Premier exercice : fonctionnement aléatoire d'un ordinateur (6 *pt.*)

Un conseiller de la caisse d'épargne de Crémieu[1] arrive, un mardi matin à huit heures trente, à son agence. Une fois installé dans son bureau, il met en route son ordinateur. Il sait que le système d'exploitation de ce dernier, Windaube Céveune, moins robuste que son ancêtre Windaube Nelty[2], ne démarre correctement qu'avec une probabilité p, où p est un réel strictement compris entre $0,9$ et 1, et que, dans ce cas, le délai exprimé en minutes d'apparition du premier dysfonctionnement est une variable aléatoire obéissant à la loi exponentielle $\mathcal{E}(\lambda)$ de paramètre λ, avec λ un réel strictement positif. Le fait que le système d'exploitation ne démarre pas correctement étant évidemment un dysfonctionnement, l'on désigne alors par X la variable aléatoire égale à l'intervalle temporel, exprimé en minutes, entre l'instant où le conseiller met en route son ordinateur et celui où apparaît le premier dysfonctionnement.

1. *(2 pt.)* Donner l'expression de la fonction de répartition F_X de la variable aléatoire X, puis en déduire que celle-ci admet pour densité la distribution f_X définie par

$$\forall x \in \mathbb{R}, \quad f_X(x) = (1-p)\delta_0(x) + \lambda p\, \mathrm{e}^{-\lambda x} \mathbb{1}_{[0,+\infty[}(x)$$

L'on a, clairement, $F_X(x) = 0$ pour tout réel x strictement négatif. Soit donc x un réel positif ou nul, soit E l'événement « le système d'exploitation de l'ordinateur démarre correctement », et soit F_{X_0} la fonction de répartition de la variable aléatoire X_0 égale au délai, exprimé en minutes, d'apparition du premier dysfonctionnement lorsque cet événement se réalise. Alors,

$$\begin{aligned} F_X(x) &= P\{X \leq x\} \\ &= \underbrace{P(E)}_{p}\,\underbrace{P(\{X \leq x\}/E)}_{F_{X_0}(x)} + \underbrace{P(\bar{E})}_{1-p}\,\underbrace{P(\{X \leq x\}/\bar{E})}_{1} \\ &= p(1-\mathrm{e}^{-\lambda x}) + (1-p) \\ &= 1 - p\mathrm{e}^{-\lambda x} \end{aligned}$$

de sorte que la fonction de répartition F_X de la variable aléatoire X a pour expression

$$\forall x \in \mathbb{R}, \quad F_X(x) = (1 - p\mathrm{e}^{-\lambda x})\mathbb{1}_{[0,+\infty[}(x)$$

En dérivant cette fonction au sens des distributions, l'on obtient ensuite la densité f_X de la variable aléatoire X définie par

$$\forall x \in \mathbb{R}, \quad f_X(x) = (1-p)\delta_0(x) + \lambda p\, \mathrm{e}^{-\lambda x} \mathbb{1}_{[0,+\infty[}(x)$$

2. *(2 pt.)* Pour chaque entier naturel n, donner l'expression du moment m_n d'ordre n de la variable aléatoire X, puis en déduire son espérance et sa variance.

[1] Nord Isère, France métropolitaine.

[2] Acronyme pour NeandErthaL TechnologY.

L'on a, évidemment, $m_0 = \mathbb{E}X^0 = \mathbb{E}1 = 1$, et, si n est un entier naturel non nul,

$$\begin{aligned}
m_n &= \mathbb{E}X^n \\
&= \int_{\mathbb{R}} x^n f_X(x)\,\mathrm{d}x \\
&= \int_{\mathbb{R}} x^n((1-p)\delta_0(x) + \lambda p\,\mathrm{e}^{-\lambda x}\mathbb{1}_{[0,+\infty[}(x))\,\mathrm{d}x \\
&= (1-p)\underbrace{\int_{\mathbb{R}} x^n \delta_0(x)\,\mathrm{d}x}_{0} + p\int_0^{+\infty} x^n \lambda\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x \qquad \text{(on fait le changement de variable } u = \lambda x\text{)} \\
&= p\int_0^{+\infty} \frac{u^n}{\lambda^n}\mathrm{e}^{-u}\,\mathrm{d}u \\
&= \frac{p}{\lambda^n}\underbrace{\int_0^{+\infty} u^n\mathrm{e}^{-u}\,\mathrm{d}u}_{\Gamma(n+1)} \\
&= \frac{pn!}{\lambda^n}
\end{aligned}$$

Pour $n = 1$ l'on a alors $m_1 = \mathbb{E}X = \frac{p1!}{\lambda^1} = \frac{p}{\lambda}$, tandis que si $n = 2$ l'on obtient dans ce cas $m_2 = \mathbb{E}X^2 = \frac{p2!}{\lambda^2} = \frac{2p}{\lambda^2}$, et l'on en déduit alors $\mathbb{V}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{2p}{\lambda^2} - \frac{p^2}{\lambda^2} = \frac{p(2-p)}{\lambda^2}$

3. *(1 pt.)* La politique de gestion du parc informatique des agences de caisse d'épargne consistant à remplacer l'ordinateur si celui-ci ne fonctionne pas correctement plus d'une heure après sa mise en route, l'on désigne par Y la variable aléatoire égale à 1 si l'ordinateur doit être remplacé, et à 0 sinon. Quelle est la loi de cette variable aléatoire ? Que vaut son espérance ?

   La variable aléatoire Y obéit à la loi de Bernoulli $\mathcal{B}(p_1)$, de paramètre

   $$p_1 = P\{X \leq 60\} = F_X(60) = 1 - p\mathrm{e}^{-60\lambda}$$

   qui n'est autre que son espérance.

4. *(1 pt.)* Pour chaque entier naturel n non nul, l'on prélève, au hasard, n ordinateurs dans les agences de caisse d'épargne. À chacun de ces prélèvements est alors associée une réalisation du n-échantillon aléatoire $(Y_i)_{1\leq i\leq n}$, où les variables aléatoires $(Y_i)_{1\leq i\leq n}$ sont indépendantes et de même loi que Y, ainsi que de la fréquence aléatoire

   $$F_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

   d'ordinateurs devant être remplacés. Justifier qu'alors la suite $(F_n)_{n\geq 1}$ converge presque sûrement et en moyenne quadratique vers une constante que l'on précisera.

   Les variables aléatoires $(Y_i)_{1\leq i\leq n}$ sont deux à deux indépendantes, de même moyenne $m = p_1$ et de même écart-type $\sigma = \sqrt{p_1(1-p_1)}$, si bien que la loi forte des grands nombres nous permet d'affirmer que la suite $(F_n)_{n\geq 1}$ converge presque sûrement et en moyenne quadratique vers cette moyenne, c'est-à-dire vers la constante $p_1 = 1 - p\mathrm{e}^{-60\lambda}$.

Deuxième exercice : dépendance ou indépendance d'un couple aléatoire uniformément réparti (5 *pt.*)

Un couple $X = (X_1, X_2)$ de variables aléatoires est uniformément réparti sur le triangle du plan euclidien, de sommets les points de coordonnées $(-1, 0)$, $(1, 0)$ et $(0, 1)$.

1. *(2 pt.)* Déterminer la densité du couple $X = (X_1, X_2)$, puis en déduire celle des variables aléatoires X_1 et X_2.
   Soit T le triangle du plan euclidien, de sommets les points de coordonnées $(-1, 0)$, $(1, 0)$ et $(0, 1)$. La densité f_X du couple aléatoire $X = (X_1, X_2)$ est alors donnée par
   $$\forall (x_1, x_2) \in \mathbb{R}^2, \quad f_X(x_1, x_2) = \frac{1}{\underbrace{\mathrm{mes}(T)}_{1}} \mathbb{1}_T(x_1, x_2) = \mathbb{1}_T(x_1, x_2)$$
   L'on en déduit ensuite l'expression de la densité f_{X_1} de la variable aléatoire X_1, définie par
   $$\forall x_1 \in \mathbb{R}, \quad f_{X_1}(x_1) = \int_{\mathbb{R}} f_X(x_1, x_2)\, \mathrm{d}x_2 = \int_{\mathbb{R}} \mathbb{1}_T(x_1, x_2)\, \mathrm{d}x_2 = (1 - |x_1|)\mathbb{1}_{[-1,1]}(x_1)$$
   puis celle f_{X_2} de la variable aléatoire X_2, donnée par
   $$\forall x_2 \in \mathbb{R}, \quad f_{X_2}(x_2) = \int_{\mathbb{R}} f_X(x_1, x_2)\, \mathrm{d}x_1 = \int_{\mathbb{R}} \mathbb{1}_T(x_1, x_2)\, \mathrm{d}x_1 = 2(1 - x_2)\mathbb{1}_{[0,1]}(x_2)$$

2. *(2 pt.)* Calculer l'espérance puis la variance des variables aléatoires X_1 et X_2.
   La densité f_{X_1} de la variable aléatoire X_1 étant une fonction paire, l'on a clairement $\mathbb{E}X_1 = 0$. L'on a alors $\mathbb{V}(X_1) = \mathbb{E}X_1^2$, ce qui donne
   $$\begin{aligned} \mathbb{V}(X_1) &= \int_{\mathbb{R}} x_1^2 f_{X_1}(x_1)\, \mathrm{d}x_1 \\ &= \int_{\mathbb{R}} x_1^2 (1 - |x_1|)\mathbb{1}_{[-1,1]}(x_1)\, \mathrm{d}x_1 \\ &= \int_{-1}^{1} x_1^2(1 - |x_1|)\, \mathrm{d}x_1 \\ &= 2\int_0^1 (x_1^2 - x_1^3)\, \mathrm{d}x_1 \\ &= 2(\tfrac{1}{3} - \tfrac{1}{4}) \\ &= \frac{1}{6} \end{aligned}$$
   L'on a ensuite
   $$\mathbb{E}X_2 = \int_{\mathbb{R}} x_2 f_{X_2}(x_2)\, \mathrm{d}x_2 = \int_{\mathbb{R}} x_2\, 2(1 - x_2)\mathbb{1}_{[0,1]}(x_2)\, \mathrm{d}x_2 = 2\int_0^1 (x_2 - x_2^2)\, \mathrm{d}x_2 = 2(\tfrac{1}{2} - \tfrac{1}{3}) = \frac{1}{3}$$
   ainsi que
   $$\mathbb{E}X_2^2 = \int_{\mathbb{R}} x_2^2 f_{X_2}(x_2)\, \mathrm{d}x_2 = \int_{\mathbb{R}} x_2^2\, 2(1 - x_2)\mathbb{1}_{[0,1]}(x_2)\, \mathrm{d}x_2 = 2\int_0^1 (x_2^2 - x_2^3)\, \mathrm{d}x_2 = 2(\tfrac{1}{3} - \tfrac{1}{4}) = \frac{1}{6}$$
   et l'on en déduit alors $\mathbb{V}(X_2) = \mathbb{E}X_2^2 - (\mathbb{E}X_2)^2 = \frac{1}{6} - \frac{1}{9} = \frac{1}{18}$.

3. *($\frac{1}{2}$ pt.)* Calculer la covariance des variables aléatoires X_1 et X_2.
   L'application partielle $x_1 \mapsto f_X(x_1, x_2)$ étant, pour tout réel x_2, une fonction paire, l'on a clairement $\mathbb{E}(X_1X_2) = 0$, et donc,
   $$\mathrm{Cov}(X_1, X_2) = \underbrace{\mathbb{E}(X_1X_2)}_{0} - \underbrace{\mathbb{E}X_1}_{0}\, \mathbb{E}X_2 = 0$$

4. *($\frac{1}{2}$ pt.)* Les variables aléatoires X_1 et X_2 sont-elles indépendantes ?
   Comme la densité f_X du couple aléatoire $X = (X_1, X_2)$ n'est pas égale au produit des densités f_{X_1} et f_{X_2} des variables aléatoires X_1 et X_2 qui sont ses composantes, ces dernières ne sont pas indépendantes.

Troisième exercice : quand la parité n'engendre pas la symétrie *(6 pt. + bonus)*

L'on suppose qu'en France il y a autant d'hommes que de femmes. L'on suppose par ailleurs que la taille des hommes, exprimée en centimètres, est une variable aléatoire X_1 obéissant à la loi normale $\mathcal{N}(m_1, \sigma_1)$, de moyenne $m_1 = 175$ et d'écart-type $\sigma_1 = 15$, tandis que celle des femmes, exprimée également en centimètres, est pour sa part une variable aléatoire X_2, indépendante de X_1, et régie par la loi normale $\mathcal{N}(m_2, \sigma_2)$, de moyenne $m_2 = 165$ et d'écart-type $\sigma_2 = 10$. L'on désigne enfin par X la variable aléatoire égale à la taille, exprimée en centimètres, des adultes en France.

1. *(2 pt.)* À l'aide des fonctions de répartition F_{X_1} et F_{X_2} des variables aléatoires X_1 et X_2, donner l'expression de la fonction de répartition F_X de la variable aléatoire X.
2. *(1 pt.)* De l'expression de F_X obtenue à la question 1, déduire que la variable aléatoire X est absolument continue, puis exprimer sa densité f_X en fonction des densités f_{X_1} et f_{X_2} des variables aléatoires X_1 et X_2.
3. *(1 pt.)* Montrer que la variable aléatoire X a pour moyenne $m = 170$ cm. S'agit-il d'une variable gaussienne ?
   Indication Comparer $f_X(165)$ et $f_X(175)$.
4. *(2 pt.)* Ayant choisi au hasard un adulte dans la polulation française, l'on constate qu'il mesure plus de 175 cm. Quelle est la probabilité qu'il s'agisse d'une femme ?
   Indication La fonction de répartition ϕ de la loi normale centrée réduite vaut $0,8413$ en $x = 1$.
5. **Bonus***(2 pt.)* Comparer la demi-somme des fonctions f_{X_1} et f_{X_2} à leur produit de convolution. Que peut-on en déduire ?

Dans tout ce qui suit, l'on désigne par E l'événement « l'adulte est un homme ». Comme l'on suppose qu'en France il y a autant d'hommes que de femmes, l'on a donc $P(E) = P(\overline{E}) = \frac{1}{2}$.

1. L'on a, pour tout réel x,
$$F_X(x) = P\{X \le x\} = \underbrace{P(E)}_{\frac{1}{2}}\underbrace{P(\{X \le x\}/E)}_{F_{X_1}(x)} + \underbrace{P(\bar{E})}_{\frac{1}{2}}\underbrace{P(\{X \le x\}/\bar{E})}_{F_{X_2}(x)} = \frac{1}{2}(F_{X_1}(x)+F_{X_2}(x))$$
2. Comme F_{X_1} et F_{X_2}, la fonction de répartition F_X de la variable aléatoire X est indéfiniment dérivable, ce qui montre que cette dernière est absolument continue et a pour densité $f_X = F'_X = \frac{1}{2}(F'_{X_1} + F'_{X_2}) = \frac{1}{2}(f_{X_1} + f_{X_2})$.
3. L'on a
$$\begin{aligned}
\mathbb{E}X &= \int_{\mathbb{R}} x f_X(x)\,\mathrm{d}x \\
&= \int_{\mathbb{R}} x \tfrac{1}{2}(f_{X_1}(x) + f_{X_2}(x))\,\mathrm{d}x \\
&= \underbrace{\tfrac{1}{2}\int_{\mathbb{R}} x f_{X_1}(x)\,\mathrm{d}x}_{\mathbb{E}X_1} + \underbrace{\tfrac{1}{2}\int_{\mathbb{R}} x f_{X_2}(x)\,\mathrm{d}x}_{\mathbb{E}X_2} \\
&= \tfrac{1}{2}(175 + 165) \\
&= 170
\end{aligned}$$

Comme l'on a, pour tout réel x, $f_{X_1}(x) = \frac{1}{\sqrt{2\Pi}\sigma_1} e^{-\frac{1}{2}(\frac{x-m_1}{\sigma_1})^2}$ avec $m_1 = 175$ et $\sigma_1 = 15$, ainsi que $f_{X_2}(x) = \frac{1}{\sqrt{2\Pi}\sigma_2} e^{-\frac{1}{2}(\frac{x-m_2}{\sigma_2})^2}$ avec cette fois $m_2 = 165$ et $\sigma_1 = 10$, l'on a $f_{X_1}(165) = \frac{1}{15\sqrt{2\Pi}} e^{-\frac{2}{9}}$, $f_{X_2}(165) = \frac{1}{10\sqrt{2\Pi}}$, $f_{X_1}(175) = \frac{1}{15\sqrt{2\Pi}}$ et $f_{X_2}(175) = \frac{1}{10\sqrt{2\Pi}} e^{-\frac{1}{2}}$, et l'on en déduit alors

$$f_X(165) = \tfrac{1}{2}(f_{X_1}(165) + f_{X_2}(165)) = \tfrac{1}{60\sqrt{2\Pi}}(3 + 2e^{-\frac{2}{9}})$$

ainsi que

$$f_X(175) = \tfrac{1}{2}(f_{X_1}(175) + f_{X_2}(175)) = \tfrac{1}{60\sqrt{2\Pi}}(2 + 3e^{-\frac{1}{2}})$$

Comme ces deux grandeurs ne sont pas égales et que la densité d'une variable gaussienne est symétrique par rapport à sa moyenne, la variable aléatoire X n'est pas gaussienne.

4. L'on a

$$\begin{aligned} P(\overline{E}/\{X \geq 175\}) &= \frac{P(\overline{E} \cap \{X \geq 175\})}{P\{X \geq 175\}} \\ &= \frac{\overbrace{P(\overline{E})}^{\frac{1}{2}}\,\overbrace{P(\{X \geq 175\}/\overline{E})}^{1-F_{X_2}(175)}}{\underbrace{P\{X \geq 175\}}_{1-F_X(175)}} \\ &= \frac{1 - F_{X_2}(175)}{2 - F_{X_1}(175) - F_{X_2}(175)} \end{aligned}$$

Or, l'on a, en désignant par ϕ la fonction de répartition d'une variable aléatoire U obéissant à la loi normale centrée réduite,

$$F_{X_1}(175) = P\{X_1 \leq 175\} = P\left\{\frac{X_1 - m_1}{\sigma_1} \leq \frac{175 - 175}{15}\right\} = P\{U \leq 0\} = \phi(0) = 0,5$$

ainsi que

$$F_{X_2}(175) = P\{X_2 \leq 175\} = P\left\{\frac{X_2 - m_2}{\sigma_2} \leq \frac{175 - 165}{10}\right\} = P\{U \leq 1\} = \phi(1) = 0,8413$$

de sorte que l'on a, finalement $P(\overline{E}/\{X \geq 175\}) = \frac{1-0{,}8413}{2-0{,}5-0{,}8413} = \frac{0{,}1587}{0{,}6587} \approx 0,24$

5. Nous savons que la demi-somme des fonctions f_{X_1} et f_{X_2} n'est autre que la densité de la variable aléatoire X. Le produit de convolution de ces mêmes fonctions est quant à lui la densité f_S de la somme $S = X_1 + X_2$ des variables aléatoires indépendantes X_1 et X_2. Il s'agit donc, du fait de la stabilité par l'indépendance de variables gaussiennes, de la densité d'une variable aléatoire gaussienne. Ainsi, la demi-somme $Z = \frac{1}{2}S$ des variables aléatoires X_1 et X_2 est elle aussi une variable aléatoire gaussienne, si bien que sa densité f_Z donnée, pour tout réel z, par $f_Z(z) = 2f_S(2z)$, ne peut être égale à celle de la variable aléatoire X qui n'est pas de nature gaussienne. Ainsi, cette variable aléatoire, de densité f_X égale à la demi-somme des densités des variables aléatoires X_1 et X_2, n'est en revanche ni égale ni égale en loi à leur demi-somme.

Quatrième exercice : convergence *(3 pt.)*

Soit, pour chaque entier naturel n non nul, X_n la variable aléatoire discrète de loi donnée par $P\{X_n = 0\} = 1 - \frac{1}{n}$ et $P\{X_n = n\} = \frac{1}{n}$, et soit X la variable aléatoire égale à 0 presque sûrement.

1. *(1 pt.)* Montrer que la suite de variables aléatoires $(X_n)_{n\geq 1}$ converge en loi vers X.

   Soit F_X la fonction de répartition de la variable aléatoire X et soit, pour chaque entier naturel n non nul, F_{X_n} celle de la variable aléatoire X_n. L'on a alors, pour tout réel x, $F_X(x) = \mathbb{1}_{[0,+\infty[}(x)$ ainsi que $F_{X_n}(x) = (1-\frac{1}{n})\mathbb{1}_{[0,n[}(x) + \frac{1}{n}\mathbb{1}_{[n,+\infty[}(x)$. Ainsi, si x est un réel strictement négatif l'on a $F_{X_n}(x) = 0$ pour tout entier naturel n non nul, tandis que si ce réel x est strictement positif l'on a cette fois $F_{X_n}(x) = 1$ pour tout entier naturel n inférieur ou égal à $[x]$, et $F_{X_n}(x) = 1 - \frac{1}{n}$ pour tout entier naturel n supérieur ou égal à $[x]+1$, de sorte que pour tout réel x non nul la suite $(F_{X_n}(x))_{n\geq 1}$ converge vers $F_X(x)$ lorsque n tend vers $+\infty$. Ainsi, la suite de variables aléatoires $(X_n)_{n\geq 1}$ converge en loi vers X.

2. *(1 pt.)* Montrer que la suite de variables aléatoires $(X_n)_{n\geq 1}$ converge en probabilité vers X.

   Soit ε un réel strictement positif. L'on a alors $P\{X_n - X > \varepsilon\} = P\{X_n > \varepsilon\} = 0$ pour tout entier naturel n inférieur ou égal à $[\varepsilon]$, et $P\{X_n - X > \varepsilon\} = P\{X_n > \varepsilon\} = \frac{1}{n}$ pour tout entier naturel n supérieur ou égal à $[\varepsilon]+1$, de sorte que

   $$\lim_{n\to+\infty} P\{X_n - X > \varepsilon\} = 0$$

   ce qui montre que la suite de variables aléatoires $(X_n)_{n\geq 1}$ converge en probabilité vers X.

3. *(1 pt.)* Soit p un entier naturel non nul. La suite de variables aléatoires $(X_n)_{n\geq 1}$ converge-t-elle en moyenne d'ordre p vers X ?

   L'on a, pour tout entier naturel n non nul,

   $$\mathbb{E}(X_n - X)^p = \mathbb{E}X_n^p = (1-\tfrac{1}{n})0^p + \tfrac{1}{n}n^p = n^{p-1}$$

   Ainsi, la suite de variables aléatoires $(X_n)_{n\geq 1}$ ne converge pas en moyenne d'ordre p vers X.