ISFA- 2ème année (M1)

STATISTIQUE INFÉRENTIELLE

FICHE DE TD N° 1

Exercice 1.

Dans le cadre d'un modèle paramétrique réel d'échantillonnage, on considère un n-échantillon $X = (X_1, \dots, X_n)$, de loi de probabilité P_θ.

1. Pour un estimateur T_n de θ donné, montrer les propriétés suivantes :
   a. $R(T_n, \theta) = B^2(T_n) + \mathrm{Var}(T_n)$.
   b. Si T_n est un estimateur sans biais de θ et $\mathrm{Var}(T_n) \to 0$ quand $n \to 0$, alors T_n est un estimateur convergent pour θ.
2. Si P_θ est une loi de poisson de paramètre $\theta > 0$, écrire le modèle statistique correspondant. On considère l'estimateur de θ :
$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$
   N.B : cette notation de la moyenne empirique sera utilisée dans l'ensemble du TD.

   Calculer la moyenne et la variance de $\overline{X}$. Cet estimateur est-il sans biais ? Convergent ?
3. Si P_θ est une loi de Bernoulli de paramètre $\theta \in [0,1]$, écrire le modèle statistique, puis vérifier que l'estimateur :
$$T_n = \overline{X}\left(1 - \overline{X}\right)$$
   de $\theta(1-\theta)$ est biaisé. Donner un estimateur sans biais de $\theta(1-\theta)$.

Exercice 2.

Soit $X = (X_1, \dots, X_n)$, un n-échantillon de loi uniforme sur $[0, \theta]$. Écrire le modèle statistique. On considère l'estimateur de θ :
$$X_{\max} := \max_{1 \le i \le n} X_i.$$

1. Déterminer la loi de $X_{\max}$ et calculer son espérance et sa variance.
2. Comparer cet estimateur avec l'estimateur sans biais de θ construit avec $\overline{X}$.

Exercice 3.

Soit un modèle paramétrique réel d'échantillonnage $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_\theta, \theta \in \mathbb{R}))^n$, tel que la loi de probabilité P_θ admette pour densité :
$$f(x, \theta) = e^{-(x-\theta)} \mathbf{1}_{x \ge \theta}, \text{ avec } \theta \in \mathbb{R}.$$
Soit le vecteur aléatoire $X = (X_1, \dots, X_n)$, associé à ce modèle.

1. Vérifier que $U_1 = X_1 - \theta$ suit sous P_θ une loi exponentielle de paramètre 1.
2. Comparer les estimateurs :

$$Y_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - 1) \text{ et } X_{\min} = \min_{1\leq i\leq n} X_i.$$

Exercice 4.

Soit $X = (X_1, \ldots, X_n)$ un n-échantillon de loi normale $\mathcal{N}(m, \sigma^2)$. Pour estimer σ^2, on pose :

$$T_n = c(n)\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2.$$

1. Quelle est la loi de $\sigma^{-2}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$?
2. calculer $B(T_n)$, et donner un estimateur sans biais de σ^2.
3. Quelle est la fonction $c(n)$ qui minimise le risque quadratique de T_n ?
4. On suppose que $\sigma = 1$ et $m \in [0,1]$, et on définit l'estimateur :

$$U_n = \begin{cases} 0 & \text{si } \overline{X} < 0 \\ \overline{X} & \text{si } 0 \leq \overline{X} \leq 1 \\ 1 & \text{si } \overline{X} > 1 \end{cases}.$$

   a. Montrer que U_n est un estimateur de m, strictement meilleur que $\overline{X}$, en calculant la différence entre les risques quadratiques.

   b. Montrer que

$$\mathrm{E}(U_n) = \frac{1}{\sqrt{2\pi}}\int_{-\sqrt{n}\theta}^{\sqrt{n}(1-\theta)}\left(\theta + \frac{t}{\sqrt{n}}\right)e^{-\frac{t^2}{2}}dt + \mathrm{P}\left(\overline{X} > 1\right).$$

   c. En déduire que $E_m(U_n) \to \theta$ lorsque $n \to \infty$.

Exercice 5.

Afin de diminuer les sinistres de ses clients, un assureur décide de financer la construction d'une digue destinée à empêcher les inondations provoquées par les crues d'une rivière. La construction d'une digue de hauteur h coûtera $c_1 h$ à la compagnie d'assurance. En cas de crue, il n'y aura aucun sinistre si la hauteur H de la crue est inférieure à h et un sinistre évalué à $c_2(H-h)$ si $H > h$. La hauteur de la crue H est une variable aléatoire de loi exponentielle de paramètre $1/\theta$. On suppose que $c_2 > c_1 > 0$.

1. On choisit comme fonction de perte :

$$L(\theta, h) = c_1 h + c_2\mathrm{E}_\theta\left[(H-h)\,\mathbf{1}_{H>h}\right].$$

   Calculer $L(\theta, h)$ si θ est connu, quelle hauteur choisir ?

2. Ayant observé n crues indépendantes de hauteurs $H_1, \dots, H_n$, l'assureur qui ne connaît pas θ, décide de fixer la hauteur de la digue à $d_n = k\overline{H}$

   a. Évaluer la limite (quand $n \to 0$) du risque de cette décision.

   b. Quelle valeur donneriez-vous à k ?

Exercice 6.

Soient $X = (X_1, \dots, X_n)$ un n-échantillon de loi de bernoulli de paramètre $\theta \in [0,1]$, a et b deux constantes positives. On considère les estimateurs de θ :

$$T_{a,b} = \frac{n\overline{X} + a}{n + a + b}.$$

1. Calculer $R(T_{a,b}, \theta)$.
2. Comparer à l'aide des risques quadratiques, les estimateurs $T_{\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2}}$ et $T_{0,0}$ selon les valeurs de θ. Le critère du risque quadratique est-il approprié pour comparer ces estimateurs ?
3. Quel est parmi ces deux estimateurs celui qui minimise le maximum du risque quadratique ? *Cet estimateur est dit meilleur au sens mini-max.*
4. On suppose maintenant que θ est une variable aléatoire suivant une loi uniforme sur $[0,1]$. Calculer r_1 et r_2 définis par :

$$r_1 = \mathrm{E}\left(R(T_{0,0}, \theta)\right) \quad \text{et} \quad r_2 = \mathrm{E}\left(R(T_{\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2}}, \theta)\right).$$

Exercice 1.

Dans le cadre d'un modèle paramétrique réel d'échantillonnage, on considère un n-échantillon $X = (X_1, \ldots, X_n)$, de loi de probabilité P_θ.

1. Pour un estimateur T_n de θ donné, montrer les propriétés suivantes :
   a. $R(T_n, \theta) = B^2(T_n) + \mathrm{Var}(T_n)$.
   b. Si T_n est un estimateur sans biais de θ et $\mathrm{Var}(T_n) \to 0$ quand $n \to 0$, alors T_n est un estimateur convergent pour θ.
2. Si P_θ est une loi de poisson de paramètre $\theta > 0$, écrire le modèle statistique correspondant. On considère l'estimateur de θ :
$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$
N.B : cette notation de la moyenne empirique sera utilisée dans l'ensemble du TD.
Calculer la moyenne et la variance de $\overline{X}$. Cet estimateur est-il sans biais ? Convergent ?
3. Si P_θ est une loi de Bernoulli de paramètre $\theta \in [0, 1]$, écrire le modèle statistique, puis vérifier que l'estimateur :
$$T_n = \overline{X}\,(1 - \overline{X})$$
de $\theta(1-\theta)$ est biaisé. Donner un estimateur sans biais de $\theta(1-\theta)$.

Exercice 1 :

1) $X_1, \ldots, X_n$ iid tq $X_i \sim P_\theta$, T_n stat de θ

a) $R(T_n, \theta) = E[(T_n - \theta)^2]$
$= E[T_n^2] - 2\theta E[T_n] + \theta^2$
$= E[T_n^2] - E[T_n]^2 + E[T_n]^2 - 2\theta E[T_n] + \theta^2$
$= V(T_n) + (E[T_n] - \theta)^2$

ccl : $R(T_n, \theta) = B^2(T_n) + V(T_n)$

b) Inégalité de

$$P(|T_n - E[T_n]| > \varepsilon) \leq \frac{V(T_n)}{\varepsilon^2}$$

Sans biais $E[T_n] = \theta$ et $V(T_n) \xrightarrow[n\to\infty]{} 0$

Donc $P(|T_n - \theta| > \varepsilon) \xrightarrow[n\to\infty]{} 0$

D'où T_n converge pour θ

2) $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mathcal{P}_\theta, \theta \in \mathbb{R}_+^*)^n$

$E[\bar{X}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n} \cdot n E[X_1]$

$\Rightarrow E[\bar{X}] = \theta$. $\bar{X}$ sans biais

$V(\bar{X}) = \frac{1}{n^2} \cdot n E[X_1] = \frac{\theta}{n} \xrightarrow[n\to\infty]{} 0$. $\bar{X}$ converge pour θ

3) $(\{0,1\}, \mathcal{P}(\{0,1\}), \mathcal{B}(\theta)\ \theta \in [0;1])^n$

$E[T_n] = E[\bar{X}] - E[\bar{X}^2]$
$= \theta - (V(\bar{X}) + E[\bar{X}]^2)$
$= \theta - \left(\frac{\theta(1-\theta)}{n} + \theta^2\right)$

$$\mathbb{E}[T_n] = \theta(1-\theta) - \frac{\theta(1-\theta)}{n} = \frac{n-1}{n}\,\theta(1-\theta)$$

$$B_{\theta(1-\theta)}(T_n) = \mathbb{E}[T_n] - \theta(1-\theta)$$

$$B_{\theta(1-\theta)}(T_n) = -\frac{\theta(1-\theta)}{n}$$ T_n est biaisé.

$$T_n^* = \frac{n}{n-1}\,T_n \quad , \quad \mathbb{E}[T_n^*] = \frac{n}{n-1}\,\mathbb{E}[T_n] = \theta(1-\theta)$$

T_n^* est un estimateur sans biais de $\theta(1-\theta)$.

<u>Rq sur 1b</u> : Si $B_\theta(T_n) \longrightarrow \theta$ et $V(T_n) \longrightarrow 0$ alors $T_n \xrightarrow{\mathbb{P}} 0$

<u>Rappel</u> : $X_n \xrightarrow{\mathcal{L}^2} X \implies X_n \xrightarrow{\mathcal{L}^1} X \implies X_n \xrightarrow{\mathbb{P}} X$

$\mathbb{E}[|X_n - X|^2] \longrightarrow 0$

<u>Ici</u>, $\mathbb{E}[|T_n - \theta|^2] = R(T_n,\theta) = \underbrace{B_\theta(T_n)}_{\downarrow 0}{}^2 + \underbrace{V(T_n)}_{\downarrow 0} \longrightarrow 0 \implies T_n \xrightarrow{\mathcal{L}^2} \theta$

$\implies T_n \xrightarrow{\mathbb{P}} \theta$.

Exercice 2.

Soit $X = (X_1, \dots, X_n)$, un n-échantillon de loi uniforme sur $[0, \theta]$. Écrire le modèle statistique. On considère l'estimateur de θ :

$$X_{\max} := \max_{1 \le i \le n} X_i.$$

1. Déterminer la loi de $X_{\max}$ et calculer son espérance et sa variance.
2. Comparer cet estimateur avec l'estimateur sans biais de θ construit avec $\overline{X}$.

espace des résultats là où vivent les X_i — Tribu — famille de probas indexées par $\theta \in \Theta$ — espace des paramètres = valeurs possibles de θ — dimens° de θ — taille de l'échantillon

$$(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_\theta, \theta \in \Theta \subseteq \mathbb{R}^p)^n$$

Exercice 2

$X_{max} = \max_{1 \le i \le n} X_i$. Modèle statistique : $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{U}[0;\theta], \theta \in \mathbb{R}_+^*)^n$

1) • **Loi de X_{max}** : Soit $x \in [0;\theta]$, $\mathbb{P}(X_{max} \le x) \underset{X_i \text{ iid}}{=} \mathbb{P}(X \le x)^n = \left(\frac{x}{\theta}\right)^n$

• **Densité** : $f_{X_{max}}(x) = \frac{n}{\theta^n} x^{n-1} \mathbb{1}_{[0;\theta]}(x)$

$\Rightarrow$ • **Espérance** : $\mathbb{E}[X_{max}] = \int x f_{X_{max}}(x)dx = \int_0^\theta x \frac{n}{\theta^n} x^{n-1} dx = \int_0^\theta \frac{n}{\theta^n} x^n dx = \frac{n}{n+1}\theta$.

• **Variance** : $V(X_{max}) = \mathbb{E}[X_{max}^2] - \mathbb{E}[X_{max}]^2$

$$= \int x^2 f_{X_{max}}(x)dx - \left(\frac{n}{n+1}\theta\right)^2$$

$$= \int_0^\theta \frac{n}{\theta^n} x^{n+1} dx - \frac{n^2}{(n+1)^2}\theta^2 = \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2$$

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

Donc $V(X_{max}) = \frac{n\theta^2}{(n+1)^2(n+2)}$

2) **Rappel** : Estimateur ou Statistique $T_n = $ fct° de $X_1, \dots, X_n$

Critère de qualité de T_n :

- **Biais** : $B_\theta(T_n) = \mathbb{E}[T_n] - \theta$
- **Convergence** : $T_n \xrightarrow{\mathbb{P}} \theta$ ie $\forall \varepsilon > 0, \mathbb{P}(|X_n - \theta| > \varepsilon) \longrightarrow 0$

  ou $B_\theta(T_n) \longrightarrow 0$ et $V(T_n) \longrightarrow 0$

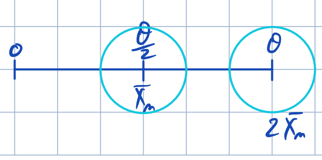
- **Risque quadratique** : $R(T_n, \theta) = \mathbb{E}[(T_n - \theta)^2] = B_\theta(T_n)^2 + V(T_n)$

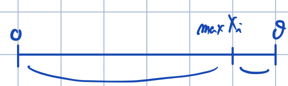
On a 2 façons de voir θ :

- $\theta = 2\mathbb{E}[X]$ donc comme $\overline{X}_n$ est un estimateur sans biais et convergent pour $\frac{\theta}{2}$, on peut estimer θ par $2\overline{X}_n$

- θ est la valeur max théorique de la distribut°, on peut l'estimer par $\max_{1 \le i \le n} X_i$

$T_n = 2\overline{X}_n$ $\qquad\qquad$ $X_{max} = \max_{i \in [1,n]} X_i$





Rappel : Pour comparer des estimateurs, on utilise le risque quadratique et on "préfère" celui qui a le + petit.

Etude de T_n

Biais $\mathbb{E}[T_n] = \mathbb{E}[2\bar{X}_n] = 2\,\mathbb{E}[\bar{X}_n] \underset{\bar{X}_n \text{ SB pour } \theta}{=} 2\,\frac{\theta}{2} = \theta \implies B_\theta(T_n) = 0$

Variance : $V(T_n) = 4\,V(\bar{X}_n) = 4\,\frac{V(X)}{n} = 4 \times \frac{\theta^2}{12n} = \frac{\theta^2}{3n} \longrightarrow 0$ donc convergent pour θ.

Risque quadratique : $\mathcal{R}(T_n, \theta) = B_\theta(T_n)^2 + V(T_n) = \frac{\theta^2}{3n}$

Etude de X_{max} :

- **Loi de X_{max} :** $\mathbb{P}(X_{max} \leq x) = \left(\frac{x}{\theta}\right)^n \quad x \in [0;\theta]$
- **Densité :** $f_{X_{max}}(x) = \frac{n x^{n-1}}{\theta^n}\,\mathbb{1}_{[0;\theta]}(x)$
- **Biais de X_{max} :**

$$\mathbb{E}[X_{max}] = \frac{n}{n+1}\,\theta$$

$$B_\theta(X_{max}) = \mathbb{E}[X_{max}] - \theta = -\frac{\theta}{n+1}$$

- **Variance de X_{max} :** $V(X_{max}) = \frac{n\theta^2}{(n+1)^2(n+2)} \xrightarrow[n\to\infty]{} 0 \implies X_{max}$ convergent pour θ.

- **Risque quadratique :**

$$\mathcal{R}(X_{max}, \theta) = B_\theta(X_{max})^2 + V(X_{max}) = \frac{\theta^2}{(n+1)^2} + \frac{n\theta^2}{(n+1)^2(n+2)}$$

$$= \frac{2\theta^2}{(n+1)(n+2)} < \mathcal{R}(T_n, \theta) = \frac{\theta^2}{3n} \quad \text{si } n \geq 3.$$

Rappel : Si T_n est tq $\mathbb{E}[T_n] = a\theta + b$ alors $\frac{T_n - b}{a}$ est SB pour θ.

Ici, on a $\mathbb{E}[X_{max}] = \frac{n}{n+1}\,\theta$ donc $X^*_{max} = \frac{n+1}{n}\max\limits_{i\in[1,n]} X_i$ est sans biais pour θ.

$$\implies \mathbb{E}[X^*_{max}] = \theta. \implies B_\theta(X^*_{max}) = 0$$

et $V(X^*_{max}) = V\left(\frac{n+1}{n}X_{max}\right) = \frac{(n+1)^2}{n^2}V(X_{max}) = \frac{\theta^2}{n(n+2)}$

$$\implies \mathcal{R}(X^*_{max}, \theta) = \underbrace{B_\theta(X^*_{max})}_{=0} + V(X^*_{max}) = V(X^*_{max}) = \frac{\theta^2}{n(n+2)}$$

Comparaison entre X_{max} et X^*_{max} : $\frac{\mathcal{R}(X_{max}, \theta)}{\mathcal{R}(X^*_{max}, \theta)} = \frac{2\cancel{\theta^2}}{(n+1)\cancel{(n+2)}} \times \frac{n\cancel{(n+2)}}{\cancel{\theta^2}} = \frac{2n}{n+1} > 1$ si $n \geq 2$

$$\Rightarrow \mathcal{R}(X_{max}, \theta) > \mathcal{R}(X^*_{max}, \theta) \quad si\ n \geq 2$$

Conclusion : Top 3 des meilleurs estimateur de θ *

* Parmi une sélect° d'estimateurs

1- X^*_{max}

2- X_{max}

3- T_n

Exercice 3.

Soit un modèle paramétrique réel d'échantillonnage $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_\theta, \theta \in \mathbb{R}))^n$, tel que la loi de probabilité P_θ admette pour densité :

$$f(x,\theta) = e^{-(x-\theta)} \mathbf{1}_{x \geq \theta}, \text{ avec } \theta \in \mathbb{R}.$$

Soit le vecteur aléatoire $X = (X_1, \dots, X_n)$, associé à ce modèle.

1. Vérifier que $U_1 = X_1 - \theta$ suit sous P_θ une loi exponentielle de paramètre 1.
2. Comparer les estimateurs :

$$Y_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - 1) \text{ et } X_{\min} = \min_{1 \leq i \leq n} X_i.$$

Exercice 3 :

On a $X_1, \dots, X_n$ iid tq $X_i \sim P_\theta$ qui admet pour densité

$$f(x,\theta) = e^{-(x-\theta)} \mathbf{1}_{\{x \geq \theta\}} \approx \text{loi exponentielle décalée de } \theta$$

θ peut être vu comme :

- quand on calcule $\mathbb{E}[X]$, ça dépend de $\theta \longrightarrow$ on estime $\hat{\theta}_n$
- θ est la valeur théorique de la distribut°, on estime par $\min_{i \in [\![1,n]\!]} X_i$

1) On pose $U_1 = X_1 - \theta$

$$F(x,\theta) = \mathbb{P}(X_1 \leq x) = \int_\theta^x e^{-(t-\theta)}\,dt = \left[-e^{-(t-\theta)}\right]_\theta^x = 1 - e^{-(x-\theta)} \quad si\ x \geq \theta$$

$$\mathbb{P}(\underset{\in \mathbb{R}_+}{U_1} \leq x) = \mathbb{P}(X_1 - \theta \leq x) = \mathbb{P}(X_1 \leq x + \theta) = \underbrace{1 - e^{-x}}_{\text{f.d.r d'une } \mathcal{E}(1)} \quad si\ x \geq 0$$

$$\Rightarrow X_1 - \theta = U_1 \sim \mathcal{E}(1)$$

2) On pose

$$\underbrace{Y_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - 1)}_{\text{estimateur des moments}} \quad et \quad \underbrace{X_{min} = \min_{i \in [\![1,n]\!]} X_i}_{\text{estimateur du max de vraisemblance}}$$

Etude de Y_n :

On a $U_1 = X_1 - \theta \sim \mathcal{E}(1)$; $\mathbb{E}[U_1] = \mathbb{E}[X_1 - \theta] = 1$; $V(U_1) = V(X_1 - \theta) = 1$

$\Rightarrow \mathbb{E}[X_1] = 1 + \theta$ et $V(X_1) = V(X_1 - \theta) = 1$

Biais :

$$\mathbb{E}[Y_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - 1)\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i - 1] = \theta \Rightarrow \text{Estimateur sans biais.}$$

Variance :

$$V(Y_n) = V\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - 1)\right) = \frac{1}{n^2}\sum_{i=1}^{n}\underbrace{V(X_i - 1)}_{=V(X_i)=1} = \frac{1}{n} \xrightarrow[\infty]{} 0 \Rightarrow \text{Convergent pour } \theta.$$

Risque quadratique :

$$R(Y_n, \theta) = B_\theta(Y_n)^2 + V(Y_n) = \frac{1}{n}$$

Etude de X_{min} :

Loi de X_{min} :

$$\mathbb{P}(X_{min} \leq x) = \mathbb{P}(\min_{i \in [\![1,n]\!]} X_i \leq x) = 1 - \mathbb{P}(\min_{i \in [\![1,n]\!]} X_i \geq x)$$

$$= 1 - \mathbb{P}(X_1 > x, \ldots, X_n > x) \underset{\text{indep}}{=} 1 - \mathbb{P}(X_1 > x) \ldots \mathbb{P}(X_n > x)$$

$$= 1 - \mathbb{P}(X_1 > x)^n = 1 - (1 - (1 - e^{-(x-\theta)}))^n$$

$$= 1 - e^{-n(x-\theta)} \quad \text{si } x \geq \theta.$$

Si on pose $V_n = \min_{i \in [\![1,n]\!]} X_i - \theta$, on a $V_n \sim \mathcal{E}(n)$

$$\Rightarrow \mathbb{E}[V_n] = \frac{1}{n} = \mathbb{E}[X_{min} - \theta] \Rightarrow \mathbb{E}[X_{min}] = \theta + \frac{1}{n} \xrightarrow[\infty]{} \theta.$$

X_{min} est asymptotiquement sans biais et $B_\theta(X_{min}) = \mathbb{E}[X_{min}] - \theta = \frac{1}{n}$

$$\Rightarrow V(V_n) = V(X_{min} - \theta) = \frac{1}{n^2} = V(X_{min})$$

$$\Rightarrow R(X_{min}, \theta) = B_\theta(X_{min})^2 + V(X_{min}) = \frac{1}{n^2} + \frac{1}{n^2} = \frac{2}{n^2}$$

Si on pose $X_{min}^* = \min_{i \in [\![1,n]\!]} X_i - \frac{1}{n}$

On a :

$$\mathbb{E}[X_{min}^*] = \mathbb{E}[X_{min}] - \frac{1}{n} = \theta \Rightarrow B_\theta(X_{min}^*) = 0$$

et

$$V(X_{min}^*) = V(X_{min}) = \frac{1}{n^2}$$

$$\Rightarrow R(X_{min}^*, \theta) = \underbrace{B_\theta(X_{min}^*)}_{=0}^2 + V(X_{min}^*) = \frac{1}{n^2}$$

$$\Rightarrow R(X_{min}^*, \theta) \leq R(X_{min}, \theta) \leq R(Y_n, \theta)$$

$\Rightarrow$ On préfèrera X_{min}^* au sens du risque quadratique.

Exercice 4.

Soit $X = (X_1, \ldots, X_n)$ un n-échantillon de loi normale $\mathcal{N}(m, \sigma^2)$. Pour estimer σ^2, on pose :

$$T_n = c(n) \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

1. Quelle est la loi de $\sigma^{-2} \sum_{i=1}^{n} (X_i - \overline{X})^2$?
2. calculer $B(T_n)$, et donner un estimateur sans biais de σ^2.
3. Quelle est la fonction $c(n)$ qui minimise le risque quadratique de T_n ?
4. On suppose que $\sigma = 1$ et $m \in [0,1]$, et on définit l'estimateur :

$$U_n = \begin{cases} 0 & \text{si } \overline{X} < 0 \\ \overline{X} & \text{si } 0 \leq \overline{X} \leq 1 \\ 1 & \text{si } \overline{X} > 1 \end{cases}.$$

   a. Montrer que U_n est un estimateur de m, strictement meilleur que $\overline{X}$, en calculant la différence entre les risques quadratiques.

   b. Montrer que

$$\mathrm{E}(U_n) = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{n}\theta}^{\sqrt{n}(1-\theta)} \left(\theta + \frac{t}{\sqrt{n}}\right) e^{-\frac{t^2}{2}} dt + \mathrm{P}(\overline{X} > 1).$$

   c. En déduire que $E_m(U_n) \to \theta$ lorsque $n \to \infty$.

Exercice 4 :

On a $X_1, \ldots, X_n$ iid tq $X_i \sim \mathcal{N}(m, \sigma^2)$

On veut estimer σ^2 par $T_n = c_n \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$

1) Loi de $\sigma^{-2} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$:

On sait que $\sigma^{-2} \sum_{i=1}^{n} (X_i - m)^2 = \sum_{i=1}^{n} \overbrace{\Big(\underbrace{\frac{X_i - m}{\sigma}}_{\sim \mathcal{N}(0,1)}\Big)^2}^{\sim \chi^2_1} \sim \chi^2_n$

On part de :

$$\sigma^{-2} \sum (X_i - m)^2 = \sigma^{-2} \sum_{i=1}^{n} (X_i - \overline{X}_n + \overline{X}_n - m)^2$$

$$= \sigma^{-2} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 + \sigma^{-2} \sum_{i=1}^{n} (\overline{X}_n - m)^2 + \underbrace{2\sigma^{-2} \sum_{i=1}^{n} (\overline{X}_n - m)(X_i - \overline{X}_n)}_{= (\overline{X}_n - m) \sum_{i=1}^{n} (X_i - \overline{X}_n)} = 0$$

$\sum_{i=1}^{n} (X_i - \overline{X}_n) = \sum X_i - n \frac{1}{n} \sum X_i$

$\sim \chi^2_1$: $n\sigma^{-2} (\overline{X}_n - m)^2$ or $\overline{X}_n \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$ $\quad \left(\frac{\overline{X}_n - m}{\sqrt{\sigma^2/n}}\right) \sim \mathcal{N}(0,1)$

On a :

$$\underbrace{\sigma^{-2} \sum_{i=1}^{n} (X_i - m)^2}_{\sim \chi^2_n} = \underbrace{\sigma^{-2} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2}_{\sim \chi^2_{n-1}} + \underbrace{\left(\frac{\overline{X}_n - m}{\sigma/\sqrt{n}}\right)^2}_{\sim \chi^2_1}$$

Si $\sigma^{-2} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$ indep $\left(\frac{\overline{X}_n - m}{\sigma/\sqrt{n}}\right)^2$

On veut mq $X_i - \overline{X}_n$ est indep de $\overline{X}_n$ ($\Rightarrow (X_i - \overline{X}_n)^2$ indep de $(\overline{X}_n - m)^2 \ldots$)

On a $(X_1 - \overline{X}_n, X_2 - \overline{X}_n, \ldots, X_n - \overline{X}_n, \overline{X}_n)$ est un vecteur Gaussien.

Car toute C-L de composantes du vecteur peuvent s'écrire comme une CL de $X_1, \ldots, X_n$ et donc est Gaussienne.

$\Rightarrow$ Comme on est dans un vecteur Gaussien, on a indép $\Leftrightarrow$ Cov $= 0$

On a $\mathrm{Cov}(X_i - \overline{X}_n, \overline{X}_n) = \mathrm{Cov}(X_i, \overline{X}_n) - \mathrm{Cov}(\overline{X}_n, \overline{X}_n)$

$$= \underbrace{\frac{1}{n} \sum_{j=1}^{n} \underbrace{\mathrm{Cov}(X_i, X_j)}_{= \begin{cases} 0 & \text{si } i \neq j \\ \sigma^2 & \text{si } i = j \end{cases}}}_{= \frac{\sigma^2}{n}} - \underbrace{V(\overline{X}_n)}_{= \frac{\sigma^2}{n}}$$

$\Rightarrow \mathrm{Cov}(X_i - \overline{X}_n, \overline{X}_n) = 0$

$$\Rightarrow \sigma^{-2} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 \sim \chi^2_{n-1}$$

2) On pose $T_n = c_n \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$

On a $\sigma^{-2} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \sim \chi^2_{n-1}$ $\rightarrow$ $\mathbb{E}(\sigma^{-2} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2) = n-1$

$\rightarrow$ $V(\sigma^{-2} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2) = 2(n-1)$

$\Rightarrow \mathbb{E}[\sum_{i=1}^{n} (X_i - \bar{X}_n)^2] = (n-1)\sigma^2$ et $V(\sum_{i=1}^{n} (X_i - \bar{X}_n)^2) = 2(n-1)\sigma^4$

$\Rightarrow \mathbb{E}[T_n] = \mathbb{E}[c_n \sum_{i=1}^{n} (X_i - \bar{X}_n)^2] = c_n(n-1)\sigma^2 = \sigma^2$ ssi $c_n = \frac{1}{n-1}$

$\Rightarrow \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ est sans biais pour σ^2

et $B_\sigma(T_n) = \mathbb{E}[T_n] - \sigma = c_n(n-1)\sigma^2 - \sigma^2 = \sigma^2(c_n(n-1)-1)$

3. On veut le T_n qui minimise le risque quadratique.

$V(T_n) = V(c_n \sum_{i=1}^{n} (X_i - \bar{X})^2 = 2c_n^2(n-1)\sigma^4$

$R(T_n, \sigma^2) = B_\sigma(T_n)^2 + V(T_n) = \sigma^4(c_n(n-1)-1)^2 + 2c_n^2(n-1)\sigma^4$
$= \sigma^4(c_n^2(n-1)^2 + 1 - 2c_n(n-1)) + 2c_n^2(n-1)\sigma^4$

On pose $g(c_n) = c_n^2(n-1)^2 + 1 - 2c_n(n-1) + 2c_n^2(n-1)$

On calcule $g'(c_n) = 2(n-1)^2 c_n - 2(n-1) + 4c_n(n-1)$

On cherche c_n tq $(n-1)c_n - 1 + 2c_n = 0$

$\Rightarrow c_n = \frac{1}{n+1} \Rightarrow \frac{1}{n+1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ minimise le risque quadratique (parmi les estimateurs de la forme T_n)

4) $X_1, \dots, X_n$ iid tq $X_i \sim \mathcal{N}(m, \sigma^2)$

On a $\sigma = 1$ et $m \in [0,1]$

On sait que $\bar{X}_n$ est un ESB convergent (efficace et exhaustif) de m.

On pose
$$U_n = \begin{cases} 0 & \text{si } \bar{X}_n < 0 \\ \bar{X}_n & \text{si } 0 \le \bar{X}_n \le 1 \\ 1 & \text{si } \bar{X}_n > 1 \end{cases}$$
on oblige $\bar{X}_n$ à être dans $[0,1]$

a) On veut montrer que $R(U_n, m) < R(\bar{X}_n, m)$

$R(\bar{X}_n, m) - R(U_n, m) = \mathbb{E}[(\bar{X}_n - m)^2] - \mathbb{E}[(U_n - m)^2]$.

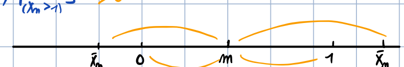
$= \mathbb{E}[(\bar{X}_n - m)^2 1_{(\bar{X}_n < 0)}] + \mathbb{E}[(\bar{X}_n - m)^2 1_{(0 \le \bar{X}_n \le 1)}] + \mathbb{E}[(\bar{X}_n - m)^2 1_{(\bar{X}_n > 1)}]$

$- \mathbb{E}[(\underbrace{U_n}_{0} - m)^2 1_{(\bar{X}_n < 0)}] + \mathbb{E}[(\underbrace{\bar{X}_n}_{\bar{X}_n} - m)^2 1_{(0 \le \bar{X}_n \le 1)}] + \mathbb{E}[(\underbrace{\bar{X}_n}_{1} - m)^2 1_{(\bar{X}_n > 1)}]$

$R(\bar{X}_n, m) - R(U_n, m) = \mathbb{E}[((\bar{X}_n - m)^2 - (0-m)^2) 1_{(\bar{X}_n < 0)}]$ >0

$+ \mathbb{E}[((\bar{X}_n - m)^2 - (1-m)^2) 1_{(\bar{X}_n > 1)}]$ >0

$\Rightarrow R(\bar{X}_n, m) > R(U_n, m)$



b) $\mathbb{E}[U_n] = \mathbb{E}[\underbrace{U_n}_{0} 1_{(\bar X_n < 0)}] + \mathbb{E}[\underbrace{U_n}_{\bar X_n} 1_{(0 \leq \bar X_n \leq 1)}] + \mathbb{E}[\underbrace{U_n}_{1} 1_{(\bar X_n > 1)}]$.

$$= \mathbb{E}[\bar X_n 1_{(0 \leq \bar X_n \leq 1)}] + \mathbb{E}[1_{(\bar X_n \geq 1)}]$$

Rappel : $\mathbb{E}[1_A] = \mathbb{P}_A$

Or comme $X_1, \dots, X_n$ iid tq $X_i \sim \mathcal{N}(m, 1)$ alors $\bar X_n \sim \mathcal{N}(m, \frac{1}{n})$

$$\mathbb{E}[U_n] = \int_0^1 x \times \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1/n}} \exp\left(-\frac{n}{2}(x-m)^2\right) dx + \mathbb{P}(\bar X_n > 1)$$

On pose $t = \sqrt{n}(x-m)$ $\quad dt = \sqrt{n}\, dx$

$x = m + \frac{t}{\sqrt{n}}$

$$\mathbb{E}[U_n] = \int_{-m\sqrt{n}}^{\sqrt{n}(1-m)} \left(m + \frac{t}{\sqrt{n}}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt + \mathbb{P}(\bar X_n > 1)$$

$$\mathbb{E}[U_n] = \underbrace{m \int_{-m\sqrt{n}}^{\sqrt{n}(1-m)} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt}_{m \times 1} + \underbrace{\frac{1}{\sqrt{n}} \int_{-m\sqrt{n}}^{\sqrt{n}(1-m)} t \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}}_{\mathcal{N}(0,1)} dt}_{\to 0} + \underbrace{\mathbb{E}[1_{(\bar X_n > 1)}]}_{\to 0}.$$

$\bar X_n \sim \mathcal{N}(m, \frac{1}{n})$

$\bar X_n \xrightarrow{\mathbb{P}} m$

$d\mathbb{P}_{\bar X_n} \longrightarrow \delta_m$

$$\int 1_{(x>1)} d\mathbb{P}_{\bar X_n}(x) \longrightarrow \underbrace{\int 1_{(x>1)} d\delta_m(x)}_{= 1_{(m>1)}}.$$

$\mathbb{E}[U_n] \longrightarrow m$ qd $n \to \infty$