Corrigé final - SQ20 - P2018

Exercice 1 : 5 questions indépendantes

1. $P_A(B) = \dfrac{P(A\cap B)}{P(A)} = \dfrac{P(A)+P(B)-P(A\cup B)}{P(A)} = \dfrac{0,45+0,5-0,7}{0,45} = \dfrac{5}{9}$
2. $P([X=3]) = \dfrac{3\times 2\times 1}{3\times 3\times 3} = \dfrac{2}{9}$
3. $E(Z) = \displaystyle\int_0^4 \sqrt{t}\frac{1}{4}\,\mathrm{d}t = \frac{4}{3}$
4. D'après la loi faible des grands nombres, Z est constante égale à 1.
5. $[19.61\,,21.65]$

Exercice 2 : test d'hypothèse

1. On veut effectuer un **test de conformité sur la proportion** p des patients allergiques guéris grâce au médicament. On choisit comme hypothèse nulle $H_0\ :\ p = 90\%$ car **on fait confiance a priori au fabricant**.

   L'hypothèse nulle notée H_0 est l'hypothèse que l'on souhaite contrôler (la différence observée entre la proportion annoncée et la proportion observée sur un échantillon est-elle significative ou due aux fluctuations d'échantillonnage ?) C'est l'hypothèse nulle qui est soumise au test et toute la démarche du test s'effectue en considérant cette hypothèse comme vraie.

   On a choisi comme hypothèse alternative $H_1\ :\ p < 90\%$ au lieu de $p \neq 90\%$ car **on ne peut en vouloir au fabricant de produire un médicament qui guérirait plus de 90% des patients.**

2. On se place sous l'hypothèse H_0. Soit X la variable aléatoire qui, à tout échantillon aléatoire de $n = 200$ patients atteints par l'allergie, associe le nombre de personnes guéries par le médicament.

   Alors, en posant $p_0 = 90\%$, X suit la loi binomiale $\mathscr{B}(n, p_0)$ d'espérance $E(X) = np_0$ et de variance $V(X) = np_0(1-p_0)$.

   Puisque $n > 30$, $n\,p_0 \geqslant 15$ et $n\,p_0(1-p_0) > 5$, le théorème central limite permet de remplacer la loi de X par la loi normale $\mathcal{N}\left(n\,p_0\,,\ \sqrt{n\,p_0(1-p_0)}\right)$. Or $F = \dfrac{X}{n}$.

   Donc $\boxed{F \text{ suit approximativement la loi normale } \mathcal{N}\left(p_0\,,\ \sqrt{\dfrac{p_0(1-p_0)}{n}}\right)}$.

3. • Variable de test : on pose $Z = \dfrac{F-p_0}{\sqrt{p_0(1-p_0)/n}}$.

   Alors Z suit approximativement la loi normale centrée réduite $\mathcal{N}(0\,,1)$.

   • On cherche le réel z_α tel que $P(Z \geqslant z_\alpha) = 0,99$ c'est-à-dire $P(Z \leqslant z_\alpha) = 0,01$ c'est-à-dire $\boxed{z_\alpha = \Phi^{-1}(0,01)}$

   Avec Scilab, en entrant dans la console `z=cdfnor("X",0,1,0.01,0.99)` , on obtient $z_\alpha \approx -2.326$.

4. Puisque $Z \geqslant z_\alpha \iff F \geqslant p_0 + z_\alpha\sqrt{\dfrac{p_0(1-p_0)}{n}}$, on obtient

   $$1-\alpha = 0,99 = P(Z \geqslant z_\alpha) = P\left(F \geqslant p_0 + z_\alpha\sqrt{\frac{p_0(1-p_0)}{n}}\right)$$

   On en déduit le domaine d'acceptation $D_0 = \left[p_0 + z_\alpha\sqrt{\dfrac{p_0(1-p_0)}{n}}\,,\,1\right]$

   $\boxed{D_0 = [85\%\,,\,100\%]}$

5. On calcule la valeur prise par F dans l'échantillon $f_e = \dfrac{169}{200} = 84,5\%$

   On voit que $f_e \notin D_0$. **Conclusion** : on rejette l'hypothèse H_0 avec un risque d'erreur de $\alpha = 1\%$. L'affirmation du fabricant est mensongère.

Exercice 3 : estimation ponctuelle

Soit $a \in \mathbb{R}^{+*}$. La fonction f est définie sur $\mathbb{R}$ par :

$$f(t) = \frac{3\,a^3}{t^4} \quad \text{si } t \geqslant a \qquad \text{et} \qquad f(t) = 0 \quad \text{si } t < a$$

1. (a) • Soit x un réel supérieur à a.

   $$\int_a^x \frac{1}{t^4}\,\mathrm{d}t = \int_a^x t^{-4}\,\mathrm{d}t = \left[\frac{t^{-4+1}}{-4+1}\right]_{t=a}^{t=x} = \left[-\frac{1}{3\,t^3}\right]_{t=a}^{t=x} = -\frac{1}{3x^3} + \frac{1}{3\,a^3}$$

   Or $\dfrac{1}{3x^3} \underset{(x\to+\infty)}{\longrightarrow} 0$. Donc $\displaystyle\lim_{x\to+\infty}\int_a^x \frac{1}{t^4}\,\mathrm{d}t = \frac{1}{3\,a^3}$

   ce qui signifie que l'intégrale généralisée $\displaystyle\int_a^{+\infty}\frac{1}{t^4}\,\mathrm{d}t$ est convergente et que

   $$\boxed{\int_a^{+\infty}\frac{1}{t^4}\,\mathrm{d}t = \frac{1}{3\,a^3}}$$

• On aurait pu aussi remarquer que $\displaystyle\int_a^{+\infty}\frac{1}{t^4}\,\mathrm{d}t$ était une intégrale de Riemann convergente car $4>1$, et écrire directement

$$\int_a^{+\infty}\frac{1}{t^4}\,\mathrm{d}t=\int_a^{+\infty}t^{-4}\,\mathrm{d}t=\left[\frac{t^{-3}}{-3}\right]_a^{+\infty}=\left[\frac{-1}{3\,t^3}\right]_a^{+\infty}=-0+\frac{1}{3\,a^3}$$

(b) • La fonction f est strictement positive sur $[a\,,\,+\infty[$ et nulle sur $]-\infty\,,\,a[$.
• f est continue sur $\mathbb{R}\setminus\{a\}$.
• D'après la question précédente, l'intégrale généralisée $\displaystyle\int_a^{+\infty}\frac{1}{t^4}\,\mathrm{d}t$ est convergente ainsi que $\displaystyle\int_a^{+\infty}\frac{3\,a^3}{t^4}\,\mathrm{d}t$. D'où

$$\int_{-\infty}^{+\infty}f(t)\,\mathrm{d}t=\int_a^{+\infty}f(t)\,\mathrm{d}t=\int_a^{+\infty}\frac{3\,a^3}{t^4}\,\mathrm{d}t=3\,a^3\int_a^{+\infty}\frac{1}{t^4}\,\mathrm{d}t=1$$

On en déduit f est une densité de probabilité.

2. (a) On admet que $\displaystyle\int_a^{+\infty}\frac{1}{t^3}\,\mathrm{d}t$ converge et vaut $\dfrac{1}{2\,a^2}$.
Il en résulte que l'intégrale généralisée $\displaystyle\int_a^{+\infty}\frac{3\,a^3}{t^3}\,\mathrm{d}t$ est convergente et vaut $\displaystyle 3\,a^3\int_a^{+\infty}\frac{1}{t^3}\,\mathrm{d}t=\frac{3\,a^3}{2\,a^2}=\frac{3}{2}a.$
Donc X admet une espérance donnée par :

$$E(X)=\int_{-\infty}^{+\infty}t\,f(t)\,\mathrm{d}t=\int_a^{+\infty}t\,f(t)\,\mathrm{d}t=\int_a^{+\infty}\frac{3\,a^3}{t^3}\,\mathrm{d}t=\frac{3}{2}a$$

(b) Soit x un réel supérieur à a.

$$\int_a^x t^2\,f(t)\,\mathrm{d}t=\int_a^x\frac{3\,a^3}{t^2}\,\mathrm{d}t=3\,a^3\int_a^x\frac{1}{t^2}\,\mathrm{d}t=3\,a^3\left[-\frac{1}{t}\right]_a^x=3a^3\left(-\frac{1}{x}+\frac{1}{a}\right)$$

D'où $\displaystyle\int_a^x t^2\,f(t)\,\mathrm{d}t=3\,a^2-\frac{3\,a^3}{x}\underset{(x\to+\infty)}{\longrightarrow}3\,a^2$

Donc l'intégrale généralisée $\displaystyle\int_a^{+\infty}t^2\,f(t)\,\mathrm{d}t$ est convergente et vaut $3\,a^2$ ce qui signifie que X admet un moment d'ordre 2 égal à $E(X^2)=3\,a^2$.
On en déduit d'après la **formule de Kœnig-Huygens**, que X admet une variance égale à :

$$\boxed{V(X)=E(X^2)-E(X)^2}=3\,a^2-\left(\frac{3}{2}a\right)^2=3\,a^2-\frac{9}{4}a^2=\frac{3\,a^2}{4}$$

3. On rappelle que pour tout réel x, $\displaystyle F(x)=P\big([X\leqslant x]\big)=\int_{-\infty}^x f(t)\,\mathrm{d}t.$

(a) Si $x<a$ alors $F(x)=0$ car $\forall\,t\in]-\infty,a[,\ f(t)=0$.

(b) Supposons $x\geqslant a$.
Alors $\displaystyle F(x)=\int_a^x f(t)\,\mathrm{d}t=3\,a^3\int_a^x\frac{1}{t^4}\,\mathrm{d}t\overset{\text{d'après 1.(a)}}{=}3a^3\left(-\frac{1}{3x^3}+\frac{1}{3\,a^3}\right).$
Donc $\forall\,x\geqslant a$, $\boxed{F(x)=1-\left(\frac{a}{x}\right)^3}$

(c) $$P_{[X>2a]}\big(\,[X>4a]\,\big)=\frac{P\big([X>4a]\cap[X>2a]\big)}{P\big([X>2a)\big)}$$
Or l'événement $[X>4a]$ «implique» l'événement $[X>2a]$ c'est-à-dire $[X>4a]\subset[X>2a]$.
Donc $[X>4a]\cap[X>2a]=[X>4a]$ et $P_{[X>2a]}\big([X>4a]\big)=\dfrac{P(X>4a)}{P(X>2a)}$

D'où $$P_{[X>2a]}\big(\,[X>4a]\,\big)=\frac{1-F(4a)}{1-F(a)}=\frac{1-\left(1-\left(\frac{a}{4a}\right)^3\right)}{1-\left(1-\left(\frac{a}{2a}\right)^3\right)}=\frac{(1/4)^3}{(1/2)^3}$$

Ainsi $\boxed{P_{[X>2a]}\big(\,[X>4a]\,\big)=\dfrac{1}{8}}$

4. $\forall\,n\in\mathbb{N}^*$, $\displaystyle\overline{X_n}=\frac{1}{n}\sum_{k=1}^n X_k$

(a) • Calculons l'espérance de $\overline{X_n}$.

Par linéarité de l'espérance, $\displaystyle E(\overline{X_n})=E\left(\frac{1}{n}\sum_{k=1}^n X_k\right)=\frac{1}{n}\sum_{k=1}^n E(X_k)$
$\displaystyle=\frac{1}{n}\sum_{k=1}^n E(X_1)=E(X_1)=\frac{3}{2}a$. Donc $\overline{X_n}$ un estimateur **biaisé** de a.

• Posons pour tout entier $n\geqslant 1$, $\boxed{T_n=\dfrac{2}{3}\overline{X_n}}$

On voit que $\displaystyle E(T_n)=\frac{2}{3}E\left(\overline{X_n}\right)=\frac{2}{3}\times\frac{3}{2}a=a$
T_n est bien un estimateur sans biais du paramètre a.

(b) Pour montrer que T_n est un estimateur convergent de a, T_n étant sans biais, il suffit de prouver que $\lim\limits_{n\to+\infty} V(T_n) = 0$.

$$V(T_n) = V\left(\frac{2}{3}\overline{X_n}\right) = \frac{4}{9}V\left(\overline{X_n}\right) = \frac{4}{9}V\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{4}{9n^2}V\left(\sum_{k=1}^{n} X_k\right)$$

Or les variables aléatoires X_1, X_2, ..., X_n sont indépendantes.

Donc $V(T_n) = \dfrac{4}{9n^2}\displaystyle\sum_{k=1}^{n} V(X_k) = \dfrac{4}{9n^2}\sum_{k=1}^{n} V(X_1) = \dfrac{4}{9n^2}\, n\, V(X_1) = \dfrac{4}{9n}\,\dfrac{3\,a^2}{4}$

Ainsi $\boxed{V(T_n) = \dfrac{a^2}{3n}}$ et $V(T_n) \underset{(n\to+\infty)}{\longrightarrow} 0$

(c) On prendra comme estimation ponctuelle de a, la réalisation de la variable aléatoire T_n sur l'échantillon observé en prenant ici $n = 100$.

$$\widehat{a} = T_{100}(\omega) = \frac{2}{3}\overline{x_e} = \frac{2}{3}\,\frac{1}{100}\sum_{k=1}^{100} x_k = \frac{2}{3} \times 6,03 = 4,02$$

Attention : il est inexact d'écrire $a = 4,02$ car le paramètre a reste inconnu même après avoir étudié un échantillon.