Corrigé final - SQ20 - P2017

Exercice 1 : intervalle de confiance

Dans cet exercice, la variable parente X suit une loi normale $\mathcal{N}(\mu,\sigma)$ d'espérance connue $\mu = 20$. On a ici $\alpha = 1\%$ et $n = 100$.

1. On prend comme estimateur de σ^2 la variance empirique d'échantillon
$$T = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$$

2. On prend comme variable de confiance $Z = \dfrac{n\,T}{\sigma^2} = \displaystyle\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$ qui suit la loi du χ^2_n à n degrés de liberté.

3. On cherche des réels positifs a et b tels que $P\,(a \leqslant Z \leqslant b) = 1-\alpha$.

   Contrairement aux lois normale et de Student, la densité de la loi du χ^2 n'est pas une fonction paire. Mais on a l'habitude de répartir le risque de façon symétrique en prenant :
$$P\,(Z \leqslant a) = \frac{\alpha}{2} \quad \text{et} \quad P\,(Z \leqslant b) = 1-\frac{\alpha}{2}$$
   Avec Scilab : `a=cdfchi("X",100,0.01/2,1-0.01/2)`
   et `b=cdfchi("X",100,1-0.01/2,0.01/2)`
   On obtient : $a \approx 67{,}33$ et $b \approx 140{,}17$.

4. $0{,}99 = 1-\alpha = P\left(a \leqslant \dfrac{n\,T}{\sigma^2} \leqslant b\right) = P\left(\dfrac{n\,T}{b} \leqslant \sigma^2 \leqslant \dfrac{n\,T}{a}\right)$

   On en déduit un intervalle de confiance aléatoire à 99% pour la variance σ^2 :
$$I_{99\%} = \left[\frac{n\,T}{b}\,,\,\frac{n\,T}{a}\right]$$

5. Nous avons $0{,}99 = P\left(\dfrac{n\,T}{b} \leqslant \sigma^2 \leqslant \dfrac{n\,T}{a}\right) = P\left(\sqrt{\dfrac{n\,T}{b}} \leqslant \sigma \leqslant \sqrt{\dfrac{n\,T}{a}}\right)$

   D'où un intervalle de confiance observé pour l'**écart-type** σ au niveau 99% :
$$J_{99\%\ obs} = \left[\sqrt{\frac{n\,s_e^2}{b}}\,,\,\sqrt{\frac{n\,s_e^2}{a}}\right]$$
   avec $s_e^2 = \widehat{\sigma}^2 = \left(\dfrac{1}{n}\displaystyle\sum_{n}^{i=1} x_i^2\right) - \mu^2 = \dfrac{40\,011}{100} - 20^2 = 0{,}11.$

   Ainsi $J_{99\%\ obs} = [0{,}280\ ;\ 0{,}404]$.

Exercice 2 : les 2 questions sont indépendantes

1. Soit $n \in \mathbb{N}$ tel que $n \geqslant 3$.

   (a) Soit X_n la variable aléatoire qui compte le nombre de «Pile» obtenus au cours d'une partie de n lancers. Alors X_n suit la loi binomiale $\mathscr{B}\,(n\,,\,1/2)$.
   Notons A l'événement : «une partie comporte un gagnant».
   Alors $A = [X_n = 1] \cup [X_n = n-1]$. Donc $p_n = P(A)$
   $= P\big([X_n = 1] \cup [X_n = n-1]\big) = P\big([X_n = 1]\big) + P\big([X_n = n-1]\big)$
   $= \dbinom{n}{1}(1/2)^1\,(1/2)^{n-1} + \dbinom{n}{n-1}(1/2)^{n-1}\,(1/2)^1$
   $= \dfrac{n}{2^n} + \dfrac{n}{2^n} = \dfrac{2n}{2^n}$
$$\text{Ainsi} \quad \boxed{p_n = \frac{n}{2^{n-1}}}$$

   (b) La variable aléatoire G_n suit la loi géométrique $\mathscr{G}(p_n)$.
   On sait alors que $E(G_n) = \dfrac{1}{p_n} = \dfrac{2^{n-1}}{n}$

2. Pour tout $n \in \mathbb{N}^*$, $X_n \hookrightarrow \mathcal{N}\left(\mu\,,\,\dfrac{\sigma}{\sqrt{n}}\right)$.

   (a) Soit $x \in \mathbb{R}$. $F_n(x) = P\big(X_n \leqslant x\big) = P\left(\dfrac{X_n-\mu}{\sigma/\sqrt{n}} \leqslant \dfrac{x-\mu}{\sigma/\sqrt{n}}\right)$

   Or la variable aléatoire $X_n^* = \dfrac{X_n-\mu}{\sigma/\sqrt{n}}$ suit la loi normale centrée réduite $\mathcal{N}(0\,,\,1)$. Donc $\boxed{F_n(x) = \Phi\left(\dfrac{x-\mu}{\sigma/\sqrt{n}}\right) = \Phi\left(\dfrac{x-\mu}{\sigma}\sqrt{n}\right)}$

   (b) • Supposons $x > \mu$. Alors $\dfrac{x-\mu}{\sigma} > 0$ d'où $\displaystyle\lim_{n\to+\infty} \frac{x-\mu}{\sigma}\sqrt{n} = +\infty$.

   Or, en tant que fonction de répartition, $\displaystyle\lim_{t\to+\infty} \Phi(t) = 1$.

   Donc, par composition, $F_n(x) = \Phi\left(\dfrac{x-\mu}{\sigma}\sqrt{n}\right) \underset{(n\to+\infty)}{\longrightarrow} 1$

- Supposons $x < \mu$. Alors $\dfrac{x-\mu}{\sigma} < 0$ d'où $\lim\limits_{n\to+\infty} \dfrac{x-\mu}{\sigma}\sqrt{n} = -\infty$.

  Or, en tant que fonction de répartition, $\lim\limits_{t\to-\infty} \Phi(t) = 0$.

  Donc, par composition, $F_n(x) = \Phi\left(\dfrac{x-\mu}{\sigma}\sqrt{n}\right) \underset{(n\to+\infty)}{\longrightarrow} 0$
- Dans le cas où $x = \mu$, $\quad F_n(\mu) = \Phi(0) = 1/2$.

Finalement $\boxed{\lim\limits_{n\to+\infty} F_n(x) = \begin{cases} 0 & \text{si} \quad x < \mu \\ 1/2 & \text{si} \quad x = \mu \\ 1 & \text{si} \quad x > \mu \end{cases}}$

Notons que cette fonction limite n'est pas la fonction de répartition d'une variable aléatoire, car elle n'est pas continue à droite en μ

(c) Soit Y une variable aléatoire définie sur le même espace probabilisé que les X_n et qui est constante égale à μ. Alors la fonction de répartition de Y est définie par $F_Y(x) = \begin{cases} 0 & \text{si} \quad x < \mu \\ 1 & \text{si} \quad x \geqslant \mu \end{cases}$

Puisque cette fonction F_Y est continue sur $\mathbb{R}\setminus\{\mu\}$
et que $\forall x \in \mathbb{R}\setminus\{\mu\}$, $\lim\limits_{n\to+\infty} F_n(x) = F_Y(x)$, on peut conclure que la suite

$\boxed{(X_n)_{n\in\mathbb{N}^*} \text{ converge en loi vers la variable aléatoire constante égale à } \mu}$.

Exercice 3 : loi de Pareto

Soit $a > 1$ et f la fonction définie par : $\forall t \in \mathbb{R}$, $f(t) = \begin{cases} \dfrac{a}{t^{a+1}} & \text{si } t \geqslant 1 \\ 0 & \text{sinon} \end{cases}$

1. Soit $B \geqslant 1$.

   (a) On rappelle que $a > 1$, a fortiori $a > 0$.
   $I(B) = \displaystyle\int_1^B \frac{a}{t^{a+1}}\,\mathrm{d}t = \left[-\frac{1}{t^a}\right]_{t=1}^{t=B} = -\frac{1}{B^a} + 1.$
   Comme $\dfrac{1}{x^a} \underset{(x\to+\infty)}{\longrightarrow} 0$, on obtient $\boxed{\lim\limits_{B\to+\infty} I(B) = 1}$

   (b)
   - La fonction f est strictement positive sur $[1\,,+\infty[$ et nulle sur $]-\infty\,,1[$.
   - f est continue sur $\mathbb{R}\setminus\{1\}$.
   - D'après la question précédente, l'intégrale généralisée $\displaystyle\int_1^{+\infty} \frac{a}{t^{a+1}}\,\mathrm{d}t$ est convergente et

   $$\int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t = \int_1^{+\infty} f(t)\,\mathrm{d}t = \int_1^{+\infty} \frac{a}{t^{a+1}}\,\mathrm{d}t = \lim_{B\to+\infty} I(B) = 1$$

   On en déduit f est une densité de probabilité.

2. (a) Soit $B \geqslant 1$. On rappelle que $a > 1$ donc $a - 1 > 0$.
   $\displaystyle\int_1^B \frac{a}{t^a}\,\mathrm{d}t = \frac{a}{a-1}\int_1^B \frac{a-1}{t^a}\,\mathrm{d}t = \frac{a}{a-1}\left[-\frac{1}{t^{a-1}}\right]_{t=1}^{t=B}$
   $= \dfrac{a}{a-1}\left(-\dfrac{1}{B^{a-1}} + 1\right)$. $\quad \boxed{\displaystyle\int_1^B \frac{a}{t^a}\,\mathrm{d}t = \frac{a}{a-1}\left(1 - \frac{1}{B^{a-1}}\right)}$

   (b) $\displaystyle\lim_{B\to+\infty}\int_1^B \frac{a}{t^a}\,\mathrm{d}t = \frac{a}{a-1}$
   donc l'intégrale généralisée $\displaystyle\int_1^{+\infty} \frac{a}{t^a}\,\mathrm{d}t$ est convergente et vaut :

   $$\int_1^{+\infty} \frac{a}{t^a}\,\mathrm{d}t = \frac{a}{a-1} = \int_1^{+\infty} t\,f(t)\,\mathrm{d}t$$

   Ainsi l'espérance de X existe et

   $$E(X) = \int_{-\infty}^{+\infty} t\,f(t)\,\mathrm{d}t = \int_1^{+\infty} t\,f(t)\,\mathrm{d}t. \qquad \boxed{E(X) = \frac{a}{a-1}}$$

3. Pour tout réel x, $F(x) = P(X \leqslant x) = \displaystyle\int_{-\infty}^x f(t)\,\mathrm{d}t$.

   - Si $x < 1$ alors $F(x) = 0$.
   - Supposons $x \geqslant 1$. Alors $F(x) = \displaystyle\int_1^x f(t)\,\mathrm{d}t = I(x) = 1 - \frac{1}{x^a}$ d'après la question 1.(a)

4. On pose $Y = \ln(X)$ de fonction de répartition G.

   (a) Soit $x \in \mathbb{R}$.
   $G(x) = P(Y \leqslant x) = P(\ln(X) \leqslant x) = P\left(\mathrm{e}^{\ln(X)} \leqslant \mathrm{e}^x\right) = P(X \leqslant \mathrm{e}^x)$
   $= F(\mathrm{e}^x)$

   (b)
   - Supposons $x \geqslant 0$. Alors $\mathrm{e}^x \geqslant 1$. Donc d'après la question 3,
     $G(x) = F(\mathrm{e}^x) = 1 - \dfrac{1}{(\mathrm{e}^x)^a} = 1 - \dfrac{1}{\mathrm{e}^{a\,x}} = 1 - \mathrm{e}^{-a\,x}$

- Supposons $x < 0$. Alors $\mathrm{e}^x < 1$. Donc d'après la question 3, $G(x) = 0$.

Ainsi $\boxed{G(x) = \begin{cases} 1 - \mathrm{e}^{-a\,x} & \text{si} \quad x \geqslant 0 \\ 0 & \text{si} \quad x < 0 \end{cases}}$

(c) La fonction de répartition G est continue sur $\mathbb{R}$ et de classe $\mathcal{C}^1$ sur $\mathbb{R}^*$. Donc Y est une variable aléatoire à densité, dont une densité g vérifie pour tout réel x non nul :

$$g(x) = G'(x) = \begin{cases} a\,\mathrm{e}^{-a\,x} & \text{si } x > 0 \\ 0 & \text{si } x < 0 \end{cases}$$

On peut poser $g(0) = a$. On reconnaît une densité de la loi exponentielle de paramètre a.

Ainsi $\boxed{Y \hookrightarrow \mathscr{E}(a)}$ d'où $E(Y) = \dfrac{1}{a}$ et $V(Y) = \dfrac{1}{a^2}$.

5. Pour tout entier $i \in \mathbb{N}^*$, posons $Y_i = \ln(X_i)$. Alors $(Y_i)_{i \in \mathbb{N}^*}$ est une suite de variables aléatoires mutuellement indépendantes, de même loi exponentielle de paramètre a.

Donc pour tout entier $n \geqslant 2$, $\quad T_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} Y_i$

Pour tout entier $n \geqslant 2$,

$$\begin{aligned} E(T_n) &= E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} E(Y_i) \quad \text{par linéarité de l'espérance} \\ &= \frac{1}{n}\sum_{i=1}^{n} E(Y) \quad \text{car chaque } Y_i \text{ suit la même loi que } Y \\ &= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{a} = \frac{1}{n}\,na \\ &= \frac{1}{a} \end{aligned}$$

T_n est donc un estimateur sans biais de $\dfrac{1}{a}$.

Pour tout entier $n \geqslant 2$,

$$\begin{aligned} V(T_n) &= V\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) \\ &= \frac{1}{n^2}\, V\left(\sum_{i=1}^{n} Y_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} V(Y_i) \quad \text{par indépendance des } Y_i \\ &= \frac{1}{n^2}\sum_{i=1}^{n} V(Y) \quad \text{car chaque } Y_i \text{ suit la même loi que } Y \\ &= \frac{1}{n^2}\sum_{i=1}^{n} \frac{1}{a^2} = \frac{1}{n^2} \times n\frac{1}{a^2} = \frac{1}{n\,a^2} \end{aligned}$$

T_n est un estimateur sans biais de $\dfrac{1}{a}$ avec $\displaystyle\lim_{n\to+\infty} V(T_n) = 0$.

On peut conclure que T_n est un estimateur convergent de $\dfrac{1}{a}$.

6. Dans cette question $n = 100$. La commande `mean(log(X))` renvoie la moyenne $\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \ln(x_i)$

On dispose d'une réalisation de la variable aléatoire T_n : $\quad T_n(\omega) = 0,33$ qui peut être choisie comme estimation ponctuelle de $1/a$.

On prendra comme estimation ponctuelle du paramètre a : $\widehat{a} = \dfrac{1}{T_n(\omega)} \approx 3,03$.