Feuille TD 3

Exercice 1

① ⓐ $X_1, \dots, X_n$ iid de loi $\mathcal{G}(\theta)$, $\theta \in [0,1 ; 0,5]$. Ici, θ représente la proportion de vis défectueuses. Donc le paramètre d'intérêt est $\frac{1}{\theta}$.

D'après la LGN, $\overline{X}_n \simeq \mathbb{E}(X_1)$ pour n grand. Or $X_1 \sim \mathcal{G}(\theta)$ donc $\mathbb{E}(X_1) = \frac{1}{\theta}$ et $\overline{X}_n \simeq \frac{1}{\theta}$ pour n grand. De ce fait, $\overline{X}_n$ est un estimateur naturel de $\frac{1}{\theta}$.

Est-il sans biais ? $\Leftrightarrow$ a-t-on $\mathbb{E}(\overline{X}_n) = \frac{1}{\theta}$?

$$\mathbb{E}(\overline{X}_n) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\theta}, \text{ car } X_i \sim \mathcal{G}(\theta) \Rightarrow \mathbb{E}(X_i) = \frac{1}{\theta}$$

$$= \frac{1}{\theta}$$

Donc l'estimateur $\overline{X}_n$ est sans biais.

ⓑ Risque quadratique de $\overline{X}_n$:

$$R_{\overline{X}_n}(\theta) = \mathbb{E}\Big(\big(\overline{X}_n - \underbrace{\tfrac{1}{\theta}}_{= \mathbb{E}(\overline{X}_n)}\big)^2\Big)$$: représente une notion de distance entre l'estimateur $\overline{X}_n$ et la cible $\frac{1}{\theta}$.

$$= \mathrm{var}(\overline{X}_n)$$

$$= \mathrm{var}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big)$$

$$= \frac{1}{n^2}\,\mathrm{var}\Big(\sum_{i=1}^{n} X_i\Big)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}(X_i)$$ $X_1, \dots, X_n$ indép.

Rappels

- $a \in \mathbb{R}$, Z va : $\mathrm{var}(aZ) = a^2\,\mathrm{var}(Z)$
- si Z_1, Z_2 va indép, $\mathrm{var}(Z_1 + Z_2) = \mathrm{var}(Z_1) + \mathrm{var}(Z_2)$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\frac{1-\theta}{\theta^2}$$, car $X_i \sim \mathcal{G}(\theta) \overset{\text{TD1}}{\Longrightarrow} \mathrm{var}(X_i) = \frac{1-\theta}{\theta^2}$

$$= \frac{1}{n}\,\frac{1-\theta}{\theta^2}$$

Pour la majoration, on pose $f(\theta) = \frac{1-\theta}{\theta^2}$, $\theta \in [0{,}1 ; 0{,}5]$

$$f'(\theta) = \frac{-\theta^{\cancel{2}\,1} - (1-\theta)\times 2\cancel{\theta}}{\theta^{\cancel{4}\,3}} = \frac{-\theta - 2 + 2\theta}{\theta^3} = \frac{\theta - 2}{\theta^3} < 0 \text{ car } \theta \in [0{,}1 ; 0{,}5]$$

Donc f est $\searrow$ sur $[0{,}1 ; 0{,}5]$, ce qui entraîne

$$f(\theta) = \frac{1-\theta}{\theta^2} \leq \frac{1-0{,}1}{0{,}1^2} = \frac{0{,}9}{\frac{1}{100}} = 90 \quad \forall \theta \in [0{,}1 ; 0{,}5]$$

Ainsi,

$$\boxed{R_{\overline{X}_n}(\theta) = var(\overline{X}_n) \leq \frac{90}{n}}$$

ⓒ D'après l'inégalité de BT, on a $\forall a > 0$:

$$\mathbb{P}\Big(\big|\overline{X}_n - \underbrace{\tfrac{1}{\theta}}_{\mathbb{E}(\overline{X}_n)}\big| \geq a\Big) = \mathbb{P}\big(|\overline{X}_n - \mathbb{E}(\overline{X}_n)| \geq a\big) \overset{BT}{\leq} \frac{var(\overline{X}_n)}{a^2} \overset{\text{par } Q1b}{\leq} \frac{90}{na^2}$$

$$\Leftrightarrow 1 - \mathbb{P}\Big(\big|\overline{X}_n - \tfrac{1}{\theta}\big| < a\Big) \leq \frac{90}{na^2} \Leftrightarrow \mathbb{P}\Big(\big|\overline{X}_n - \tfrac{1}{\theta}\big| < a\Big) \geq 1 - \frac{90}{na^2}$$

On cherche a pour que $1-\frac{90}{na^2}=0,95 \iff \frac{90}{na^2}=0,05$

$\iff a^2=\frac{90}{0,05\,n}=\frac{1800}{n}=\frac{2\times 30^2}{n}$ donc $a=\frac{30\sqrt{2}}{\sqrt{n}}$

Ainsi,

$$\mathbb{P}\left(\left|\overline{X}_n-\frac{1}{\theta}\right|<\frac{30\sqrt{2}}{\sqrt{n}}\right)\geq 0,95$$

Rappel

$|x-a|<r$

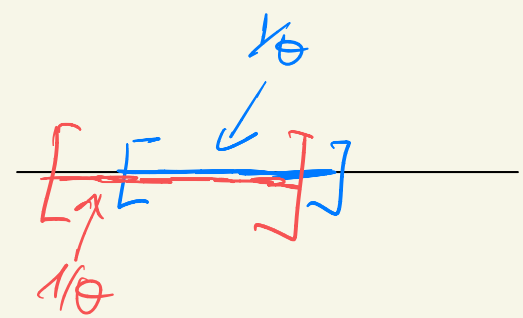
$\iff x\in\,]a-r,a+r[$

$$\iff \mathbb{P}\left(\frac{1}{\theta}\in\left]\overline{X}_n-\frac{30\sqrt{2}}{\sqrt{n}},\overline{X}_n+\frac{30\sqrt{2}}{\sqrt{n}}\right[\right)\geq 0,95$$

Donc l'IC par excès au niveau 95% pour $\frac{1}{\theta}$ est

$$\left]\overline{X}_n-\frac{30\sqrt{2}}{\sqrt{n}},\overline{X}_n+\frac{30\sqrt{2}}{\sqrt{n}}\right[$$

On sait que $0,1\leq\theta\leq 0,5 \iff 2\leq\frac{1}{\theta}\leq 10$



Donc

$$\frac{1}{\theta}\in\left[\max\left(2,\overline{X}_n-\frac{30\sqrt{2}}{\sqrt{n}}\right);\min\left(10,\overline{X}_n+\frac{30\sqrt{2}}{\sqrt{n}}\right)\right]$$ avec confiance $\geq 0,95$

② On sait que

$$\max\left(2, \bar{X}_n - \frac{30\sqrt{2}}{\sqrt{n}}\right) \leq \frac{1}{\theta} \leq \min\left(10, \bar{X}_n + \frac{30\sqrt{2}}{\sqrt{n}}\right) \quad \text{avec proba} \geq 0{,}95$$

$$\Leftrightarrow \quad \frac{1}{\min\left(10, \bar{X}_n + \frac{30\sqrt{2}}{\sqrt{n}}\right)} \leq \theta \leq \frac{1}{\max\left(2, \bar{X}_n - \frac{30\sqrt{2}}{\sqrt{n}}\right)} \quad \text{avec proba} \geq 0{,}95$$

Exercice 2

① ⓐ $\bar{X}_n$ estime $m(\theta)$ sans biais $\Leftrightarrow \mathbb{E}(\bar{X}_n) = m(\theta)$

On calcule donc

$$\mathbb{E}(\bar{X}_n) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n} m(\theta) \text{ par hypothèse}$$

$$= m(\theta)$$

Donc $\bar{X}_n$ est un estimateur sans biais de $m(\theta)$.

ⓑ "$\bar{X}_n \approx m(\theta) \Rightarrow \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \approx \frac{1}{n}\sum_{i=1}^{n}(X_i - m(\theta))^2 \overset{\text{LGN}}{\approx} \mathbb{E}\left((X_1 - m(\theta))^2\right)$ "

$$\mathbb{E}\left((X_1 - m(\theta))^2\right) = \mathrm{var}(X_1) = \sigma^2(\theta)$$

$\hat{\sigma}^2$ estime $\sigma^2(\theta)$ sans biais $\iff \mathbb{E}(\hat{\sigma}^2) = \sigma^2(\theta)$

$$\mathbb{E}(\hat{\sigma}^2) = \mathbb{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\right) = \frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}\left((X_i-\overline{X}_n)^2\right) = \frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}\left(X_i^2 + \overline{X}_n^2 - 2X_i\overline{X}_n\right)$$

$$= \frac{1}{n-1}\underbrace{\sum_{i=1}^{n}\mathbb{E}(X_i^2)}_{\substack{= var(X_i) + \mathbb{E}(X_i)^2 \\ = \sigma^2(\theta) + m(\theta)^2}} + \underbrace{\frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}\left(\overline{X}_n^2\right)}_{\frac{n}{n-1}\mathbb{E}(\overline{X}_n^2)} - \frac{2}{n-1}\underbrace{\sum_{i=1}^{n}\mathbb{E}\left(X_i\overline{X}_n\right)}_{= \mathbb{E}\left(\sum_{i=1}^{n}X_i\overline{X}_n\right) = \mathbb{E}\left(n\overline{X}_n\overline{X}_n\right) = n\mathbb{E}(\overline{X}_n^2)}$$

Rappel

$\mathbb{E}(Z^2) = var(Z) + \mathbb{E}(Z)^2$

$$= \frac{n}{n-1}\left(\sigma^2(\theta) + m(\theta)^2\right) + \frac{n}{n-1}\mathbb{E}\left(\overline{X}_n^2\right) - \frac{2n}{n-1}\mathbb{E}\left(\overline{X}_n^2\right)$$

$$= \frac{n}{n-1}\left(\sigma^2(\theta) + m(\theta)^2\right) - \frac{n}{n-1}\mathbb{E}\left(\overline{X}_n^2\right)$$

Or,

$$\mathbb{E}\left(\overline{X}_n^2\right) = \mathbb{E}\left(\frac{1}{n^2}\left(\sum_{i=1}^{n}X_i\right)^2\right) = \frac{1}{n^2}\mathbb{E}\left(\left(\sum_{i=1}^{n}X_i\right)^2\right) = \frac{1}{n^2}\left(var\left(\sum_{i=1}^{n}X_i\right) + \left(\mathbb{E}\left(\sum_{i=1}^{n}X_i\right)\right)^2\right)$$

$$\mathbb{E}(\bar{X}_n^2) = \frac{1}{n^2}\left(\sum_{i=1}^{n} \mathrm{var}(X_i) + \left(\sum_{i=1}^{n} \mathbb{E}(X_i)\right)^2\right) = \frac{1}{n^2}\left(n\,\sigma^2(\theta) + (n\,m(\theta))^2\right)$$

car $X_1, \dots, X_n$ indep.

$$= \frac{1}{n}\sigma^2(\theta) + m(\theta)^2$$

Donc,

$$\mathbb{E}(\hat{\sigma}^2) = \frac{n}{n-1}\left(\sigma^2(\theta) + m(\theta)^2\right) - \frac{n}{n-1}\left(\frac{1}{n}\sigma^2(\theta) + m(\theta)^2\right)$$

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

$$= \frac{n}{n-1}\,\frac{n-1}{n}\,\sigma^2(\theta) = \sigma^2(\theta),$$

ce qui montre que $\hat{\sigma}^2$ estime $\sigma^2(\theta)$ sans biais.

② ⓐ Rappel Le risque quadratique de l'estimateur $\hat{g}$ en θ est

$$R_{\hat{g}}(\theta) = E\left[(\hat{g} - g(\theta))^2\right],$$ si le paramètre d'intérêt est $g(\theta)$

On a vu en cours que $\hat{\theta}_1$ est sans biais, i.e. $E(\hat{\theta}_1) = \theta$. D'après la décomposition biais-variance :

$$R_{\hat{\theta}_1}(\theta) = var(\hat{\theta}_1) + (\underbrace{B_{\hat{\theta}_1}(\theta)}_{=0})^2 = var(\hat{\theta}_1) = var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n^2} var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} var(X_i),$$ car $X_1, \dots, X_n$ indep.

$$= \frac{1}{n^2}\sum_{i=1}^{n}\theta(1-\theta),$$ car $X_i \sim \mathcal{B}(\theta)$ donc $var(X_i) = \theta(1-\theta)$

$$\boxed{R_{\hat{\theta}_1}(\theta) = \frac{\theta(1-\theta)}{n}}$$

Pour $\hat{\theta}_2$, il suffit de remplacer n par m, d'où $\boxed{R_{\hat{\theta}_2}(\theta) = \frac{\theta(1-\theta)}{m}}$

Puis, pour $\hat{\theta}_3 = \frac{\hat{\theta}_1 + \hat{\theta}_2}{2}$. $\hat{\theta}_1$ et $\hat{\theta}_2$ sont sans biais, d'où

$$\mathbb{E}(\hat{\theta}_3) = \frac{1}{2}\left(\mathbb{E}(\hat{\theta}_1) + \mathbb{E}(\hat{\theta}_2)\right) = \frac{1}{2}(\theta + \theta) = \theta \quad : \quad \hat{\theta}_3 \text{ est sans biais}$$

Alors, par la décomposition biais - variance :

$$R_{\hat{\theta}_3}(\theta) = \operatorname{var}(\hat{\theta}_3) + \underbrace{\left(B_{\hat{\theta}_3}(\theta)\right)^2}_{=0} = \operatorname{var}(\hat{\theta}_3)$$

⚠ $\hat{\theta}_1$ et $\hat{\theta}_2$ ne sont pas des va indep. !

$$= \operatorname{var}\left(\frac{\hat{\theta}_1 + \hat{\theta}_2}{2}\right) = \frac{1}{4} \operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{m}\sum_{j=1}^{m} X_j\right)$$

$$= \frac{1}{4} \operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{m} X_i + \frac{1}{n}\sum_{i=m+1}^{n} X_i + \frac{1}{m}\sum_{j=1}^{m} X_j\right)$$

$$= \frac{1}{4} \operatorname{var}\left(\left(\frac{1}{n} + \frac{1}{m}\right)\underbrace{\sum_{i=1}^{m} X_i}_{\text{f° de } X_1, \dots, X_m} + \frac{1}{n}\underbrace{\sum_{i=m+1}^{n} X_i}_{\text{f° de } X_{m+1}, \dots, X_n}\right)$$

↪ les 2 sommes sont indep. car $X_1, \dots, X_m$ indep. de $X_{m+1}, \dots, X_n$

$$= \frac{1}{4}\left(\mathrm{var}\left(\left(\frac{1}{n}+\frac{1}{m}\right) \sum_{i=1}^{m} X_i \right) + \mathrm{var}\left(\frac{1}{n} \sum_{i=m+1}^{n} X_i \right) \right)$$ par indépendance

$$= \frac{1}{4}\left(\left(\frac{1}{n}+\frac{1}{m}\right)^2 m\,\theta(1-\theta) + \frac{1}{n^2} \sum_{i=m+1}^{n} \theta(1-\theta) \right)$$ par indépendance

Nb de termes de $(m+1)$ à n ?

$$\sum_{i=m+1}^{n} 1 = \sum_{i=m+1+0}^{m+1+(n-m-1)} 1 = n-m$$

Donc

$$R_{\hat{\theta}_3}(\theta) = \frac{1}{4}\left(\left(\frac{1}{n}+\frac{1}{m}\right)^2 m\,\theta(1-\theta) + \frac{n-m}{n^2}\,\theta(1-\theta) \right)$$

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

$$\boxed{R_{\hat{\theta}_3}(\theta) = \frac{\theta(1-\theta)}{4}\left(\frac{1}{m} + \frac{3}{n} \right)}$$

$$R_{\hat\theta_1}(\theta) = \frac{\theta(1-\theta)}{n}, \quad R_{\hat\theta_2} = \frac{\theta(1-\theta)}{m}, \quad R_{\hat\theta_3} = \frac{\theta(1-\theta)}{4}\left(\frac{1}{m}+\frac{3}{n}\right)$$

Comme $m < n$, $\frac{1}{n} < \frac{1}{m}$ donc

$$R_{\hat\theta_1}(\theta) < R_{\hat\theta_2}(\theta) \text{ et } R_{\hat\theta_3}(\theta) = \frac{\theta(1-\theta)}{4}\left(\frac{1}{m}+\frac{3}{n}\right) > \frac{\theta(1-\theta)}{4}\left(\frac{1}{n}+\frac{3}{n}\right) = R_{\hat\theta_1}(\theta)$$

<u>Conclusion</u> $\hat\theta_1$ est préférable à $\hat\theta_2$ et $\hat\theta_3$.

(b) On a $\mathbb{E}\left(a\hat\theta_1 + b\hat\theta_2\right) = a\mathbb{E}(\hat\theta_1) + b\mathbb{E}(\hat\theta_2) = a\theta + b\theta = (a+b)\theta$, donc l'estimateur $a\hat\theta_1 + b\hat\theta_2$ est sans biais ssi $\boxed{a+b=1}$.

On considère donc les estimateurs $a\hat\theta_1 + (1-a)\hat\theta_2 \overset{\text{not.}}{=} \hat g_a$. On a

$$\hat g_a = \frac{a}{n}\sum_{i=1}^{n} X_i + \frac{1-a}{m}\sum_{i=1}^{m} X_i = \frac{a}{n}\left(\sum_{i=1}^{m} X_i + \sum_{i=m+1}^{n} X_i\right) + \frac{1-a}{m}\sum_{i=1}^{m} X_i$$

$$= \underbrace{\left(\frac{a}{n}+\frac{1-a}{m}\right)\sum_{i=1}^{m} X_i}_{} + \underbrace{\frac{a}{n}\sum_{i=m+1}^{n} X_i}_{}$$

les va sont indep.

$$R_{\hat{g}_a}(\theta) = \mathrm{var}(\hat{g}_a) \quad \text{car } \hat{g}_a \text{ est sans biais}$$

$$= \mathrm{var}\left(\left(\frac{a}{n} + \frac{1-a}{m}\right) \sum_{i=1}^{m} X_i + \frac{a}{n} \sum_{i=m+1}^{n} X_i \right)$$

$$= \mathrm{var}\left(\left(\frac{a}{n} + \frac{1-a}{m}\right) \sum_{i=1}^{m} X_i \right) + \mathrm{var}\left(\frac{a}{n} \sum_{i=m+1}^{n} X_i \right) \quad \text{par indep.}$$

$$= \left(\frac{a}{n} + \frac{1-a}{m}\right)^2 m\,\theta(1-\theta) + \left(\frac{a}{n}\right)^2 (n-m)\,\theta(1-\theta) \quad \text{par indep.}$$

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

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

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

On pose $f(a) = \left(\frac{1}{m} - \frac{1}{n}\right) a^2 + \left(\frac{2}{n} - \frac{2}{m}\right) a + \frac{1}{m}$,

$$f'(a) = 2\Big(\underbrace{\tfrac{1}{m}-\tfrac{1}{n}}_{\geq 0}\Big)a + 2\Big(\tfrac{1}{n}-\tfrac{1}{m}\Big) . \quad \text{Donc } f'(a) \geq 0 \iff a \geq 1$$

a	$-\infty$		1		$+\infty$
f'		$-$	0	$+$	
f		$\searrow$		$\nearrow$	

Donc $f(1) \leq f(a) \quad \forall a \in \mathbb{R}$.

$$\Rightarrow \boxed{R_{\hat{g}_1}(\theta) \leq R_{\hat{g}_a}(\theta) \quad \forall a \in \mathbb{R}}$$

Ainsi, $\hat{g}_1$ est l'estimateur préférable parmi tous les estimateurs $\hat{g}_a$

$\hat{g}_1 = \hat{\theta}$!

<u>Exercice 3</u> On a vu à l'ex. 2, TD2 que le modèle stat est constitué de tous les échantillons de va indép. et de même loi $X_1, \dots, X_n$ de loi $\mathcal{B}(\alpha+\beta-2\alpha\beta)$

avec $\begin{cases} \alpha = \text{proba que pièce de A donne pile} \\ \beta = \quad " \quad\quad " \quad\quad B \quad\quad " \end{cases}$

$\alpha+\beta-2\alpha\beta$ est la proba d'avoir un vainqueur = proba qu'une seule pièce donne pile.

① D'après BT, $\forall a>0$: $\qquad p(\alpha,\beta)=\alpha+\beta-2\alpha\beta$.

$$\mathbb{P}\left(|\overline{X}_n-\underbrace{p(\alpha,\beta)}_{=\mathbb{E}(\overline{X}_n)}|\geq a\right)\leq\frac{\mathrm{Var}(\overline{X}_n)}{a^2}\underset{\substack{\text{par indep.}\\ \text{de } X_1,\dots,X_n}}{=}\frac{\mathrm{Var}(X_1)}{na^2}\underset{X_1\sim\mathcal{B}(p(\alpha,\beta))}{=}\frac{p(\alpha,\beta)(1-p(\alpha,\beta))}{na^2}\underset{\substack{x(1-x)\leq\frac{1}{4}\\ \text{si } x\in[0,1]\text{ (cf cours)}}}{\leq}\frac{1}{4na^2}$$

Donc

$$\mathbb{P}\left(|\overline{X}_n-p(\alpha,\beta)|<a\right)\geq 1-\frac{1}{4na^2}$$

On choisit a pour que $1-\frac{1}{4na^2}=0{,}95 \iff \frac{1}{4na^2}=0{,}05$

$$\iff a^2=\frac{1}{0{,}2n}=\frac{5}{n}\iff a=\sqrt{\frac{5}{n}}$$

Alors, $\mathbb{P}\left(p(\alpha,\beta) \in \left]\overline{X}_n - \sqrt{\frac{5}{n}}, \overline{X}_n + \sqrt{\frac{5}{n}}\right[\right) = \mathbb{P}\left(|\overline{X}_n - p(\alpha,\beta)| < \sqrt{\frac{5}{n}}\right) \geq 0,95.$

$\Rightarrow$ l'IC pour $p(\alpha,\beta)$ ~~par~~ excés au niveau 95% est

$$\left]\overline{X}_n - \sqrt{\frac{5}{n}}, \overline{X}_n + \sqrt{\frac{5}{n}}\right[$$

2) a) Si $x = y$, $f(x,y) = f(x,x) = 2x(1-x) \leq 2 \times \frac{1}{4} = \frac{1}{2}$ $\forall x \in [0,1]$.

Avec la contraposée :

$$f(x,y) > \frac{1}{2} \Rightarrow x \neq y.$$

b) On note tout d'abord que

$$f(\alpha,\beta) = \alpha(1-\beta) + \beta(1-\alpha) = \alpha + \beta - 2\alpha\beta = p(\alpha,\beta)$$

Avec l'expérience, on a donc

$$f(\alpha,\beta) = p(\alpha,\beta) \in \left]\overline{x}_{500} - \sqrt{\frac{5}{500}}, \overline{x}_{500} + \sqrt{\frac{5}{500}}\right[= \left]\frac{350}{500} - \frac{1}{10}, \frac{350}{500} + \frac{1}{10}\right[= \left]\frac{6}{10}, \frac{8}{10}\right[$$

avec une confiance $\geq 0,95$

Donc, avec une confiance $\geq 0{,}95$, $f(\alpha,\beta) > \frac{1}{2} \Rightarrow \alpha \neq \beta$: les pièces de A et B n'ont pas la m̄ proba. de donner pile.

Exercice 4

Rappel Z suit une loi de Poisson de paramètre $\theta > 0$, notée $\mathcal{P}(\theta)$, si

$\mathbb{P}(Z=k) = e^{-\theta}\frac{\theta^k}{k!}$, $k \in \mathbb{N}$. On a vu (cf feuille TD1) que $\mathbb{E}(Z) = \mathrm{Var}(Z) = \theta$.

①

X_i = v.a. qui modélise le nb d'accidents survenus au cours de la i-ème semaine.

Dans la suite $X_1, \dots, X_n$ va i.i.d de loi $\mathcal{P}(\theta)$, avec $\theta > 0$ (inconnue).

↪ Modèle statistique

Paramètre d'intérêt $= g(\theta) = \mathbb{P}(X_1 = 0)$ = proba. qu'aucun accident n'a eu lieu en 1 semaine.

$= e^{-\theta}$

② Par la LGN : $\overline{X}_n \approx \mathbb{E}(X_1) = \theta$, car $X_1 \sim \mathcal{P}(\theta)$.

Par insertion : $e^{-\overline{X}_n}$ est un estimateur raisonnable de $e^{-\theta}$

Considérons un autre estimateur :

$$\hat{g} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i = 0\}} \approx \mathbb{E}\left(\mathbb{1}_{\{X_1 = 0\}}\right) \text{ par la LGN.}$$

$$\mathbb{P}(X_1 = 0) = e^{-\theta}, \text{ car } X_1 \sim \mathcal{P}(\theta).$$

Donc $\hat{g} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i = 0\}}$ est un estimateur naturel de $e^{-\theta}$.

$$B_{\hat{g}}(\theta) \overset{\text{def}}{=} \mathbb{E}(\hat{g}) - g(\theta) = \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i = 0\}}\right) - e^{-\theta}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left(\mathbb{1}_{\{X_i = 0\}}\right) - e^{-\theta} = \frac{1}{n} \sum_{i=1}^{n} \underbrace{\mathbb{P}(X_i = 0)}_{= e^{-\theta}, \text{ car } X_i \sim \mathcal{P}(\theta)} - e^{-\theta}$$

$$= e^{-\theta} - e^{-\theta} = 0 \; : \; \boxed{\hat{g} \text{ est sans biais}}$$

$$R_{\hat g}(\theta) = \mathbb{E}\left(\left(\hat g - g(\theta)\right)^2\right) = \mathbb{E}\left(\left(\hat g - e^{-\theta}\right)^2\right) = \operatorname{var}(\hat g)$$, car $\hat g$ sans biais

(th. de décomposition biais-variance)

$$= \operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i=0\}}\right) = \frac{1}{n^2}\operatorname{var}\left(\sum_{i=1}^{n} \mathbb{1}_{\{X_i=0\}}\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{var}\left(\mathbb{1}_{\{X_i=0\}}\right)$$, car $\mathbb{1}_{\{X_1=0\}}, \dots, \mathbb{1}_{\{X_n=0\}}$ indép.

$$= \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{var}\left(\mathbb{1}_{\{X_1=0\}}\right)$$, car " " de même loi.

$$= \frac{1}{n}\operatorname{var}\left(\mathbb{1}_{\{X_1=0\}}\right) = \frac{1}{n}\left(\mathbb{E}\left(\mathbb{1}^{\cancel{2}}_{\{X_1=0\}}\right) - \underbrace{\mathbb{E}\left(\mathbb{1}_{\{X_1=0\}}\right)}_{=\mathbb{P}(X_1=0)=e^{-\theta}}{}^2\right)$$

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

③ D'après l'inégalité BT, $\forall a>0$:

$$\mathbb{P}\left(|\hat{g}-e^{-\theta}|\geq a\right) \underset{\hat{g} \text{ sans biais}}{=} \mathbb{P}\left(|\hat{g}-\mathbb{E}(\hat{g})|\geq a\right) \overset{BT}{\leq} \frac{\mathrm{Var}(\hat{g})}{a^2} \overset{\text{car } \hat{g} \text{ sans biais}}{=} \frac{R_{\hat{g}}(\theta)}{a^2} \underset{Q2}{=} \frac{e^{-\theta}(1-e^{-\theta})}{n a^2}$$

Or, $e^{-\theta}\in\,]0,1[$ donc (cf cours) : $e^{-\theta}(1-e^{-\theta})\leq \frac{1}{4}$ $\forall \theta>0$. Alors,

$$\mathbb{P}\left(|\hat{g}-e^{-\theta}|\geq a\right)\leq \frac{1}{4na^2} \iff 1-\mathbb{P}\left(|\hat{g}-e^{-\theta}|<a\right)\leq \frac{1}{4na^2}$$

$$\iff \mathbb{P}\left(|\hat{g}-e^{-\theta}|<a\right)\geq 1-\frac{1}{4na^2}$$

On choisit $a>0$ t.q.

$$1-\frac{1}{4na^2}=1-\alpha \iff a^2=\frac{1}{4n\alpha} \iff a=\frac{1}{\sqrt{4n\alpha}}$$

Donc

$$\mathbb{P}\left(|\hat{g} - e^{-\theta}| < \frac{1}{\sqrt{4n\alpha}}\right) \geq 1-\alpha$$

$$\Leftrightarrow \mathbb{P}\left(e^{-\theta} \in \left]\hat{g} - \frac{1}{\sqrt{4n\alpha}}, \hat{g} + \frac{1}{\sqrt{4n\alpha}}\right[\right) \geq 1-\alpha$$

Par suite, l'IC par excès au niveau $1-\alpha$ pour $e^{-\theta}$ est

$$\boxed{\left]\hat{g} - \frac{1}{\sqrt{4n\alpha}}, \hat{g} + \frac{1}{\sqrt{4n\alpha}}\right[}$$

AN $n=250$: on a relevé les accidents pendant 25 semaines, la fréquence des semaines qui n'ont pas vu d'accidents est 20%.

Prenons $\alpha = 0,1$.

La réalisation de $\hat{g}$ est $0,2$ donc, avec une confiance $\geq 90\%$, la proba de ne pas avoir d'accident dans une semaine est dans l'intervalle $]0,2-0,1\,;\,0,2+0,1[=]0,1\,;\,0,3[$

Exercice 5

Rappel Si $Z \sim \mathcal{U}(\{1, \dots, N\})$ avec $N \in \mathbb{N}^*$.

$$\mathbb{E}(Z) = \sum_k k\,\mathbb{P}(Z = k) = \sum_{k=1}^{N} \frac{k}{N} = \frac{1}{N} \sum_{k=1}^{N} k = \frac{1}{N} \frac{N(N+1)}{2} = \frac{N+1}{2}$$

$$\text{var}(Z) = \frac{N^2 - 1}{12}$$

① $B_{\hat{k}_1}(k) = \mathbb{E}(\hat{k}_1) - k = \mathbb{E}(2\overline{X}_n - 1) - k = 2\,\mathbb{E}(\overline{X}_n) - 1 - k$

$$= 2\,\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - k - 1 = \frac{2}{n}\sum_{i=1}^{n} \mathbb{E}(X_i) - k - 1$$

$= \frac{2}{n} \sum_{i=1}^{n} \frac{k+1}{2} - k - 1$, car $X_i \sim \mathcal{U}(\{1, \dots, k\})$ donc $\mathbb{E}(X_i) = \frac{k+1}{2}$

$$= k + 1 - k - 1 = 0$$

Donc $\boxed{\hat{k}_1 \text{ est sans biais}}$

$$R_{\hat{k}_1}(k) = \mathrm{var}(\hat{k}_1)$$, car $\hat{k}_1$ est sans biais (th. de décomposition biais - variance)

$$= \mathrm{var}(2\bar{X}_n - 1) = 4\,\mathrm{var}(\bar{X}_n) = 4\,\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{4}{n^2}\sum_{i=1}^{n}\mathrm{var}(X_i)$$, car $X_1, \dots, X_n$ indep.

$$= \frac{4}{n^2}\sum_{i=1}^{n}\frac{k^2-1}{12} = \frac{k^2-1}{3n}$$

② Soit $N \in \mathbb{N}^*$ connu. On suppose, dans le modèle stat, que $\boxed{k \leq N}$

D'après BT, $\forall a > 0$:

$$\mathbb{P}(|\hat{k}_1 - k| \geq a) = \mathbb{P}(|\hat{k}_1 - \mathbb{E}(\hat{k}_1)| \geq a) \underset{BT}{\leq} \frac{\mathrm{var}(\hat{k}_1)}{a^2} \underset{\hat{k}_1 \text{ sans biais}}{=} \frac{R_{\hat{k}_1}(k)}{a^2} \underset{Q1}{=} \frac{k^2-1}{3n a^2}$$

Comme $k \leq N$ (connu) :

$$\mathbb{P}(|\hat{k}_1 - k| \geq a) \leq \frac{N^2}{3n a^2} \iff 1 - \mathbb{P}(|\hat{k}_1 - k| < a) \leq \frac{N^2}{3n a^2}$$

$$\mathbb{P}\left(|\hat{k}_1 - k| < a\right) \geq 1 - \frac{N^2}{3na^2}$$

On choisit $a>0$ t.q.

$$1 - \frac{N^2}{3na^2} = 0{,}95 \iff \frac{N^2}{3na^2} = 0{,}05 \iff a^2 = \frac{N^2}{3n\,0{,}05}$$

$$\iff a = \frac{N}{\sqrt{0{,}15n}}$$

Ainsi,

$$\mathbb{P}\left(|\hat{k}_1 - k| < \frac{N}{\sqrt{0{,}15n}}\right) \geq 0{,}95$$

$$\iff \mathbb{P}\left(k \in \left]\hat{k}_1 - \frac{N}{\sqrt{0{,}15n}}, \hat{k}_1 + \frac{N}{\sqrt{0{,}15n}}\right[\right) \geq 0{,}95$$

L'IC par excès au niveau 95% pour k est donc

$$\left]\hat{k}_1 - \frac{N}{\sqrt{0{,}15n}}, \hat{k}_1 + \frac{N}{\sqrt{0{,}15n}}\right[$$