M1 · MAS

Correction Examen Statistique du 21/12/2023

Exercice 1 - Soit la v.a. X_i représentant la note sur 20 du i^e étudiant. Les X_i sont iid, avec $X_i \sim \mathcal{N}(\mu, \sigma^2)$.

On observe $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = 11,5$; et $\hat{\sigma}_n^2 = 9$ la variance empirique.

On désire mettre en place le test suivant : (au niveau $\alpha = 5\%$) :

$$H_0 : \mu \leq \underbrace{11,2}_{\mu_0} \quad \text{contre} \quad H_1 : \mu > \underbrace{11,2}_{\mu_0}.$$

D'après le cours, on sait que la région de rejet d'un tel test vaut

$\mathcal{R} = \left\{ \overline{X}_n > \mu_0 - t_{n-1,\alpha} \frac{S_n'}{\sqrt{n}} \right\}$ avec $t_{n-1,\alpha}$ le quantile d'ordre α de la loi de Student à $(n-1)$ degrés de liberté, S_n' la racine carré de la variance débiaisée, $\mu_0 = 11,2$, et $n = 29$ étudiants.

Exercice 2

Soit X une v.a. continue, de densité

$$f_\theta(x) = \begin{cases} b - \frac{\theta}{2} & \text{si } -2 < x \leq 0 \\ b + \frac{\theta}{2} & \text{si } 0 < x < 2 . \end{cases}$$

θ est inconnu, mais $\theta \in \left]-\frac{1}{2}, \frac{1}{2}\right[$.

① - $$\int_{-\infty}^{+\infty} f_\theta(x)\,dx = 1 \iff \int_{-2}^{0} \left(b - \frac{\theta}{2}\right) dx + \int_{0}^{2} \left(b + \frac{\theta}{2}\right) dx = 1$$

donc $$\left(b - \frac{\theta}{2}\right)[x]_{-2}^{0} + \left(b + \frac{\theta}{2}\right)[x]_{0}^{2} = 1 \iff 2\left(b - \frac{\theta}{2}\right) + 2\left(b + \frac{\theta}{2}\right) = 1$$

$$2b \cancel{-\theta} + 2b \cancel{+\theta} = 1 \Rightarrow \boxed{b = 1/4 .}$$

② - $$E[X] = \int_{-\infty}^{+\infty} x f_\theta(x)\,dx = \int_{-\infty}^{+\infty} x \left[\left(\frac{1}{4} - \frac{\theta}{2}\right) \mathbb{1}_{]-2,0]}(x) + \left(\frac{1}{4} + \frac{\theta}{2}\right) \mathbb{1}_{]0,2[}(x) \right] dx$$

$$= \int_{-2}^{0} x \left(\frac{1}{4} - \frac{\theta}{2}\right) dx + \int_{0}^{2} x \left(\frac{1}{4} + \frac{\theta}{2}\right) dx = \left(\frac{1}{4} - \frac{\theta}{2}\right) \left[\frac{x^2}{2}\right]_{-2}^{0} + \left(\frac{1}{4} + \frac{\theta}{2}\right) \left[\frac{x^2}{2}\right]_{0}^{2}$$

$$= \left(\frac{1}{4} - \frac{\theta}{2}\right)\left(0 - \frac{4}{2}\right) + \left(\frac{1}{4} + \frac{\theta}{2}\right)\left(\frac{4}{2} - 0\right) = \left(\frac{1}{4} - \frac{\theta}{2}\right) \times (-2) + \left(\frac{1}{4} + \frac{\theta}{2}\right) \times 2$$

$$= -\cancel{\frac{1}{2}} + \theta + \cancel{\frac{1}{2}} + \theta = 2\theta .$$

Par la loi des grands nombres, on sait que $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow[n \to +\infty]{p.s.} E[X]$ si les X_i sont iid d'espérance finie (ce qui est le cas puisque $\theta \in \left]-\frac{1}{2}, \frac{1}{2}\right[$).

Ainsi, sachant que $\theta = \frac{E[X]}{2}$, on propose

$$\boxed{\hat{\theta}_n^{(1)} = \frac{1}{2} \overline{X}_n .}$$

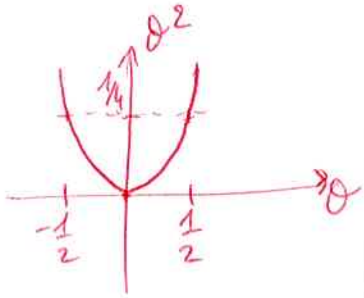
3 - • $Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

On a $\mathbb{E}[X^2] = \int_{-\infty}^{+\infty} x^2 f_\theta(x)\,dx = \int_{-2}^{0} x^2\left(\frac{1}{4}-\frac{\theta}{2}\right)dx + \int_{0}^{2} x^2\left(\frac{1}{4}+\frac{\theta}{2}\right)dx$

$= \left(\frac{1}{4}-\frac{\theta}{2}\right)\left[\frac{x^3}{3}\right]_{-2}^{0} + \left(\frac{1}{4}+\frac{\theta}{2}\right)\left[\frac{x^3}{3}\right]_{0}^{2} = \left(\frac{1}{4}-\frac{\theta}{2}\right)\left(0-\frac{-8}{3}\right) + \left(\frac{1}{4}+\frac{\theta}{2}\right)\left(\frac{8}{3}-0\right)$

$= \left(\frac{1}{4}-\frac{\theta}{2}\right)\frac{8}{3} + \left(\frac{1}{4}+\frac{\theta}{2}\right)\frac{8}{3} = \frac{8}{12} - \cancel{\frac{8\theta}{6}} + \frac{8}{12} + \cancel{\frac{8\theta}{6}} = \frac{16}{12} = \frac{4}{3}$

Ainsi $Var(X) = \frac{4}{3} - (2\theta)^2 = \frac{4}{3} - 4\theta^2 = 4\left(\frac{1}{3}-\theta^2\right)$



Ainsi si $\theta \in]-\frac{1}{2}, \frac{1}{2}[$ alors $\theta^2 < \frac{1}{4}$ et donc $\frac{1}{3} - \theta^2 > 0$

Donc oui $Var(X) > 0 \quad \forall \theta \in]-\frac{1}{2}, \frac{1}{2}[$.

• $Var(\hat{\theta}_n^{(1)}) = Var\left(\frac{1}{2}\bar{X}_n\right) = \frac{1}{4} Var(\bar{X}_n) = \frac{1}{4} Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \overset{X_i \text{ iid}}{=} \frac{1}{4n^2}\, n\, Var(X_i)$

$= \frac{1}{\cancel{4}n} \times \cancel{4}\left(\frac{1}{3}-\theta^2\right) = \boxed{\frac{1}{n}\left(\frac{1}{3}-\theta^2\right)}$

4 - • $B(\hat{\theta}_n^{(1)}) = \mathbb{E}[\hat{\theta}_n^{(1)}] - \theta = \mathbb{E}\left[\frac{1}{2}\bar{X}_n\right] - \theta = \frac{1}{2}\mathbb{E}\left[\frac{1}{n}\sum X_i\right] - \theta \overset{X_i \text{ iid}}{=}$

$\overset{X_i \text{ iid}}{=} \left(\frac{1}{2} \times \frac{1}{\cancel{n}}\cancel{n}\, \mathbb{E}[X_i]\right) - \theta = \frac{1}{\cancel{2}}\cancel{2}\theta - \theta = 0 \Rightarrow \boxed{\hat{\theta}_n^{(1)} \text{ est non-biaisé}}$

• $R(\hat{\theta}_n^{(1)}) = B^2(\hat{\theta}_n^{(1)}) + Var(\hat{\theta}_n^{(1)}) = 0 + \frac{1}{n}\left(\frac{1}{3}-\theta^2\right) = \frac{1}{n}\left(\frac{1}{3}-\theta^2\right)$

avec $R(\hat{\theta}_n^{(1)}) = \mathbb{E}\left[(\hat{\theta}_n^{(1)} - \theta)^2\right] \xrightarrow[n\to+\infty]{} 0 \Rightarrow \hat{\theta}_n^{(1)}$ CV en moyenne quadratique vers θ

$\Rightarrow \hat{\theta}_n^{(1)}$ CV en proba $\Rightarrow \hat{\theta}_n^{(1)}$ est consistant.

⑤. On souhaite calculer $\mathbb{P}(0 < X < 2)$:

$$\mathbb{P}(0 < X < 2) = \int_0^2 f_\theta(x)dx = \int_0^2 \left(\frac{1}{4}+\frac{\theta}{2}\right)dx = \left(\frac{1}{4}+\frac{\theta}{2}\right)\left[x\right]_0^2$$

Ainsi $\mathbb{P}(0 < X < 2) = \left(\frac{1}{4}+\frac{\theta}{2}\right) \times 2 = \boxed{\frac{1}{2}+\theta}$

⑥. Soit $Y = \mathbb{1}_{0<X<2}$; avec le n-échantillon $Y_i = \mathbb{1}_{0<X_i<2}$ $(i=1,\dots,n)$.

Quelle est la loi de Y ?

Y est une loi de Bernoulli, avec $Y \sim \mathcal{B}(\mathbb{E}[Y])$ et

$\mathbb{E}[Y] = \mathbb{E}[\mathbb{1}_{0<X<2}] = \mathbb{P}(0<X<2) = \frac{1}{2}+\theta$. Ainsi, $\boxed{Y \sim \mathcal{B}\left(\frac{1}{2}+\theta\right)}$

⑦. Calculons $\mathbb{P}(-2<X<2)$:

$$\mathbb{P}(-2<X<2) = \int_{-2}^2 f_\theta(x)dx = \int_{-2}^0 \left(\frac{1}{4}-\frac{\theta}{2}\right)dx + \int_0^2 \left(\frac{1}{4}+\frac{\theta}{2}\right)dx$$

$$= \left(\frac{1}{4}-\frac{\theta}{2}\right)\left[x\right]_{-2}^0 + \left(\frac{1}{4}+\frac{\theta}{2}\right)\left[x\right]_0^2 = \left(\frac{1}{4}-\frac{\theta}{2}\right)\times(-2) + \left(\frac{1}{4}+\frac{\theta}{2}\right)\times 2$$

$= -\cancel{\frac{1}{2}} + \theta + \cancel{\frac{1}{2}} + \theta = 2\theta$, or $\mathbb{P}(-2<X<2) = 1$ et cette probabilité est aussi égale à l'espérance de la v.a.

⑧. On déduit que la relation qui lie les 2 v.a. est :

$$\boxed{\mathbb{1}_{-2<X\leq 0} = 1 - \mathbb{1}_{0<X<2}}$$

9 - On remarque que : $f_\theta(x) = \left(\frac{1}{4}-\frac{\theta}{2}\right)^{\mathbb{1}_{-2<x\leq 0}} \left(\frac{1}{4}+\frac{\theta}{2}\right)^{\mathbb{1}_{0<x<2}}$

La vraisemblance vaut :

$$L(\theta; X_1,\dots,X_n) = f_{(X_1,\dots,X_n)}((x_1,\dots,x_n);\theta) \overset{X_i \text{ iid}}{=} \prod_{i=1}^{n} f_\theta(x_i)$$

$$= \prod_{i=1}^{n} \left(\frac{1}{4}-\frac{\theta}{2}\right)^{\mathbb{1}_{-2<x_i\leq 0}} \left(\frac{1}{4}+\frac{\theta}{2}\right)^{\mathbb{1}_{0<x_i<2}} = \boxed{\left(\frac{1}{4}-\frac{\theta}{2}\right)^{n-S_n} \left(\frac{1}{4}+\frac{\theta}{2}\right)^{S_n}}$$

avec $S_n = \sum_{i=1}^{n} \mathbb{1}_{0<X_i<2}$

10 - On cherche $\hat{\theta}_n = \underset{\theta}{\operatorname{argmax}}\, L(\theta; x)$.

Introduisons la log-vraisemblance :

$$\mathcal{L}(\theta) = \ln L(\theta;x) = \ln\left(\left(\frac{1}{4}-\frac{\theta}{2}\right)^{n-S_n}\left(\frac{1}{4}+\frac{\theta}{2}\right)^{S_n}\right) = (n-S_n)\ln\left(\frac{1}{4}-\frac{\theta}{2}\right) + S_n \ln\left(\frac{1}{4}+\frac{\theta}{2}\right)$$

$$\frac{\partial \mathcal{L}(\theta)}{\partial\theta} = (n-S_n)\left(-\frac{1}{2}\right)\frac{1}{\frac{1}{4}-\frac{\theta}{2}} + S_n\left(\frac{1}{2}\right)\frac{1}{\frac{1}{4}+\frac{\theta}{2}} = \frac{-(n-S_n)}{\frac{1}{2}-\theta} + \frac{S_n}{\frac{1}{2}+\theta}$$

$$= \frac{(S_n-n)\left(\frac{1}{2}+\theta\right) + S_n\left(\frac{1}{2}-\theta\right)}{\left(\frac{1}{2}-\theta\right)\left(\frac{1}{2}+\theta\right)} = \frac{\frac{S_n}{2} + \cancel{\theta S_n} - \frac{n}{2} - n\theta + \frac{S_n}{2} - \cancel{\theta S_n}}{\frac{1}{4}-\theta^2}$$

$$= \frac{S_n - n\left(\theta+\frac{1}{2}\right)}{\frac{1}{4}-\theta^2}$$

Or $\dfrac{S_n - n\left(\hat{\theta}+\frac{1}{2}\right)}{\frac{1}{4}-\hat{\theta}^2} = 0 \iff S_n - n\left(\hat{\theta}+\frac{1}{2}\right) = 0 \iff \boxed{\hat{\theta} = \frac{S_n}{n} - \frac{1}{2}}$

(11) - Soit $\hat{\theta}_n^{(2)} = \frac{S_n}{n} - \frac{1}{2}$.

- $\mathbb{E}[\hat{\theta}_n^{(2)}] = \mathbb{E}\left[\frac{S_n}{n} - \frac{1}{2}\right] = \frac{1}{n}\mathbb{E}[S_n] - \frac{1}{2} = \frac{1}{n}\mathbb{E}\Big[\underbrace{\sum_{i=1}^{n} \mathbb{1}_{0<X_i<2}}_{S_n}\Big] - \frac{1}{2}$

avec S_n une loi binomiale (somme de Bernoulli), telle que $S_n \sim \mathcal{B}(n, \frac{1}{2}+\theta)$

donc $\mathbb{E}[\hat{\theta}_n^{(2)}] = \frac{1}{\cancel{n}}\left(\cancel{n}\left(\frac{1}{2}+\theta\right)\right) - \frac{1}{2} = \frac{1}{2} + \theta - \frac{1}{2} = \theta \Rightarrow \boxed{\hat{\theta}_n^{(2)} \text{ est non-biaisé}}$

- $R(\hat{\theta}_n^{(2)}) = Var(\hat{\theta}_n^{(2)}) = Var\left(\frac{S_n}{n} - \frac{1}{2}\right) = Var\left(\frac{S_n}{n}\right) = \frac{1}{n^2} Var(S_n)$ avec $S_n \sim \mathcal{B}(n, \frac{1}{2}+\theta)$

donc $R(\hat{\theta}_n^{(2)}) = \frac{1}{n^{\cancel{2}}}\left[\cancel{n}\left(\frac{1}{2}+\theta\right)\left(1 - \left(\frac{1}{2}+\theta\right)\right)\right] = \frac{1}{n}\left[\left(\frac{1}{2}+\theta\right)\left(\frac{1}{2}-\theta\right)\right]$

$= \frac{1}{n}\left(\frac{1}{4} - \theta^2\right) \xrightarrow[n\to+\infty]{} 0 \Rightarrow \boxed{\hat{\theta}_n^{(2)} \text{ est } \begin{cases}\text{consistant} \\ \text{convergent vers } \theta\end{cases}}$

- $Eff(\hat{\theta}_n^{(2)}) = ?$ $\hat{\theta}_n^{(2)}$ est sans biais, et sera donc efficace si $Var(\hat{\theta}_n^{(2)}) = \frac{1}{I_n(\theta)}$

Or $Var(\hat{\theta}_n^{(2)}) = \frac{1}{4n}(1 - 4\theta^2)$. On calcule ensuite avec $I_n(\theta) = nI_1(\theta)$ et $I_1(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ln(f_\theta(X))\right]$ ... $\boxed{\text{Finalement, on trouve que } \hat{\theta}_n^{(2)} \text{ est efficace!}}$

(12) - On a vu que : $\begin{cases} R(\hat{\theta}_n^{(1)}) = \frac{1}{n}\left(\frac{1}{3} - \theta^2\right) \\ R(\hat{\theta}_n^{(2)}) = \frac{1}{n}\left(\frac{1}{4} - \theta^2\right) \end{cases}$

$\Rightarrow$ Comme $\theta \in]-\frac{1}{2}, \frac{1}{2}[$, qu'on ne connait pas sa valeur, il semblerait qu'il faille $\boxed{\text{privilégier } \hat{\theta}_n^{(2)} \text{ car ce dernier a un risque plus faible!}}$

13 - On rappelle que $Y_i = \mathbb{1}_{0<X_i<2}$, $i=1,\dots,n$.

Les v.a. Y_i sont iid, de moyenne et de variance finie. Ainsi en appliquant le TCL, il vient : (sachant que $Y_i \sim \mathcal{B}(\frac{1}{2}+\theta)$)

$$\sum_{i=1}^{n} Y_i \underset{n\to+\infty}{\overset{\mathcal{L}}{\sim}} \mathcal{N}\left(n\,\mathbb{E}[Y_i],\ n\,\mathrm{Var}(Y_i)\right)$$

$$\Rightarrow \quad S_n \underset{n\to+\infty}{\overset{\mathcal{L}}{\sim}} \mathcal{N}\left(n\left(\frac{1}{2}+\theta\right),\ n\left(\frac{1}{2}+\theta\right)\left(1-\left(\frac{1}{2}+\theta\right)\right)\right)$$

$$\underset{n\to+\infty}{\overset{\mathcal{L}}{\sim}} \mathcal{N}\left(\frac{n}{2}+n\theta,\ \frac{n}{4}-n\theta^2\right)$$

Ainsi, on obtient : $IC_{1-\alpha}(\theta)$ est tel que $\mathbb{P}(I \leq \theta \leq S) = 1-\alpha$.

Ici, on sait que $\underbrace{\overset{\bar{Y}_n}{\frac{S_n}{n}} - \frac{1}{2}}_{=\hat{\theta}_n^{(2)}} \sim \mathcal{N}\left(\theta,\ \frac{\frac{1}{4}-\theta^2}{n}\right)$, en utilisant le lemme de Slutsky : $\frac{1}{4}-\left(\hat{\theta}_n^{(2)}\right)^2 \xrightarrow{p.s.} \frac{1}{4}-\theta^2$

$$\Rightarrow IC_{1-\alpha}(\theta) = \left[\hat{\theta}_n^{(2)} - \underset{q_{1-\frac{\alpha}{2}}^{\mathcal{N}(0,1)}}{1{,}96}\sqrt{\frac{\frac{1}{4}-\left(\hat{\theta}_n^{(2)}\right)^2}{n}},\ \hat{\theta}_n^{(2)} + 1{,}96\sqrt{\frac{\frac{1}{4}-\left(\hat{\theta}_n^{(2)}\right)^2}{n}}\right]$$

$$\Rightarrow \boxed{IC_{1-\alpha}(\theta) = \left[\hat{\theta}_n^{(2)} - q_{1-\frac{\alpha}{2}}^{\mathcal{N}(0,1)}\sqrt{\frac{\frac{1}{4}-\left(\hat{\theta}_n^{(2)}\right)^2}{n}}\ ;\ \hat{\theta}_n^{(2)} + q_{1-\frac{\alpha}{2}}^{\mathcal{N}(0,1)}\sqrt{\frac{\frac{1}{4}-\left(\hat{\theta}_n^{(2)}\right)^2}{n}}\right]}$$

Exercice 3

On a $X=(X_1,\dots,X_n)$ un échantillon t.q. $X_i \sim \mathcal{P}(\Theta)$, avec $\Theta \sim \mathrm{Gamma}(p,\lambda)$ avec $p,\lambda>0$.

① - On cherche la loi a posteriori $\Theta | X$:

On a $\begin{cases} X | \Theta = \theta \sim \mathcal{P}(\theta) \\ \Theta \sim \text{Gamma}(p, \lambda) \end{cases}$

D'après la formule de Bayes, $f_{\Theta | X = x}(\theta) = \dfrac{\mathbb{P}(\Theta = \theta, X = x)}{\mathbb{P}(X = x)}$ avec $\begin{cases} X = (X_1, \dots, X_n) \\ x = (x_1, \dots, x_n) \end{cases}$

Ainsi $f_{\Theta | X = x}(\theta) = \dfrac{\mathbb{P}(X = x | \Theta = \theta) f_{\Theta}(\theta)}{\mathbb{P}(X_1 = x_1, \dots, X_n = x_n)} \propto \mathbb{P}(X = x | \Theta = \theta) f(\theta)$

↳ même comportement

D'où $f_{\Theta | X = x}(\theta) \overset{X_i \text{ iid}}{\propto} \prod_{i=1}^{n} e^{-\theta} \dfrac{\theta^{x_i}}{x_i!} \theta^{p-1} e^{-\lambda \theta} \times \dfrac{1}{\Gamma(p) \lambda^{-p}} \mathbb{1}_{]0;+\infty[}(\theta)$

$$\propto \frac{1}{\Gamma(p) \lambda^{-p}} \theta^{p-1} e^{-\lambda \theta} e^{-n\theta} \theta^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \frac{1}{x_i!} \times \mathbb{1}_{]0;+\infty[}(\theta)$$

$$\propto \frac{1}{\Gamma(p) \lambda^{-p}} \prod_{i=1}^{n} \frac{1}{x_i!} \times e^{-\theta(\lambda + n)} \theta^{\sum_{i=1}^{n} x_i + p - 1} \mathbb{1}_{]0;+\infty[}(\theta)$$

$\Rightarrow$ A un facteur de normalisation près, on reconnait la densité d'une loi Gamma. Ainsi, $\boxed{\Theta | X = x \sim \text{Gamma}\left(\sum_{i=1}^{n} x_i + p, \lambda + n\right)}$.

② - D'après le cours, l'estimateur bayesien associé à la perte quadratique est défini par : $\boxed{T^*((X_1, \dots, X_n)) = \mathbb{E}[\Theta | X] = \dfrac{\sum_{i}^{n} X_i + p}{\lambda + n}}$

③ - Montrons que : $R(\theta, T^*) = \dfrac{n\theta + \lambda^2 \left(\theta - \frac{p}{\lambda}\right)^2}{(n + \lambda)^2}$

On commence par s'intéresser au biais de l'estimateur $T^*(X)$:

$$\mathbb{E}_X\Big[\underbrace{\mathbb{E}[\Theta\,|\,X]}_{T^*(X)}\Big] = \mathbb{E}\left[\frac{\sum X_i + p}{\lambda+n}\right] = \frac{1}{\lambda+n}\left(\mathbb{E}\Big[\sum_{i=1}^{n} X_i\Big] + p\right)$$

$$= \frac{1}{\lambda+n}\Big(n\,\mathbb{E}[X_i] + p\Big) = \frac{1}{\lambda+n}\Big(n\times\theta + p\Big) = \frac{n\theta+p}{\lambda+n}$$

• $$\mathbb{E}\Big[\big(T^*(X)-\theta\big)^2\Big] = \mathbb{E}_X\Big[\big(\mathbb{E}[\Theta|X]-\theta\big)^2\Big]$$

$$= \Big[B\big(T^*(X)\big)\Big]^2 + Var\big(T^*(X)\big)$$

avec $B\big(T^*(X)\big) = \mathbb{E}\big[T^*(X)\big] - \theta = \dfrac{n\theta+p}{\lambda+n} - \theta = \dfrac{n\theta+p-\lambda-n}{\lambda+n}$

et $Var\big(T^*(X)\big) = Var\left(\dfrac{\sum_{i=1}^{n} X_i + p}{\lambda+n}\right) = \dfrac{1}{(\lambda+n)^2}\, n \underbrace{Var(X_i)}_{=\theta} = \dfrac{n\theta}{(\lambda+n)^2}$

Ainsi $R\big(\theta, T^*(X)\big) = \left(\dfrac{n\theta+p-\lambda-n}{\lambda+n}\right)^2 + \dfrac{n\theta}{(\lambda+n)^2} = \left(\dfrac{n\theta+p}{\lambda+n}\right)^2 + \theta^2 - 2\,\dfrac{\theta(n\theta+p)}{\lambda+n} + \dfrac{n\theta}{(\lambda+n)^2}$

$$= \frac{n^2\theta^2+p^2+2np\theta}{(\lambda+n)^2} + \theta^2 - \frac{2n\theta^2+2\theta p}{\lambda+n} + \frac{n\theta}{(\lambda+n)^2}$$

$$= \frac{n^2\theta^2+p^2+2np\theta+\theta^2(\lambda+n)^2-(2n\theta^2+2\theta p)(\lambda+n)+n\theta}{(\lambda+n)^2}$$

$$= \frac{n^2\theta^2+p^2+\cancel{2np\theta}+\theta^2(\lambda+n)^2 2n\theta^2\lambda-2n^2\theta^2-2\theta\lambda p-\cancel{2\theta np}+n\theta}{(\lambda+n)^2}$$

$$= \frac{n^2\theta^2+p^2+\theta^2(\lambda+n)^2-2n\theta^2\lambda-2n^2\theta^2-2\theta\lambda p+n\theta}{(\lambda+n)^2}$$

$$= \frac{\cancel{n^2\theta^2} + p^2 + \theta^2\lambda^2 + \cancel{2n\lambda\theta^2} + \cancel{n^2\theta^2} - \cancel{2n\theta^2\lambda} - \cancel{2n^2\theta^2} + 2\theta\lambda p + n\theta}{(n+\lambda)^2}$$

$$\left(\overset{?}{=} \frac{n\theta + \lambda^2\left(\theta^2 + \frac{p^2}{\lambda^2} - 2\theta\frac{p}{\lambda}\right)}{(n+\lambda)^2} \right)$$

$$= \frac{n\theta + p^2 + \theta^2\lambda^2 - 2\theta\lambda p}{(n+\lambda)^2} = \boxed{\frac{n\theta + \lambda^2\left(\theta - \frac{p}{\lambda}\right)^2}{(n+\lambda)^2} = R(\theta, T^*)}$$

(4) - $R^\beta = E_\Theta\left[R(\Theta, T^*)\right] = E_\Theta\left[\frac{n\theta + \lambda^2\left(\theta - \frac{p}{\lambda}\right)^2}{(n+\lambda)^2}\right]$

$$= \frac{1}{(n+\lambda)^2}\left[n\, E_\Theta(\Theta) + \lambda^2\left(\underbrace{E[\Theta^2]}_{= Var(\Theta) + (E(\Theta))^2} - \frac{2p}{\lambda}E[\Theta] + \frac{p^2}{\lambda^2}\right)\right]$$

$$= \frac{1}{(n+\lambda)^2}\left[n \times \frac{p}{\lambda} + \lambda^2\left[\left(\frac{p}{\lambda^2} + \frac{p^2}{\lambda^2}\right) - \frac{2p}{\lambda}\frac{p}{\lambda} + \frac{p^2}{\lambda^2}\right]\right]$$

$$= \frac{1}{(n+\lambda)^2}\left[\frac{np}{\lambda} + \lambda^2\left(\frac{p}{\lambda^2} + \cancel{\frac{p^2}{\lambda^2}} - \cancel{\frac{2p^2}{\lambda^2}} + \cancel{\frac{p^2}{\lambda^2}}\right)\right]$$

$$= \frac{1}{(n+\lambda)^2}\left[\frac{np}{\lambda} + \frac{p\cancel{\lambda^2}}{\cancel{\lambda^2}}\right] = \frac{1}{(n+\lambda)^{\cancel{2}}}\left(\frac{p\cancel{(n+\lambda)}}{\lambda}\right) = \boxed{\frac{p}{\lambda(n+\lambda)}}$$

5 - On considère maintenant la perte :

$$\ell(\theta,T) = \theta^3 e^{-2\theta} (\theta-T)^2 .$$

Quel est l'estimateur de Bayes associé ?

On cherche un minimiseur $T'(X)$ du risque a posteriori, qui s'écrit :

$$R'(\theta, T'(X)) = \int_\Theta \theta^3 e^{-2\theta} (\theta-T)^2 \, d\Pi(\theta|X)$$ où $\Pi(\theta|X)$ désigne la loi a posteriori.

On sait que $\Theta|X \sim \text{Gamma}(p+n\bar{X}_n, \lambda+n)$, donc on obtient :

$$R'(\theta, T'(X)) = \int_0^{+\infty} \theta^3 e^{-2\theta} (\theta-T)^2 \frac{\theta^{p+n\bar{X}_n-1} e^{-(\lambda+n)\theta}}{\Gamma(p+n\bar{X}_n)(\lambda+n)^{-(p+n\bar{X}_n)}} d\theta .$$

$$= \frac{(\lambda+n)^{p+n\bar{X}_n}}{\Gamma(p+n\bar{X}_n)} \int_0^{+\infty} (\theta-T')^2 \, \theta^{p+n\bar{X}_n+2} e^{-(\lambda+n-2)\theta} d\theta$$

En étudiant la fonction $T' \longmapsto R'(\theta, T'(X))$, on se rend compte qu'elle est proportionnelle à $\mathbb{E}[(\tilde{\theta}-T')^2|X]$ où la loi de

$\tilde{\theta}|X \sim \text{Gamma}(\underbrace{p+n\bar{X}_n+3}_{=\tilde{p}}, \underbrace{\lambda+n+2}_{=\tilde{\lambda}})$.

Comme pour toute v.a. de carré intégrable, la fonction $a \mapsto \mathbb{E}[(Z-a)^2]$ est minimale en $a = \mathbb{E}[Z]$, donc on obtient l'estimateur :

$$\boxed{T'(X) = \frac{p+n\bar{X}_n+3}{\lambda+n+2} = \frac{\tilde{p}}{\tilde{\lambda}}}$$

Année universitaire 2023-2024

Site : ☐ Luminy ☒ St-Charles ☐ St-Jérôme ☐ Cht-Gombert ☐ Aix-Montperrin ☐ Aubagne-SATIS
Sujet Session de ☒ 1er semestre - ☐ 2ème semestre - ☒ Session unique Durée de l'épreuve : 3h
Examen de ☐ L1/☐ L2/☐ L3 - ☒ M1/☐ M2 - ☐ LP - ☐ DU Nom Diplôme : MAS
Code Apogée du module : SMSAU02C Libellé du module : **Statistique.**
Document autorisé : ☒ OUI-☐ NON Calculatrice autorisée ☒ OUI-☐ NON

Examen final du jeudi 21 décembre 2023.

Les documents, notes de cours et calculatrices sont interdits, à l'exception d'une feuille A4 recto-verso. Le soin apporté à la rédaction sera un élément d'appréciation. Le barème est indicatif.

Exercice 1 (4 points). Une classe de 29 étudiants vient de passer un examen de Logiciel R. Les notes des étudiants varient de 5/20 à 18/20, et on suppose qu'elles sont des réalisations d'une loi Normale. La moyenne de la classe vaut 11,5. La variance est égale à 9. Peut-on affirmer, avec un risque de 5%, que la moyenne de cet examen est strictement plus grande que 11,2 ?

N.B. : On rappelle les quantiles suivants, où $t_{n,\alpha}$ désigne le quantile d'ordre α de la loi de Student à n degrés de liberté, et $q_\alpha^{N(0,1)}$ celui de la loi gaussienne centrée réduite.

α	95%	97,5%
$q_\alpha^{N(0,1)}$	1,64	1,96
$t_{28,\alpha}$	1,701	2,048

Exercice 2 (10 points). Soit une variable aléatoire continue X, de densité :

$$f_\theta(x) = \begin{cases} b - \frac{\theta}{2} & \text{si} \quad -2 < x \leq 0 \\ b + \frac{\theta}{2} & \text{si} \quad 0 < x < 2 \end{cases}$$

avec b une constante à déterminer. Par ailleurs, le paramètre θ est inconnu, avec $\theta \in \left]-\frac{1}{2}, \frac{1}{2}\right[$.
On considère un n-échantillon $(X_1, \ldots, X_n)$ de cette loi.

1. (0,5 point) Trouver la valeur de b.
2. (0,5) Calculer l'espérance de la variable aléatoire X. En déduire un estimateur de θ par la méthode des moments, qui sera noté $\hat{\theta}_n^{(1)}$ dans la suite.
3. (1) Calculer $Var_\theta(X)$. Est-elle > 0 pour tout $\theta \in \left]-\frac{1}{2}, \frac{1}{2}\right[$? En déduire $Var_\theta(\hat{\theta}_n^{(1)})$.
4. (0,75) Etudier maintenant le biais et la convergence de $\hat{\theta}_n^{(1)}$.
5. (0,5) Calculer la probabilité $P_\theta(0 < X < 2)$.
6. (0,5) On introduit la variable aléatoire $Y = 1_{0<X<2}$, ainsi que le n-échantillon correspondant $Y_i = 1_{0<X_i<2}$ pour $i = 1, \ldots, n$. Donner la loi de Y.

7. (0,5) Calculer la probabilité $P_\theta(-2 < X < 2)$, et déduisez quelles sont les valeurs possibles pour la variable aléatoire X.
8. (0,5) En déduire la relation entre les variables aléatoires $1_{-2<X\leq 0}$ et $1_{0<X<2}$
9. (0,75) On remarque que la densité de la variable aléatoire X peut aussi s'écrire

$$f_\theta(x) = \left(\frac{1}{4} - \frac{\theta}{2}\right)^{1_{-2<X\leq 0}} \left(\frac{1}{4} + \frac{\theta}{2}\right)^{1_{0<X<2}}.$$

Utilisez cette expression pour écrire la vraisemblance du n-échantillon $X_1, \ldots, X_n$, en fonction de $S_n = \sum_{i=1}^n 1_{0<X_i<2}$.

10. (1) En déduire l'estimateur du maximum de vraisemblance de θ.
11. (1,5) Considérons l'estimateur : $\hat{\theta}_n^{(2)} = \dfrac{S_n}{n} - \dfrac{1}{2}$. En étudier biais, lconvergence et l'efficacité.
12. (0,5) Quel estimateur faut-il privilégier entre $\hat{\theta}_n^{(1)}$ et $\hat{\theta}_n^{(2)}$, et pourquoi ?
13. (1,5) Ecrire le théorème central limite pour la suite de variables aléatoires $Y_1, \ldots, Y_n$. En utilisant les résultats des questions précédentes, et en considérant $\hat{\theta}_n^{(2)}$ comme estimateur ponctuel de θ, proposer un intervalle de confiance de seuil asymptotique α (ou de niveau de confiance asymptotique $1-\alpha$) pour le paramètre θ.

Exercice 3 (6 points). On s'intéresse à un portefeuille d'assurance automobile, et plus particulièrement au nombre d'accidents de cette population assurée. On a observé $\mathbf{x} = (x_1, \cdots, x_n)$ le nombre d'accidents de n conducteurs, et on suppose que $\mathbf{x}$ est une réalisation d'un n-échantillon $\mathbf{X} = (X_1, \cdots, X_n)$ de loi de Poisson, dont le paramètre Θ est lui-même une variable aléatoire de loi Gamma $\Gamma(p, \lambda)$ avec $p, \lambda > 0$. Cela permet de représenter la diversité des comportements conducteur (et le risque associé). On rappelle que

— la fonction de masse d'une variable aléatoire X de loi Poisson $\mathcal{P}(\theta)$ est donnée par (pour x à valeurs dans $\mathbb{N}$)

$$P_\theta(X = x) = \exp(-\theta)\frac{\theta^x}{x!}.$$

On a alors $E_\theta(X) = Var_\theta(X) = \theta$.

— la densité d'une loi Gamma, $\Gamma(p, \lambda)$, est donnée par (pour θ à valeurs dans $\mathbb{R}^{+*}$)

$$f(\theta; p, \lambda) = \frac{\theta^{p-1}\exp(-\lambda\theta)}{\Gamma(p)\lambda^{-p}}, \text{ avec } \Gamma(.) \text{ la fonction Gamma.}$$

On rappelle que $E[\Theta] = p/\lambda$ et que $Var(\Theta) = p/\lambda^2$.

1. (1) Quelle est la loi a posteriori de $\Theta \,|\, \mathbf{X}$?
2. (1,5) Donner un estimateur de Bayes associé à la perte quadratique, noté ensuite $T^\star(\mathbf{X})$.
3. (1,5) Montrer que pour tout $\theta > 0$, le risque quadratique de cet estimateur vaut

$$R(\theta, T^\star) = \frac{n\theta + \lambda^2(\theta - \frac{p}{\lambda})^2}{(n+\lambda)^2}.$$

4. (1) En déduire que le risque bayésien vaut $R^B = \dfrac{p}{\lambda(n+\lambda)}$.
5. (1) Montrer qu'un estimateur de Bayes associé à la perte $l(\theta, T) = \theta^3\exp(-2\theta)(\theta - T)^2$ vaut

$$\tilde{T}(\mathbf{X}) = \frac{n\bar{X}_n + p + 3}{\lambda + n + 2}.$$