb{R}$ la fonction définie de la manière suivante :

$$f(x) = 0 \text{ si } x \in [0\,; \tfrac{1}{5}[, \quad f(x) = 2 \text{ si } x \in [\tfrac{1}{5}\,; \tfrac{7}{10}[, \quad f(x) = 5 \text{ si } x \in [\tfrac{7}{10}\,; \tfrac{9}{10}[, \quad f(x) = 10 \text{ si } x \in [\tfrac{9}{10}\,; 1].$$

i) Soit U une variable aléatoire de loi uniforme sur $[0\,;1]$. La variable aléatoire $f(U)$ est à valeurs dans $\{0, 2, 5, 10\}$ et :

$$\mathbb{P}\big(f(U) = 0\big) = \mathbb{P}\big(0 \leqslant U < \tfrac{1}{5}\big) = \frac{1}{5}; \quad \mathbb{P}\big(f(U) = 2\big) = \mathbb{P}\big(\tfrac{1}{5} \leqslant U < \tfrac{7}{10}\big) = \tfrac{7}{10} - \tfrac{1}{5} = \tfrac{5}{10} = \tfrac{1}{2};$$

$$\mathbb{P}\big(f(U) = 5\big) = \mathbb{P}\big(\tfrac{7}{10} \leqslant U < \tfrac{9}{10}\big) = \tfrac{9}{10} - \tfrac{7}{10} = \tfrac{2}{10} = \tfrac{1}{5}; \quad \mathbb{P}\big(f(U) = 10\big) = \mathbb{P}\big(\tfrac{9}{10} \leqslant U \leqslant 1\big) = 1 - \tfrac{9}{10} = \tfrac{1}{10}.$$

On a bien démontré que la variable aléatoire $f(U)$ suit la même loi que X_i.

ii) En application de ce qui précède, on simule une variable aléatoire de même loi que X_i en simulant $f(U)$:

```
import numpy.random as rd

def X():
    U = rd.rand()
    if U < 1/5:
        return 0
    elif U < 7/10:
        return 2
    elif U < 9/10:
        return 5
    else:
        return 10
```

Après n lancers de fléchettes, le score du joueur est $S_n = \displaystyle\sum_{i=1}^{n} X_i$.

d) Avec les valeurs obtenues précédemment, on a ici $\mu = 3$, $\sigma = \sqrt{8} = 2\sqrt{2}$, et :

$$Z_n = \sqrt{n}\Big(\frac{\overline{X_n} - 3}{2\sqrt{2}}\Big) = \sqrt{n}\Big(\frac{\frac{1}{n}S_n - 3}{2\sqrt{2}}\Big) = \frac{S_n - 3n}{2\sqrt{2n}}.$$

e) Un joueur lance $n = 200$ fléchettes. On a alors $Z_{200} = \dfrac{S_{200} - 600}{2\sqrt{400}} = \dfrac{S_{200} - 600}{40}$, et :

$$\mathbb{P}(S_{200} \leqslant 500) = \mathbb{P}(S_{200} - 600 \leqslant -100) = \mathbb{P}\Big(\frac{S_{200} - 600}{40} \leqslant -2.5\Big)$$

Or le théorème central limite assure que pour tout réel a : $\displaystyle\lim_{n \to +\infty} \mathbb{P}(Z_n \leqslant a) = \Phi(a)$.

On peut alors dire que : $\mathbb{P}(S_{200} \leqslant 500\big) \approx \Phi(-2.5)$.

4. a) Soit N un entier naturel fixé supérieur ou égal à 1. Le théorème central limite assure que pour tout $k \in \{1, \dots, 2N-1\}$, $\lim\limits_{n \to +\infty} \mathbb{P}(Z_n \leqslant x_k) = \Phi(x_k)$ et par définition de la convergence d'une suite réelle (avec $\varepsilon = \frac{1}{2N}$), cela signifie qu'il existe, pour chacun de ces entiers k, un entier n_k tel que pour tout $n \geqslant n_k$, $\left|\mathbb{P}(Z_n \leqslant x_k) - \Phi(x_k)\right| \leqslant \dfrac{1}{2N}$.

Comme on a affaire à un nombre fini d'entiers k : en posant $n_0 = \max(n_1, \dots, n_{2N-1})$, on a :

$$\forall n \geqslant n_0, \quad \forall k \in \{1, \dots, 2N-1\}, \ \left|\mathbb{P}(Z_n \leqslant x_k) - \Phi(x_k)\right| \leqslant \frac{1}{2N}$$

Et comme $\left|\mathbb{P}(Z_n \leqslant x_0) - \Phi(x_0)\right| = |0-0|$ et $\left|\mathbb{P}(Z_n \leqslant x_{2N}) - \Phi(x_{2N})\right| = |1-1| = 0$, alors :

$$\forall n \geqslant n_0, \ \forall k \in \{0, 1, \dots, 2N\}, \quad \left|\mathbb{P}(Z_n \leqslant x_k) - \Phi(x_k)\right| \leqslant \frac{1}{2N}$$

$$\Longrightarrow \forall n \geqslant n_0, \ \max_{k \in \{0, \dots, 2N\}} \left|\mathbb{P}(Z_n \leqslant x_k) - \Phi(x_k)\right| \leqslant \frac{1}{2N}.$$

On divise l'ensemble des réels en intervalles $I_k =]x_{k-1}\,; x_k]$ pour $k \in \{1, \dots, 2N-1\}$, avec par convention $I_1 =]-\infty\,; x_1]$ et $]x_{2N-1}\,; +\infty[$.

b) Soit $k \in \{1, \dots, 2N\}$ et x un réel quelconque tel que $x \in I_k$. Soit $n \geqslant n_0$.

i) Puisque $x \in I_k$, alors $x \leqslant x_k$, donc $[Z_n \leqslant x] \subset [Z_n \leqslant x_k]$ et par croissance de la probabilité :

$$\mathbb{P}(Z_n \leqslant x) \leqslant \mathbb{P}(Z_n \leqslant x_k).$$

On a aussi $x_{k-1} \leqslant x$, donc $[Z \leqslant x_{k-1}] \subset [Z \leqslant x]$ et

$$\mathbb{P}(Z \leqslant x_{k-1}) \leqslant \mathbb{P}(Z \leqslant x) \iff \Phi(x_{k-1}) \leqslant \Phi(x) \iff -\Phi(x) \leqslant -\Phi(x_{k-1}),$$

de sorte qu'en effet :

$$\mathbb{P}(Z_n \leqslant x) - \Phi(x) \leqslant \mathbb{P}(Z_n \leqslant x_k) - \Phi(x_{k-1}).$$

ii) D'après le résultat de **4**.a) :

$$\mathbb{P}(Z_n \leqslant x_k) - \Phi(x_k) \leqslant \frac{1}{2N} \iff \mathbb{P}(Z_n \leqslant x_k) \leqslant \Phi(x_k) + \frac{1}{2N}$$

$$\Longrightarrow \mathbb{P}(Z_n \leqslant x_k) - \Phi(x_{k-1}) \leqslant \Phi(x_k) - \Phi(x_{k-1}) + \frac{1}{2N},$$

et donc, par transitivité de l'inégalité :

$$\mathbb{P}(Z_n \leqslant x) - \Phi(x) \leqslant \Phi(x_k) - \Phi(x_{k-1}) + \frac{1}{2N} = \frac{k}{2N} - \frac{k-1}{2N} + \frac{1}{2N} = \frac{2}{2N} = \frac{1}{N}.$$

iii) De même que précédemment : $\Phi(x) \leqslant \Phi(x_k)$ et $\mathbb{P}(Z_n \leqslant x_{k-1}) \leqslant \mathbb{P}(Z_n \leqslant x)$, donc :

$$\Phi(x) - \mathbb{P}(Z_n \leqslant x) \leqslant \Phi(x_k) - \mathbb{P}(Z_n \leqslant x_{k-1}),$$

et toujours d'après **4**.a) :

$$-\frac{1}{2N} \leqslant \mathbb{P}(Z_n \leqslant x_{k-1} - \Phi(x_{k-1}), \ \text{donc} \ -\mathbb{P}(Z_n \leqslant x_{k-1}) \leqslant \tfrac{1}{2N} - \Phi(x_{k-1}),$$

et ainsi :

$$\Phi(x_k) - \Phi(x_{k-1}) \leqslant \Phi(x_k) - \Phi(x_{k-1}) + \tfrac{1}{2N},$$

donc par transitivité de l'inégalité :

$$\phi(x) - \mathbb{P}(Z_n \leqslant x) \leqslant \Phi(x_k) - \Phi(x_{k-1}) + \tfrac{1}{2N} = \tfrac{1}{N}.$$

c) Les intervalles $(I_k)_{1\leqslant k\leqslant 2N-1}$ forment une partition de $\mathbb{R}$: pour tout réel x, il existe un entier $k \in \{1, \ldots, 2N-1\}$ tel que $x \in I_k$, et pour tout $n \geqslant n_0$:

$$\Phi(x) - \mathbb{P}(Z_n \leqslant x) \leqslant \frac{1}{N} \text{ et } \mathbb{P}(Z_n \leqslant x) - \Phi(x) \leqslant \frac{1}{N}.$$

On en conclut que :

$$\forall x \in \mathbb{R}, \quad -\frac{1}{N} \leqslant \mathbb{P}(Z_n \leqslant x) - \Phi(x) \leqslant \frac{1}{N} \iff \big|\mathbb{P}(Z_n \leqslant x) - \Phi(x)\big| \leqslant \frac{1}{N}.$$

d) Soit $(M_n)_{n\geqslant 1}$ une suite de majorants des fonctions D_n définies pour $x \in \mathbb{R}$ par :

$$D_n(x) = \mathbb{P}(Z_n \leqslant x) - \Phi(x).$$

On a vu que pour tout $N \in \mathbb{N}^*$, il existe un entier $n_0(N)$ (qu'on note ainsi pour bien montrer qu'il dépend de N) tel que : $\forall n \geqslant n_0(N),\ \forall x \in \mathbb{R},\ \big|\mathbb{P}(Z_n \leqslant x) - \Phi(x)\big| \leqslant \dfrac{1}{N}$.

On peut aussi s'arranger pour avoir : $\forall N \in \mathbb{N}^*,\ n_0(N+1) > n_0(N)$.

On peut alors poser : $\forall n \in [\![1\,;n_0(1)]\!],\ M_1 = 2$ car $\big|\mathbb{P}(Z_n \leqslant x) - \Phi(x)\big| \leqslant \mathbb{P}(Z_n \leqslant x) + \Phi(x) \leqslant 2$; ensuite : $\forall N \in \mathbb{N}^*,\ \forall n \in [\![n_0(N)\,;n_0(N+1)]\!],\ M_n = \dfrac{1}{N}$; d'après ce qui précède, et puisque N prend successivement toutes les valeurs entières non nulles, on a bien $\lim\limits_{n\to+\infty} M_n = 0$.

5. Soit x un réel fixé.

a) Soit $(x_n)_{n\in\mathbb{N}^*}$ une suite de réels telle que $\lim\limits_{n\to+\infty} x_n = x$.

i) La fonction Φ est continue sur $\mathbb{R}$, donc $\lim\limits_{n\to+\infty} \Phi(x_n) = \Phi(x)$.

ii) Avec la suite de majorants $(M_n)_{n\in\mathbb{N}^*}$ définie à la question **4**.d) : on a $\Big|\mathbb{P}(Z_n \leqslant x) - \Phi(x)\Big| \leqslant M_n$ pour tout réel x, donc en particulier $\big|\mathbb{P}(Z_n \leqslant x_n) - \Phi(x_n)\big| \leqslant M_n$; le théorème d'encadrement (une valeur absolue est toujours positive) assure alors que $\lim\limits_{n\to+\infty} \big|\mathbb{P}(Z_n \leqslant x_n) - \Phi(x_n)\big| = 0$.

iii) Pour tout $n \in \mathbb{N}^*$, d'après l'inégalité triangulaire :

$$0 \leqslant \big|\mathbb{P}(Z_n \leqslant x_n) - \Phi(x)\big| = \big|\mathbb{P}(Z_n \leqslant x_n) - \Phi(x_n) + \Phi(x_n) - \Phi(x)\big| \leqslant \big|\mathbb{P}(Z_n \leqslant x_n) - \Phi(x_n)\big| + \big|\Phi(x_n) - \Phi(x)\big|.$$

Comme $\lim\limits_{n\to+\infty} \big|\mathbb{P}(Z_n \leqslant x_n) - \Phi(x_n)\big| = \lim\limits_{n\to+\infty} \big|\Phi(x_n) - \Phi(x)\big|$ d'après les deux questions précédentes, le théorème d'encadrement permet finalement de conclure :

$$\lim_{n\to+\infty} \mathbb{P}(Z_n \leqslant x_n) = \Phi(x).$$

b) i) Pour tout entier $n \geqslant 1$: $x - \frac{1}{n} < x$, donc $\big[Z_n \leqslant x - \frac{1}{n}\big] \subset [Z_n < x] \subset [Z_n \leqslant x]$, donc $\mathbb{P}\big(Z_n \leqslant x - \frac{1}{n}\big) \leqslant \mathbb{P}(Z_n < x) \leqslant \mathbb{P}(Z_n \leqslant x)$.

ii) Puisque $\lim\limits_{n\to+\infty} x - \frac{1}{n} = x$, alors d'après le résultat de **5**.a)iii) avec $x_n = x - \frac{1}{n}$:

$$\lim_{n\to+\infty} \mathbb{P}\big(Z_n \leqslant x - \tfrac{1}{n}\big) = \Phi(x).$$

Et comme $\lim\limits_{n\to+\infty} \mathbb{P}(Z_n \leqslant x) = \Phi(x)$ aussi (c'est le théorème central limite !), alors le théorème d'encadrement s'applique encore, qui donne :

$$\lim_{n\to+\infty} \mathbb{P}(Z_n < x) = \Phi(x).$$

c) Pour tous réels a et b qui vérifient $a < b$:

$$\mathbb{P}\big(Z_n \in [a\,;b]\big) = \mathbb{P}(Z_n \leqslant a) - \mathbb{P}(Z_n < b),$$

donc d'après **2**.a) avec $x = a$ et **5**.b)ii) avec $x = b$:

$$\lim_{n\to+\infty} \mathbb{P}\big(Z_n \in [a\,;b]\big) = \Phi(b) - \Phi(a)$$

Deuxième Partie : Applications en statistique.

6. a) L'énoncé demande de redémontrer ces résultats de cours :

$$E(X_i) = 0 \cdot \mathbb{P}(X_i = 0) + 1 \cdot \mathbb{P}(X_i = 1) = p \text{ et } E(X_i^2) = 0^2 \cdot \mathbb{P}(X_i = 0) + 1^2 \cdot \mathbb{P}(X_i = 1) = p,$$

donc d'après la formule de Koenig-Huygens : $V(X_i) = E(X_i^2) - E(X_i)^2 = p - p^2 = p(1-p)$.

b) On note dans la suite $\sigma = \sqrt{p(1-p)}$. Dans cette question très classique, on rappelle que $p(1-p) = p - p^2$ est une fonction trinôme du second degré de la variable p, de racines évidentes 0 et 1. Comme le coefficient dominant est négatif, cette fonction admet sur $[0\,;1]$ un maximum dont l'abscisse est le milieu des deux racines ; en $p = \frac{1}{2}$, $p(1-p) = \frac{1}{4}$ donc :

$$\forall p \in [0\,;1], \quad p(1-p) \leqslant \tfrac{1}{4} \Longrightarrow \sigma = \sqrt{p(1-p)} \leqslant \tfrac{1}{2}.$$

c) Par linéarité de l'espérance : $E(\overline{X_n}) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} E(X_i) = \dfrac{1}{n}\sum_{i=1}^{n} p = \dfrac{1}{n} \times np = p.$

d) D'après les propriétés de la variance, et par mutuelle indépendance des (X_i) :

$$\mathrm{Var}(\overline{X_n}) = \mathrm{Var}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{1}{n^2} \times n\sigma^2 = \frac{1}{n}\sigma^2.$$

7. a) Dans le cas de variables de Bernoulli, on a $Z_n = \sqrt{n}\Big(\dfrac{\overline{X_n} - p}{\sigma}\Big)$, donc d'après le résultat de la question **5.**c) :

$$\lim_{n\to+\infty} \mathbb{P}(Z_n \in [-a\,;a]) = \Phi(a) - \Phi(-a) \iff \lim_{n\to+\infty} \mathbb{P}\Big(\frac{\sqrt{n}}{\sigma}(\overline{X_n} - p) \in [-a\,;a]\Big) = \Phi(a) - \Phi(-a).$$

b) On a vu à la partie I que : $\Phi(-a) = 1 - \Phi(a)$, donc $\Phi(a) - \Phi(-a) = \Phi(a) - 1 + \Phi(a) = 2\Phi(a) - 1$.

D'autre part :

$$\begin{aligned}
\frac{\sqrt{n}}{\sigma}(\overline{X_n} - p) \in [-a\,;a] &\iff -a \leqslant \frac{\sqrt{n}}{\sigma}(\overline{X_n} - p) \leqslant a \\
&\iff -a\frac{\sigma}{\sqrt{n}} \leqslant \overline{X_n} - p \leqslant a\frac{\sigma}{\sqrt{n}} \\
&\iff \overline{X_n} - a\frac{\sigma}{\sqrt{n}} \leqslant p \leqslant \overline{X_n} + a\frac{\sigma}{\sqrt{n}}
\end{aligned}$$

Le résultat de la question précédente se réécrit donc bien :

$$\lim_{n\to+\infty} \mathbb{P}\Big(p \in \Big[\overline{X_n} - a\frac{\sigma}{\sqrt{n}}\,; \overline{X_n} + a\frac{\sigma}{\sqrt{n}}\Big]\Big) = 2\Phi(a) - 1.$$

c) Une table de valeurs de Φ donne $\Phi(1,96) \approx 0,975 \iff 2\Phi(1,96) - 1 \approx 1,95 - 1 = 0,95$.
Dnc pour $a = 1,96$ et pour n grand, on a :

$$\mathbb{P}\Big(p \in \Big[\overline{X_n} - 1,96\frac{\sigma}{\sqrt{n}}\,; \overline{X_n} + 1,96\frac{\sigma}{\sqrt{n}}\Big]\Big) \approx 0,95.$$

Ce résultat signifie bien que pour n grand, le paramètre p a approximativement 95% de chances d'appartenir à l'intervalle $\Big[\overline{X_n} - 1,96\frac{\sigma}{\sqrt{n}}\,; \overline{X_n} + 1,96\frac{\sigma}{\sqrt{n}}\Big]$.

d) Comme on l'a vu à la question **6**.b) : $\sigma \leqslant \frac{1}{2}$, donc $1,96\frac{\sigma}{\sqrt{n}} \leqslant \frac{0,98}{\sqrt{n}}$. Par conséquent :

$$\left[\overline{X_n} - 1,96\frac{\sigma}{\sqrt{n}}\,;\overline{X_n} + 1,96\frac{\sigma}{\sqrt{n}}\right] \subset \left[\overline{X_n} - \frac{0,98}{\sqrt{n}}\,;\overline{X_n} + \frac{0,98}{\sqrt{n}}\right],$$

et donc :

$$\mathbb{P}\Big(p \in \left[\overline{X_n} - \frac{0,98}{\sqrt{n}}\,;\overline{X_n} + \frac{0,98}{\sqrt{n}}\right]\Big) \geqslant \mathbb{P}\Big(p \in \left[\overline{X_n} - 1,96\frac{\sigma}{\sqrt{n}}\,;\overline{X_n} + 1,96\frac{\sigma}{\sqrt{n}}\right]\Big) \approx 0,95.$$

8. On pose, pour $n \geqslant 1$, $V_n = \overline{X_n}(1-\overline{X_n}) + \frac{1}{n}$. Soit $\varepsilon > 0$.

a) En développant le membre de droite de la relation demandée :

$$(\overline{X_n} - p)(1-\overline{X_n} - p) = \overline{X_n} - p - \overline{X_n}^2 + p\overline{X_n} - p\overline{X_n} + p^2 = \overline{X_n}(1-\overline{X_n}) - p(1-p) = V_n - \tfrac{1}{n} - \sigma^2.$$

b) De ce qui précède, on déduit que :

$$|V_n - \sigma^2| = |(\overline{X_n} - p)(1-\overline{X_n} - p) + \frac{1}{n}| \leqslant |\overline{X_n} - p| \cdot |1 - \overline{X_n} - p| + \frac{1}{n}.$$

Or $|1-\overline{X_n} - p| \leqslant |1-p| + |\overline{X_n}|$, avec $|1-p| = 1-p \leqslant 1$; d'autre part, puisque $S_n = \sum_{i=1}^{n} X_i$ est une somme de n variables de Bernoulli qui valent chacune 0 ou 1, alors $S_n \leqslant n \iff \overline{X_n} \leqslant 1$, de sorte qu'en effet : $|1-p-\overline{X_n}| \leqslant 2$, donc $|V_n - \sigma^2| \leqslant 2|\overline{X_n} - p| + \frac{1}{n}$.

c) De ce qui précède, on déduit que : $|V_n - \sigma^2| > \varepsilon$ implique $2|\overline{X_n} - p| + \frac{1}{n} > \varepsilon$.

On a ainsi : $\mathbb{P}\big(|V_n - \sigma^2| > \varepsilon\big) \leqslant \mathbb{P}\big(2|\overline{X_n} - p| + \frac{1}{n} > \varepsilon\big) = \mathbb{P}\big(|\overline{X_n} - p| > \frac{\varepsilon}{2} - \frac{1}{2n}\big)$.

d) Puisque $\lim\limits_{n\to+\infty} \frac{1}{2n} = 0$, alors pour n assez grand, $\frac{\varepsilon}{2} - \frac{1}{2n} > \frac{\varepsilon}{4}$, et alors $|\overline{X_n} - p| > \frac{\varepsilon}{2} - \frac{1}{2n}$ implique $|\overline{X_n} - p| > \frac{\varepsilon}{4}$, donc :

$$\mathbb{P}\big(|\overline{X_n} - p| > \tfrac{\varepsilon}{2} - \frac{1}{2n}\big) \leqslant \mathbb{P}\big(|\overline{X_n} - p| > \tfrac{\varepsilon}{4}\big).$$

Ainsi pour tout $\varepsilon > 0$: pour n assez grand, on a $0 \leqslant \mathbb{P}(|V_n - \sigma^2| > \varepsilon) \leqslant \mathbb{P}\big(|\overline{X_n} - p| > \frac{\varepsilon}{4}\big)$.

La loi faible des grands nombres rappelée à la question **2**.a) qui donne $\lim\limits_{n\to+\infty} \mathbb{P}\big(|\overline{X_n} - p| > \frac{\varepsilon}{4}\big) = 0$, assure alors par encadrement, que :

$$\forall \varepsilon > 0, \quad \lim_{n\to+\infty} \mathbb{P}(|V_n - \sigma^2| > \varepsilon) = 0.$$

9. On pose maintenant $W_n = \frac{\sqrt{n}}{\sqrt{V_n}}(\overline{X_n} - p) = \frac{\sigma}{\sqrt{V_n}} Z_n$. On fixe un réel x.

a) Soit $\varepsilon > 0$.

i) Puisque $\sigma > 0$ et $\sqrt{V_n}$ est une variable aléatoire à valeurs positives, $\mathbb{P}(W_n \leqslant x) = \mathbb{P}\big(\frac{\sigma}{\sqrt{V_n}} Z_n \leqslant x\big) = \mathbb{P}\big(Z_n \leqslant \frac{\sqrt{V_n}}{\sigma} x\big)$.

On utilise alors la formule des probabilités totales avec le système complet des deux événements contraires $\big[\frac{\sqrt{V_n}}{\sigma} \leqslant 1+\varepsilon\big]$ et $\big[\frac{\sqrt{V_n}}{\sigma} > 1+\varepsilon\big]$:

$$\mathbb{P}\big(Z_n \leqslant \frac{\sqrt{V_n}}{\sigma} x\big) = \mathbb{P}\big(\big[Z_n \leqslant \frac{\sqrt{V_n}}{\sigma} x\big] \cap \big[\frac{\sqrt{V_n}}{\sigma} \leqslant 1+\varepsilon\big]\big) + \mathbb{P}\big(\big[Z_n \leqslant \frac{\sqrt{V_n}}{\sigma} x\big] \cap \big[\frac{\sqrt{V_n}}{\sigma} > 1+\varepsilon\big]\big),$$

où : $\big[Z_n \leqslant \frac{\sqrt{V_n}}{\sigma} x\big] \cap \big[\frac{\sqrt{V_n}}{\sigma} \leqslant 1+\varepsilon\big]$ implique $[Z_n \leqslant (1+\varepsilon)x]$ et $\big[Z_n \leqslant \frac{\sqrt{V_n}}{\sigma} x\big] \cap \big[\frac{\sqrt{V_n}}{\sigma} > 1+\varepsilon\big] \subset \big[\frac{\sqrt{V_n}}{\sigma} > 1+\varepsilon\big]$ (l'inclusion traduisant aussi une implication), ce qui implique que :

$$\mathbb{P}\big(\big[Z_n \leqslant \frac{\sqrt{V_n}}{\sigma} x\big] \cap \big[\frac{\sqrt{V_n}}{\sigma} \leqslant 1+\varepsilon\big]\big) \leqslant \mathbb{P}\big(Z_n \leqslant (1+\varepsilon)x\big) \text{et } \mathbb{P}\big(\big[Z_n \leqslant \frac{\sqrt{V_n}}{\sigma} x\big] \cap \big[\frac{\sqrt{V_n}}{\sigma} > 1+\varepsilon\big]\big) \leqslant \mathbb{P}\big(\frac{\sqrt{V_n}}{\sigma} > 1+\varepsilon\big),$$

ce qui donne bien :

$$\mathbb{P}(W_n \leqslant x) \leqslant \mathbb{P}\big(Z_n \leqslant (1+\varepsilon)x\big) + P\big(\tfrac{\sqrt{V_n}}{\sigma} > 1+\varepsilon\big).$$

ii) Puisque $\sigma > 0$, alors :

$$\mathbb{P}\left(\frac{\sqrt{V_n}}{\sigma} > 1+\varepsilon\right) = \mathbb{P}\left(\sqrt{V_n} > (1+\varepsilon)\sigma\right) = \mathbb{P}\left(V_n > (1+\varepsilon)^2\sigma^2\right)$$
$$= \mathbb{P}\left(V_n - \sigma^2 > (\varepsilon^2+2\varepsilon)\sigma^2\right) \leqslant \mathbb{P}\left(|V_n-\sigma^2| > (\varepsilon^2+2\varepsilon)\sigma^2\right).$$

Or d'après **8**.c) : $\lim_{n\to+\infty} \mathbb{P}\left(|V_n-\sigma^2| > (\varepsilon^2+2\varepsilon)\sigma^2\right) = 0$, donc par encadrement (une probabilité est toujours positive) :

$$\lim_{n\to+\infty} \mathbb{P}\left(\frac{\sqrt{V_n}}{\sigma} > 1+\varepsilon\right) = 0.$$

iii) Il s'agit ici simplement de reprendre le résultat de **2**.b) pour le réel $(1+\varepsilon)x$:

$$\lim_{n\to+\infty} \mathbb{P}\left(Z_n \leqslant (1+\varepsilon)x\right) = \Phi\left((1+\varepsilon)x\right).$$

iv) Des deux résultats précédents on déduit que

$$\lim_{n\to+\infty} \mathbb{P}\left(Z_n \leqslant (1+\varepsilon)x\right) + P\left(\tfrac{\sqrt{V_n}}{\sigma} > 1+\varepsilon\right) = \Phi\left((1+\varepsilon)x\right),$$

et donc que pour $\varepsilon > 0$ fixé, il existe (par définition de la convergence d'une suite) un entier n_ε tel que pour tout $n \geqslant n_\varepsilon$,

$$\Phi\left((1+\varepsilon)x\right) - \varepsilon \leqslant \mathbb{P}\left(Z_n \leqslant (1+\varepsilon)x\right) + P\left(\tfrac{\sqrt{V_n}}{\sigma} > 1+\varepsilon\right) \leqslant \Phi\left((1+\varepsilon)x\right) + \varepsilon,$$

ce qui implique bien, par transitivité de l'inégalité et d'après **9**.a)i), que :

$$\forall n \geqslant n_\varepsilon, \quad \mathbb{P}(W_n \leqslant x) \leqslant \Phi\left((1+\varepsilon)x\right) + \varepsilon.$$

b) L'énoncé admettait que de manière symétrique, pour tout $\varepsilon > 0$ il existe un n_ε tel que pour tout $n \geqslant n_\varepsilon$ on a $\mathbb{P}(W_n \leqslant x) \geqslant \Phi\left((1-\varepsilon)x\right) - \varepsilon$.
Remarque : on peut considérer que cet entier n_ε est le même que le précédent, sinon en cas de différence il suffit de prendre le plus grand des deux pour n_ε.

Notons que pour tout $\varepsilon > 0$, par croissance de Φ sur $\mathbb{R}$ on a toujours :

$$\Phi\left((1-\varepsilon)x\right) - \varepsilon < \Phi\left((1-\varepsilon)x\right) = \Phi(x-\varepsilon x) \leqslant \Phi(x) \leqslant \Phi(x+\varepsilon x) = \Phi\left((1+\varepsilon)x\right) < \Phi\left((1+\varepsilon)x\right) + \varepsilon.$$

Une conclusion possible est donc la suivante : l'intervalle $\left[\Phi\left((1-\varepsilon)x\right) - \varepsilon\,;\Phi\left((1+\varepsilon)x\right) + \varepsilon\right]$ est de longueur arbitrairement petite (puisque $\lim_{\varepsilon\to 0^+} \Phi\left((1-\varepsilon)x\right) - \varepsilon = \lim_{\varepsilon\to 0^+} \Phi\left((1+\varepsilon)x\right) + \varepsilon = \Phi(x)$ par continuité de Φ sur $\mathbb{R}$), mais cet intervalle contient toujours $\Phi(x)$, et contient tous les réels $\mathbb{P}(W_n \leqslant x)$ à partir d'un certain rang n_ε.
On peut donc bien en conclure, par définition de la convergence d'une suite, que :

$$\lim_{n\to+\infty} \mathbb{P}(W_n \leqslant x) = \Phi(x),$$

et ce pour tout réel x. La suite de variables aléatoires $(W_n)_{n\in\mathbb{N}^*}$ converge donc en loi vers une variable aléatoire qui suit la loi normale centrée, réduite.

10. a) Il s'agit simplement ici de réécrire l'événement $[W_n \leqslant x]$, sachant que $W_n = \frac{\sqrt{n}}{\sqrt{V_n}}(\overline{X_n} - p)$:

$$\left[\tfrac{\sqrt{n}}{\sqrt{V_n}}(\overline{X_n} - p) \leqslant x\right] = \left[\overline{X_n} - p \geqslant x\tfrac{\sqrt{V_n}}{\sqrt{n}}\right] = \left[\overline{X_n} - x\frac{\sqrt{n}}{\sqrt{V_n}} \leqslant p\right],$$

Donc le résultat de la question précédente se réécrit en effet :

$$\forall x \in \mathbb{R}, \quad \lim_{n\to+\infty} \mathbb{P}\left(p \geqslant \overline{X_n} - x\tfrac{\sqrt{V_n}}{\sqrt{n}}\right) = \Phi(x).$$

b) Le candidat A remporte effectivement l'élection si on a $p \geqslant \frac{1}{2}$. Pour $n = 1000$ considéré assez grand, on cherche donc un réel x tel que $\overline{X_n} - x\frac{\sqrt{V_n}}{\sqrt{n}} = \frac{1}{2}$, de sorte que $\mathbb{P}\big(p \geqslant \frac{1}{2}\big) \approx \Phi(x)$ sera une bonne approximation de la probabilité que A remporte l'élection :

$$\overline{X_n} - x\frac{\sqrt{V_n}}{\sqrt{n}} = \frac{1}{2} \iff \overline{X_n} - \frac{1}{2} = x\tfrac{\sqrt{V_n}}{\sqrt{n}} \iff x = \tfrac{\sqrt{n}}{\sqrt{V_n}}\big(\overline{X_n} - \tfrac{1}{2}\big)$$

Avec les valeurs fournies par l'énoncé pour la réalisation de $\overline{X_n}$, on prend donc $x = \frac{\sqrt{1000}}{\sqrt{0{,}2506}}(0,52-\frac{1}{2})$ et $\Phi(x) = \Phi\big(\frac{\sqrt{1000}}{\sqrt{0{,}2506}}(0,52 - \frac{1}{2})\big)$ est une valeur approchée de la probabilité que A remporte l'élection.

11. L'énoncé définit les variables aléatoires $Y_i = X_i + (1 - X_i)T_i$, où T_i est elle-même une variable de Bernoulli, de paramètre q cette fois.

a) Pour toute issue ω de l'univers Ω des possibles, il y a différents cas à étudier :

- Si $X_i(\omega) = 1$, alors $Y_i(\omega) = 1 + (1 - 1)T_i(\omega) = 1 = X_i(\omega)$ et en fait $T_i(\omega) = 0$ dans ce cas, c'est-à-dire que la personne interrogée vote vraiment pour A, comme elle l'avait déclaré.
- Si $X_i(\omega) = 0$, alors $Y_i(\omega) = T_i(\omega)$ qui peut être égal à 1 si la personne a changé d'avis et voté pour B alors qu'elle avait déclaré voter pour A, ou à 0 si la personne a voté, comme elle l'avait déclaré, pour B.

En définitive, Y_i prend bien uniquement les valeurs 0 ou 1, donc Y_i est une variable de Bernoulli, avec :

$$\mathbb{P}(Y_i = 0) = \mathbb{P}\big([X_i = 0] \cap [T_i = 0]\big) = \mathbb{P}(X_i = 0) \times \mathbb{P}(T_i = 0) = (1 - p)(1 - q)$$
$$\text{par indépendance de } X_i \text{ et } T_i$$

$$\mathbb{P}(Y_i = 1) = \mathbb{P}(X_i = 1) + \mathbb{P}\big([X_i = 0] \cap [T_i = 1]\big) = p + \mathbb{P}(X_i = 0) \times \mathbb{P}(T_i = 1) = p + (1 - p)q$$

Ainsi Y_i est une variable de Bernoulli de paramètre $r = p + (1 - p)q$.

b) On travaille encore directement à partir de la probabilité considérée dans cette question :

$$\mathbb{P}\Big(\frac{\sqrt{n}}{\sqrt{U_n}}(\overline{Y_n} - r) \leqslant x\Big) = \mathbb{P}\Big(\overline{Y_n} - p - (1 - p)q \leqslant x\frac{\sqrt{U_n}}{\sqrt{n}}\Big) = \mathbb{P}\Big(\overline{Y_n} - p(1 - q) - q \leqslant x\frac{\sqrt{U_n}}{\sqrt{n}}\Big)$$
$$= \mathbb{P}\Big(\overline{Y_n} - x\frac{\sqrt{U_n}}{\sqrt{n}} - q \leqslant p(1 - q)\Big) \overset{1-q>0}{=} \mathbb{P}\Big(p \geqslant \tfrac{1}{1-q}\big(\overline{Y_n} - x\frac{\sqrt{U_n}}{\sqrt{n}} - q\big)\Big).$$

On a bien :

$$\lim_{n\to+\infty} \mathbb{P}\Big(p \geqslant \tfrac{1}{1-q}\big(\overline{Y_n} - x\frac{\sqrt{U_n}}{\sqrt{n}} - q\big)\Big) = \lim_{n\to+\infty} \mathbb{P}\Big(\frac{\sqrt{n}}{\sqrt{U_n}}(\overline{Y_n} - r) \leqslant x\Big) = \Phi(x).$$

c) On suppose que le sondage sur n personnes a donné à $\overline{Y_n}$ la valeur $\overline{y_n}$, et donc pour U_n la valeur $u_n = \overline{y_n}(1 - \overline{y_n}) + \frac{1}{n}$.

La probabilité que le candidat A remporte effectivement l'élection est alors $\Phi(x)$ pour le réel x tel que :

$$\tfrac{1}{1-q}\big(\overline{y_n} - x\tfrac{\sqrt{u_n}}{\sqrt{n}} - q\big) = \frac{1}{2} \iff \overline{y_n} - x\frac{\sqrt{u_n}}{\sqrt{n}} - q = \frac{1 - q}{2} \iff \overline{y_n} - q - \frac{1 - q}{2} = x\frac{\sqrt{u_n}}{\sqrt{n}}$$
$$\iff \overline{y_n} - \frac{1 + q}{2} = x\frac{\sqrt{u_n}}{\sqrt{n}} \iff x = \frac{\sqrt{n}}{\sqrt{u_n}}\big(\overline{y_n} - \tfrac{1}{2}(1 + q)\big), \text{ CQFD.}$$

d) Pour le sondage de $n = 1000$ personne, on a $\overline{y_n} = 0,52$. Si $q = 0,04$ (c'est-à-dire que 4% seulement des personnes ont changé d'avis en faveur de B entre le sondage et le moment de voter), alors :

$$\overline{y_n} - \tfrac{1}{2}(1 + q) = 0,52 - \tfrac{1}{2} \times 1,04 = 0,52 - 0,52 = 0$$

donc la probabilité que A remporte l'élection est bien $\Phi(0) = \frac{1}{2}$ dans ce cas (incertitude complète !).

Troisième partie : Démonstration du théorème limite central.

Dans cette partie, l'énoncé admettait que les variables aléatoires X_i admettent un moment d'ordre 3 (et donc d'après le cours, tous les moments d'ordres inférieurs), pour démontrer le théorème limite central dans ce cas spécifique.

On suppose aussi que $\mu = E(X_i) = 0$ et que $\sigma^2 = \text{Var}(X_i) = 1$, et donc $Z_n = \sqrt{n}\overline{X_i} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$.

Soit x un réel fixé pour toute la suite de cette partie.

12. a) i) On procède comme demandé par intégrations par parties successives ; les fonctions concernées sont polynômiales, donc toujours de classe $\mathcal{C}^1$ sur $[0\,;1]$:

$$\int_0^1 u^3(1-u)^3\mathrm{d}u = \underbrace{\left[\frac{u^4}{4}(1-u)^3\right]_0^1}_{=0} - \int_0^1 \frac{u^4}{4}\cdot(-3)(1-u)^2\mathrm{d}u \qquad \begin{cases} a(u) = (1-u)^3 & \longrightarrow \quad a'(u) = -3(1-u)^2 \\ b'(u) = u^3 & \longrightarrow \quad b(u) = \frac{u^4}{4} \end{cases}$$

$$= \frac{3}{4}\int_0^1 u^4(1-u)^2\mathrm{d}u$$

$$= \frac{3}{4}\underbrace{\left[\frac{u^5}{5}(1-u)^3\right]_0^1}_{=0} - \frac{3}{4}\int_0^1 \frac{u^5}{5}\cdot(-2)(1-u)\mathrm{d}u \qquad \begin{cases} a(u) = (1-u)^2 & \longrightarrow \quad a'(u) = -2(1-u) \\ b'(u) = u^4 & \longrightarrow \quad b(u) = \frac{u^5}{5} \end{cases}$$

$$= \frac{3}{10}\int_0^1 u^5(1-u)\mathrm{d}u$$

$$= \frac{3}{10}\underbrace{\left[\frac{u^6}{6}(1-u)\right]_0^1}_{=0} - \frac{3}{10}\int_0^1 \frac{u^6}{6}\cdot(-1)\mathrm{d}u \qquad \begin{cases} a(u) = (1-u) & \longrightarrow \quad a'(u) = -1 \\ b'(u) = u^5 & \longrightarrow \quad b(u) = \frac{u^6}{6} \end{cases}$$

$$= \frac{1}{20}\int_0^1 u^6\mathrm{d}u.$$

Remarque : on aurait pu demander aussi un calcul directe après avoir développé $(1-u)^3$ avec la formule du binôme de Newton :

$$\int_0^1 u^3(1-u)^3\mathrm{d}u = \int_0^1 u^3(1-2u+3u^2-u^3)\mathrm{d}u.$$

ii) On termine le calcul :

$$\int_0^1 u^3(1-u)^3\mathrm{d}u = \frac{1}{20}\int_0^1 u^6\mathrm{d}u = \frac{1}{20}\left[\frac{u^7}{7}\right]_0^1 = \frac{1}{140}.$$

b) La fonction h est continue sur $]-\infty\,;0[$ et sur $]1\,;+\infty[$, puisqu'elle est constante sur chacun de ces deux intervalles. Sur $]0\,;1[$, h est définie sous forme intégrale comme la primitive de la fonction polynômiale $u \mapsto u^3(1-u)^3$: elle est donc elle-même polytnômiale, et par conséquent continue sur cet intervalle.
Enfin, toujours par continuité de $u \mapsto u^3(1-u)^3$ sur $\mathbb{R}$, donc sur $[0\,;1]$:

$$\lim_{z\to 0^+} 140\int_0^z u^3(1-u)^3\mathrm{d}u = 140\int_0^0 u^3(1-u)^3\mathrm{d}u = 0 = h(0) = \lim_{z\to 0^-} h(z)$$

$$\lim_{z\to 0^-} 140\int_0^z u^3(1-u)^3\mathrm{d}u = 140\int_0^1 u^3(1-u)^3\mathrm{d}u = \frac{140}{140} = 1 = h(1) = \lim_{z\to 1^+} h(z),$$

Donc h est aussi continue en 0 et en 1, et finalement sur $\mathbb{R}$ tout entier.

c) Il est clair que $0 \leqslant h(z) \leqslant 1$ est vrai pour tout réel z appartenant à l'un des deux intervalles $]-\infty\,;0]$ ou $[1\,;+\infty[$.

Sur $[0\,;1]$, la fonction $u \mapsto u^3(1-u)^3$ est positive et continue, donc $h:\ z \mapsto 140\displaystyle\int_0^z u^3(1-u)^3\mathrm{d}u$ est croissante, et par conséquent a pour minimum $h(0)=0$ et pour maximum $h(1)=1$. Ainsi, en effet :

$$\forall z \in \mathbb{R}, \quad 0 \leqslant h(z) \leqslant 1.$$

L'énoncé admettait ensuite que h', h'' et h''' sont aussi continues et bornées sur $\mathbb{R}$.

13. On pose $a_n = n^{-1/12}$ et $g_n(z) = 1 - h\big(\frac{1}{a_n}(z-x)\big)$.

a) i) La fonction h étant continue sur $\mathbb{R}$, de même que la fonction affine $z \mapsto \frac{1}{a_n}(z-x)$, la fonction g_n est bien continue sur $\mathbb{R}$ comme composée de fonctions qui le sont.

De l'inégalité obtenue en **12**.c), on déduit aussi :

$$\forall z \in \mathbb{R}, \quad 0 \leqslant h\big(\tfrac{1}{a_n}(z-x)\big) \leqslant 1 \iff 1 \geqslant 1 - h\big(\tfrac{1}{a_n}(z-x)\big) \geqslant 0 \iff 1 \geqslant g_n(z) \geqslant 0.$$

L'énoncé admettait ensuite que g_n', g_n'' et g_n''' sont aussi continues et bornées sur $\mathbb{R}$.

ii) Pour tout $z \in \mathbb{R}$, grâce à la formule de dérivation d'une fonction composée :

$$g_n'(z) = -\tfrac{1}{a_n}\cdot h'\big(\tfrac{1}{a_n}(z-x)\big), \quad g_n''(z) = -\tfrac{1}{a_n^2}h''\big(\tfrac{1}{a_n}(z-x)\big) \text{ et } g_n'''(z) = -\tfrac{1}{a_n^3}h'''\big(\tfrac{1}{a_n}(z-x)\big).$$

On a alors : $\quad \forall z \in \mathbb{R},\ |g_n'''(z)| = \frac{1}{a_n^3}\big|h'''\big(\frac{1}{a_n}(z-x)\big)\big| \leqslant \frac{1}{(n^{-1/12})^3}M_{h'''} = n^{1/4}M_{h'''}$,

Donc $n^{1/4}M_{h'''}$ est un majorant de $|g_n'''|$ sur $\mathbb{R}$ et on peut choisir un majorant $M_{g_n'''}$ tel que $M_{g_n'''} \leqslant n^{1/4}M_{h'''}$.

b) i) Par définition de h, on sait que :

$$h\big(\tfrac{1}{a_n}(z-x)\big) = 0 \text{ si } \tfrac{1}{a_n}(z-x) \leqslant 0 \iff z \leqslant x, \text{ donc } g_n(z) = 1 \text{ si } z \leqslant x$$

$$h\big(\tfrac{1}{a_n}(z-x)\big) = 1 \text{ si } \tfrac{1}{a_n}(z-x) > 1 \iff z-x \geqslant a_n, \text{ donc } g_n(z) = 0 \text{ si } z > x + a_n$$

Pour un événement A, on définit la variable aléatoire $\mathbb{1}_A$ par $\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$.

ii) Pour tout $\omega \in \Omega$, d'après ce qui précède :

$$X(\omega) \leqslant x \Longrightarrow \Big(\mathbb{1}_{[X\leqslant x]}(\omega) = 1 = g\big(X(\omega)\big) \text{ et } \mathbb{1}_{[X\leqslant x+a_n]}(\omega) = 1\Big)$$

$$X(\omega) > x + a_n \Longrightarrow \Big(\int [X \leqslant x+a_n](\omega) = 0 = g\big(X(\omega)\big) = \mathbb{1}_{[X\leqslant x+a_n]}(\omega),$$

et sinon :

$$x < X(\omega) \leqslant x + a_n \Longrightarrow \Big(\mathbb{1}_{[X\leqslant x]}(\omega) = 0,\ 0 \leqslant g\big(X(\omega)\big) \text{ et } \mathbb{1}_{[X\leqslant x+a_n]}(\omega) = 1\Big)$$

Dans tous les cas, on a bien : $\forall \omega \in \Omega,\ \mathbb{1}_{[X\leqslant x]}(\omega) \leqslant g\big(X(\omega)\big) \leqslant \mathbb{1}_{[X\leqslant x+a_n]}(\omega)$, ce qui est la définition même de la double inégalité entre variables aléatoires :

$$\mathbb{1}_{[X\leqslant x]} \leqslant g(X) \leqslant \mathbb{1}_{[X\leqslant x+a_n]}.$$

c) D'une part, la première inégalité précédente écrite pour Z_n : $\quad \mathbb{1}_{[Z_n\leqslant x]} \leqslant g(Z_n)$ et la propriété de croissance de l'espérance (les espérances existent car les deux variables aléatoires sont des fonctions bornées de Z_n), donnent :

$$E\big(\mathbb{1}_{[Z_n\leqslant x]}\big) \leqslant E\big(g_n(X)\big) \iff \mathbb{P}(Z_n \leqslant x) \leqslant E\big(g_n(Z_n)\big).$$

En effet : $E(\mathbb{1}_{[Z_n \leqslant x]}) = 0 \cdot \mathbb{P}(Z_n > x) + 1 \cdot \mathbb{P}(Z_n \leqslant x) = \mathbb{P}(Z_n \leqslant x)$.
La deuxième inégalité de la question précédente écrite avec $X = Z_n - a_n$ et $x - 2a_n$ à la place de x et la croissance de l'espérance, donnent :

$$E\big(g_n(Z_n - a_n)\big) \leqslant E\big(\mathbb{1}_{[Z_n - a_n \leqslant x - 2a_n + a_n]}\big) \iff E\big(g_n(Z_n - a_n)\big) \leqslant \mathbb{P}(Z_n - a_n \leqslant x - a_n)$$
$$\iff E\big(g_n(Z_n - a_n)\big) \leqslant \mathbb{P}(Z_n \leqslant x)$$

14. Soit g une fonction de classe $\mathcal{C}^3$ sur $\mathbb{R}$.

a) On procède à nouveau par intégrations par parties successives : pour tous réels z et u fixés,

$$\frac{1}{2}\int_z^{z+u} (z+u-t)^2 g'''(t)\mathrm{d}t = \frac{1}{2}\Big[(z+u-t)^2 g''(t)\Big]_z^{z+u} - \frac{1}{2}\int_z^{z+u} -2(z+u-t)g''(t)\mathrm{d}t$$
$$\begin{pmatrix} a(t) = (z+u-t)^2 & \longrightarrow & a'(t) = -2(z+u-t) \\ b'(t) = g'''(t) & \longrightarrow & b(t) = g''(t) \end{pmatrix}$$
$$= -\tfrac{1}{2}u^2 g''(z) + \int_z^{z+u} (z+u-t)g''(t)\mathrm{d}t$$
$$= -\tfrac{1}{2}u^2 g''(z) + \Big[(z+u-t)g'(t)\Big]_z^{z+u} - \int_z^{z+u} (-1)\cdot g'(t)\mathrm{d}t$$
$$= -\tfrac{1}{2}u^2 g''(z) - ug'(z) + \int_z^{z+u} g'(t)\mathrm{d}t$$
$$= -\tfrac{1}{2}u^2 g''(z) - ug'(z) + g(z+u) - g(z), \quad \text{CQFD.}$$

b) La relation précédente donne alors, pour tous réels u et z :

$$g(z+u) = g(z) + ug'(z) + \tfrac{1}{2}u^2 g'(z) + \tfrac{1}{2}\int_z^{z+u} (z+u-t)^2 g'''(t)\mathrm{d}t.$$

En remplaçant partout u par v (réel quelconque lui aussi), on a :

$$g(z+v) = g(z) + vg'(z) + \tfrac{1}{2}v^2 g'(z) + \tfrac{1}{2}\int_z^{z+v} (z+v-t)^2 g'''(t)\mathrm{d}t.$$

Par conséquent, et en effectuant la soustraction membre à membre des deux égalités précédente, on en déduit :

$$g(z+u) - g(z+v) = g(z) - g(z) + g'(z)(u-v) + \tfrac{1}{2}g''(z)(u^2 - v^2) + R(z,u,v),$$

où $R(z,u,v) = \frac{1}{2}\displaystyle\int_z^{z+u} (z+u-t)^2 g'''(t)\mathrm{d}t - \frac{1}{2}\int_z^{z+v} (z+v-t)^2 g'''(t)\mathrm{d}t$.

c) On suppose ici g''' bornée ; l'inégalité triangulaire pour les intégrales permet alors d'écrire, en supposant dans un premier temps que $u \geqslant 0$ (de sorte que $z + u \geqslant z$:

$$\left|\int_z^{z+u} (z+u-t)^2 g'''(t)\mathrm{d}t\right| \leqslant \int_z^{z+u} (z+u(t)^2 |g'''(t)|\mathrm{d}t$$

où : $\forall t \in [z\,;z+u]$, $|g'''(t)| \leqslant M_{g'''}$, donc $(z+u-t)^2|g'''(t)| \leqslant (z+u-t)^2$.
Les fonctions concernées sont continues sur $\mathbb{R}$, et $z + u \geqslant z$, donc par croissance de l'intégrale :

$$\int_z^{z+u} (z+u-t)^2|g'''(t)|\mathrm{d}t \leqslant \int_z^{z+u} (z+u-t)^2 M_{g'''}\mathrm{d}t = M_{g'''}\cdot\Big[-\frac{(z+u-t)^3}{3}\Big] = \tfrac{1}{3}M_{g'''}\cdot u^3.$$

Si $u \leqslant 0$, alors :

$$\left|\int_z^{z+u} (z+u-t)^2 g'''(t)\mathrm{d}t\right| \leqslant \int_{z+u}^{z} (z+u-t)^2|g'''(t)|\mathrm{d}t \leqslant M_{g'''}\cdot\int_{z+u}^{u} (z+u-t)^2\mathrm{d}t = M_{g'''}\cdot\frac{(-u)^3}{3} = \tfrac{1}{3}M_{g'''}\cdot|u|^3.$$

Les mêmes inégalités s'appliquent à l'autre intégrale, de sorte qu'en commençant par utiliser l'inégalité triangulaire pour la valeur absolue de la somme :

$$\begin{aligned}\big|R(z,u,v)\big| &\leqslant \tfrac{1}{2}\left|\int_z^{z+u}(z+u-t)^2 g'''(t)\mathrm{d}t\right| + \tfrac{1}{2}\left|\int_z^{z+v}(z+v-t)^2 g'''(t)\mathrm{d}t\right| \\ &\leqslant \tfrac{1}{2}\times\tfrac{1}{3}M_{g'''}|u|^3 + \tfrac{1}{2}\cdot\tfrac{1}{3}M_{g'''}|v|^3 = \tfrac{1}{6}M_{g'''}\big(|u|^3+|v|^3\big).\end{aligned}$$

15. Soit $(Y_i)_{i\geqslant 1}$ une suite de variables aléatoires de loi normale $\mathcal{N}(0,1)$, mutuellement indépendantes entre elles et indépendantes des variables X_j.

a) Soit un entier $n \geqslant 2$.

i) C'est du cours : la somme de variables mutuellement indépendantes suivant des lois normales, suit encore une loi normale, où l'espérance de la somme est la somme des espérances, et la variance est la somme des variances.

Ici, la variable aléatoire $\displaystyle\sum_{i=1}^{n} Y_i$ suit donc la loi normale $\mathcal{N}(0,n)$.

ii) On pose $T_n = \frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n} Y_i$: la loi normale est stable par linéarité, c'est-à-dire que T_n suit la loi normale $\mathcal{N}\big(\frac{1}{\sqrt{n}}\times 0, \big(\frac{1}{\sqrt{n}}\big)^2\times n\big) = \mathcal{N}(0,1)$, ce qui signifie que T_n suit la loi normale centrée, réduite.

b) Pour $k\in\{2,\dots,n-1\}$, on pose $W_k = \frac{1}{\sqrt{n}}(Y_1+\dots+Y_{k-1}+X_{k+1}+\dots+X_n)$, avec $W_1 = \frac{1}{\sqrt{n}}\sum\limits_{i=2}^{n} X_i$ et $W_n = \frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n-1} Y_i$.

i) Pour tout $k\in\{2,\dots,n-1\}$:

$$\begin{aligned}W_k + \tfrac{1}{\sqrt{n}}Y_k &= \frac{1}{\sqrt{n}}(Y_1+\dots+Y_{k-1}+Y_k+X_{k+1}+X_{k+2}+\dots+X_n) \\ &= \frac{1}{\sqrt{n}}(Y_1+\dots+Y_k+X_{k+2}+\dots+X_n) + \frac{1}{\sqrt{n}}X_{k+1} \\ &= W_{k+1} + \frac{1}{\sqrt{n}}X_{k+1}.\end{aligned}$$

Cette relation est encore vraie lorsque $k=1$: dans ce cas,

$$W_1 + \frac{1}{\sqrt{n}}Y_1 = \frac{1}{\sqrt{n}}(Y_1+X_2+\dots+X_n) = \frac{1}{\sqrt{n}}(Y_1+X_3+\dots+X_n) + \frac{1}{\sqrt{n}}W_2 + \frac{1}{\sqrt{n}}X_2,$$

donc la relation est aussi vraie pour $k=1$, et finalement pour tout $k\in\{1,2,\dots,n-1\}$.

ii) La relation précédente permet d'en déduire :

$$\begin{aligned}\sum_{k=1}^{n-1}\big(g_n(W_k+\tfrac{1}{\sqrt{n}}X_k) - g_n(W_k+\tfrac{1}{\sqrt{n}}Y_k)\big) &= \sum_{k=1}^{n-1}\big(g_n(W_k+\tfrac{1}{\sqrt{n}}X_k) - g_n(W_{k+1}+\tfrac{1}{\sqrt{n}}X_{k+1})\big) \\ &= g_n(W_1+\tfrac{1}{\sqrt{n}}X_1) - g_n(W_n+\tfrac{1}{\sqrt{n}}X_n) \\ &= g_n\big(\tfrac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i\big) - g_n\big(\tfrac{1}{\sqrt{n}}\sum_{i=1}^{n}Y_i\big) = g_n(Z_n) - g_n(T_n).\end{aligned}$$

c) i) La variable aléatoire W_k est définie en fonction de $Y_1,\dots,Y_{k-1},X_{k+1},\dots,X_n$ donc d'après le lemme des coalitions, W_k est indépendante de X_k et de Y_k.

Par conséquent :

$$E\big((X_k-Y_k)g_n'(W_k)\big) = E(X_k-Y_k)\times E\big(g_n'(W_k)\big) = \big(E(X_k)-E(Y_k)\big)\times E\big(g_n'(W_k)\big) = 0$$

puisque $E(X_k)=E(Y_k)=0$.

ii) Les mêmes arguments assurent que W_k est indépendante de X_k^2 et de Y_k^2, donc :

$$E\big((X_k^2 - Y_k^2)g_n''(W_k)\big) = \big(E(X_k^2) - E(Y_k^2)\big) \times E\big(g_n''(W_k)\big) = 0$$

puisque $E(X_k^2) = \mathrm{Var}(X_k) + E(X_k)^2 = 1 = \mathrm{Var}(Y_k) + E(Y_k)^2 = E(Y_k^2)$, donc $E(X_k^2) - E(Y_k^2) = 0$.

iii) En reprenant la relation obtenue en 14.b) avec la fonction g_n (et avec $z = W_k(\omega)$, $u = \frac{1}{\sqrt{n}}X_k(\omega)$, $v = \frac{1}{\sqrt{n}}Y_k(\omega)$ pour $\omega \in \Omega$, on peut écrire l'égalité de variables aléatoires :

$$g_n(W_k+\tfrac{1}{\sqrt{n}}X_k)-g_n(W_k+\tfrac{1}{\sqrt{k}}Y_k) = g_n'(W_k)(X_k-Y_k)+\tfrac{1}{2}g_n''(W_k)(X_k^2-Y_k^2)+R(W_k, \tfrac{1}{\sqrt{n}}X_k, \tfrac{1}{\sqrt{n}}Y_k).$$

La linéarité de l'espérance et les résultats précédents permettent donc d'obtenir :

$$\begin{aligned} &E\big(g_n(W_k + \tfrac{1}{\sqrt{n}}X_k) - g_n(W_k + \tfrac{1}{\sqrt{n}}Y_k)\big) \\ &= E\big((X_k - Y_k)g_n'(W_k)\big) + \tfrac{1}{2}E\big((X_k^2 - Y_k^2)g_n''(W_k)\big) + E\big(R(W_k, \tfrac{1}{\sqrt{n}}X_k, \tfrac{1}{\sqrt{n}}Y_k)\big) \\ &= E\big(R(W_k, \tfrac{1}{\sqrt{n}}X_k, \tfrac{1}{\sqrt{n}}Y_k)\big) \end{aligned}$$

d) En calculant l'espérance des deux membres de la relation obtenue en **15**.b)ii), et toujours par linéarité de l'intégrale :

$$E\big(g_n(Z_n)\big)-E\big(g_n(T_n)\big) = \sum_{k=1}^{n-1}\Big(E\big(g_n(W_k+\tfrac{1}{\sqrt{n}}X_k)\big)-g_n(W_k+\tfrac{1}{\sqrt{n}}Y_k)\big)\Big) = \sum_{k=1}^{n-1}E\big(R(W_k, \tfrac{1}{\sqrt{n}}X_k, \tfrac{1}{\sqrt{n}}Y_k)\big),$$

donc :

$$\begin{aligned} \big|E\big(g_n(Z_n)\big) - E\big(g_n(T_n)\big)\big| &= \Big|\sum_{k=1}^{n-1}E\big(R(W_k, \tfrac{1}{\sqrt{n}}X_k, \tfrac{1}{\sqrt{n}}Y_k)\big)\Big| \leqslant \sum_{k=1}^{n-1}\big|E\big(R(W_k, \tfrac{1}{\sqrt{n}}X_k, \tfrac{1}{\sqrt{n}}Y_k)\big)\big| \\ &\leqslant \sum_{k=1}^{n}E\big(|R(W_k, \tfrac{1}{\sqrt{n}}X_k, \tfrac{1}{\sqrt{n}}Y_k)|\big) \quad \text{(on a toujours } \big|E(X)\big| \leqslant E\big(|X|\big)\text{)} \end{aligned}$$

e) Le résultat de **14**.c) permet d'écrire :

$$\big|R(W_k, \tfrac{1}{\sqrt{n}}X_k, \tfrac{1}{\sqrt{n}}Y_k)\big| \leqslant \tfrac{1}{6}M_{g_n'''}\Big(\big(\tfrac{1}{\sqrt{n}}\big)^3|X_k|^3 + \big(\tfrac{1}{\sqrt{n}}\big)^3|Y_k|^3\Big) = \frac{1}{6n\sqrt{n}}M_{g_n'''}\big(|X_k|^3 + |Y_k|^3\big),$$

donc par croissance et linéarité de l'espérance :

$$E\Big(\big|R(W_k, \tfrac{1}{\sqrt{n}}X_k, \tfrac{1}{\sqrt{n}}Y_k)\big|\Big) \leqslant \frac{1}{6n\sqrt{n}}M_{g_n'''}\Big(E\big(|X_k|^3\big)+E\big(|Y_k|^3\big)\Big) = \frac{1}{6n\sqrt{n}}M_{g_n'''}\Big(E\big(|X_1|^3\big)+E\big(|Y_1|^3\big)\Big),$$

puisque les $(X_k)_{k\geqslant 1}$ suivent la même loi (que X_1 donc), et qu'il en est de même pour les $(Y_k)_{k\geqslant 1}$.
Enfin, d'après le résultat de la question précédente :

$$\begin{aligned} \big|E\big(g_n(Z_n)\big) - E\big(g_n(T_n)\big)\big| &\leqslant \sum_{k=1}^{n}E\big(|R(W_k, \tfrac{1}{\sqrt{n}}X_k, \tfrac{1}{\sqrt{n}}Y_k)|\big) \leqslant \sum_{k=1}^{n}\underbrace{\frac{1}{6n\sqrt{n}}M_{g_n'''}\Big(E\big(|X_1|^3\big) + E\big(|Y_1|^3\big)\Big)}_{\text{ne dépend pas de } k} \\ &\leqslant n \times \frac{1}{6n\sqrt{n}}M_{g_n'''}\Big(E\big(|X_1|^3\big) + E\big(|Y_1|^3\big)\Big) \\ &\leqslant \frac{1}{6\sqrt{n}}M_{g_n'''}\Big(E\big(|X_1|^3\big) + E\big(|Y_1|^3\big)\Big) \end{aligned}$$

f) Une valeur absolue est toujours positive, donc d'après le résultat précédent et vu le fait que $\lim\limits_{n \to +\infty} \dfrac{1}{6\sqrt{n}} M_{g_n'''} \Big(E\big(|X_1|^3\big) + E\big(|Y_1|^3\big) \Big) = 0$, on peut effectivement conclure que :

$$\lim_{n \to +\infty} \big| E\big(g_n(Z_n)\big) - E\big(g_n(T_n)\big) \big| = 0.$$

16. a) En reprenant ici le résultat de la question **13**.b)ii) avec la variable aléatoire $X = T_n$, pour tout réel x :

$$\mathbb{1}_{[T_n \leqslant x]} \leqslant g_n(T_n) \leqslant \mathbb{1}_{[T_n \leqslant x + a_n]},$$

et par croissance de l'espérance :

$$E\big(\mathbb{1}_{[T_n \leqslant x]}\big) \leqslant E\big(g_n(T_n)\big) \leqslant E\big(\mathbb{1}_{[T_n \leqslant x + a_n]}\big),$$

ce qui donne bien :

$$\mathbb{P}(T_n \leqslant x) \leqslant E\big(g_n(T_n)\big) \leqslant \mathbb{P}(T_n \leqslant x + a_n),$$

Puisqu'on rappelle (comme vu en **13**.c)) que $\mathbb{1}_A$ est, pour tout événement A, une variable de Bernoulli qui a pour espérance :

$$E(\mathbb{1}_A) = 0 \times \mathbb{P}(\mathbb{1}_A = 0) + 1 \times \mathbb{P}(\mathbb{1}_A = 1) = 0 + \mathbb{P}(A) = \mathbb{P}(A).$$

b) On a vu en **15**.a)ii) que T_n suit déjà la loi normale centrée, réduite donc $\mathbb{P}(T_n \leqslant x) = \Phi(x)$ pour tout réel x.

On a aussi : $\forall x \in \mathbb{R},\ \mathbb{P}(T_n \leqslant x + a_n) = \Phi(x + a_n)$, où $\lim\limits_{n \to +infty} a_n = \lim\limits_{n \to +\infty} n^{-1/12} = 0$; sachant que Φ est continue sur $\mathbb{R}$, on peut donc conclure que $\lim\limits_{n \to +\infty} \Phi(x + a_n) = \Phi(x) = \lim\limits_{n \to +\infty} \mathbb{P}(T_n \leqslant x + a_n)$.

On conclut donc une fois de plus avec le théorème d'encadrement, appliqué à la double inégalité obtenue à la question précédente :

$$\lim_{n \to +\infty} E\big(g_n(T_n)\big) = \Phi(x).$$

c) On termine en mettant en lien les résultats de **15**.f) et de **16**.b) qui expriment d'une part que $E\big(g_n(Z_n)\big)$ et $E\big(g_n(T_n)\big)$ se rapprochent indéfiniment l'un de l'autre, et par ailleurs que $E\big(g_n(T_n)\big)$ se rapproche indéfiniment de $\Phi(x)$, pour aboutir à la conclusion logique que $E\big(g_n(Z_n)\big)$ se rapproche indéfiniment de $\Phi(x)$.

La preuve formelle travaille une dernière fois sur la notion de distance :

$$\begin{aligned} 0 \leqslant \big| E\big(g_n(Z_n)\big) - \Phi(x) \big| &= \big| E\big(g_n(Z_n)\big) - E\big(g_n(T_n)\big) + E\big(g_n(T_n)\big) - \Phi(x) \big| \\ &\leqslant \big| E\big(g_n(Z_n)\big) - E\big(g_n(T_n)\big) \big| + \big| E\big(g_n(T_n)\big) - \Phi(x) \big| \end{aligned}$$

Et comme $\lim\limits_{n \to +\infty} \big| E\big(g_n(Z_n)\big) - E\big(g_n(T_n)\big) \big| = 0 = \lim\limits_{n \to +\infty} \big| E\big(g_n(T_n)\big) - \Phi(x) \big| = 0$, alors par encadrement :

$$\lim_{n \to +\infty} \big| E\big(g_n(Z_n)\big) - \Phi(x) \big| = 0 \iff \lim_{n \to +\infty} E\big(g_n(Z_n)\big) = \Phi(x).$$

★ ★ ★ **FIN DU SUJET** ★ ★ ★