# ESSEC I 2017

Est-il possible que le marketing digital pose des problèmes de sécurité des données personnelles ? De récents travaux mettant en cause les outils de mesure de performance en temps réel des différentes campagnes de publicité sur internet, démontrent que certaines données très sensibles (préférences religieuses, sexuelles, etc.) peuvent être obtenues par des segmentations précises des audiences et sans aucune action de la part de l'utilisateur.

Dans ce problème, nous nous intéressons à une méthode proposée pour protéger ces données, méthode baptisée **confidentialité différentielle**.

Les parties I et II sont totalement indépendantes. Vous trouverez une aide **Scilab** en fin de sujet.

On considère un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$ sur lequel sont définies les variables aléatoires qui apparaissent dans l'énoncé.

## Partie I - Lois de Laplace - propriétés et simulation

Soit $\alpha \in \mathbb{R}$ et $\beta > 0$. On dit qu'une variable aléatoire réelle à densité suit une loi de Laplace de paramètre $(\alpha, \beta)$, notée $\mathcal{L}(\alpha, \beta)$, si elle admet comme densité la fonction $f$ donnée par :

$$\forall t \in \mathbb{R},\ f(t) = \frac{1}{2\beta}\ \exp\left(-\frac{|t-\alpha|}{\beta}\right)$$

***1.*** Vérifier que $f$ est bien une densité de probabilité d'une variable aléatoire réelle.

*Démonstration.*

- La fonction $f$ est continue sur $\mathbb{R}$ car elle est la composée $f = f_2 \circ f_1$ où :
  - × $f_1 : t \mapsto \dfrac{-1}{\beta}\ |t-\alpha|$ est :
    - continue sur $\mathbb{R}$ car la fonction valeur absolue l'est,
    - telle que : $f_1(\mathbb{R}) \subset \mathbb{R}$.
  - × $f_2 : t \mapsto \dfrac{1}{2\beta}\ \exp(t)$ est continue sur $\mathbb{R}$.

  La fonction $f$ est continue sur $\mathbb{R}$.

- Tout d'abord, $\beta > 0$, donc on a : $\dfrac{1}{2\beta} > 0$. De plus, pour tout $u \in \mathbb{R}$, $\exp(u) > 0$.

  Ainsi, pour tout $t \in \mathbb{R}$, $f(t) = \dfrac{1}{2\beta}\ \exp\left(-\dfrac{|t-\alpha|}{\beta}\right) \geqslant 0$.

- Montrons que l'intégrale $\displaystyle\int_{-\infty}^{+\infty} f(t)\ dt$ converge et vaut 1.
  - Sous réserve de convergence, commençons par effectuer le changement de variable $u = t - \alpha$.

    $u = t - \alpha$ (et donc $t = u + \alpha$)
    $\hookrightarrow du = dt$
    • $t = -\infty \ \Rightarrow\ u = -\infty$
    • $t = +\infty \ \Rightarrow\ u = +\infty$

– Ce changement de variable est valide car $\varphi : u \mapsto u + \alpha$ est de classe $\mathcal{C}^1$ sur $]-\infty, +\infty[$.
On obtient alors :

$$\int_{-\infty}^{+\infty} \frac{1}{2\beta} \exp\left(-\frac{|t-\alpha|}{\beta}\right) dt = \int_{-\infty}^{+\infty} \frac{1}{2\beta} \exp\left(-\frac{|u|}{\beta}\right) du$$

– L'intégrale impropre $\displaystyle\int_{-\infty}^{+\infty} \frac{1}{2\beta} \exp\left(-\frac{|u|}{\beta}\right) du$ est convergente si et seulement si les intégrales impropres $\displaystyle\int_{-\infty}^{0} \frac{1}{2\beta} \exp\left(-\frac{|u|}{\beta}\right) du$ et $\displaystyle\int_{0}^{+\infty} \frac{1}{2\beta} \exp\left(-\frac{|u|}{\beta}\right) du$ le sont.

– La fonction $g : u \mapsto \dfrac{1}{2\beta} \exp\left(-\dfrac{|u|}{\beta}\right)$ est continue sur $[0, +\infty[$.

Ainsi, pour tout $A \in [0, +\infty[$, l'intégrale $\displaystyle\int_0^A g(u)du$ est bien définie. De plus :

$$\begin{aligned}
\int_0^A g(u)\, du &= \int_0^A \frac{1}{2\beta} \exp\left(-\frac{|u|}{\beta}\right) du = \frac{1}{2\beta} \int_0^A \exp\left(-\frac{u}{\beta}\right) du \\
&= \frac{-\cancel{\beta}}{2\cancel{\beta}} \int_0^A \frac{-1}{\beta} \exp\left(-\frac{u}{\beta}\right) du \\
&= \frac{-1}{2} \left[ \mathrm{e}^{-\frac{u}{\beta}} \right]_0^A \\
&= \frac{-1}{2} \left( \mathrm{e}^{-\frac{A}{\beta}} - \mathrm{e}^{-\frac{0}{\beta}} \right) = \frac{1}{2} \left( 1 - \mathrm{e}^{-\frac{A}{\beta}} \right) \underset{A \to +\infty}{\longrightarrow} \frac{1}{2}
\end{aligned}$$

En effet, comme $-\frac{A}{\beta} \underset{A \to +\infty}{\longrightarrow} -\infty$, alors $\mathrm{e}^{-\frac{A}{\beta}} \underset{A \to +\infty}{\longrightarrow} 0$.

L'intégrale impropre $\displaystyle\int_0^{+\infty} g(u)\, du$ est convergente. De plus : $\displaystyle\int_0^{+\infty} g(u)\, du = \frac{1}{2}$.

– On remarque enfin que la fonction $g$ est impaire :

$$\forall u \in \mathbb{R},\ g(-u) = \frac{1}{2\beta} \exp\left(-\frac{|-u|}{\beta}\right) = \frac{1}{2\beta} \exp\left(-\frac{|u|}{\beta}\right) = g(u)$$

Ainsi, en posant le changement de variable $v = -u$, on obtient :

$$\int_0^{+\infty} g(u)\, du = \int_0^{-\infty} g(-v)(-dv) = -\int_0^{-\infty} g(v)\, dv = \int_{-\infty}^{0} g(v)\, dv$$

On en déduit que l'intégrale impropre $\displaystyle\int_{-\infty}^{0} g(v)\, dv$ est convergente.

Il en est de même de $\displaystyle\int_{-\infty}^{+\infty} f(t)\, dt$. Et :

$$\int_{-\infty}^{+\infty} f(t)\, dt = \int_{-\infty}^{+\infty} g(u)\, du = \int_{-\infty}^{0} g(u)\, du + \int_0^{+\infty} g(u)\, du = \frac{1}{2} + \frac{1}{2} = 1$$

Ainsi, $\displaystyle\int_{-\infty}^{+\infty} f(t)\, dt = 1$.

La fonction $f$ est bien la densité d'une variable aléatoire réelle.

**Commentaire**

- Le programme officiel précise que « les changements de variables **non affines** ne seront pratiqués qu'avec des intégrales sur un segment ». Il est donc autorisé, sous réserve de convergence, d'effectuer un changement de variable affine sur une intégrale généralisée (ce qui est fait dans cette question).
- Le changement de variable $\boxed{u = t - \alpha}$ permet de bénéficier de la parité de la fonction $g$. Il est aussi possible de travailler directement avec $f$ comme ci-après.
  - L'intégrale impropre $\displaystyle\int_{-\infty}^{+\infty} f(t)\,dt$ est convergente si et seulement si les intégrales impropres $\displaystyle\int_{-\infty}^{\alpha} f(t)\,dt$ et $\displaystyle\int_{\alpha}^{+\infty} f(t)\,dt$ le sont.
  - La fonction $f$ est continue sur $[\alpha, +\infty[$.
    Ainsi, pour tout $A \in [\alpha, +\infty[$, l'intégrale $\displaystyle\int_{\alpha}^{A} f(t)\,dt$ est bien définie. De plus :

$$\begin{aligned}
\int_{\alpha}^{A} f(t)\,dt &= \int_{\alpha}^{A} \frac{1}{2\beta}\exp\left(-\frac{|t-\alpha|}{\beta}\right)\,dt \\
&= \frac{-\cancel{\beta}}{2\cancel{\beta}} \int_{\alpha}^{A} \frac{-1}{\beta}\exp\left(-\frac{t-\alpha}{\beta}\right)\,dt \\
&= \frac{-1}{2}\left[\,\mathrm{e}^{-\frac{t-\alpha}{\beta}}\,\right]_{\alpha}^{A} \\
&= \frac{-1}{2}\left(\mathrm{e}^{-\frac{A-\alpha}{\beta}} - \mathrm{e}^{0}\right) = \frac{1}{2}\left(1 - \mathrm{e}^{-\frac{A-\alpha}{\beta}}\right) \underset{A\to+\infty}{\longrightarrow} \frac{1}{2}
\end{aligned}$$

  - De même, $f$ est continue sur $]-\infty, \alpha]$.
    Ainsi, pour tout $B \in ]-\infty, \alpha]$, l'intégrale $\displaystyle\int_{B}^{\alpha} f(t)\,dt$ est bien définie. De plus :

$$\begin{aligned}
\int_{B}^{\alpha} f(t)\,dt &= \int_{B}^{\alpha} \frac{1}{2\beta}\exp\left(-\frac{|t-\alpha|}{\beta}\right)\,dt = \frac{\cancel{\beta}}{2\cancel{\beta}} \int_{B}^{\alpha} \frac{1}{\beta}\exp\left(\frac{t-\alpha}{\beta}\right)\,dt \\
&= \frac{1}{2}\left[\,\mathrm{e}^{\frac{t-\alpha}{\beta}}\,\right]_{B}^{\alpha} = \frac{1}{2}\left(\mathrm{e}^{0} - \mathrm{e}^{\frac{B-\alpha}{\beta}}\right) = \frac{1}{2}\left(1 - \mathrm{e}^{\frac{B-\alpha}{\beta}}\right) \underset{B\to-\infty}{\longrightarrow} \frac{1}{2}
\end{aligned}$$

  Ainsi, l'intégrale impropre $\displaystyle\int_{-\infty}^{+\infty} f(t)\,dt$ est convergente et $\displaystyle\int_{-\infty}^{+\infty} f(t)\,dt = 1$.

□

***2.*** Déterminer la fonction de répartition, notée $\Psi$, de la loi $\mathcal{L}(0,1)$.

*Démonstration.*

- On considère ici le cas où $\alpha = 0$ et $\beta = 1$. On note alors $h : t \mapsto \dfrac{1}{2}\mathrm{e}^{-|t|}$.
  Par définition, pour tout $x \in \mathbb{R}$ :

$$\Psi(x) = \int_{-\infty}^{x} h(t)\,dt = \lim_{B\to-\infty} \int_{B}^{x} h(t)\,dt$$

*(cette limite existe et est finie d'après la question précédente)*

- Soit $x \in \mathbb{R}$. Deux cas se présentent :
  - × si $x \leqslant 0$. Soit $B \in ]-\infty, 0[$.

$$\int_B^x h(t)\,dt = \int_B^x \frac{1}{2}\,\mathrm{e}^{-|t|}\,dt = \frac{1}{2}\int_B^x \mathrm{e}^t\,dt = \frac{1}{2}\left[\,\mathrm{e}^t\,\right]_B^x = \frac{1}{2}\left(\mathrm{e}^x - \mathrm{e}^B\right) \underset{B\to-\infty}{\longrightarrow} \frac{1}{2}\,\mathrm{e}^x$$

Ainsi, pour tout $x \leqslant 0$, $\Psi(x) = \dfrac{1}{2}\,\mathrm{e}^x$.

  - × si $x > 0$

$$\Psi(x) = \int_{-\infty}^x h(t)\,dt = \int_{-\infty}^0 h(t)\,dt + \int_0^x h(t)\,dt = \Psi(0) + \int_0^x h(t)\,dt$$

avec $\Psi(0) = \dfrac{1}{2}\,\mathrm{e}^0 = \dfrac{1}{2}$ (d'après le calcul précédent) et :

$$\int_0^x h(t)\,dt = \int_0^x \frac{1}{2}\,\mathrm{e}^{-|t|}\,dt = \int_0^x \frac{1}{2}\,\mathrm{e}^{-t}\,dt = \frac{1}{2}\left[\,-\mathrm{e}^{-t}\,\right]_0^x = \frac{-1}{2}\left(\mathrm{e}^{-x} - \mathrm{e}^{-0}\right) = \frac{1}{2}\left(1 - \mathrm{e}^{-x}\right)$$

Ainsi, pour tout $x > 0$, $\Psi(x) = \dfrac{1}{2} + \dfrac{1}{2}\,(1 - \mathrm{e}^{-x}) = 1 - \dfrac{1}{2}\,\mathrm{e}^{-x}$.

□

***3.*** On suppose que $X$ suit la loi $\mathcal{L}(0,1)$.

***a)*** Montrer que $\beta X + \alpha$ suit la loi $\mathcal{L}(\alpha, \beta)$.

*Démonstration.*

- Notons $Y = \beta X + \alpha$. Soit $x \in \mathbb{R}$.

$$\begin{aligned} F_Y(x) &= \mathbb{P}([Y \leqslant x]) = \mathbb{P}([\beta X + \alpha \leqslant x]) \\ &= \mathbb{P}([\beta X \leqslant x - \alpha]) \\ &= \mathbb{P}\left(\left[X \leqslant \tfrac{x-\alpha}{\beta}\right]\right) && \textit{(car } \beta > 0\textit{)} \\ &= \Psi\left(\tfrac{x-\alpha}{\beta}\right) \end{aligned}$$

Deux cas se présentent.

× Si $\frac{x-\alpha}{\beta} \leqslant 0$ autrement dit si $x \leqslant \alpha$. Alors :

$$\Psi\left(\frac{x-\alpha}{\beta}\right) = \frac{1}{2}\,\mathrm{e}^{\frac{x-\alpha}{\beta}}$$

× Si $\frac{x-\alpha}{\beta} > 0$ autrement dit si $x > \alpha$. Alors :

$$\Psi\left(\frac{x-\alpha}{\beta}\right) = 1 - \frac{1}{2}\,\mathrm{e}^{-\frac{x-\alpha}{\beta}}$$

En résumé : $F_Y : x \mapsto \begin{cases} \dfrac{1}{2}\,\mathrm{e}^{\frac{x-\alpha}{\beta}} & \text{si } x \leqslant \alpha \\ 1 - \dfrac{1}{2}\,\mathrm{e}^{-\frac{x-\alpha}{\beta}} & \text{si } x > \alpha \end{cases}$.

- La fonction $F_Y$ est continue sur $]-\infty, \alpha[$ car la fonction $x \mapsto \mathrm{e}^{\frac{x-\alpha}{\beta}}$ l'est comme la composée $F_2 \circ F_1$ où :
  - $\times$ $F_1 : x \mapsto \frac{x-\alpha}{\beta}$ est :
    - continue sur $]-\infty, \alpha[$ car polynomiale,
    - telle que : $F_1(]-\infty, \alpha[) \subset \mathbb{R}$.
  - $\times$ $F_2 : x \mapsto \mathrm{e}^x$ est continue sur $\mathbb{R}$.

  On démontre de même que $F_Y$ est continue sur $]\alpha, +\infty[$.

  Enfin :

  $\times$ $\lim\limits_{x \to \alpha^-} F_Y(x) = \lim\limits_{x \to \alpha} \dfrac{1}{2}\, \mathrm{e}^{\frac{x-\alpha}{\beta}} = \dfrac{1}{2}\mathrm{e}^0 = \dfrac{1}{2}$,

  $\times$ $\lim\limits_{x \to \alpha^+} F_Y(x) = \lim\limits_{x \to \alpha} \left(1 - \dfrac{1}{2}\, \mathrm{e}^{-\frac{x-\alpha}{\beta}}\right) = 1 - \dfrac{1}{2}\, \mathrm{e}^0 = \dfrac{1}{2}$.

  D'où : $\lim\limits_{x \to \alpha^-} F_Y(x) = F_Y(\alpha) = \lim\limits_{x \to \alpha^+} F_Y(x)$.

  Ainsi, $F_Y$ est continue sur $\mathbb{R}$.

  On démontre de même que $F_Y$ est de classe $\mathcal{C}^1$ sur $]-\infty, \alpha[$ et sur $]\alpha, +\infty[$.

- Ainsi, $Y$ est une variable à densité. On obtient une densité $f_Y$ en dérivant sur les intervalles ouverts :

$$f_Y : x \mapsto \begin{cases} \dfrac{1}{2\beta}\, \mathrm{e}^{\frac{x-\alpha}{\beta}} & \text{si } x < \alpha \\ \dfrac{1}{2\beta}\, \mathrm{e}^{-\frac{x-\alpha}{\beta}} & \text{si } x > \alpha \end{cases}$$

  que l'on complète en posant $f_Y(\alpha) = \dfrac{1}{2\beta}$.

  Ainsi : $\forall x \in \mathbb{R},\ f_Y(x) = \dfrac{1}{2\beta}\, \mathrm{e}^{-\frac{|x-\alpha|}{\beta}} = f(x)$.

  On en conclut : $Y = \beta X + \alpha \hookrightarrow \mathcal{L}(\alpha, \beta)$.

**Commentaire**

Le programme officiel précise que les candidats doivent savoir retrouver une densité de $aX + b$ (où $a \neq 0$), ce qui explique la rédaction choisie à cette question.
Cependant, on peut penser qu'un candidat utilisant directement l'expression de la densité d'une transformée affine se verrait attribuer la totalité des points.
On rappelle que, pour $X$ une v.a.r. à densité et $Y = aX + b$ avec $a \neq 0$, on obtient :

$$\forall x \in \mathbb{R},\ f_Y(x) = \frac{1}{|a|}\, f_X\left(\frac{x-b}{a}\right)$$

□

***b)*** En déduire la fonction de répartition de la loi $\mathcal{L}(\alpha, \beta)$.

*Démonstration.*
Dans la question précédente, on a déterminé la fonction de répartition $F_Y$ d'une v.a.r. $Y$ et on a démontré : $Y \hookrightarrow \mathcal{L}(\alpha, \beta)$.

On en déduit que la fonction de répartition de la loi $\mathcal{L}(\alpha, \beta)$ est :

$$F : x \mapsto \begin{cases} \dfrac{1}{2}\,\mathrm{e}^{\frac{x-\alpha}{\beta}} & \text{si } x \leqslant \alpha \\ 1 - \dfrac{1}{2}\,\mathrm{e}^{-\frac{x-\alpha}{\beta}} & \text{si } x > \alpha \end{cases}$$

□

***4.*** *Espérance et variance.*

***a)*** On suppose que $X$ suit la loi $\mathcal{L}(0, 1)$.
Montrer que $\mathbb{E}(X)$ et $\mathbb{V}(X)$ existent et valent respectivement 0 et 2.

*Démonstration.*

- La v.a.r. $X$ admet une espérance si l'intégrale impropre $\displaystyle\int_{-\infty}^{+\infty} t\ h(t)\ dt$ est absolument convergente ce qui équivaut à démontrer sa convergence pour des calculs de moment de type $\displaystyle\int_{-\infty}^{+\infty} t^n\ h(t)\ dt$.
- Remarquons alors que la fonction $t \mapsto t\ h(t)$ est impaire puisque $h$ est paire :

$$\forall t \in \mathbb{R},\ h(-t) = \frac{1}{2}\,\mathrm{e}^{-|-t|} = \frac{1}{2}\,\mathrm{e}^{-|t|} = h(t)$$

- Ainsi, l'intégrale $\displaystyle\int_{-\infty}^{+\infty} t\ h(t)\ dt$ est convergente si l'intégrale impropre $\displaystyle\int_{0}^{+\infty} t\ h(t)\ dt$ l'est.

  En cas de convergence, à l'aide du changement de variable $\boxed{u = -t}$, on démontre :

$$\int_{0}^{+\infty} t\ h(t)\ dt = -\int_{-\infty}^{0} u\ h(u)\ du$$

  Et, dans ce cas, on peut alors conclure à l'aide de la relation de Chasles :

$$\int_{-\infty}^{+\infty} t\ h(t)\ dt = \int_{-\infty}^{0} t\ h(t)\ dt + \int_{0}^{+\infty} t\ h(t)\ dt = 0$$

- Il reste à démontrer que l'intégrale impropre $\displaystyle\int_{0}^{+\infty} t\ h(t)\ dt$ est convergente.

  Or : $\forall t \in [0, +\infty[,\ \mathrm{e}^{-|t|} = \mathrm{e}^{-t}$. Ainsi on reconnnaît, à une constante multiplicative près, l'espérance d'une variable aléatoire suivant la loi $\mathcal{E}(1)$.
  On en déduit que cette intégrale impropre est convergente.

La v.a.r. $X$ admet une espérance et $\mathbb{E}(X) = \displaystyle\int_{-\infty}^{+\infty} t\ h(t)\ dt = 0$.

- Pour les mêmes raisons que celles qui précèdent, la v.a.r. $X$ admet une variance si l'intégrale impropre $\displaystyle\int_{-\infty}^{+\infty} t^2\ h(t)\ dt$ est convergente.
- La fonction $t \mapsto t^2\ h(t)$ est paire.

  Ainsi, l'intégrale $\displaystyle\int_{-\infty}^{+\infty} t^2\ h(t)\ dt$ est convergente si l'intégrale impropre $\displaystyle\int_{0}^{+\infty} t^2\ h(t)\ dt$ l'est.

  Et, en cas de convergence :

$$\int_{-\infty}^{+\infty} t^2\ h(t)\ dt = 2\int_{0}^{+\infty} t^2\ h(t)\ dt = \int_{0}^{+\infty} 2\ t^2\ h(t)\ dt$$

- Or : $\forall t \in [0, +\infty[,\ 2\ t^2\ h(t) = t^2\ \mathrm{e}^{-|t|} = t^2\ \mathrm{e}^{-t}$.
  Ainsi on reconnnaît, le moment d'ordre 2 d'une variable aléatoire $Z$ telle que : $Z \hookrightarrow \mathcal{E}(1)$.

  On en déduit que l'intégrale impropre $\displaystyle\int_{0}^{+\infty} 2t^2\ h(t)\ dt$ est convergente.
- Déterminons alors $\mathbb{E}(Z^2)$. D'après la formule de Kœnig-Huygens :

$$\begin{array}{ccccc} \mathbb{V}(Z) & = & \mathbb{E}\big(Z^2\big) & - & \big(\mathbb{E}(Z)\big)^2 \\ \| & & & & \| \\ \dfrac{1}{1^2} & & & & \left(\dfrac{1}{1}\right)^2 \end{array}$$

  D'où : $\mathbb{E}\big(Z^2\big) = 1 + 1 = 2$.
- On en conclut que la v.a.r. $X$ admet un moment d'ordre 2 et :

$$\mathbb{E}\big(X^2\big) = \int_{-\infty}^{+\infty} t^2\ h(t)\ dt = \int_{0}^{+\infty} 2t^2\ h(t)\ dt = \mathbb{E}\big(Z^2\big) = 2$$

Enfin, par la formule de Kœnig-Huygens : $\mathbb{V}(X) = \mathbb{E}\big(X^2\big) - \cancel{\big(\mathbb{E}(X)\big)^2} = 2$.

**Commentaire**

- On rappelle que la plupart des résultats sur les intégrales généralisées (comme la relation de Chasles ou les changements de variable affines) exigent la convergence des intégrales étudiées.
- Dans cette question on s'est ramené à l'étude d'une v.a.r. $Z$ suivant une loi classique, ce qui permet de s'affranchir de longs calculs (on renvoie au cours pour ce qui est du calcul de $\mathbb{E}(Z)$ et $\mathbb{E}\big(Z^2\big)$). Évidemment, faire tous ces calculs n'est pas sanctionné le jour J. Mais la perte de temps qui en découle provoque, à terme, une perte de points. Il faut donc veiller à prendre du recul sur les questions posées.

□

***b)*** En déduire l'existence et les valeurs de l'espérance et de la variance d'une variable aléatoire réelle qui suit la loi $\mathcal{L}(\alpha, \beta)$.

*Démonstration.*

- Soit $X$ une v.a.r. telle que : $X \hookrightarrow \mathcal{L}(0, 1)$. D'après la question ***3.a)*** :

$$\beta X + \alpha \hookrightarrow \mathcal{L}(\alpha, \beta)$$

- La v.a.r. $Y = \beta X + \alpha$ admet une espérance et une variance comme transformée affine de la v.a.r. $X$ qui admet une espérance et une variance. De plus :
  - × par linéarité de l'espérance : $\mathbb{E}(Y) = \mathbb{E}(\beta X + \alpha) = \beta\ \mathbb{E}(X) + \alpha = \alpha$.
  - × par propriété de la variance : $\mathbb{V}(Y) = \mathbb{V}(\beta X + \alpha) = \beta^2\ \mathbb{V}(X) = 2\beta^2$.

$$\boxed{\mathbb{E}(Y) = \alpha \quad \text{et} \quad \mathbb{V}(Y) = 2\beta^2}$$

**Commentaire**

- Le résultat de la question précédente étant donné par l'énoncé, il n'était pas nécessaire de la traiter pour pouvoir résoudre cette question. C'est une remarque très générale : même si l'on ne parvient pas à traiter complètement une question, il faut s'atteler à la suivante et celle qui suit encore. Il faut alors être vigilant et bien récupérer toutes les informations de l'énoncé.
- La manière de procéder ici est à rapprocher de celle du cours lors de l'étude des lois normales. On démontre (résultat à connaître) que pour tout $(a, b) \in \mathbb{R}^* \times \mathbb{R}$ :

$$X \hookrightarrow \mathcal{N}\left(m, \sigma^2\right) \ \Leftrightarrow\ aX + b \hookrightarrow \mathcal{N}\left(am + b, a^2\sigma^2\right)$$

  En particulier : $X \hookrightarrow \mathcal{N}\left(m, \sigma^2\right) \ \Leftrightarrow\ X^* = \dfrac{X - m}{\sigma} \hookrightarrow \mathcal{N}(0, 1)$.
  Ainsi, on peut déterminer l'espérance d'une v.a.r. $X$ telle que $X \hookrightarrow \mathcal{N}\left(m, \sigma^2\right)$ à l'aide de l'espérance et de la variance de la loi $\mathcal{N}(0, 1)$ en remarquant :
  $X = \sigma X^* + m$.

□

***5.*** *Simulation à partir d'une loi exponentielle.*
Soit $U$ une variable aléatoire qui suit la loi exponentielle de paramètre 1 et $V$ une variable aléatoire qui suit la loi de Bernoulli de paramètre $\frac{1}{2}$ et indépendante de $U$.

***a)*** En utilisant le système complet naturellement associé à $V$, montrer que $X = (2V - 1)\ U$ suit la loi $\mathcal{L}(0, 1)$.

*Démonstration.*

- Soit $x \in \mathbb{R}$. Déterminons la fonction de répartition de $X$ :

$$F_X(x) \ = \ \mathbb{P}([X \leqslant x]) \ = \ \mathbb{P}([(2V - 1)\ U \leqslant x])$$

- La famille $([V = 0], [V = 1])$ est un système complet d'événements.
  Ainsi, d'après la formule des probablités totales :

$$\begin{aligned}
& \mathbb{P}([(2V - 1)\ U \leqslant x]) \\
= \ & \mathbb{P}([V = 0] \cap [(2V - 1)\ U \leqslant x]) + \mathbb{P}([V = 1] \cap [(2V - 1)\ U \leqslant x]) \\
= \ & \mathbb{P}([V = 0] \cap [-U \leqslant x]) + \mathbb{P}([V = 1] \cap [U \leqslant x]) \\
= \ & \mathbb{P}([V = 0]) \times \mathbb{P}([-U \leqslant x]) + \mathbb{P}([V = 1]) \times \mathbb{P}([U \leqslant x]) && \textit{(car } U \textit{ et } V \textit{ sont indépendantes)} \\
= \ & \frac{1}{2}\ \mathbb{P}([-U \leqslant x]) + \frac{1}{2}\ \mathbb{P}([U \leqslant x]) && \textit{(car } V \hookrightarrow \mathcal{B}\left(\tfrac{1}{2}\right)\textit{)} \\
= \ & \frac{1}{2}\ \mathbb{P}([U \geqslant -x]) + \frac{1}{2}\ F_U(x) \\
= \ & \frac{1}{2}\ (1 - \mathbb{P}([U < -x])) + \frac{1}{2}\ F_U(x) \ = \ \frac{1}{2}\ (1 - F_U(-x)) + \frac{1}{2}\ F_U(x)
\end{aligned}$$

- Deux cas se présentent alors.
  - × Si $x \leqslant 0$ alors $F_U(x) = 0$ et $F_U(-x) = 1 - \mathrm{e}^{-(-x)} = 1 - \mathrm{e}^{x}$. Dans ce cas :

$$F_X(x) = \frac{1}{2}\,(1 - F_U(-x)) + \frac{1}{2}\,F_U(x) = \frac{1}{2}\,(\cancel{1} - (\cancel{1} - \mathrm{e}^{x})) + \frac{1}{2}\,0 = \frac{1}{2}\,\mathrm{e}^{x}$$

  - × Si $x \geqslant 0$ alors $F_U(x) = 1 - \mathrm{e}^{-x}$ et $F_U(-x) = 0$. Dans ce cas :

$$F_X(x) = \frac{1}{2}\,(1 - F_U(-x)) + \frac{1}{2}\,F_U(x) = \frac{1}{2}\,(1 - 0) + \frac{1}{2}\,(1 - \mathrm{e}^{-x}) = 1 - \frac{1}{2}\,\mathrm{e}^{-x}$$

  On reconnaît la fonction de répartition $\Psi$ associée à la loi $\mathcal{L}(0,1)$.
  Or, la fonction de répartition caractérise la loi.

| On en conclut : $X = (2V-1)U \hookrightarrow \mathcal{L}(0,1)$. |
|---|

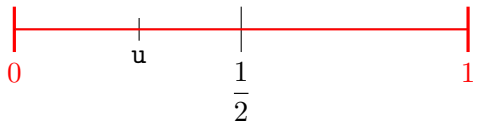
□

***b)*** Compléter la définition **Scilab** ci-dessous pour que la fonction ainsi définie réalise la simulation d'une variable aléatoire qui suit la loi $\mathcal{L}(\alpha, \beta)$ :

```
1  function r = Laplace(alpha, beta)
2      if --- <= 1/2
3          V = 1
4      else
5          V = 0
6      end
7      X = (2 * V - 1) * grand(1, 1, 'exp', 1)
8      r = ---
9  endfunction
```

*Démonstration.*

- La fonction `rand` permet de simuler une v.a.r. $U$ telle que $U \hookrightarrow \mathcal{U}([0,1])$. Afin de simuler une v.a.r. $V$ telle que $V \hookrightarrow \mathcal{B}\left(\frac{1}{2}\right)$ à l'aide de `rand`, on procède comme suit. L'appel `rand()` renvoie un réel `u` choisit aléatoirement dans $[0,1]$ :




  Le réel `u` appartient à l'intervalle $[0, \frac{1}{2}]$ avec probabilité :

$$\mathbb{P}\left(\left[U \in [0, \tfrac{1}{2}]\right]\right) = \mathbb{P}\left(\left[U \leqslant \frac{1}{2}\right]\right) = \frac{1}{2} = \mathbb{P}([V = 1])$$

  Le réel `u` appartient à l'intervalle $]\frac{1}{2}, 1]$ avec probabilité :

$$\mathbb{P}\left(\left[U \in ]\tfrac{1}{2}, 1]\right]\right) = \mathbb{P}\left(\left[U > \frac{1}{2}\right]\right) = \frac{1}{2} = \mathbb{P}([V = 0])$$

  C'est ce que réalise le programme demandé en complétant la ligne 2 comme suit :

```
2      if rand() <= 1/2
```

- L'instruction `grand(1, 1, 'exp', 1)` permet de simuler une v.a.r. $U$ telle que $U \hookrightarrow \mathcal{E}(1)$.
- D'après la question qui précède, l'instruction `X = (2 * V - 1) * grand(1, 1, 'exp', 1)` permet de simuler une v.a.r. $X$ telle que $X \hookrightarrow \mathcal{L}(0,1)$.
- Enfin, d'après la question ***3.a)***, la v.a.r. $\beta X + \alpha \hookrightarrow \mathcal{L}(\alpha, \beta)$.
  On en déduit la ligne 8 du programme à compléter :

```
8        r = beta * X + alpha
```

**Commentaire**

Afin de permettre une bonne compréhension des mécanismes en jeu dans cette question, on a détaillé sa réponse. Cependant, compléter correctement le programme **Scilab** démontre la bonne compréhension de la simulation demandée et permet certainement d'obtenir tous les points alloués à cette question.

□

## Partie II - Lois $\varepsilon$-différentielles

Soit $\varepsilon > 0$. On dit que $(X, Y)$, un couple de variables aléatoires, est un couple $\varepsilon$-différentiel si, pour tout intervalle $I$ de $\mathbb{R}$ :

$$e^{-\varepsilon}\ \mathbb{P}([X \in I]) \leqslant \mathbb{P}([Y \in I]) \leqslant e^{\varepsilon}\ \mathbb{P}([X \in I])$$

Intuitivement, les lois de $X$ et $Y$ seront d'autant plus proches que le plus petit $\varepsilon$ tel que $(X, Y)$ soit un couple $\varepsilon$-différentiel est proche de 0.

***6.*** Soit $(X, Y, Z)$ un triplet de variables aléatoires réelles.

***a)*** Montrer que si $(X, Y)$ est $\varepsilon$-différentiel alors $(Y, X)$ l'est aussi.

*Démonstration.*
Soit $I$ un intervalle de $\mathbb{R}$.

- Comme $(X, Y)$ est $\varepsilon$-différentiel :

$$e^{-\varepsilon}\ \mathbb{P}([X \in I]) \leqslant \mathbb{P}([Y \in I]) \leqslant e^{\varepsilon}\ \mathbb{P}([X \in I])$$

- En multipliant l'inégalité de gauche par $e^{\varepsilon} > 0$, on obtient :

$$\mathbb{P}([X \in I]) \leqslant e^{\varepsilon}\ \mathbb{P}([Y \in I])$$

- En multipliant l'inégalité de droite par $e^{-\varepsilon} > 0$, on obtient :

$$e^{-\varepsilon}\ \mathbb{P}([Y \in I]) \leqslant \mathbb{P}([X \in I])$$

- Ainsi, en combinant ces deux résultats :

$$e^{-\varepsilon}\ \mathbb{P}([Y \in I]) \leqslant \mathbb{P}([X \in I]) \leqslant e^{\varepsilon}\ \mathbb{P}([Y \in I])$$

Cette inégalité étant vérifiée pour tout $I \subset \mathbb{R}$, on en déduit que $(Y, X)$ est $\varepsilon$-différentiel. □

***b)*** Montrer que si $(X,Y)$ est $\varepsilon$-différentiel et $(Y,Z)$ est $\varepsilon'$-différentiel alors $(X,Z)$ est $(\varepsilon+\varepsilon')$-différentiel.

*Démonstration.*
Soit $I$ intervalle de $\mathbb{R}$.

- Tout d'abord, comme $(Y,Z)$ est $\varepsilon'$-différentiel :

$$\mathrm{e}^{-\varepsilon'}\ \mathbb{P}([Y\in I]) \ \leqslant\ \mathbb{P}([Z\in I]) \ \leqslant\ \mathrm{e}^{\varepsilon'}\ \mathbb{P}([Y\in I])$$

- Comme $(X,Y)$ est $\varepsilon$-différentiel : $\mathrm{e}^{-\varepsilon}\ \mathbb{P}([X\in I]) \leqslant \mathbb{P}([Y\in I])$. En multipliant par $\mathrm{e}^{-\varepsilon'}>0$ :

$$\mathrm{e}^{-\varepsilon}\ \mathrm{e}^{-\varepsilon'}\ \mathbb{P}([X\in I]) \ \leqslant\ \mathrm{e}^{-\varepsilon'}\ \mathbb{P}([Y\in I])$$

- Comme $(X,Y)$ est $\varepsilon$-différentiel : $\mathbb{P}([Y\in I]) \leqslant \mathrm{e}^{\varepsilon}\ \mathbb{P}([X\in I])$. En multipliant par $\mathrm{e}^{\varepsilon'}>0$ :

$$\mathrm{e}^{\varepsilon'}\ \mathbb{P}([Y\in I]) \ \leqslant\ \mathrm{e}^{\varepsilon'}\ \mathrm{e}^{\varepsilon}\ \mathbb{P}([X\in I])$$

- En combinant ces résultats, on obtient :

$$\mathrm{e}^{-(\varepsilon+\varepsilon')}\ \mathbb{P}([X\in I]) \ \leqslant\ \mathrm{e}^{-\varepsilon'}\ \mathbb{P}([Y\in I]) \ \leqslant\ \mathbb{P}([Z\in I]) \ \leqslant\ \mathrm{e}^{\varepsilon'}\ \mathbb{P}([Y\in I]) \ \leqslant\ \mathrm{e}^{\varepsilon+\varepsilon'}\ \mathbb{P}([X\in I])$$

Cette inégalité étant vérifiée pour tout $I\subset\mathbb{R}$, on en déduit que $(X,Z)$ est $(\varepsilon+\varepsilon')$-différentiel. □

***7.*** Soit $(X,Y)$ un couple de variables aléatoires réelles discrètes.
On suppose que $X(\Omega)\ \cup\ Y(\Omega)=\{z_n \mid n\in J\}$ où $J$ est un sous-ensemble non vide de $\mathbb{N}$.

Montrer que $(X,Y)$ est $\varepsilon$-différentiel si et seulement si

$$\forall n\in J,\ \mathrm{e}^{-\varepsilon}\ \mathbb{P}([X=z_n]) \ \leqslant\ \mathbb{P}([Y=z_n]) \ \leqslant\ \mathrm{e}^{\varepsilon}\ \mathbb{P}([X=z_n])$$

*Démonstration.*
On procède par double implication.

($\Rightarrow$) On suppose que $(X,Y)$ est $\varepsilon$-différentiel. Soit $n\in J$.
En notant $I=[z_n,z_n]\subset\mathbb{R}$, on obtient :

$$\begin{array}{ccccccc}
\mathrm{e}^{-\varepsilon} & \mathbb{P}([X\in I]) & \leqslant & \mathbb{P}([Y\in I]) & \leqslant & \mathrm{e}^{\varepsilon} & \mathbb{P}([X\in I]) \\
& \shortparallel & & \shortparallel & & & \shortparallel \\
& \mathbb{P}([X=z_n]) & & \mathbb{P}([Y=z_n]) & & & \mathbb{P}([X=z_n])
\end{array}$$

Ce qui démontre le résultat souhaité.

($\Leftarrow$) Soit $I$ un intervalle de $\mathbb{R}$. Alors :

$$[Y\in I] \ =\ \bigcup_{\substack{y\in Y(\Omega)\\ y\in I}} [Y=y] \ =\ \bigcup_{\substack{n\in J\\ z_n\in I}} [Y=z_n]$$

Par incompatibilité des événements de la réunion :

$$\mathbb{P}\left([Y\in I]\right) \ =\ \mathbb{P}\Big(\bigcup_{\substack{n\in J\\ z_n\in I}} [Y=z_n]\Big) \ =\ \sum_{\substack{n\in J\\ z_n\in I}} \mathbb{P}([Y=z_n])$$

*(ces égalités sont aussi vérifiées pour la v.a.r. $X$)*

Or, par hypothèse :

$$\forall n \in J,\ \mathrm{e}^{-\varepsilon}\ \mathbb{P}([X=z_n]) \leqslant \mathbb{P}([Y=z_n]) \leqslant \mathrm{e}^{\varepsilon}\ \mathbb{P}([X=z_n])$$

En sommant ces inégalités membre à membre, on obtient :

$$\begin{array}{ccccccc} \mathrm{e}^{-\varepsilon} & \displaystyle\sum_{\substack{n\in J\\ z_n\in I}} \mathbb{P}([X=z_n]) & \leqslant & \displaystyle\sum_{\substack{n\in J\\ z_n\in I}} \mathbb{P}([Y=z_n]) & \leqslant & \mathrm{e}^{\varepsilon} & \displaystyle\sum_{\substack{n\in J\\ z_n\in I}} \mathbb{P}([X=z_n]) \\ & \| & & \| & & & \| \\ & \mathbb{P}([X=I]) & & \mathbb{P}([Y=I]) & & & \mathbb{P}([X=I]) \end{array}$$

On a donc bien démontré que pour tout $I$ intervalle de $\mathbb{R}$ :

$$\mathrm{e}^{-\varepsilon}\ \mathbb{P}([X=I]) \leqslant \mathbb{P}([Y=I]) \leqslant \mathrm{e}^{\varepsilon}\ \mathbb{P}([X=I])$$

Autrement dit, $(X,Y)$ est bien $\varepsilon$-différentiel.

| $(X,Y)$ est $\varepsilon$-différentiel ssi $\forall n \in J,\ \mathrm{e}^{-\varepsilon}\ \mathbb{P}([X=z_n]) \leqslant \mathbb{P}([Y=z_n]) \leqslant \mathrm{e}^{\varepsilon}\ \mathbb{P}([X=z_n])$. |
|---|

**Commentaire**

- Pour pouvoir traiter cette question, il faut avoir compris que le caractère $\varepsilon$-différentiel est une propriété qui doit être démontrée pour tout intervalle $I$ de $\mathbb{R}$. Cette question revient à démontrer que, dans le cas où $X$ et $Y$ sont des v.a.r. discrètes, le caractère $\varepsilon$-différentiel est une propriété qui doit être démontrée seulement pour les intervalles $I$ du type $I=[z_n,z_n]$ (ce qui permet d'écrire $[Y\in I]=[Y=z_n]$).
- Le sens direct est très abordable : si la propriété est vraie pour tout intervalle $I$, elle l'est en particulier pour des intervalles du type $I=[z_n,z_n]$.
- Le sens réciproque est bien plus complexe. En faisant l'hypothèse que la propriété est vraie pour des intervalles d'un type particulier (ceux qui s'écrivent $I=[z_n,z_n]$), on doit en déduire la propriété pour tout intervalle $I$ de $\mathbb{R}$.

□

***8.*** *Premier exemple.*
Dans cette question, on suppose que $X$ suit la loi géométrique de paramètre $\frac{1}{2}$, $Z$ suit la loi de Bernoulli de paramètre $p\in\,]0,1[$ et elles sont indépendantes. On pose $Y=X+Z$.

***a)*** Déterminer la loi de $Y$.

*Démonstration.*

- Tout d'abord, $X(\Omega)=\mathbb{N}^*$ et $Z(\Omega)=\{0,1\}$.

| Ainsi : $Y(\Omega)\subset\mathbb{N}^*$ |
|---|

- Soit $k\in\mathbb{N}^*$.
  La famille $([Z=0],[Z=1])$ est un système complet d'événements.
  D'après la formule des probabilités totatles :

$$\begin{array}{rcll} \mathbb{P}([Y=k]) &=& \mathbb{P}([X+Z=k]) & \\ &=& \mathbb{P}([Z=0]\cap[X+Z=k]) + \mathbb{P}([Z=1]\cap[X+Z=k]) & \\ &=& \mathbb{P}([Z=0]\cap[X=k]) + \mathbb{P}([Z=1]\cap[X+1=k]) & \\ &=& \mathbb{P}([Z=0])\times\mathbb{P}([X=k]) + \mathbb{P}([Z=1])\times\mathbb{P}([X=k-1]) & \textit{(car } X \textit{ et } Z \\ & & & \textit{sont indépendantes)} \\ &=& (1-p)\ \mathbb{P}([X=k]) + p\ \mathbb{P}([X=k-1]) & \end{array}$$

Deux cas se présentent.

× Si $k = 1$ alors $[X = k-1] = [X = 0] = \varnothing$.

$$\mathbb{P}([Y=1]) = (1-p)\,\mathbb{P}([X=1]) + p\,\cancel{\mathbb{P}([X=0])} = \frac{1-p}{2}$$

× Si $k \geqslant 2$ :

$$\begin{aligned}\mathbb{P}([Y=k]) &= (1-p)\,\mathbb{P}([X=k]) + p\,\mathbb{P}([X=k-1]) \\ &= (1-p)\left(\frac{1}{2}\right)^{k-1}\frac{1}{2} + p\left(\frac{1}{2}\right)^{k-2}\frac{1}{2} \\ &= \left(\frac{1}{2}\right)^{k}((1-p)+2p) = \left(\frac{1}{2}\right)^{k}(1+p)\end{aligned}$$

En résumé : $\mathbb{P}([Y=k]) = \begin{cases} \dfrac{1-p}{2} & \text{si } k=1 \\ \left(\dfrac{1}{2}\right)^{k}(1+p) & \text{si } k \geqslant 2 \end{cases}$.

□

***b)*** Établir que pour tout $k \in \mathbb{N}^*$, $1-p \leqslant \dfrac{\mathbb{P}([Y=k])}{\mathbb{P}([X=k])} \leqslant \dfrac{1}{1-p}$.

*Démonstration.*

Soit $k \in \mathbb{N}^*$. Deux cas se présentent.

- Si $k = 1$

$$\frac{\mathbb{P}([Y=1])}{\mathbb{P}([X=1])} = \frac{\frac{1-p}{2}}{\frac{1}{2}} = 1-p$$

Dans ce cas, on a bien : $1-p \leqslant \dfrac{\mathbb{P}([Y=1])}{\mathbb{P}([X=1])}$.

On raisonne par équivalence pour la deuxième inégalité. Comme $1-p \in ]0,1[$ :

$$1-p \leqslant \frac{1}{1-p} \Leftrightarrow (1-p)^2 \leqslant 1$$

Cette dernière inégalité est vérifiée car $1-p \in ]0,1[$. Il en est donc de même de la première.

$$\boxed{1-p \leqslant \frac{\mathbb{P}([Y=1])}{\mathbb{P}([X=1])} \leqslant \frac{1}{1-p}}$$

- Si $k \geqslant 2$

$$\frac{\mathbb{P}([Y=k])}{\mathbb{P}([X=k])} = \frac{\cancel{\left(\frac{1}{2}\right)^{k}}(1+p)}{\cancel{\left(\frac{1}{2}\right)^{k-1}}\frac{1}{2}} = 1+p$$

Raisonnons par équivalence :

$$1-p \leqslant 1+p \Leftrightarrow 0 \leqslant 2p \Leftrightarrow 0 \leqslant p$$

Cette dernière inégalité est vérifiée. Il en est donc de même de la première.

De même :

$$1+p \leqslant \frac{1}{1-p} \Leftrightarrow 1-p^2 \leqslant 1 \Leftrightarrow 0 \leqslant p^2$$

Cette dernière inégalité est vérifiée. Il en est donc de même de la première.

Ainsi, pour tout $k \geqslant 2$, $1-p \leqslant \dfrac{\mathbb{P}([Y=k])}{\mathbb{P}([X=k])} \leqslant \dfrac{1}{1-p}$.

□

***c)*** En déduire que $(X,Y)$ est $-\ln(1-p)$-différentiel.

*Démonstration.*

- Rappelons tout d'abord : $X(\Omega)=\mathbb{N}^*$ et $Y(\Omega)\subset\mathbb{N}^*$. Ainsi, $X(\Omega)\cup Y(\Omega)=\mathbb{N}^*$.
- D'après la question ***7.***, pour démontrer que $(X,Y)$ est $-\ln(1-p)$-différentiel, on doit vérifier :

$$\forall k\in\mathbb{N}^*,\quad \mathrm{e}^{-(-\ln(1-p))}\,\mathbb{P}([X=k])\leqslant\mathbb{P}([Y=k])\leqslant\mathrm{e}^{-\ln(1-p)}\,\mathbb{P}([X=k])$$

Ce qui équivaut, en divisant de part et d'autre par $\mathbb{P}([X=k])>0$, à :

$$\forall k\in\mathbb{N}^*,\quad \mathrm{e}^{-(-\ln(1-p))}\leqslant\frac{\mathbb{P}([Y=k])}{\mathbb{P}([X=k])}\leqslant\mathrm{e}^{-\ln(1-p)}$$

- Or :

$$\mathrm{e}^{-(-\ln(1-p))}=\mathrm{e}^{\ln(1-p)}=1-p\quad\text{et}\quad\mathrm{e}^{-\ln(1-p)}=\mathrm{e}^{\ln\left((1-p)^{-1}\right)}=(1-p)^{-1}=\frac{1}{1-p}$$

Ainsi, le couple $(X,Y)$ est $-\ln(1-p)$-différentiel si l'inégalité de la question précédente est vérifiée.

| Le couple $(X,Y)$ est bien $-\ln(1-p)$-différentiel. |
|---|

□

***d)*** Que ce passe-t-il lorsque $p$ s'approche de 0 ou lorsqu'il s'approche de 1 ? Était-ce prévisible ?

*Démonstration.*

- Lorsque $p$ s'approche de 0, $-\ln(1-p)$ s'approche de $-\ln(1)=0$. Or les lois de $X$ et $Y$ sont d'autant plus proches que le plus petit $\varepsilon$ rendant $(X,Y)$ $\varepsilon$-différentiel est proche de 0.

| Les lois de $X$ et $Y$ sont donc d'autant plus proches que $p$ s'approche de 0. |
|---|

On pouvait s'attendre à ce résultat puisque, par définition : $Y=X+Z$ avec $Z\hookrightarrow\mathcal{B}(p)$.
Ainsi, si $p$ s'approche de 0, alors la probabilité $\mathbb{P}([X=Y])=\mathbb{P}([Z=0])=1-p$ se rapproche de 1.

- Lorsque $p$ s'approche de 1, $-\ln(1-p)$ s'approche de $+\infty$.
L'inégalité obtenue en ***7.b)*** fournit une faible information : $\dfrac{\mathbb{P}([Y=k])}{\mathbb{P}([X=k])}\in[0,+\infty[$.
Cela provient essentiellement du cas $k=1$ pour lequel : $\dfrac{\mathbb{P}([Y=1])}{\mathbb{P}([X=1])}=1-p$.
Le réel $\varepsilon=-\ln(1-p)$ est alors le meilleur choix (celui qui réalise l'égalité) pour assurer :

$$\mathrm{e}^{-\varepsilon}\leqslant\frac{\mathbb{P}([Y=1])}{\mathbb{P}([X=1])}$$

On aurait pu s'attendre à cette difficulté amenée par le cas $k=1$. En effet, si $p$ s'approche de 0, alors la probabilité $\mathbb{P}([X=Y])=\mathbb{P}([Z=0])=1-p$ se rapproche de 0. □

***9.*** On suppose que $X$ et $Y$ sont deux variables à densité de densités respectives $f$ et $g$ et de fonction de répartition $F$ et $G$.

***a)*** On suppose que pour tout $t\in\mathbb{R}$ : $\mathrm{e}^{-\varepsilon}f(t)\leqslant g(t)\leqslant\mathrm{e}^{\varepsilon}f(t)$.
Montrer que $(X,Y)$ est $\varepsilon$-différentiel.

*Démonstration.*

- Soit $I$ un intervalle de $\mathbb{R}$.
On note $a$ la borne inférieure de cet intervalle et $b$ la borne supérieure (on a notamment $a\leqslant b$).
Afin de couvrir tous les cas possibles, $a$ et $b$ sont des éléments de $\overline{\mathbb{R}}=\mathbb{R}\cup\{-\infty,+\infty\}$.

- Par hypothèse :
$$\forall t \in \mathbb{R},\ \mathrm{e}^{-\varepsilon} f(t) \ \leqslant\ g(t) \ \leqslant\ \mathrm{e}^{\varepsilon} f(t)$$
Par croissance de l'intégrale, les bornes étant dans l'ordre croissant ($a \leqslant b$) :
$$\begin{array}{ccccc} \displaystyle\int_a^b \mathrm{e}^{-\varepsilon} f(t)\ dt & \leqslant & \displaystyle\int_a^b g(t)\ dt & \leqslant & \displaystyle\int_a^b \mathrm{e}^{\varepsilon} f(t)\ dt \\ || & & || & & || \\ \mathrm{e}^{-\varepsilon}\ \mathbb{P}([X \in I]) & = & \mathrm{e}^{-\varepsilon} \displaystyle\int_a^b f(t)\ dt \qquad \mathbb{P}([Y \in I]) & & \mathrm{e}^{\varepsilon} \displaystyle\int_a^b f(t)\ dt \ = \ \mathrm{e}^{\varepsilon}\ \mathbb{P}([X \in I]) \end{array}$$
Ainsi, $(X, Y)$ est $\varepsilon$-différentiel. □

***b)*** On suppose dans la suite de cette question que $(X, Y)$ est $\varepsilon$-différentiel.
Soit $h > 0$ et $t \in \mathbb{R}$ où $f$ et $g$ sont continues.

Montrer que :
$$\mathrm{e}^{-\varepsilon}\ \frac{F(t+h) - F(t)}{h} \ \leqslant\ \frac{G(t+h) - G(t)}{h} \ \leqslant\ \mathrm{e}^{\varepsilon}\ \frac{F(t+h) - F(t)}{h}$$
En conclure que : $\mathrm{e}^{-\varepsilon} f(t) \ \leqslant\ g(t) \ \leqslant\ \mathrm{e}^{\varepsilon}\ f(t)$.

*Démonstration.*

- Comme $X$ est une variable à densité :
$$F(t+h) - F(t) \ =\ \mathbb{P}([t \leqslant X \leqslant t+h]) \ =\ \int_t^{t+h} f(t)\ dt \qquad (\star)$$
*(on obtient une égalité analogue pour la v.a.r. $Y$)*
- Comme $(X, Y)$ est $\varepsilon$-différentiel, avec $I = [t, t+h]$, on obtient :
$$\mathrm{e}^{-\varepsilon}\mathbb{P}([t \leqslant X \leqslant t+h]) \leqslant \mathbb{P}([t \leqslant Y \leqslant t+h]) \leqslant \mathrm{e}^{\varepsilon}\mathbb{P}([t \leqslant X \leqslant t+h])$$
Soit :
$$\mathrm{e}^{-\varepsilon}\ \int_t^{t+h} f(t)\ dt \ \leqslant\ \int_t^{t+h} g(t)\ dt \ \leqslant\ \mathrm{e}^{\varepsilon}\ \int_t^{t+h} f(t)\ dt$$
En multipliant de part et d'autre par $\dfrac{1}{h} > 0$ et en appliquant l'égalité $(\star)$ :
$$\mathrm{e}^{-\varepsilon}\ \frac{F(t+h) - F(t)}{h} \ \leqslant\ \frac{G(t+h) - G(t)}{h} \ \leqslant\ \mathrm{e}^{\varepsilon}\ \frac{F(t+h) - F(t)}{h}$$
- La fonction $f$ étant continue en $t$, $F$ est dérivable en $t$.
On en déduit que le taux d'accroissement $\dfrac{F(t+h) - F(t)}{h}$ admet une limite finie lorsque $h \to 0$. Plus précisément :
$$\lim_{h \to 0} \frac{F(t+h) - F(t)}{h} \ =\ F'(t) \ =\ f(t)$$
*(on obtient une formule analogue pour les fonctions $G$ et $g$)*

Ainsi, par passage à la limite dans l'encadrement précédent :

$$\begin{array}{ccccc} \mathrm{e}^{-\varepsilon}\ \dfrac{F(t+h)-F(t)}{h} & \leqslant & \dfrac{G(t+h)-G(t)}{h} & \leqslant & \mathrm{e}^{\varepsilon}\ \dfrac{F(t+h)-F(t)}{h} \\ \Big\downarrow{\scriptstyle h\to 0} & & \Big\downarrow{\scriptstyle h\to 0} & & \Big\downarrow{\scriptstyle h\to 0} \\ \mathrm{e}^{-\varepsilon}\ f(t) & \leqslant & g(t) & \leqslant & \mathrm{e}^{\varepsilon}\ f(t) \end{array}$$

$$\boxed{\mathrm{e}^{-\varepsilon}\ f(t) \leqslant g(t) \leqslant \mathrm{e}^{\varepsilon}\ f(t)}$$

□

***10.*** *Deuxième exemple : lois de Cauchy.*

***a)*** Montrer que $\displaystyle\int_{-\infty}^{+\infty} \frac{1}{t^2+1}\,dt$ converge. On admet que cette intégrale est égale à $\pi$.

*Démonstration.*

- L'intégrale impropre $\displaystyle\int_{-\infty}^{+\infty} \frac{1}{t^2+1}\,dt$ est convergente si les intégrales $\displaystyle\int_{-\infty}^{0} \frac{1}{t^2+1}\,dt$ et $\displaystyle\int_{0}^{+\infty} \frac{1}{t^2+1}\,dt$ le sont.
- La fonction $f : t \mapsto \dfrac{1}{1+t^2}$ est continue sur $[0, +\infty[$.
  - × $f(t) = \dfrac{1}{1+t^2} \underset{t\to+\infty}{\sim} \dfrac{1}{t^2}$
  - × $\forall t \in [1, +\infty[,\ \dfrac{1}{1+t^2} \geqslant 0$ et $\dfrac{1}{t^2} \geqslant 0$
  - × L'intégrale $\displaystyle\int_{1}^{+\infty} \frac{1}{t^2}\,dt$ est convergente en tant qu'intégrale de Riemann impropre en $+\infty$, d'exposant 2 ($2 > 1$).

  Ainsi, par critère de convergence des intégrales généralisées de fonctions continues positives, l'intégrale $\displaystyle\int_{1}^{+\infty} \frac{1}{1+t^2}\,dt$ est convergente.

  De plus, comme la fonction $f$ est continue sur $[0, 1]$, l'intégrale $\displaystyle\int_{0}^{1} f(t)\,dt$ est bien définie.

  $$\boxed{\text{L'intégrale } \int_{0}^{+\infty} \frac{1}{1+t^2}\,dt \text{ est convergente.}}$$

- La fonction $f$ est paire sur $\mathbb{R}$. Ainsi, à l'aide du changement de variable $\boxed{u = -t}$ :

  $$\int_{-\infty}^{0} f(t)\,dt = \int_{0}^{+\infty} f(u)\,du$$

  L'intégrale impropre $\displaystyle\int_{-\infty}^{0} f(t)\,dt$ est donc elle aussi convergente.

  $$\boxed{\text{L'intégrale impropre } \int_{-\infty}^{+\infty} \frac{1}{1+t^2}\,dt \text{ est convergente.}}$$

□

***b)*** On définit, pour $a > 0$, la fonction $f_a$ sur $\mathbb{R}$ par, pour tout $t \in \mathbb{R}$, $f_a(t) = \dfrac{a}{\pi\ (t^2 + a^2)}$.
Montrer que $f_a$ est une densité de probabilité d'une variable aléatoire à densité.

*Démonstration.*

- La fonction $f_a$ est continue sur $\mathbb{R}$.
- $\forall t \in \mathbb{R},\ f_a(t) \geqslant 0$.
- Sous réserve de convergence, effectuons le changement de variable $\boxed{u = \frac{1}{a}\ t}$.

$$\left|\begin{array}{l} u = \frac{1}{a}\ t \quad (\text{et donc } t = au) \\ \hookrightarrow du = \frac{1}{a}\ dt \quad \text{et} \quad dt = a\ du \\ \bullet\ t = -\infty \ \Rightarrow\ u = -\infty \\ \bullet\ t = +\infty \ \Rightarrow\ u = +\infty \end{array}\right.$$

Ce changement de variable est valide car $\varphi : u \mapsto au$ est de classe $\mathcal{C}^1$ sur $]-\infty, +\infty[$.
On obtient alors :

$$\int_{-\infty}^{+\infty} \frac{a}{\pi\ (t^2 + a^2)}\ dt \ =\ \int_{-\infty}^{+\infty} \frac{a}{\pi\ ((au)^2 + a^2)}\ (a\ du) \ =\ \int_{-\infty}^{+\infty} \frac{\cancel{a^2}}{\pi\ \cancel{a^2}\ (u^2+1)}\ du$$

On reconnaît, à une constante multiplicative près l'intégrale impropre de la question précédente. Ainsi, $\displaystyle\int_{-\infty}^{+\infty} f_a(t)\ dt$ est convergente et :

$$\int_{-\infty}^{+\infty} f_a(t)\ dt \ =\ \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1}{(u^2+1)}\ du \ =\ \frac{1}{\pi}\ \pi \ =\ 1$$

On en déduit que $f_a$ peut être considérée comme une densité de probabilité. □

***c)*** On suppose que $X$ et $Y$ sont deux variables aléatoires admettant comme densités respectives $f_1$ et $f_a$ avec $a > 1$.
Montrer que $(X, Y)$ est $\ln(a)$-différentiel.

*Démonstration.*

- D'après la question ***9.a)***, pour démontrer que $(X, Y)$ est $\ln(a)$-différentiel, il suffit de démontrer :
$$\forall t \in \mathbb{R},\ \mathrm{e}^{-\ln(a)}\ f_1(t) \ \leqslant\ f_a(t) \ \leqslant\ \mathrm{e}^{\ln(a)}\ f_1(t)$$
avec $\mathrm{e}^{-\ln(a)} = \mathrm{e}^{\ln(a^{-1})} = \dfrac{1}{a}$ et $\mathrm{e}^{\ln(a)} = a$.
- Soit $t \in \mathbb{R}$. Raisonnons par équivalence pour démontrer l'inégalité de gauche :

$$\begin{aligned} \frac{1}{a}\ f_1(t) \ \leqslant\ f_a(t) \ &\Leftrightarrow\ \frac{1}{a}\ \frac{1}{\pi\ (1+t^2)} \ \leqslant\ \frac{a}{\pi\ (a^2+t^2)} \\ &\Leftrightarrow\ \cancel{a^2} + t^2 \ \leqslant\ a^2\ (\cancel{1} + t^2) \qquad \textit{(en multipliant par } a\pi(1+t^2)(a^2+t^2) > 0\textit{)} \\ &\Leftrightarrow\ t^2 \ \leqslant\ a^2\ t^2 \\ &\Leftrightarrow\ 0 \ \leqslant\ (a^2-1)\ t^2 \end{aligned}$$

Or, comme $a > 1$ alors $a^2 > 1$ et ainsi : $(a^2-1)\ t^2 \geqslant 0$. On en déduit : $\mathrm{e}^{-\ln(a)}\ f_1(t) \ \leqslant\ f_a(t)$.

- Démontrons de même l'inégalité de droite :

$$\begin{aligned} f_a(t) \leqslant a\, f_1(t) &\Leftrightarrow \frac{a}{\pi\,(a^2+t^2)} \leqslant a\,\frac{1}{\pi\,(1+t^2)} \\ &\Leftrightarrow 1+t^2 \leqslant a^2+t^2 && \text{(en multipliant par } \tfrac{1}{a}\,\pi(1+t^2)(a^2+t^2)>0\text{)} \\ &\Leftrightarrow 1 \leqslant a^2 \end{aligned}$$

Cette dernière inégalité étant vérifiée, on en déduit : $f_a(t) \leqslant a\, f_1(t) = \mathrm{e}^{\ln(a)}\, f_1(t)$.

L'inégalité souhaitée étant démontrée, on en conclut que $(X,Y)$ est $\ln(a)$-différentiel. □

***11.*** *Une première interprétation.*
On suppose que $(X,Y)$ est un couple $\varepsilon$-différentiel et que $U$ est une variable de Bernoulli de paramètre $p \in ]0,1[$ indépendante de $X$ et $Y$.
On définit la variable aléatoire $Z$ par :

$$\forall \omega \in \Omega,\ Z(\omega) = \begin{cases} X(\omega) & \text{si } U(\omega)=1 \\ Y(\omega) & \text{sinon.} \end{cases}$$

***a)*** Soit $I$ un intervalle de $\mathbb{R}$ telle que $\mathbb{P}([Z\in I]) \neq 0$.

Montrer que :

$$\mathbb{P}_{[Z\in I]}([U=1]) = p\,\frac{\mathbb{P}([X\in I])}{p\,\mathbb{P}([X\in I]) + (1-p)\,\mathbb{P}([Y\in I])}$$

En déduire que :

$$\frac{p}{p+(1-p)\,\mathrm{e}^{\varepsilon}} \leqslant \mathbb{P}_{[Z\in I]}([U=1]) \leqslant \frac{p}{p+(1-p)\,\mathrm{e}^{-\varepsilon}}$$

*Démonstration.*

- Comme $\mathbb{P}([Z\in I]) \neq 0$, par définition des probabilités conditionnelles :

$$\mathbb{P}_{[Z\in I]}([U=1]) = \frac{\mathbb{P}([Z\in I]\cap[U=1])}{\mathbb{P}([Z\in I])}$$

- Par définition de la v.a.r. $Z$ :

$$\begin{aligned} \mathbb{P}([Z\in I]\cap[U=1]) &= \mathbb{P}([X\in I]\cap[U=1]) \\ &= \mathbb{P}([U=1])\,\mathbb{P}([X\in I]) && \text{(car } X \text{ et } U \text{ sont indépendantes)} \\ &= p\,\mathbb{P}([X\in I]) \end{aligned}$$

- La famille $([U=1],[U=0])$ est un système complet d'événements.
Ainsi, par la formule des probabilités totales :

$$\begin{aligned} \mathbb{P}([Z\in I]) &= \mathbb{P}([U=1]\cap[Z\in I]) + \mathbb{P}([U=0]\cap[Z\in I]) \\ &= \mathbb{P}([U=1]\cap[X\in I]) + \mathbb{P}([U=0]\cap[Y\in I]) && \text{(par définition de } Y\text{)} \\ &= \mathbb{P}([U=1])\,\mathbb{P}([X\in I]) + \mathbb{P}([U=0])\,\mathbb{P}([Y\in I]) && \text{(car } U \text{ est indépendante de } X \text{ et } Y\text{)} \\ &= p\,\mathbb{P}([X\in I]) + (1-p)\,\mathbb{P}([Y\in I]) \end{aligned}$$

En combinant ces résultats : $\mathbb{P}_{[Z\in I]}([U=1]) = p\,\dfrac{\mathbb{P}([X\in I])}{p\,\mathbb{P}([X\in I]) + (1-p)\,\mathbb{P}([Y\in I])}$.

- Comme $(X,Y)$ est supposé $\varepsilon$-différentiel, $\mathrm{e}^{-\varepsilon}\ \mathbb{P}([X\in I])\leqslant\mathbb{P}([Y\in I])$. Ainsi :

$$\begin{aligned} p\ \mathbb{P}([X\in I])+(1-p)\ \mathbb{P}([Y\in I]) &\geqslant p\ \mathbb{P}([X\in I])+(1-p)\ \mathrm{e}^{-\varepsilon}\ \mathbb{P}([X\in I]) \\ &= \\ &(p+(1-p)\ \mathrm{e}^{-\varepsilon})\ \mathbb{P}([X\in I]) \end{aligned}$$

  Ainsi, par décroissance de la fonction inverse sur $]0,+\infty[$ :

$$\frac{1}{p\ \mathbb{P}([X\in I])+(1-p)\ \mathbb{P}([Y\in I])}\leqslant\frac{1}{(p+(1-p)\ \mathrm{e}^{-\varepsilon})\ \mathbb{P}([X\in I])}$$

  Par multiplication par $p\ \mathbb{P}([X\in I])\geqslant 0$, on obtient :

$$\mathbb{P}_{[Z\in I]}([U=1])=\frac{p\ \mathbb{P}([X\in I])}{p\ \mathbb{P}([X\in I])+(1-p)\ \mathbb{P}([Y\in I])}\leqslant\frac{p\ \cancel{\mathbb{P}([X\in I])}}{(p+(1-p)\ \mathrm{e}^{-\varepsilon})\ \cancel{\mathbb{P}([X\in I])}}$$

- En raisonnant de même, comme $\mathbb{P}([Y\in I])\leqslant\mathrm{e}^{\varepsilon}\ \mathbb{P}([X\in I])$ :

$$p\ \mathbb{P}([X\in I])+(1-p)\ \mathbb{P}([Y\in I])\leqslant(p+(1-p)\ \mathrm{e}^{\varepsilon})\ \mathbb{P}([X\in I])$$

  puis :

$$\mathbb{P}_{[Z\in I]}([U=1])=\frac{p\ \mathbb{P}([X\in I])}{p\ \mathbb{P}([X\in I])+(1-p)\ \mathbb{P}([Y\in I])}\geqslant\frac{p\ \cancel{\mathbb{P}([X\in I])}}{(p+(1-p)\ \mathrm{e}^{\varepsilon})\ \cancel{\mathbb{P}([X\in I])}}$$

$$\boxed{\frac{p}{p+(1-p)\ \mathrm{e}^{\varepsilon}}\leqslant\mathbb{P}_{[Z\in I]}([U=1])\leqslant\frac{p}{p+(1-p)\ \mathrm{e}^{-\varepsilon}}}$$

□

***b)*** Si $\varepsilon$ est proche de zéro, le fait de disposer d'une information sur la valeur de $Z$ change-t-il notablement le paramètre de la loi de $U$ et par conséquent la probabilité d'en déduire la valeur prise par $U$ ?

*Démonstration.*

- Si $\varepsilon$ est proche de 0 alors $\mathrm{e}^{\varepsilon}$ et $\mathrm{e}^{-\varepsilon}$ sont proches de 1.
  On peut alors faire les approximations : $p+(1-p)\ \mathrm{e}^{\varepsilon}\simeq 1$ et $p+(1-p)\ \mathrm{e}^{-\varepsilon}\simeq 1$.
- On obtient alors, à l'aide de l'inégalité précédente :

$$p\simeq\frac{p}{p+(1-p)\ \mathrm{e}^{\varepsilon}}\leqslant\mathbb{P}_{[Z\in I]}([U=1])\leqslant\frac{p}{p+(1-p)\ \mathrm{e}^{-\varepsilon}}\simeq p$$

  Ainsi, $\mathbb{P}_{[Z\in I]}([U=1])$ est proche de $p$.
  Or, par définition de $U$ : $\mathbb{P}([U=1])=p$.

On en conclut que disposer d'une information sur la valeur de $Z$ (savoir que $[Z\in I]$ est réalisé) ne permet pas de déduire la valeur prise par $U$. □

## Partie III - Confidentialité différentielle

- Soit $d \in \mathbb{N}^*$. On considère $D = [\![0, d]\!]$ et $n$ un entier naturel plus grand que 2.
- On dira que deux éléments de $D^n$, $a$ et $b$, sont voisins s'ils ne différent que d'une composante au plus. On note $\mathcal{V}$ l'ensemble des couples de voisins.
- On considère $q$ une application de $D^n$ dans $\mathbb{R}$.

Concrètement, un élément de $D^n$ représente une table d'une base de données et $q$ une requête sur cette base. Étant donné $a = (a_1, \ldots, a_n)$, on s'intéresse au problème de la confidentialité de certains des $a_i$ lorsque les autres $a_i$ sont connus, ainsi que $D$, $q$ et $q(a)$.

***12.*** Dans cette question on suppose que $a_2, \ldots, a_n$ sont connus et on cherche à protéger $a_1$.

***a)*** Quelle est probabilité d'obtenir la bonne valeur de $a_1$ si l'on choisit une valeur au hasard dans $[\![0, d]\!]$ ?

*Démonstration.*
L'entier $a_1$ est un élément de $[\![0, d]\!]$, ensemble de cardinal $d + 1$.
Un choix au hasard dans $D$ correspond à effectuer un tirage uniforme dans $D$.

La probabilité d'obtenir la bonne valeur de $a_1$ en choisissant une valeur au hasard dans $D$ est donc de $\dfrac{1}{d+1}$.

**Commentaire**

- On peut s'interroger sur la pertinence de l'interprétation du problème présentée dans l'énoncé. Tout d'abord, il est fait mention de « base de données », de « requête » et de « table ». Ces notions sont absentes du programme de voie ECE / ECS (elles sont étudiées en prépa scientifique). Précisons ces termes :
  - × **une base de données** est un moyen de stocker des informations appelées données. Par exemple, on peut penser à une enseigne de grande distribution qui récolterait des informations via les cartes de fidélité.
  - × **une table** est un regroupement d'informations du même type.
    En reprenant l'exemple précédent, on peut penser à une table qui regroupe les informations sur les clients : c'est un tableau dont chaque ligne (appelée **enregistrement**) contient le nom, le prénom, l'adresse, le numéro de téléphone, la date de naissance et le nom du magasin fréquenté par le porteur de carte.
    Une telle base de données contiendrait aussi une table sur les informations des magasins : un enregistrement d'une telle table pourrait contenir le nom du magasin, son adresse, ses horaires d'ouverture.
  - × **une requête** est une interrogation de la base de données. Par exemple, si un magasin particulier organise un événement promotionnel, l'enseigne souhaitera recueillir tous les numéros de téléphone des clients fréquentant ce magasin afin de pouvoir communiquer par sms sur cet événement.
- Revenons à l'énoncé. On considère que les données apparaissent sous forme de nombres. Un élément $(a_1, \ldots, a_n) \in D^n$ serait alors un enregistrement (plutôt qu'une table) et une table serait un sous ensemble de $D^n$. Une fonction $q : D^n \to \mathbb{R}$ peut être considérée comme permettant la sélection de certains enregistrements ($q$ peut par exemple être à valeur dans $\{0, 1\}$ et on sélectionne tous les enregistrements $a \in D^n$ pour lesquels $q(a) = 1$).

□

***b)*** Dans cette question $q(a_1, \ldots, a_n) = \sum_{i=1}^{n} a_i$.

Montrer que si $q(a)$ est publique alors on sait déterminer la valeur de $a_1$.

*Démonstration.*

- On rappelle que l'on suppose connus $a_2$, …, $a_n$. Ainsi, $\sum_{i=2}^{n} a_i$ est connu.
- On suppose en plus dans cette question que $q(a) = \sum_{i=1}^{n} a_i$ est connu. Or :

$$a_1 = q(a) - \sum_{i=2}^{n} a_i$$

Ainsi, si $q(a)$ et $a_2$, …, $a_n$ sont connus, alors $a_1$ est connu.

**Commentaire**

Le terme « publique » présent dans l'énoncé doit être compris comme un synonyme de « connu ».

□

On dit que l'on dispose d'un procédé de $\varepsilon$-confidentialité de $D^n$ pour $q$ si :

(c1) pour tout $a \in D^n$, on dispose d'une variable aléatoire réelle $X_a$ ;

(c2) pour tout $(a, b) \in \mathcal{V}$, $(X_a, X_b)$ est $\varepsilon$-différentiel.

(c3) pour tout $a \in D^n$, $\mathbb{E}(X_a) = q(a)$.

***13.*** *Majoration de la probabilité de trouver $a_1$.*

Dans cette question, nous allons justifier en partie la terminologie.

On suppose à nouveau que $a_2, \ldots, a_n$ sont connus, que l'on cherche à protéger $a_1$ et que :

× le public connaît des d'intervalles $I_0, \ldots, I_d$ disjoints de réunion $\mathbb{R}$ tels qu'avec les valeurs fixées de $a_2, \ldots, a_n$, si $q(a) \in I_j$ alors $a_1 = j$. Cela signifie que si $q(a)$ est publique alors $a_1$ aussi.

× on dispose d'un procédé de $\varepsilon$-confidentialité de $D^n$ pour $q$ et que l'on rend $X_a$ publique à la place de $q(a)$.

On considère alors que l'expérience aléatoire modélisée par $(\Omega, \mathscr{A}, \mathbb{P})$ comporte comme première étape le choix au hasard de $a_1$ dans $[\![0, d]\!]$ et on définit :

× $A_1$ la variable aléatoire associée à ce choix ;

× pour tout $j \in [\![0, d]\!]$, $Y_j = X_{(j, a_2, \ldots, a_n)}$.
On suppose que $A_1$ et $Y_j$ sont indépendantes pour tout $j \in D$.

× la variable aléatoire réelle $R$ par :

$$\forall \omega \in \Omega, \text{ si } A_1(\omega) = j \text{ alors on détermine l'unique } k \text{ tel que } Y_j(\omega) \in I_k \text{ et on pose } R(\omega) = k$$

× $\theta = \mathbb{P}([R = A_1])$.

***a)*** Montrer que $\theta = \sum_{j=0}^{d} \mathbb{P}([Y_j \in I_j] \cap [A_1 = j])$.

*Démonstration.*

- Par défintion, $A_1(\Omega) = [\![0, d]\!]$.
  Ainsi, $([A_1 = j])_{j \in [\![0,d]\!]}$ est un système complet d'événements.
  D'où, par la formule des probabilités totales :

$$\begin{aligned}
\mathbb{P}([R = A_1]) &= \sum_{j=0}^{d} \mathbb{P}([A_1 = j] \cap [R = A_1]) \\
&= \sum_{j=0}^{d} \mathbb{P}([A_1 = j] \cap [R = j]) \\
&= \sum_{j=0}^{d} \mathbb{P}([A_1 = j] \cap [Y_j \in I_j]) \quad \textit{(par définition de } R\textit{)}
\end{aligned}$$

- Détaillons la dernière égalité. D'après l'énoncé, si $(j, k) \in [\![0, d]\!]^2$ alors, pour tout $\omega \in \Omega$ :

$$R(\omega) = k \;\Leftrightarrow\; A_1(\omega) = j \;\text{ et }\; Y_j(\omega) \in I_k$$

  On en déduit : $[R = k] = [A_1 = j] \cap [Y_j \in I_k]$. Et en particulier : $[R = j] = [A_1 = j] \cap [Y_j \in I_j]$

$$\boxed{\theta = \mathbb{P}([R = A_1]) = \sum_{j=0}^{d} \mathbb{P}([A_1 = j] \cap [Y_j \in I_j])}$$

□

***b)*** En déduire que $\theta = \dfrac{1}{d+1} \sum_{j=0}^{d} \mathbb{P}([Y_j \in I_j])$.

*Démonstration.*
D'après ce qui précède :

$$\begin{aligned}
\theta &= \sum_{j=0}^{d} \mathbb{P}([A_1 = j] \cap [Y_j \in I_j]) \\
&= \sum_{j=0}^{d} \mathbb{P}([A_1 = j]) \times \mathbb{P}([Y_j \in I_j]) && \textit{(car } A_1 \textit{ et } Y_j \textit{ sont indépendantes)} \\
&= \sum_{j=0}^{d} \frac{1}{d+1}\, \mathbb{P}([Y_j \in I_j]) && \textit{(car } A_1 \hookrightarrow \mathcal{U}([\![0, d]\!])\textit{)} \\
&= \frac{1}{d+1} \sum_{j=0}^{d} \mathbb{P}([Y_j \in I_j])
\end{aligned}$$

$$\boxed{\theta = \frac{1}{d+1} \sum_{j=0}^{d} \mathbb{P}([Y_j \in I_j])}$$

**Commentaire**

La question ***13.a)*** présente une vraie difficulté car il faut prendre l'initiative d'introduire un système complet d'événements et d'utiliser la formule des probabilités totales de manière adéquate. En contrepartie, cette question ***13.b)*** ne présente aucune difficulté. Comme le signale l'énoncé (à l'aide du « En déduire »), il s'agit d'utiliser le résultat de la question précédente ce qui permet ici de conclure de manière directe. Il est important d'apprendre à repérer ce genre de questions et d'avoir le courage de les traiter même si l'on a passé la question qui précède !

□

***c)*** En conclure que :

$$\theta \leqslant \frac{1}{d+1}\left(\mathrm{e}^{\varepsilon}-(\mathrm{e}^{\varepsilon}-1)\,\mathbb{P}([Y_0\in I_0])\right) \leqslant \frac{\mathrm{e}^{\varepsilon}}{d+1}$$

*Démonstration.*
Soit $j\in[\![1,d]\!]$.

- Tout d'abord, notons que $Y_j = X_{(j,a_2,\dots,a_n)}$ et $Y_0 = X_{(0,a_2,\dots,a_n)}$.
  Les $n$-uplets $(j,a_2,\dots,a_n)$ et $(0,a_2,\dots,a_n)$ sont voisins car ne différent que d'une composante.
  Ainsi, comme on suppose que l'on dispose d'un procédé de $\varepsilon$-confidentialité, le couple $(Y_j,Y_0)$ est $\varepsilon$-différentiel.
- On en déduit :
$$\mathrm{e}^{-\varepsilon}\,\mathbb{P}([Y_0\in I_j]) \leqslant \mathbb{P}([Y_j\in I_j]) \leqslant \mathrm{e}^{\varepsilon}\,\mathbb{P}([Y_0\in I_j])$$
- Or, d'après la question précédente :

$$\begin{aligned}(d+1)\,\theta &= \sum_{j=0}^{d}\mathbb{P}([Y_j\in I_j])\\ &= \mathbb{P}([Y_0\in I_0])+\sum_{j=1}^{d}\mathbb{P}([Y_j\in I_j])\\ &\leqslant \mathbb{P}([Y_0\in I_0])+\sum_{j=1}^{d}\mathrm{e}^{\varepsilon}\,\mathbb{P}([Y_0\in I_j]) && \textit{(d'après l'inégalité précédente)}\\ &= \mathbb{P}([Y_0\in I_0])+\mathrm{e}^{\varepsilon}\sum_{j=1}^{d}\mathbb{P}([Y_0\in I_j])\end{aligned}$$

- Par ailleurs, comme $I_0,\dots,I_d$ est une partition de $\mathbb{R}$, la famille $([Y_0\in I_j])_{i\in[\![0,d]\!]}$ est un système complet d'événements. Ainsi :
$$\sum_{j=0}^{d}\mathbb{P}([Y_0\in I_j]) = 1 \quad\text{et donc}\quad \sum_{j=1}^{d}\mathbb{P}([Y_0\in I_j]) = 1-\mathbb{P}([Y_0\in I_0])$$
- En combinant ces informations, on obtient :

$$\begin{aligned}(d+1)\,\theta &\leqslant \mathbb{P}([Y_0\in I_0])+\mathrm{e}^{\varepsilon}\left(1-\mathbb{P}([Y_0\in I_0])\right)\\ &= \mathrm{e}^{\varepsilon}-(\mathrm{e}^{\varepsilon}-1)\,\mathbb{P}([Y_0\in I_0])\\ &\leqslant \mathrm{e}^{\varepsilon} && \textit{(car } (e^{\varepsilon}-1)\,\mathbb{P}([Y_0\in I_0])\geqslant 0\textit{)}\end{aligned}$$

En divisant par $d+1>0$, on obtient bien : $\theta \leqslant \frac{1}{d+1}\left(\mathrm{e}^{\varepsilon}-(\mathrm{e}^{\varepsilon}-1)\,\mathbb{P}([Y_0\in I_0])\right) \leqslant \frac{\mathrm{e}^{\varepsilon}}{d+1}$. □

***d)*** On pose $\rho=\dfrac{1}{d+1}$ et $\tau=\dfrac{\theta-\rho}{\rho}$.
Donner une majoration de $\tau$. Que représente cette quantité ?
Qu'en déduire concernant la méthode de confidentialité présentée dans cette question lorsque $\varepsilon$ est proche de 0 ?

*Démonstration.*

- Tout d'abord :
$$\tau = \frac{\theta-\rho}{\rho} = \frac{\theta}{\rho}-1 = \frac{\theta}{\frac{1}{d+1}}-1 = (d+1)\,\theta-1 \leqslant \mathrm{e}^{\varepsilon}-1$$

  l'inégalité finale étant obtenue par la question précédente.

- Pour comprendre ce que représente la quantité $\tau$, reprenons en détail l'énoncé :
  - × le réel $a_1$ a été choisi au hasard mais est inconnu du public.
  - × la v.a.r. $X_a$ est connue du public.
  - × la partition de $\mathbb{R}$, formée des intervalles $I_0, \ldots, I_d$, est aussi connue du public.
    Ces intervalles sont choisis de telle sorte que si $q(a)$ est un jour rendu publique, alors on pourra en déduire $a_1$.
  - × on cherche à déduire $a_1$ de ces connaissances. Pour ce faire, on regarde à quel unique intervalle $I_k$ (l'unicité est une conséquence directe du partitionnement de $\mathbb{R}$) appartient $X_a$ et on définit alors la v.a.r. $R$ égale à cet entier $k$.

  La procédure de découverte consiste à décréter que la valeur de $R$ est $a_1$.
  On cherche à évaluer à quel point cette procédure est robuste. Il s'agit donc d'évaluer la probabilité $\mathbb{P}([R = A_1])$.

  Plus précisément, il s'agit de comparer la probabilité d'obtenir la valeur de $a_1$ via la procédure de l'énoncé (c'est $\theta = \mathbb{P}([R = A_1])$) à la probabilité d'obtenir $a_1$ avec la procédure basique consistant à piocher au hasard un nombre dans $[\![0, d]\!]$ (on tombe sur le bon résultat avec probabilité $\rho = \dfrac{1}{d+1}$).
  On considèrera que la procédure est robuste si $\theta$ est très grand devant $\rho$.
- On suppose que $\theta \geqslant \rho$ (si ce n'est pas le cas, le procédé de découverte n'a pas d'intérêt car il est moins bon que celui consistant à piocher une valeur au hasard dans $[\![0, d]\!]$). D'après ce qui précède :
$$0 \leqslant \frac{\theta - \rho}{\rho} \leqslant \mathrm{e}^{\varepsilon} - 1$$
  Lorsque $\varepsilon$ se rapproche de 0, $\mathrm{e}^{\varepsilon} - 1$ se rapproche de 0. Alors, par théorème d'encadrement, $\theta - \rho$ se rapproche de 0 donc $\theta$ se rapproche de $\rho$.

Lorsque $\varepsilon$ se rapproche de 0, la méthode de confidentialité est performante car on ne peut déduire la valeur $a_1$ des informations rendues publiques qu'avec une probabilité proche de celle obtenue en piochant une valeur au hasard dans $[\![0, d]\!]$. Les informations publiques ne nous renseignent donc pratiquement pas sur la valeur de $a_1$. □

On pose $\delta = \max\limits_{(a,b)\in\mathcal{V}} |q(a) - q(b)|$ et on suppose que $\delta > 0$.

***14.*** Dans cette question, pour tout $a \in D^n$, on pose $X_a = q(a) + Y$ où $Y$ suit la loi de Laplace de paramètre $(0, \beta)$.

***a)*** Pour tout $a \in D^n$, déterminer $\mathbb{E}(X_a)$ et une densité de probabilité $f_a$ de la loi de $X_a$ en fonction de $q(a)$ et de $\beta$.

*Démonstration.*
Commençons par déterminer $F_{X_a}$.

- Soit $x \in \mathbb{R}$.
$$F_{X_a}(x) = \mathbb{P}([q(a) + Y \leqslant x]) = \mathbb{P}([Y \leqslant x - q(a)]) = F_Y(x - q(a))$$
  Deux cas se présentent alors.
  - × Si $x - q(a) \leqslant 0$, *i.e.* si $x \leqslant q(a)$ :
$$F_{X_a}\big(x - q(a)\big) = \frac{1}{2}\,\mathrm{e}^{\frac{x-q(a)}{\beta}}$$
  *(d'après la fonction de répartition associée à $\mathcal{L}(0,1)$ déterminée en question **3.b**))*

× Si $x - q(a) \geqslant 0$, *i.e.* si $x > q(a)$ :

$$F_{X_a}\big(x - q(a)\big) \;=\; 1 - \frac{1}{2}\,\mathrm{e}^{-\frac{x-q(a)}{\beta}}$$

*(toujours d'après la question **3.b**)*

Ainsi : $F_{X_a} : x \mapsto \begin{cases} \dfrac{1}{2}\,\mathrm{e}^{\frac{x-q(a)}{\beta}} & \text{si } x \leqslant q(a) \\ 1 - \dfrac{1}{2}\,\mathrm{e}^{-\frac{x-q(a)}{\beta}} & \text{si } x > q(a) \end{cases}$.

On reconnaît la fonction de répartition associée à la loi $\mathcal{L}\big(q(a), \beta\big)$.

La fonction de répartition caractérisant la loi, on en déduit : $X_a \hookrightarrow \mathcal{L}\big(q(a), \beta\big)$.

En particulier, $X_a$ admet pour espérance $\mathbb{E}(X_a) = q(a)$ et est une v.a.r. à densité,

de densité $f_a : t \mapsto \dfrac{1}{2\beta}\,\exp\left(-\dfrac{|t - q(a)|}{\beta}\right)$

**Commentaire**

De manière plus générale, on peut démontrer, pour tout $(a, b) \in \mathbb{R}^* \times \mathbb{R}$ :

$$X \hookrightarrow \mathcal{L}(m, s) \;\Leftrightarrow\; aX + b \hookrightarrow \mathcal{L}(am + b, |a|s)$$

On peut encore une fois rapprocher ce résultat de celui sur les lois normales.

$$X \hookrightarrow \mathcal{N}\left(m, \sigma^2\right) \;\Leftrightarrow\; aX + b \hookrightarrow \mathcal{N}\left(am + b, a^2\sigma^2\right)$$

□

***b)*** Montrer que pour tout $t \in \mathbb{R}$ et $(a, b) \in \mathcal{V}$, $f_a(t) \;\leqslant\; \exp\left(\dfrac{\delta}{\beta}\right) f_b(t)$.

En déduire que pour tout $(a, b) \in \mathcal{V}$, $(X_a, X_b)$ est $\dfrac{\delta}{\beta}$-différentiel.

*Démonstration.*

Soit $t \in \mathbb{R}$ et soit $(a, b) \in \mathcal{V}$.

- Raisonnons par équivalence.

$$\begin{aligned}
f_a(t) \;\leqslant\; \exp\left(\frac{\delta}{\beta}\right) f_b(t) \;&\Leftrightarrow\; \cancel{\frac{1}{2\beta}}\,\exp\left(-\frac{|t - q(a)|}{\beta}\right) \;\leqslant\; \cancel{\frac{1}{2\beta}}\,\exp\left(\frac{\delta}{\beta}\right)\exp\left(-\frac{|t - q(b)|}{\beta}\right) \\
&\Leftrightarrow\; \exp\left(-\frac{|t - q(a)|}{\beta}\right) \;\leqslant\; \exp\left(\frac{\delta}{\beta} - \frac{|t - q(b)|}{\beta}\right) \\
&\Leftrightarrow\; -\frac{|t - q(a)|}{\beta} \;\leqslant\; \frac{\delta}{\beta} - \frac{|t - q(b)|}{\beta} \qquad \textit{(car ln est strictement croissante sur } ]0, +\infty[\textit{)} \\
&\Leftrightarrow\; \frac{|t - q(b)| - |t - q(a)|}{\cancel{\beta}} \;\leqslant\; \frac{\delta}{\cancel{\beta}} \qquad \textit{(avec } \beta > 0\textit{)}
\end{aligned}$$

- Or, d'après l'inégalité triangulaire :

$$|t - q(b)| - |t - q(a)| \;\leqslant\; \big|(\cancel{t} - q(b)) - (\cancel{t} - q(a))\big| \;=\; |q(a) - q(b)| \;\leqslant\; \delta$$

La dernière inégalité est obtenue par définition de $\delta$.

On en déduit que pour tout $t \in \mathbb{R}$ et pour tout $(a, b) \in \mathcal{V}$, $f_a(t) \;\leqslant\; \exp\left(\dfrac{\delta}{\beta}\right) f_b(t)$.

- En multipliant de part et d'autre par $\exp\left(-\frac{\delta}{\beta}\right) > 0$, on obtient pour tout $t \in \mathbb{R}$ :

$$\exp\left(-\frac{\delta}{\beta}\right) f_a(t) \leqslant f_b(t)$$

- Cette inégalité est vérifiée pour tout couple $(a,b)$ de voisins. Or, si $(a,b) \in \mathcal{V}$ alors $(b,a) \in \mathcal{V}$ et en appliquant l'inégalité en ce couple $(b,a)$, on obtient pour tout $t \in \mathbb{R}$ :

$$f_b(t) \leqslant \exp\left(\frac{\delta}{\beta}\right) f_a(t)$$

On en déduit que pour tout $t \in \mathbb{R}$ et pour tout $(a,b) \in \mathcal{V}$ :
$$\exp\left(-\frac{\delta}{\beta}\right) f_a(t) \leqslant f_b(t) \leqslant \exp\left(\frac{\delta}{\beta}\right) f_a(t).$$

Les fonctions $f_a$ et $f_b$ étant les densités respectives de $X_a$ et $X_b$, on en déduit, par la question ***9.a)***, que le couple $(X_a, X_b)$ est $\frac{\delta}{\beta}$-différentiel. □

***c)*** Comment choisir $\beta$ pour disposer alors d'un procédé de $\varepsilon$-confidentialité de $D^n$ pour $q$ ?

*Démonstration.*

- On dit que l'on dispose d'un procédé de $\varepsilon$-confidentialité de $D^n$ pour $q$ si les propriétés (c1), (c2) et (c3) sont vérifiées. Les propriétés (c1) et (c3) sont vérifiées.
Il s'agit donc de vérifier la propriété (c2).
- D'après la question précédente, pour tout $(a,b) \in \mathcal{V}$, $(X_a, X_b)$ est $\left(\frac{\delta}{\beta}\right)$-différentiel.
Pour disposer d'un procédé de $\varepsilon$-confidentialité, il suffit de choisir $\beta$ tel que $\varepsilon = \frac{\delta}{\beta}$.

En choisissant $\beta = \frac{\delta}{\varepsilon}$, on a bien un procédé de $\varepsilon$-confidentialité de $D^n$ pour $q$. □

***15.*** Dans cette question, pour tout $a = (a_1, \ldots, a_n)$ appartenant à $D^n$, $q(a) = \sum_{k=1}^{n} a_k$.

***a)*** Quelle est la valeur de $\delta$ ?

*Démonstration.*

- Commençons par rappeler la définition : $\delta = \max_{(a,b)\in\mathcal{V}} |q(a) - q(b)|$.
- Soit $(a,b) \in \mathcal{V}$. Alors $a$ et $b$ ne diffèrent que d'une composante.
Autrement dit, il existe $i_0 \in [\![1,n]\!]$ tel que :

$$\forall i \in [\![1,n]\!] \setminus \{i_0\},\ a_i = b_i \quad \text{et} \quad a_{i_0} \neq b_{i_0}$$

On en déduit :
$$\left|q(a) - q(b)\right| = \left|\sum_{i=1}^{n} a_i - \sum_{i=1}^{n} b_i\right| = \left|a_{i_0} - b_{i_0}\right| \leqslant d$$

car $a_{i_0}$ et $b_{i_0}$ sont des éléments de $D = [\![0,d]\!]$.
Ainsi, pour tout $(a,b) \in \mathcal{V}$, $\left|q(a) - q(b)\right| \leqslant d$.

On en déduit : $\delta \leqslant d$.

- D'autre part, si $a = (0, a_2, \dots, a_n)$ et $b = (d, a_2, \dots, a_n)$ (où pour tout $i \in [\![2, n]\!]$, $a_i \in [\![0, d]\!]$) alors $(a, b) \in \mathcal{V}$ et :
$$\big|q(b) - q(a)\big| \;=\; \big|d - 0\big| \;=\; d$$
Ainsi, le majorant de $\delta$ est atteint pour un couple de voisins.

On en déduit : $\delta = d$. □

On utilise dans la suite le procédé de $\varepsilon$-confidentialité tel qu'il a été défini dans la question ***14.*** mais au lieu de publier la valeur $X_a$, on procède ainsi :

× si $X_a < \frac{1}{2}$ on publie 0 ;

× si $X_a \in [k - \frac{1}{2}, k + \frac{1}{2}[$ où $k \in [\![1, nd - 1]\!]$, on publie $k$ ;

× sinon on publie $nd$.

***b)*** Montrer que la valeur aléatoire $Z_a$ publiée vérifie :
$$Z_a = \begin{cases} 0 & \text{si } X_a < \frac{1}{2} \\ \lfloor X_a + \frac{1}{2} \rfloor & \text{si } X_a \in [\frac{1}{2}, nd - \frac{1}{2}[ \\ nd & \text{si } X_a \geqslant nd - \frac{1}{2} \end{cases}$$

*Démonstration.*

- Soit $\omega \in \Omega$ et $k \in [\![1, nd - 1]\!]$. Alors :
$$\begin{aligned} X_a(\omega) \in [k - \tfrac{1}{2}, k + \tfrac{1}{2}[ \;&\Leftrightarrow\; k - \frac{1}{2} \;\leqslant\; X_a(\omega) \;<\; k + \frac{1}{2} \\ &\Leftrightarrow\; k \;\leqslant\; X_a(\omega) + \frac{1}{2} \;<\; k + 1 \\ &\Leftrightarrow\; k = \lfloor X_a(\omega) + \tfrac{1}{2} \rfloor \qquad \textit{(par définition de la fonction } \lfloor \cdot \rfloor\textit{)} \end{aligned}$$
Ainsi, dès que $X_a \in [k - \frac{1}{2}, k + \frac{1}{2}[$ avec $k \in [\![1, nd - 1]\!]$, on publie $\lfloor X_a + \frac{1}{2} \rfloor$.
Ceci étant vrai pour tout $k \in [\![1, nd - 1]\!]$, on doit publier $\lfloor X_a + \frac{1}{2} \rfloor$ dès que $X_a \in [1 - \frac{1}{2}, nd + \frac{1}{2}[$.

Le premier et dernier cas étant fournis par l'énoncé, on obtient :
$$Z_a = \begin{cases} 0 & \text{si } X_a < \frac{1}{2} \\ \lfloor X_a + \frac{1}{2} \rfloor & \text{si } X_a \in [\frac{1}{2}, nd - \frac{1}{2}[ \\ nd & \text{si } X_a \geqslant nd - \frac{1}{2} \end{cases}.$$
□

***c)*** Écrire un script qui pour $d$, $n$ et $\varepsilon$ saisis par l'utilisateur, génère une valeur aléatoire de $a \in D^n$ puis affiche $q(a)$ et $Z_a$.

*Démonstration.*

- Il s'agit tout d'abord de générer une valeur aléatoire de $a \in D^n$ (stockée dans une variable `a`) :
```
a = grand(1, n, 'uin', 0, d)
```
- La quantité $q(a) = \sum_{k=1}^{n} a_i$ est obtenue par l'appel :
```
q = sum(a)
```

- Il s'agit alors de simuler la v.a.r. $X_a = q(a) + Y$ où $Y \hookrightarrow \mathcal{L}(0, \beta)$. On a vu dans la question ***14.c)*** que pour obtenir un procédé de $\varepsilon$-confidentialité, il fallait choisir $\beta = \dfrac{\delta}{\varepsilon}$.
  De plus, d'après la question ***15.a)***, $\delta = d$. Ainsi, afin de simuler une v.a.r. qui suit la loi de Laplace, on utiliser la fonction `Laplace` de la question ***5.b)***.

```
X = q + Laplace(0, d/eps)
```

- Il s'agit enfin de simuler la v.a.r. $Z_a$. On stocke le résultat de cette simulation dans une variable `Z`. La v.a.r. $Z_a$ étant définie par cas selon les valeurs de $X_a$, la variable `Z` est définie par une structure conditionnelle dont les conditions dépendent des valeurs de la variable `X`.
  Plus précisément :

```
if X < 1/2 then
    Z = 0
elseif X <= n * d - 1/2 then
    Z = floor(X + 1/2)
else
    Z = n * d
end
```

  On rappelle que la fonction `floor` correspond à la fonction $\lfloor . \rfloor$ (partie entière par défaut).
- On obtient le programme souhaité en regroupant ces différentes instructions et en ajoutant les dialogues utilisateur et affichages demandés par l'énoncé.

```
1  d = input('Entrez un entier d plus grand que 1 : ')
2  n = input('Entrez un entier n plus grand que 2 : ')
3  eps = input('Entrez un réel epsilon strictement positif: ')
4
5  a = grand(1, n, 'uin', 0, d)
6  q = sum(a)
7  X = q + Laplace(0, d/eps)
8
9  if X < 1/2 then
10     Z = 0
11 elseif X <= n * d - 1/2 then
12     Z = floor(X + 1/2)
13 else
14     Z = n * d
15 end
16
17 disp(q)
18 disp(Z)
```

□

***d)*** Pour $n = 1000$, $d = 4$ et $\varepsilon$ choisi par l'utilisateur, écrire un script qui estime la valeur moyenne de $\dfrac{|Z_a - q(a)|}{q(a)}$ (on considèrera que $q(a)$ est toujours non nul).

**N.B.** À titre d'information, on obtient le tableau de valeurs suivant :

| $\varepsilon$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1 | 1.1 | 1.2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Moyenne | 1.91% | 1% | 0.6% | 0.5% | 0.3% | 0.3% | 0.28% | 0.2% | 0.2% | 0.19% | 0.17% | 0.16% |

*Démonstration.*

- Dans la suite, notons $T_a = \dfrac{|Z_a - q(a)|}{q(a)}$.
  L'énoncé demande d'estimer la « valeur moyenne » de la v.a.r. $T_a$. On considère, dans la suite de la question, qu'il s'agit d'estimer l'espérance de la v.a.r. $T_a$.
- L'idée naturelle pour obtenir une approximation de cette espérance est :
  - × de simuler un grand nombre de fois ($N = 10000$ par exemple) la v.a.r. $T_a$.
    Formellement, on souhaite obtenir un $N$-uplet $(t_1, \dots, t_N)$ qui correspond à l'observation d'un $N$-échantillon $(T_1, \dots, T_N)$ de la v.a.r. $T_a$.
  - × de réaliser la moyenne des résultats de cette observation.

  Cette idée est justifiée par la loi faible des grands nombres (LfGN) qui affirme :

$$\text{moyenne de l'observation} = \frac{1}{N}\sum_{i=1}^{N} t_i \simeq \mathbb{E}(T_a)$$

- Cette idée est implémentée à l'aide de la fonction suivante :

```
d = 4
n = 1000
N = 10000
eps = input('Entrez un réel epsilon strictement positif: ')
tabT = zeros(1, N)

a = grand(1, n, 'uin', 0, d)
q = sum(a)

for i = 1:N
    X = q + Laplace(0, delta/eps)
    if X < 1/2 then
        Z = 0
    elseif X <= n * d - 1/2 then
        Z = floor(X + 1/2)
    else
        Z = n * d
    end
    tabT(i) = abs(Z - q) / q
end

Moy = sum(tabT) / N
disp(Moy)
```

Détaillons les différents éléments de ce code :

× en ligne 5, on crée une matrice ligne `tabT` de taille $N$ destinée à contenir les différentes valeurs $t_i$.

× en ligne 6, on génère la valeur $a \in D^n$ permettant la bonne définition de $X_a$.

× en ligne 8, on calcul la somme $q(a)$ correspondante.

× en ligne 10 à 20, on effectue une boucle permettant d'obtenir les valeurs successives $t_1$, ..., $t_N$. Plus précisément, lors du $i^{\text{ème}}$ tour de boucle, on stocke la valeur $t_i$ dans le $i^{\text{ème}}$ coefficient de la matrice `tabT`.

× en ligne 22, on stocke dans la variable `Moy` la valeur $\frac{1}{N}\sum_{j=1}^{N} t_j$ calculée par le programme.

**Commentaire**

- L'énoncé nous incite à écrire des scripts avec des dialogues utilisateurs et des affichages. Cependant, il semble pertinent pour ces deux questions de présenter le résultat sous forme de fonctions. On rappelle qu'une fonction permet de réaliser un calcul dont le résultat est réutilisable par un autre programme. C'est d'ailleurs une des bases de la programmation de mener une réflexion sur le découpage en sous-fonctions du projet que l'on souhaite coder.
- Plus précisément, il était possible ici de factoriser certains blocs de code. Dans le deuxième programme, les lignes 12 à 18 correspondent aux lignes 9 à 15 du programme précédent. On peut donc penser à écrire une fonction, prenant en entrée les paramètres nécessaires à l'écriture de ces lignes.

```
1  function Z = simuZ(d, n, eps, a)
2      q = sum(a)
3      X = q + Laplace(0, d/eps)
4      if X < 1/2 then
5          Z = 0
6      elseif X <= n * d - 1/2 then
7          Z = floor(X + 1/2)
8      else
9          Z = n * d
10     end
11 endfunction
```

- La fonction suivante permet alors de répondre aux attentes de la question.

```
1  function Moy = approxEsperance(d, n, eps, N)
2      a = grand(1, n, 'uin', 0, d)
3      q = sum(a)
4      tabT = zeros(1, N)
5      for i=1:N
6          Z = simuZ(d, n, eps, a)
7          tabT(i) = abs(Z - q) / q
8      end
9      Moy = sum(tabT) / N
10 endfunction
```

*(on réalise son appel avec les paramètres $d = 4$, $n = 1000$, $N = 10000$ et $\varepsilon$ choisi par l'utilisateur)*

**Commentaire**

- Ajoutons enfin qu'il est plus pratique d'obtenir le tableau des résultats fournit par l'énoncé si l'on a opté pour la présentation sous forme de fonctions. Si le programme est présenté à l'aide d'un script, pour obtenir ce tableau, il faudra appeler ce script 12 fois de suite en rentrant à chaque fois la nouvelle valeur de $\varepsilon$. Si le programme est présenté sous forme de fonction, le programme suivant permet d'obtenir ce tableau des résultats :

```
tabEps = 0.1:0.1:1.2
tabMoy = zeros(1, 12) // On crée un tableau suffisamment grand
for i = 1:12
    tabMoy(i) = approxEsperance(4, 1000, tabEsp(i), 10000)
end
disp(tabMoy)
```

Ici, on écrit un script et pas une fonction car on ne souhaite pas utiliser le résultat de ce programme dans un autre programme. À titre d'information, on obtient les résultats suivants :

| $\varepsilon$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1 | 1.1 | 1.2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Moyenne | 1.97% | 1% | 0.68% | 0.51% | 0.4% | 0.32% | 0.29% | 0.24% | 0.21% | 0.21% | 0.17% | 0.17% |

Ce résultat est tout à fait comparable à celui présenté à la fin de l'énoncé.

□