Puisque la fonction g est strictement croissante, les événement $|X| \geq a$ et $g(|X|) \geq g(a)$ sont identiques. Or l'inégalité de Markov donne

$$\mathrm{E}\left(g(|X)|\right) \geq g(a)\mathrm{P}\left(g(|X|) \geq g(a)\right)$$

et donc

$$\mathrm{P}\left(|X| \geq a\right) \leq \frac{\mathrm{E}\left(g(|X|)\right)}{g(a)}$$

Rappelons

$$\mathrm{E}(X) = np \text{ et } \mathrm{V}(X) = np(1-p)$$

Considérons la variable aléatoire

$$Y = X - np = X - \mathrm{E}(X)$$

Par l'inégalité de Markov

$$\mathrm{P}\left(|Y| \geq na\right) \leq \frac{\mathrm{E}(|Y|)}{n\varepsilon}$$

donc

$$\mathrm{P}\left(\left|\frac{X}{n} - p\right| \geq \varepsilon\right) \leq \frac{\mathrm{E}(|Y|)}{n\varepsilon}$$

Or $\mathrm{E}(|Y|) \leq \sqrt{\mathrm{E}(Y^2)}$ avec $\mathrm{E}(Y^2) = \mathrm{V}(X) = np(1-p)$ donc

$$\mathrm{P}\left(\left|\frac{X}{n} - p\right| \geq \varepsilon\right) \leq \frac{\sqrt{p(1-p)}}{\varepsilon\sqrt{n}}$$

Par stricte croissance de l'exponentiel, l'événement $X - np > n\varepsilon$ équivaut à l'événement

$$\exp\left(\lambda(X - np - n\varepsilon)\right) \geq 1$$

L'inégalité de Markov appliquée à la variable $Y = \exp\left(\lambda(X - np - n\varepsilon)\right)$ permet alors de conclure

$$\mathrm{P}\left(\exp\left(\lambda(X - np - n\varepsilon)\right) \geq 1\right) \leq \frac{\mathrm{E}\left(\exp\left(\lambda(X - np - n\varepsilon)\right)\right)}{1}$$

Par l'inégalité de Bienaymé-Tchebychev

$$\mathrm{P}\left(|X - \mu| \geq \alpha\sigma\right) < \frac{\sigma^2}{(\alpha\sigma^2)} = \frac{1}{\alpha^2}$$

On conclut par considération d'évènement complémentaire.

On a

$$\mathrm{E}(Y) = \alpha^2 \mathrm{E}\left((X-\mu)^2\right) + 2\alpha \mathrm{E}(X-\mu) + \sigma^2 = (\alpha^2+1)\sigma^2$$

L'inégalité de Markov appliquée à la variable positive Y donne

$$\mathrm{P}(Y \geq a) \leq \frac{\mathrm{E}(Y)}{a}$$

Pour $a = \sigma^2(\alpha^2+1)^2$,

$$\mathrm{P}(Y \geq a) \leq \frac{1}{1+\alpha^2}$$

Or

$$(X \geq \mu + \alpha\sigma) = \left(\alpha(X-\mu) + \sigma \geq (\alpha^2+1)\sigma\right)$$

et donc

$$(X \geq \mu + \alpha\sigma) \subset (Y \geq a)$$

puis

$$\mathrm{P}\left(X \geq \mu + \alpha\sigma\right) \leq \frac{1}{1+\alpha^2}$$

1. a) $\boxed{\text{Question de cours : inégalité de Markov.}}$

Si X est une variable aléatoire positive admettant une espérance, alors pour tout réel c strictement positif : $\boxed{P([X \geqslant c]) \leqslant \dfrac{E(X)}{c}}$.

b) Soit $a \in \mathbb{R}_+$.

i) Pour tout réel strictement positif b, justifier l'inégalité :

$$\boxed{P([X \geq E(X)+a]) \leq \frac{E\big((X-E(X)+b)^2\big)}{(a+b)^2}}.$$

Soit $b > 0$.

$$P([X \geq E(X)+a]) = P([X - E(X) + b \geq a+b]) \leq P\left(\left[\left(X-E(X)+b\right)^2 \geq (a+b)^2\right]\right)$$

d'où, grâce à l'inégalité de Markov :

$$P([X \geq E(X)+a]) \leq \frac{E\big((X-E(X)+b)^2\big)}{(a+b)^2}.$$

ii) En appliquant le résultat précédent à $b = V(X)/a$, établir l'inégalité :

$$\boxed{P([X \geq E(X)+a]) \leq \frac{V(X)}{V(X)+a^2}}.$$

Pour $b = V(X)/a$, on obtient :

$$P([X \geq E(X)+a]) \leq \frac{E\big((X-E(X)+V(X)/a)^2\big)}{(a+V(X)/a)^2} = \frac{V(X)+\left(V(X)\right)^2/a^2}{a^2(1+V(X)/a^2)^2} = \frac{V(X)}{V(X)+a^2}.$$

2. On appelle *médiane* de X tout nombre réel m tel que $P([X \leq m]) = \dfrac{1}{2}$.

a) $\boxed{\text{Justifier que } X \text{ admet au moins une médiane et que l'ensemble des médianes de } X \text{ est un segment de } \mathbb{R}.}$

La fonction de répartition $F : x \longmapsto P([X \leq x])$ étant continue sur $\mathbb{R}$ (puisque X est une variable aléatoire à densité), de limite 0 en $-\infty$ et 1 en $+\infty$, le théorème des valeurs intermédiaires permet d'affirmer que, pour tout réel y strictement compris entre 0 et 1, il existe $x \in \mathbb{R}$ tel que $f(x) = y$.

En particulier, il existe $m \in \mathbb{R}$ tel que $P([X \leq m]) = \dfrac{1}{2}$.

Soit I l'ensemble (non vide) des réels x tels que $F(x) = \dfrac{1}{2}$.

• I est un *intervalle*, parce que, quels que soient $(a,b) \in I^2$, tous les réels x compris entre a et b appartiennent à I (puisque la fonction croissante F prend la même valeur en a et en b).

• $I = \{x \in \mathbb{R} \mid F(x) = 1/2\}$ est une partie *fermée* de $\mathbb{R}$, puisque F est continue.

• I est une partie *bornée* de $\mathbb{R}$, puisque, par exemple, pour q_1 et q_3 choisis tels que $F(q_1) = 1/4$ et $F(q_3) = 3/4$, l'intervalle I est inclus dans $]q_1, q_3[$, donc borné.

Par conséquent, I est un *segment* (intervalle fermé borné) de $\mathbb{R}$.

b) Dans cette question, on note m une médiane de X et on suppose que m est supérieure ou égale à $E(X)$. On note $\sigma(X)$ l'écart-type de X.

Déduire de l'inégalité prouvée en 1.b.ii, appliquée à $a = m - E(X)$, que $m - E(X)$ est inférieur ou égal à $\sigma(X)$.

Par application de l'inégalité 1.b.ii, appliquée à $a = m - E(X)$, on obtient

$$\frac{1}{2} = P([X \geq m]) \leq \frac{V(X)}{V(X) + \big(m - E(X)\big)^2}$$

d'où $V(X) + \big(m - E(X)\big)^2 \leq 2\,V(X)$, puis : $\boxed{m - E(X) \leq \sigma(X)}$.

c)

Justifier que toute médiane m de X vérifie l'inégalité : $\dfrac{\lvert m - E(X)\rvert}{\sigma(X)} \leq 1$.

D'après ce qui précède, l'inégalité demandée est vraie lorsque $m \geq E(X)$ ($\sigma(X)$ est strictement positif, puisque la variable aléatoire X possède une densité).

Dans le cas où $m \leq E(X)$, il suffit d'appliquer le résultat précédent à la variable aléatoire $-X$, dont $-m$ est une médiane, pour obtenir l'inégalité demandée.

3. Pour tout entier $n \geq 2$, on note X_n une variable aléatoire admettant pour densité la fonction f_n définie par :

$$f_n(x) = \begin{cases} 0 & \text{si } x < -1/n \\ n/2 & \text{si } -1/n \leq x < 0 \\ 1/\big(2(n-1)\big) & \text{si } 0 \leq x < 1 - 1/n \\ (n-1)/2 & \text{si } 1 - 1/n \leq x < 1 \\ 0 & \text{si } x \geq 1 \end{cases} .$$

a)

Trouver l'unique médiane de X_n.

La fonction de répartition F_n de X_n est donnée par : [1]

$$F_n(x) = \begin{cases} 0 & \text{si } x < -1/n \\ (nx+1)/2 & \text{si } -1/n \leq x < 0 \\ 1/2 + x/\big(2(n-1)\big) & \text{si } 0 \leq x < 1 - 1/n \\ 1 + (n-1)(x-1)/2 & \text{si } 1 - 1/n \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases} .$$

L'unique médiane m_n de X_n est 0.

b)

Justifier la convergence en loi de la suite $(X_n)_{n\in\mathbb{N}^*}$.

• Pour tout $x < 0$, $F_n(x)$ tend vers 0 quand n tend vers l'infini, puisque, pour n assez grand, on a $x < -1/n$ (d'où $F_n(x) = 0$).

• Pour tout $x \in [0, 1[$, $F_n(x)$ tend vers $1/2$ quand n tend vers l'infini, puisque, pour n assez grand, on a $0 \leq x < -1/n$ (d'où $F_n(x) = 1/2 + x/(2(n-1))$, qui tend vers $1/2$ quand n tend vers l'infini).

• Pour tout $x \geq 1$, $F_n(x) = 1$.

En résumé, quand n tend vers l'infini, $F_n(x) \longrightarrow G(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1/2 & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$.

La suite $(X_n)_{n\in\mathbb{N}^*}$ converge donc en loi vers une (toute) variable aléatoire suivant la loi de Bernoulli $\mathcal{B}(1/2)$. [2]

c) Quelle propriété du majorant trouvé en 2.c peut-on déduire des limites de $E(X_n)$ et $V(X_n)$ quand n tend vers l'infini ?

La limite en loi de la suite $(X_n)_{n\in\mathbb{N}^*}$ permet de faire la conjecture que $E(X_n)$ et $V(X_n)$ tendent respectivement vers $1/2$ et $1/4$ quand n tend vers l'infini. Une fois établies ces convergences, on pourra affirmer

$$\boxed{\lim_{n\to+\infty} \frac{|m_n - E(X_n)|}{\sigma(X_n)} = 1}$$

ce qui prouvera que le majorant trouvé en 2.c est optimal.

Pour déterminer effectivement les limites des moments des X_n, une méthode efficace consiste à écrire la densité f_n comme la moyenne pondérée de trois densités de lois uniformes, dont l'espérance et la variance sont connues :

$$f_n = \frac{1}{2}\, g_n + \frac{1}{2n}\, h_n + \frac{(n-1)}{2n}\, k_n$$

où g_n, h_n et k_n sont des densités respectives des lois uniformes sur $[-1/n, 0]$, $[0, 1-1/n]$ et $[1-1/n, 1]$.

On déduit de cette décomposition les expressions de $E(X_n)$ et $E({X_n}^2)$, et leurs limites.

$$E(X_n) = \frac{1}{2}\,(\frac{-1}{2n}) + \frac{1}{2n}\,(\frac{n-1}{2n}) + \frac{(n-1)}{2n}\,(\frac{2n-1}{2n})$$

tend vers $1/2$ quand n tend vers l'infini.
De même, comme les espérances des carrés des trois lois uniformes de la décomposition tendent respectivement vers 0, $1/3$ et 1, leur moyenne pondérée par les coefficients $1/2$, $1/(2n)$ et $(n-1)/(2n)$, qui est égale à $E({X_n}^2)$, tend vers $1/2$ quand n tend vers l'infini. On a donc bien :

$$\lim_{n\to+\infty} \sigma(X_n) = \sqrt{\frac{1}{2} - (\frac{1}{2})^2} = \frac{1}{2}\,.$$

Finalement, la borne supérieure des quotients $\dfrac{|m - E(X)|}{\sigma(X)}$ pour les variables aléatoires X à densité possédant un moment d'ordre deux est bien égale à 1. Il n'existe pas de majorant strictement plus petit de tous ces quotients.

Par linéarité de l'espérance,

$$\mathbb{E}(\overline{X}_n) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}(X_k) = \frac{1}{n}\sum_{k=1}^{n} \frac{\theta}{2} = \frac{\theta}{2}$$

et $\overline{X}_n$ est donc un estimateur sans biais de $\frac{\theta}{2}$.
Par suite, $T_n = 2\overline{X}_n$ est un estimateur sans biais de θ :

$$\mathbb{E}(T_n) = \theta.$$

Les X_i ont pour fonction de répartition commune :

$$F : x \in \mathbb{R} \longmapsto \begin{cases} 0 & \text{si } x < 0 \\ \frac{x}{\theta} & \text{si } 0 \leqslant x < \theta \\ 1 & \text{si } x \geqslant \theta \end{cases} .$$

Puis, pour $x \in \mathbb{R}$,

$$\begin{aligned} F_{M_n}(x) &= \mathbb{P}(M_n \leqslant x) = \mathbb{P}\left(\max(X_1, \dots, X_n) \leqslant x\right) \\ &= \mathbb{P}(X_1 \leqslant x, \dots, X_n \leqslant x) = \mathbb{P}(X_1 \leqslant x) \cdots \mathbb{P}(X_n \leqslant x) = F(x)^n \end{aligned}$$

par indépendance de $X_1, \dots, X_n$.

Ainsi M_n admet pour fonction de répartition

$$F_{M_n} : x \in \mathbb{R} \longmapsto \begin{cases} 0 & \text{si } x < 0 \\ \left(\frac{x}{\theta}\right)^n & \text{si } 0 \leqslant x < \theta \\ 1 & \text{si } x \geqslant \theta \end{cases} ,$$

continue sur $\mathbb{R}$ (même en 0 et θ) et de classe $\mathscr{C}^1$ sur $\mathbb{R} \setminus \{0, \theta\}$. La variable M_n est donc à densité donnée par

$$f_{M_n} = F'_{M_n} : x \in \mathbb{R} \longmapsto \begin{cases} \frac{n}{\theta^n} x^{n-1} & \text{si } 0 \leqslant x \leqslant \theta \\ 0 & \text{sinon} \end{cases} .$$

La variable M_n étant presque sûrement bornée, elle admet une espérance

$$\mathbb{E}(M_n) = \int_{-\infty}^{+\infty} t f_{M_n}(t)\,\mathrm{d}t = \frac{n}{\theta^n} \int_0^{\theta} t^n\,\mathrm{d}t = \frac{n}{n+1}\theta.$$

Par suite, $U_n = \frac{n+1}{n} M_n$ est un estimateur sans biais de θ :

$$\mathbb{E}(U_n) = \frac{n+1}{n}\mathbb{E}(M_n) = \theta.$$

L'estimateur T_n de θ étant non biaisé, son risque quadratique est donné par

$$r(T_n) = \mathbb{V}(2\overline{X}_n) = \mathbb{V}\left(\frac{2}{n}\sum_{k=1}^{n} X_k\right) = \frac{4}{n^2}\sum_{k=1}^{n} \mathbb{V}(X_k) = \frac{4}{n}\frac{\theta^2}{12} = \frac{\theta^2}{3n}$$

par indépendance de $X_1, \dots, X_n$. Il converge vers 0 lorsque $n \to \infty$, ce qui assure que T_n est un estimateur convergent de θ.

On peut raisonner de même pour U_n :

$$r(U_n) = \mathbb{V}\left(\frac{n+1}{n} M_n\right) = \left(\frac{n+1}{n}\right)^2 \left(\mathbb{E}(M_n^2) - \mathbb{E}(M_n)^2\right)$$

où

$$\mathbb{E}(M_n^2) = \int_{-\infty}^{+\infty} t^2 f_{M_n}(t)\,\mathrm{d}t = \frac{n}{\theta^n}\int_0^{\theta} t^{n+1}\,\mathrm{d}t = \frac{n}{n+2}\theta^2$$

si bien que

$$r(U_n) = \frac{1}{n(n+2)}\theta^2 \xrightarrow[n \to \infty]{} 0,$$

d'où l'on déduit que U_n est un estimateur convergent de θ.

Le meilleur estimateur est celui qui possède le risque quadratique le plus faible. Or un équivalent de chacun d'eux montre clairement que $r(U_n) = o\big(r(T_n)\big)$ lorsque $n \to \infty$. L'estimateur U_n est donc meilleur estimateur de θ que T_n pour n assez grand.

1. Cours.

2.a) X_n suit une loi géométrique de paramètre $1/n$.
b) $\forall t \in \mathbf{R}$, $P(U_n \leqslant t) = P(X_n \leqslant nt)$ et si $t \leqslant 0$, $P(U_n \leqslant t) = 0$.
Si $t > \dfrac{1}{n}$, $P(U_n \leqslant t) = \displaystyle\sum_{k=1}^{\lfloor nt \rfloor} P(X_n = k) = \frac{1}{n}\sum_{k=1}^{\lfloor nt \rfloor}\left(1 - \frac{1}{n}\right)^{k-1} = 1 - \left(1 - \frac{1}{n}\right)^{\lfloor nt \rfloor}$.

3. Si $t < 0$, $\lim\limits_{n\to+\infty} P(U_n \leqslant t) = 0$ et si $t > 0$, $P(U_n \leqslant t) = 1 - \exp\big(\lfloor nt \rfloor \ln(1 - 1/n)\big)$. Or,
$\lfloor nt \rfloor \underset{n\to+\infty}{\sim} nt \Longrightarrow \lfloor nt \rfloor \ln(1-1/n) \underset{n\to+\infty}{\sim} -t \Longrightarrow \lim\limits_{n\to+\infty} \lfloor nt \rfloor \ln(1-1/n) = -t \Longrightarrow \lim\limits_{n\to+\infty} P(U_n \leqslant t) = 1 - \mathrm{e}^{-t}$.
Bilan : la suite $(U_n)_{n\in\mathbf{N}^*}$ converge en loi vers une variable suivant la loi exponentielle $\mathcal{E}(1)$.

4. `n=input('n=') ; k=1 ; p=1/n ; while rand()>p ; then k=k+1 end ; U=k/n`

5.a) Soit $n \in \mathbf{N}^*$. On a : $Y_n(\Omega) = [\![2, +\infty[\![$ et pour $k \geqslant 2$, $P(Y_n = k) = (k-1)\left(\dfrac{1}{n}\right)^2\left(1 - \dfrac{1}{n}\right)^{k-2}$.

b) Si $t \leqslant 0$, $P(W_n \leqslant t) = 0$ et si $t > 0$, $P(W_n \leqslant t) = \displaystyle\sum_{k=1}^{\lfloor nt \rfloor} P(Y_n = k) = \left(\frac{1}{n}\right)^2 \sum_{k=2}^{\lfloor nt \rfloor}(k-1)\left(1 - \frac{1}{n}\right)^{k-2}$, d'où :

$P(W_n \leqslant t) = \left(\dfrac{1}{n}\right)^2 \displaystyle\sum_{j=1}^{\lfloor nt \rfloor - 1} j\left(1 - \frac{1}{n}\right)^{j-1}$ si $t > 0$.

c) On pose : $S_N(q) = \displaystyle\sum_{j=0}^{N} q^j = \frac{1 - q^{N+1}}{1-q} \Longrightarrow S'_N(q) = \sum_{j=1}^{N} j q^{j-1} = \frac{1 - N(1-q)\,q^N - q^N}{(1-q)^2}$.

d) Si $t < 0$, $\lim\limits_{n\to+\infty} P(W_n \leqslant t) = 0$. Pour $t \geqslant 0$, posons : $N = \lfloor nt \rfloor - 1$ et $q = 1 - 1/n$.

On a : $P(W_n \leqslant t) = \left(\dfrac{1}{n}\right)^2 \displaystyle\sum_{j=1}^{\lfloor nt \rfloor - 1} j\left(1 - \frac{1}{n}\right)^{j-1} = 1 - \frac{1}{n}\big(\lfloor nt \rfloor - 1\big)\left(1 - \frac{1}{n}\right)^{\lfloor nt \rfloor - 1} - \left(1 - \frac{1}{n}\right)^{\lfloor nt \rfloor - 1}$. Par suite :

$\lim\limits_{n\to+\infty} P(W_n \leqslant t) = \begin{cases} 1 - t\,\mathrm{e}^{-t} - \mathrm{e}^{-t} & \text{si } t \geqslant 0 \\ 0 & \text{si } t < 0 \end{cases}$. On vérifie que cette fonction limite est de classe C^1 sur $\mathbf{R}$,

donc W est une variable aléatoire à densité et une densité g de W est donnée par : $g(t) = \begin{cases} t\,\mathrm{e}^{-t} & \text{si } t > 0 \\ 0 & \text{si } t \leqslant 0 \end{cases}$.