Chapitre 3 : Erreur d'estimation

Dans ce chapitre, on dispose de n observations indépendantes $x_1, \dots, x_n \in \mathcal{O}$ (avec $\mathcal{O} \subset \mathcal{X}$) qui sont des réalisations de la loi $\mathcal{L}_{\theta^*}$, $\theta^* \in \Theta$.

Le modèle statistique est décrit par les échantillons $X_1, \dots, X_n$ i.i.d. de la loi $\mathcal{L}_\theta$, avec $\theta \in \Theta \subset \mathbb{R}$. Enfin, le paramètre d'intérêt est $g(\theta)$, où $g : \Theta \to \mathbb{R}$.

1- Intervalle de confiance

On veut construire un intervalle dans lequel se trouve le paramètre d'intérêt, et ceci avec un niveau de confiance élevé.

Définition (Intervalle de confiance)

Fixons $\alpha \in]0,1[$. Pour tout $\theta \in \Theta$ et $X_1,\dots,X_n$ un échantillon iid de la loi $\mathcal{L}_\theta$, un intervalle de confiance (IC) par excès de niveau de confiance $(1-\alpha)$ pour $g(\theta)$ est un intervalle aléatoire $I(X_1,\dots,X_n)$ tel que :

$$\boxed{\mathbb{P}\Big(g(\theta) \in I(X_1,\dots,X_n)\Big) \geq 1-\alpha.}$$

Les observations $x_1,\dots,x_n$ sont des réalisations de la loi $\mathcal{L}_{\theta^*}$. Ayant cet IC, on en déduit que

$$\boxed{g(\theta^*) \in I(x_1,\dots,x_n) \text{ avec un niveau de confiance } \geq 1-\alpha}$$

Remarques

(1) α vs longueur de l'IC.

2 critères de qualité pour un IC : α, qui doit être petit et la longueur de l'IC, qui doit être petite également.

Mais ces 2 critères de qualité s'opposent : + α est petit, + la longueur de l'IC est grande, et réciproquement. On doit donc réaliser un <u>compromis</u>. La plupart du temps, on prend $\alpha = 5\%$ ou 10%.

(2) Sur l'IC. L'IC est construit avec un estimateur, mais tous les estimateurs ne se valent pas, et un IC construit avec un mauvais estimateur sera un mauvais IC.

Exemple de construction d'un IC

On lance 100 fois une pièce, et on veut une approximation de θ^* = proba que la pièce tombe sur pile. En codant $\begin{matrix} 1 = \text{pile} \\ 0 = \text{face} \end{matrix}$, on a une suite $x_1, \ldots, x_{100} \in \{0,1\}$ telle que $\overline{x}_{100} = 0{,}7$. $\left(= \frac{1}{100} \sum_{i=1}^{100} x_i \right)$.

Considérons un échantillon $X_1, \ldots, X_{100}$ iid de loi $\mathcal{B}(\theta)$, $\theta \in]0,1[$. Un estimateur naturel de θ est donné par $\overline{X}_{100}$ (cf fin du chap. 2).

On a $\mathbb{E}(X_1)=\theta$ et $var(X_1)=\theta(1-\theta)$, car $X_1 \sim \mathcal{B}(\theta)$. Alors,

- $\mathbb{E}(\overline{X}_{100}) = \mathbb{E}\left(\frac{1}{100}\sum_{i=1}^{100} X_i\right) = \frac{1}{100}\sum_{i=1}^{100}\mathbb{E}(X_i) = \frac{1}{100}\sum_{i=1}^{100}\theta = \theta$

- $var(\overline{X}_{100}) = var\left(\frac{1}{100}\sum_{i=1}^{100} X_i\right) = \frac{1}{10^4}\sum_{i=1}^{100} var(X_i) = \frac{1}{10^4}\sum_{i=1}^{100}\theta(1-\theta) = \frac{\theta(1-\theta)}{100}$

car $X_1, \dots, X_n$ indep.

Alors, $\forall a > 0$:

inégalité BT

$$\mathbb{P}(|\overline{X}_{100}-\theta| \geq a) = \mathbb{P}(|\overline{X}_{100}-\mathbb{E}(\overline{X}_{100})| \geq a) \leq \frac{var(\overline{X}_{100})}{a^2} = \frac{\theta(1-\theta)}{100\, a^2}$$

Or, $\forall \theta \in \,]0,1[$, $\theta(1-\theta) \leq \frac{1}{4}$ donc

$$\mathbb{P}(|\overline{X}_{100}-\theta| \geq a) \leq \frac{1}{400\, a^2}$$

Alors, $\mathbb{P}(|\overline{X}_{100}-\theta| < a) = 1 - \mathbb{P}(|\overline{X}_{100}-\theta| \geq a) \geq 1 - \frac{1}{400\, a^2}$

Si on veut un IC par excès au niveau 95%, on choisit a t.q.

$$1 - \frac{1}{400a^2} = 0,95 \Longleftrightarrow \frac{1}{400a^2} = 0,05$$

$$\Longleftrightarrow a^2 = \frac{1}{400 \times 0,05} = \frac{1}{20} \Longleftrightarrow a = \frac{1}{\sqrt{20}}$$

On a donc

$$\mathbb{P}\left(|\overline{X}_{100} - \theta| < \frac{1}{\sqrt{20}} \right) \geq 0,95$$

$$\Longleftrightarrow \boxed{\mathbb{P}\left(\theta \in \left] \overline{X}_{100} - \frac{1}{\sqrt{20}}, \overline{X}_{100} + \frac{1}{\sqrt{20}} \right[\right) \geq 0,95}$$

Rappel

$|x - \theta| < \ell$

$\Longleftrightarrow x \in \left] \theta - \ell, \theta + \ell \right[$

Donc l'IC par excès pour θ au niveau 95% est

$$\boxed{I(X_1, \dots, X_{100}) = \left] \overline{X}_{100} - \frac{1}{\sqrt{20}}, \overline{X}_{100} + \frac{1}{\sqrt{20}} \right[}$$

Par suite,

$$\theta^* \in \left]\bar{x}_{100} - \frac{1}{\sqrt{20}}, \bar{x}_{100} + \frac{1}{\sqrt{20}}\right[= \left]0,7 - \frac{1}{\sqrt{20}} ; 0,7 + \frac{1}{\sqrt{20}}\right[= \left]0,47 ; 0,92\right[$$

avec un niveau de confiance de 95%.

2. Biais

Pour un estimateur $\hat{g}$ de $g(\theta)$, on veut que ses réalisations valeurs fluctuent autour de $g(\theta)$. De ce fait, une propriété minimale de $\hat{g}$ est qu'il se comporte en moyenne comme le paramètre d'intérêt.

Définition (Biais) Soient $\theta \in \Theta$, et $X_1, \dots, X_n$ un échantillon iid de la loi $\mathcal{L}_\theta$. Le biais en θ de l'estimateur $\hat{g} = \hat{g}(X_1, \dots, X_n)$ est la quantité

$$B_{\hat{g}}(\theta) = \mathbb{E}(\hat{g}) - g(\theta).$$

On dit que l'estimateur $\hat{g}$ est sans biais si $B_{\hat{g}}(\theta) = 0 \quad \forall \theta \in \Theta$

Exemples

1) Dans le modèle de Bernoulli décrit par les échantillons $X_1, \dots, X_n$ de va iid de loi $\mathcal{B}(\theta)$, $\theta \in]0,1[$.

→ Si le paramètre d'intérêt est θ.

Estimateur $\overline{X}_n$ (car par la LGN : $\overline{X}_n \simeq \mathbb{E}(X_1) = \theta$, car $X_1 \sim \mathcal{B}(\theta)$)

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

Donc $\boxed{B_{\overline{X}_n}(\theta) = 0 \quad \forall \theta \in]0,1[}$

→ Si le paramètre d'intérêt est $\theta(1-\theta)$ (= variance de la loi $\mathcal{B}(\theta)$)

Comme $\overline{X}_n$ est un estimateur de θ, $\hat{g} = \overline{X}_n(1-\overline{X}_n)$ est un estimateur (par insertion) de $\theta(1-\theta)$.

Calculons

$$\mathbb{E}(\hat{g}) = \mathbb{E}\left(\bar{X}_n(1-\bar{X}_n)\right) = \mathbb{E}(\bar{X}_n) - \mathbb{E}(\bar{X}_n^2) = \theta - \mathbb{E}(\bar{X}_n^2).$$

Or,

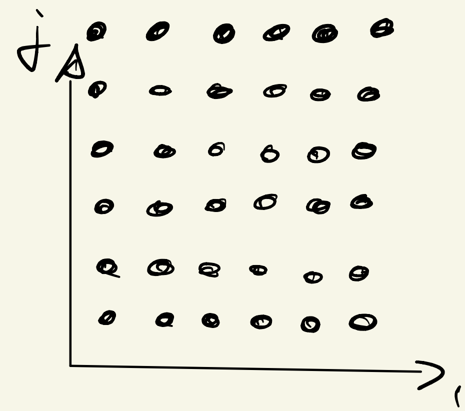
$$\mathbb{E}(\bar{X}_n^2) = \mathbb{E}\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right) = \frac{1}{n^2}\,\mathbb{E}\left(\left(\sum_{i=1}^{n} X_i\right)^2\right) = \frac{1}{n^2}\sum_{i,j=1}^{n} \mathbb{E}(X_iX_j)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{E}(X_i^2) + \frac{1}{n^2}\sum_{\substack{i,j=1\\ i\neq j}}^{n} \mathbb{E}(X_iX_j)$$

Or $X_i^2 = X_i$ car $X_i \sim \mathcal{B}(\theta)$, et si $i \neq j$, X_i indépendante de X_j d'où

$\mathbb{E}(X_iX_j) = \mathbb{E}(X_i)\,\mathbb{E}(X_j)$. Donc

$$\mathbb{E}(\bar{X}_n^2) = \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{\mathbb{E}(X_i)}_{=\theta} + \frac{1}{n^2}\sum_{\substack{i,j=1\\ i\neq j}}^{n} \underbrace{\mathbb{E}(X_i)}_{=\theta}\,\underbrace{\mathbb{E}(X_j)}_{=\theta}$$



$$= \frac{\theta}{n^2}\underbrace{\sum_{i=1}^{n} 1}_{=n} + \frac{\theta^2}{n^2}\underbrace{\sum_{\substack{i,j=1\\ i\neq j}}^{n} 1}_{=n^2-n} = \frac{\theta}{n} + \frac{\theta^2(n-1)}{n}$$

Donc $\mathbb{E}(\hat{g}) = \theta - \mathbb{E}(\bar{X}_n^2) = \theta - \frac{\theta}{n} - \frac{\theta^2(n-1)}{n}$. Par suite,

$$B_{\hat{g}}(\theta) = \mathbb{E}(\hat{g}) - \theta(1-\theta) = \theta - \frac{\theta + \theta^2(n-1)}{n} - \theta(1-\theta)$$

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

$$= \frac{\theta(\theta-1)}{n} \neq 0, \text{ donc } \boxed{\hat{g} \text{ est biaisé}}$$

② Dans le modèle décrit par les échantillons $X_1, \dots, X_n$ iid de loi $\mathcal{G}(\theta)$, avec $\theta \in]0,1[$. Supposons que le paramètre d'intérêt est θ.

Par la LGN, $\bar{X}_n \simeq \mathbb{E}(X_1) = \frac{1}{\theta}$, car $X_1 \sim \mathcal{G}(\theta)$. L'estimateur de θ par insertion est donc $\frac{1}{\bar{X}_n}$. Calculons un majorant pour son biais.

$$\left|B_{1/\bar{X}_n}(\theta)\right| = \left|\mathbb{E}\left(\frac{1}{\bar{X}_n}\right) - \theta\right| = \left|\mathbb{E}\left(\frac{1}{\bar{X}_n} - \theta\right)\right| \leq \mathbb{E}\left(\left|\frac{1}{\bar{X}_n} - \theta\right|\right)$$

Rappel $|\mathbb{E}(Z)| \le \mathbb{E}(|Z|)$

$|\mathbb{E}(Z)| \le \sqrt{\mathbb{E}(Z^2)}$

↪ inégalité de Cauchy-Schwarz

$$|B_{1/\bar{X}_n}(\theta)| \le \mathbb{E}\left(\left|\frac{1}{\bar{X}_n} - \theta\right|\right) = \mathbb{E}\left(\left|\frac{1-\theta\bar{X}_n}{\bar{X}_n}\right|\right)$$

Or, comme $X_i \sim \mathcal{G}(\theta)$, $X_i \ge 1 \ \forall i=1,\dots,n$ d'où $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \ge \frac{1}{n}\sum_{i=1}^n 1 = 1$

Alors,

$$|B_{1/\bar{X}_n}(\theta)| \le \mathbb{E}(|1-\theta\bar{X}_n|)$$ car $\bar{X}_n \ge 1 \Rightarrow \frac{1}{\bar{X}_n} \le 1$

$\Rightarrow |B_{1/\bar{X}_n}(\theta)| \le \theta\,\mathbb{E}\left(\left|\frac{1}{\theta} - \bar{X}_n\right|\right)$. Or $\mathbb{E}(\bar{X}_n) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i) = \frac{1}{n}\sum_{i=1}^n \frac{1}{\theta} = \frac{1}{\theta}$

$$\le \theta\sqrt{\mathbb{E}\left(\left(\mathbb{E}(\bar{X}_n) - \bar{X}_n\right)^2\right)} = \theta\sqrt{\mathrm{var}(\bar{X}_n)}$$

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

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

Donc $\left|B_{1/\bar{X}_n}(\theta)\right| \leq \theta\sqrt{\operatorname{var}(\bar{X}_n)} = \theta\sqrt{\frac{1-\theta}{n\theta^2}} = \frac{\sqrt{1-\theta}}{\sqrt{n}}$

3 - Risque quadratique

On veut évaluer une "distance" entre l'estimateur et son paramètre d'intérêt.

Définition (Risque quadratique) Soient $\theta \in \Theta$ et $X_1, \dots, X_n$ un échantillon iid de la loi $\mathcal{L}_\theta$. Le risque quadratique en θ de l'estimateur $\hat{g} = \hat{g}(X_1, \dots, X_n)$ est la quantité

$$R_{\hat{g}}(\theta) = \mathbb{E}\left(\left(\hat{g} - g(\theta)\right)^2\right)$$

Soit $\hat{g}'$ un autre estimateur. Si $\forall \theta \in \Theta$, $R_{\hat{g}'}(\theta) \leq R_{\hat{g}}(\theta)$, on dit que $\hat{g}'$ est préférable à $\hat{g}$

Exemples

① Modèle de Bernoulli : $X_1, \dots, X_n$ iid de loi $\mathcal{B}(\theta)$, $\theta \in]0,1[$.

Paramètre d'intérêt : θ.

Estimateurs de θ : $\overline{X}_n$, X_1

Comparons $\overline{X}_n$ à X_1 :

$$R_{X_1}(\theta) = \mathbb{E}\left(\left(X_1 - \underset{\underset{\mathbb{E}(X_1)}{\parallel}}{\theta}\right)^2\right) = \mathbb{E}\left(\left(X_1 - \mathbb{E}(X_1)\right)^2\right) = \mathrm{var}(X_1) = \theta(1-\theta)$$

car $X_1, \dots, X_n$ indep.

$$R_{\overline{X}_n}(\theta) = \mathbb{E}\left(\left(\overline{X}_n - \underset{\underset{\mathbb{E}(\overline{X}_n)}{\parallel}}{\theta}\right)^2\right) = \mathbb{E}\left(\left(\overline{X}_n - \mathbb{E}(\overline{X}_n)\right)^2\right) = \mathrm{var}(\overline{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{var}(X_i)$$

$\mathbb{E}(\overline{X}_n)$, car $\overline{X}_n$ sans biais.

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

Donc $R_{\overline{X}_n}(\theta) \leq R_{X_1}(\theta)$ $\forall \theta \in]0,1[$ $\Rightarrow$ $\boxed{\overline{X}_n \text{ est préférable à } X_1}$

② Dans le modèle décrit par le n-échantillon $X_1,\dots,X_n$ iid de loi $\mathcal{G}(\theta)$, $\theta \in]0,1[$, considérons le paramètre d'intérêt $g(\theta)=\theta$.

D'après la LGN : $\bar{X}_n \approx \mathbb{E}(X_1) = \frac{1}{\theta}$, car $X_1 \sim \mathcal{G}(\theta)$. Donc $\hat{\theta} = \frac{1}{\bar{X}_n}$ est un estimateur naturel de θ. Evaluons maintenant son risque quadratique :

$$R_{\hat{\theta}}(\theta) = \mathbb{E}\left[(\hat{\theta}-\theta)^2\right] = \mathbb{E}\left[\left(\frac{1}{\bar{X}_n}-\theta\right)^2\right] = \mathbb{E}\left[\left(\frac{1-\theta\bar{X}_n}{\bar{X}_n}\right)^2\right]$$

$\forall i$, $X_i \sim \mathcal{G}(\theta)$ donc $X_i \geq 1 \Rightarrow \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \geq \frac{1}{n}\sum_{i=1}^{n} 1 = 1$. Alors,

$$R_{\hat{\theta}}(\theta) \leq \mathbb{E}\left[(1-\theta\bar{X}_n)^2\right] = \theta^2\,\mathbb{E}\left[\left(\underbrace{\frac{1}{\theta}}_{=\,\mathbb{E}(\bar{X}_n)}-\bar{X}_n\right)^2\right] = \theta^2\,\mathrm{var}(\bar{X}_n)$$

Or, $\mathrm{var}(\bar{X}_n) \overset{\text{car } X_1,\dots,X_n \text{ indep.}}{=} \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\frac{1-\theta}{\theta^2} = \frac{1-\theta}{n\theta^2}$, d'où

$$\boxed{R_{\hat{\theta}}(\theta) \leq \theta^2\,\frac{1-\theta}{n\theta^2} = \frac{1-\theta}{n}}$$

Théorème (Décomposition biais-variance) Soient $\theta \in \Theta$, et $X_1, \dots, X_n$ un échantillon iid de loi $\mathcal{L}_\theta$. Alors, si $\hat{g} = \hat{g}(X_1, \dots, X_n)$ est un estimateur de $g(\theta)$, on a

$$\boxed{R_{\hat{g}}(\theta) = B_{\hat{g}}(\theta)^2 + var(\hat{g})}$$

Cette décomposition montre qu'abaisser le biais revient à augmenter la variance, et réciproquement.

Preuve On a

$$R_{\hat{g}}(\theta) = \mathbb{E}\left[(\hat{g} - g(\theta))^2\right] = \mathbb{E}\left[(\underbrace{\hat{g} - \mathbb{E}[\hat{g}]} + \underbrace{\mathbb{E}[\hat{g}] - g(\theta)})^2\right]$$

$$= \underbrace{\mathbb{E}\left[(\hat{g} - \mathbb{E}(\hat{g}))^2\right]}_{= var(\hat{g})} + \underbrace{\mathbb{E}\left[(\mathbb{E}(\hat{g}) - g(\theta))^2\right]}_{(\mathbb{E}(\hat{g}) - g(\theta))^2} + 2\underbrace{\mathbb{E}\left[(\hat{g} - \mathbb{E}(\hat{g}))(\mathbb{E}[\hat{g}] - g(\theta))\right]}_{= (\mathbb{E}(\hat{g}) - g(\theta))\mathbb{E}[\hat{g} - \mathbb{E}[\hat{g}]]}$$

$$= var(\hat{g}) + B_{\hat{g}}(\theta)^2 + 2(\mathbb{E}(\hat{g}) - g(\theta))(\mathbb{E}(\hat{g}) - \underbrace{\mathbb{E}(\mathbb{E}(\hat{g}))}_{= \mathbb{E}(\hat{g})})$$

$$= var(\hat{g}) + B_{\hat{g}}(\theta)^2 + 2(\mathbb{E}(\hat{g}) - g(\theta))(\underbrace{\mathbb{E}(\hat{g}) - \mathbb{E}(\hat{g})}_{= 0}) = var(\hat{g}) + B_{\hat{g}}(\theta)^2 \quad \blacksquare$$

Pour $\theta \in \Theta$ et $X_1, \dots, X_n$ iid de loi $\mathcal{L}_\theta$, on note maintenant

$$\boxed{L(\theta; y_1, \dots, y_n) = \mathbb{P}(X_1 = y_1, \dots, X_n = y_n), \quad y_1, \dots, y_n \in \mathcal{O}}$$

L = "likelihood" = vraisemblance

<u>Interlude</u> : Idée derrière la notion de vraisemblance

Par ex., $\mathcal{L}_\theta = \mathcal{B}(\theta)$ donc $X_1, \dots, X_n$ iid de loi $\mathcal{B}(\theta)$, $\theta \in]0,1[$

On a de + les observations $x_1, \dots, x_n \in \{0,1\}$ réalisations de la loi $\mathcal{B}(\theta^*)$

$$L(\theta; x_1, \dots, x_n) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$$

$\begin{cases} \text{si } \theta \text{ est loin de } \theta^*, & L(\theta; x_1, \dots, x_n) \text{ est petit} \\ \text{si } \theta \text{ est proche de } \theta^*, & \quad \text{"} \quad \text{est grande} \end{cases}$

↪ notion de vraisemblance : il est vraisemblable que si $L(\theta; x_1, \dots, x_n)$ est gd, alors θ est proche de θ^*.

↪ donne une nouvelle idée pour construire des estimateurs de θ :

$$\underset{\theta \in]0,1[}{\operatorname{argmax}} \, L(\theta; x_1, \dots, x_n)$$

On suppose que $\forall y_1, \dots, y_n \in \Theta$:

$$\begin{cases} \Theta \ni \theta \mapsto L(\theta; y_1, \dots, y_n) \text{ est de classe } \mathcal{C}^1 \\ \Theta \ni \theta \mapsto \ln L(\theta; y_1, \dots, y_n) \text{ est de classe } \mathcal{C}^1 \end{cases}$$

Hypothèses de "régularité"

↪ fonction "log-vraisemblance"

On note alors $I(\theta)$ l'information de Fisher en θ par

$$\boxed{I(\theta) = \mathrm{var}\left(\frac{\partial \ln L(\theta; X_1, \dots, X_n)}{\partial \theta}\right)}$$

Théorème de Cramer-Rao Supposons que Θ est fini et que g est dérivable sur Θ. Alors, $\forall \theta \in \Theta$, $X_1, \dots, X_n$ iid de loi $\mathcal{L}_\theta$ et pour tout estimateur sans biais $\hat{g} = \hat{g}(X_1, \dots, X_n)$, on a

$$\boxed{R_{\hat{g}}^{n}(\theta) \geq \frac{g'(\theta)^2}{I(\theta)}}$$

Cette inégalité nous donne une vitesse de convergence vers 0 de $R_{\hat{g}}(\theta)$, vitesse qu'on ne peut pas dépasser.

Preuve On a

$$\frac{\partial \ln L(\theta; y_1, \dots, y_n)}{\partial \theta} = \frac{\frac{\partial L(\theta; y_1, \dots, y_n)}{\partial \theta}}{L(\theta; y_1, \dots, y_n)}.$$

Alors,

Th. transfert

$$\mathbb{E}\left(\frac{\partial \ln L(\theta; X_1, \dots, X_n)}{\partial \theta}\right) = \sum_{y_1, \dots, y_n \in \Theta} \frac{\partial \ln L(\theta; y_1, \dots, y_n)}{\partial \theta} \mathbb{P}(X_1 = y_1, \dots, X_n = y_n)$$

$$= \sum_{y_1, \dots, y_n \in \Theta} \frac{\frac{\partial L(\theta; y_1, \dots, y_n)}{\partial \theta}}{\cancel{L(\theta; y_1, \dots, y_n)}} \cancel{\mathbb{P}(X_1 = y_1, \dots, X_n = y_n)}$$

car Θ est fini

$$= \sum_{y_1, \dots, y_n \in \Theta} \frac{\partial L(\theta; y_1, \dots, y_n)}{\partial \theta} = \frac{\partial}{\partial \theta}\left(\sum_{y_1, \dots, y_n \in \Theta} L(\theta; y_1, \dots, y_n)\right)$$

Or, $\displaystyle\sum_{y_1, \dots, y_n \in \Theta} L(\theta; y_1, \dots, y_n) = \sum_{y_1, \dots, y_n \in \Theta} \mathbb{P}(X_1 = y_1, \dots, X_n = y_n) = 1$ donc

$$\frac{\partial}{\partial \theta}\left(\sum_{y_1, \dots, y_n \in \Theta} L(\theta; y_1, \dots, y_n)\right) = 0 \implies \boxed{\mathbb{E}\left(\frac{\partial \ln L(\theta; X_1, \dots, X_n)}{\partial \theta}\right) = 0}$$

Comme $\hat{g}$ est sans biais, $\mathbb{E}(\hat{g}) = g(\theta)$ d'où

$$g'(\theta) = \frac{d}{d\theta}\mathbb{E}(\hat{g}) = \frac{d}{d\theta}\left(\sum_{y_1,\dots,y_n\in\mathcal{O}} \hat{g}(y_1,\dots,y_n)\overbrace{L(\theta;y_1,\dots,y_n)}^{=\mathbb{P}(X_1=y_1,\dots,X_n=y_n)}\right)$$

car somme finie

$$= \sum_{y_1,\dots,y_n\in\mathcal{O}} \hat{g}(y_1,\dots,y_n)\frac{\partial L(\theta;y_1,\dots,y_n)}{\partial\theta}$$

$$= \sum_{y_1,\dots,y_n\in\mathcal{O}} \hat{g}(y_1,\dots,y_n)\frac{\partial \ln L(\theta;y_1,\dots,y_n)}{\partial\theta}\underbrace{L(\theta;y_1,\dots,y_n)}_{=\mathbb{P}(X_1=y_1,\dots,X_n=y_n)}$$

$$= \mathbb{E}\left(\hat{g}(X_1,\dots,X_n)\frac{\partial \ln L(\theta;X_1,\dots,X_n)}{\partial\theta}\right) - g(\theta)\underbrace{\mathbb{E}\left(\frac{\partial \ln L(\theta;X_1,\dots,X_n)}{\partial\theta}\right)}_{=0}$$

$$= \mathbb{E}\left(\left(\hat{g}(X_1,\dots,X_n) - g(\theta)\right)\frac{\partial \ln L(\theta;X_1,\dots,X_n)}{\partial\theta}\right)$$

On utilise maintenant l'inégalité de Cauchy-Schwarz : $\mathbb{E}(z_1 z_2)^2 \leq \mathbb{E}(z_1^2)\ \mathbb{E}(z_2^2)$ pour des var z_1, z_2. Alors,

$$g'(\theta)^2 = \left(\mathbb{E}\left(\underbrace{(\hat{g}(x_1,\dots,x_n) - g(\theta))}_{= z_1} \underbrace{\frac{\partial \ln L(\theta; x_1,\dots,x_n)}{\partial \theta}}_{= z_2}\right)\right)^2$$

$$\leq \underbrace{\mathbb{E}\left[\left(\hat{g}(x_1,\dots,x_n) - g(\theta)\right)^2\right]}_{= R_{\hat{g}}(\theta)} \underbrace{\mathbb{E}\left[\left(\frac{\partial \ln L(\theta; x_1,\dots,x_n)}{\partial \theta}\right)^2\right]}_{= I(\theta)}$$

$= I(\theta)$, car $\mathbb{E}\left(\frac{\partial \ln L(\theta; x_1,\dots,x_n)}{\partial \theta}\right) = 0$.

Donc

$$g'(\theta)^2 \leq R_{\hat{g}}(\theta)\, I(\theta) \iff \boxed{R_{\hat{g}}(\theta) \geq \frac{g'(\theta)^2}{I(\theta)}}$$

<u>Corollaire</u> Dans le modèle de Bernoulli décrit par les n-échantillons $X_1, \dots, X_n$ iid de lois $\mathcal{B}(\theta)$, $\theta \in]0,1[$, et θ comme paramètre d'intérêt, la moyenne empirique $\overline{X}_n$ est préférable à tous les estimateurs sans biais.

<u>Preuve</u> Soient $\theta \in]0,1[$ et $X_1, \dots, X_n$ iid de loi $\mathcal{B}(\theta)$. Calculons

$$I(\theta) = \mathrm{var}\left(\frac{\partial \ln L(\theta; X_1, \dots, X_n)}{\partial \theta}\right)$$

Comme $X_i \sim \mathcal{B}(\theta)$, i.e. $\mathbb{P}(X_i = 1) = \theta$ et $\mathbb{P}(X_1 = 0) = 1 - \theta$. Ainsi, si $y \in \{0,1\}$:

$$\mathbb{P}(X_i = y) = \theta^y (1-\theta)^{1-y} = \begin{cases} \theta & \text{si } y = 1 \\ 1-\theta & y = 0 \end{cases}$$

Alors, pour $y_1, \dots, y_n \in \{0,1\}$:

$$L(\theta; y_1, \dots, y_n) \overset{\text{déf}}{=} \mathbb{P}(X_1 = y_1, \dots, X_n = y_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = y_i) \quad \text{car } X_1, \dots, X_n \text{ indép.}$$

$$= \prod_{i=1}^{n} \theta^{y_i} (1-\theta)^{1-y_i} = \theta^{\sum_{i=1}^{n} y_i} (1-\theta)^{n - \sum_{i=1}^{n} y_i}$$

Notons $\bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i$, d'où

$$L(\theta; y_1, \dots, y_n) = \theta^{n\bar{y}_n}(1-\theta)^{n(1-\bar{y}_n)}$$

Par suite,

$$\ln L(\theta; y_1, \dots, y_n) = n\bar{y}_n \ln\theta + n(1-\bar{y}_n)\ln(1-\theta) \quad \text{"log-vraisemblance"}$$

d'où

$$\frac{\partial \ln L(\theta; y_1, \dots, y_n)}{\partial\theta} = \frac{n\bar{y}_n}{\theta} - \frac{n(1-\bar{y}_n)}{1-\theta}$$

Rappel $\forall a, b \in \mathbb{R}$

$var(az+b) = a^2 var(z)$

On a vu que $var(\bar{X}_n) = \frac{\theta(1-\theta)}{n}$, d'où

$$I(\theta) = var\left(\frac{\partial \ln L(\theta; X_1, \dots, X_n)}{\partial\theta}\right) = var\left(\frac{n\bar{X}_n}{\theta} - \frac{n(1-\bar{X}_n)}{1-\theta}\right)$$

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

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

De plus, comme $g(\theta)=\theta$, on a $g'(\theta)=1$. Soit alors $\hat{\theta}$ un estimateur sans biais de θ. D'après le théo. de Cramer-Rao :

$$R_{\hat{\theta}}(\theta) \geq \frac{g'(\theta)^2}{I(\theta)} = \frac{1}{\frac{n}{\theta(1-\theta)}} = \frac{\theta(1-\theta)}{n}$$

Or, on a déjà vu que $R_{\bar{X}_n}(\theta)=\frac{\theta(1-\theta)}{n}$, d'où

$$\boxed{R_{\hat{\theta}}(\theta) \geq R_{\bar{X}_n}(\theta)}$$

ce qui montre que $\bar{X}_n$ est préférable à tout autre estimateur sans biais.

On vient de montrer que dans le modèle de Bernoulli, $\bar{X}_n$ est préférable à tous les estimateurs sans biais de θ. Mais existe-t-il un autre estimateur sans biais, noté $\hat{\theta}_0$, tel que

$$\hat{R}_{\hat{\theta}_0}(\theta) \leq R_{\hat{\theta}}(\theta), \text{ pour tout estimateur } \hat{\theta} \text{ sans biais ?}$$

Supposons par l'absurde, qu'un tel estimateur $\hat{\theta}_0$ existe. Pour $x \in \mathbb{R}$, notons

$$\hat{\varphi}_x = \hat{\theta}_0 + x\left(\hat{\theta}_0 - \overline{X}_n\right)$$

Alors, $\hat{\varphi}_x$ est sans biais :

$$\begin{aligned}\mathbb{E}(\hat{\varphi}_x) &= \mathbb{E}(\hat{\theta}_0) + x\left(\mathbb{E}(\hat{\theta}_0) - \mathbb{E}(\overline{X}_n)\right) \\ &= \theta + x(\theta - \theta), \text{ car } \hat{\theta}_0 \text{ et } \overline{X}_n \text{ sont sans biais} \\ &= \theta\end{aligned}$$

Par suite, par hypothèse sur $\hat{\theta}_0$:

$$R_{\hat{\theta}_0}(\theta) \leq R_{\hat{\varphi}_x}(\theta)$$

Puis, on calcule :

$$\begin{aligned}R_{\hat{\varphi}_x}(\theta) &\overset{\text{déf}}{=} \mathbb{E}\left((\hat{\varphi}_x - \theta)^2\right) = \mathbb{E}\left(\left(\hat{\theta}_0 - \theta + x(\hat{\theta}_0 - \overline{X}_n)\right)^2\right) \\ &= \underbrace{\mathbb{E}\left((\hat{\theta}_0 - \theta)^2\right)}_{= R_{\hat{\theta}_0}(\theta)} + x^2\,\mathbb{E}\left((\hat{\theta}_0 - \overline{X}_n)^2\right) + 2x\,\mathbb{E}\left((\hat{\theta}_0 - \theta)(\hat{\theta}_0 - \overline{X}_n)\right)\end{aligned}$$

On a donc

$$x^2\,\mathbb{E}\left((\hat\theta_0 - X_n)^2\right) + 2x\,\mathbb{E}\left((\hat\theta_0 - \vartheta)(\hat\theta_0 - \overline{X}_n)\right) = R_{\hat\varphi_x}(\vartheta) - R_{\hat\theta_0}(\vartheta) \geq 0 \text{ par hyp.}$$

Ce polynôme de degré 2 en x a donc 0 ou 1 racine. Son discriminant vaut

$$\Delta = 4\,\mathbb{E}\left((\hat\theta_0 - \vartheta)(\hat\theta_0 - \overline{X}_n)\right)^2$$

Comme le polynôme en x ne prend que des valeurs ≥ 0, cela impose $\Delta = 0$, autrement dit

$$\mathbb{E}\left((\hat\theta_0 - \vartheta)(\hat\theta_0 - \overline{X}_n)\right) = 0$$

On peut montrer de même, avec l'estimateur $\hat\varphi_x = \overline{X}_n + x(\overline{X}_n - \hat\theta_0)$, que

$$\mathbb{E}\left((\overline{X}_n - \vartheta)(\hat\theta_0 - \overline{X}_n)\right) = 0$$

Alors,

$$\mathbb{E}\left((\hat\theta_0 - \overline{X}_n)^2\right) = \mathbb{E}\left((\hat\theta_0 - \overline{X}_n)(\hat\theta_0 - \overline{X}_n)\right) = \mathbb{E}\left((\hat\theta_0 - \vartheta + \vartheta - \overline{X}_n)(\hat\theta_0 - \overline{X}_n)\right)$$

$$= \underbrace{\mathbb{E}\left((\hat{\theta}_0-\theta)(\hat{\theta}_0-\bar{X}_n)\right)}_{=0} - \underbrace{\mathbb{E}\left((\bar{X}_n-\theta)(\hat{\theta}_0-\bar{X}_n)\right)}_{=0}$$

Donc $\mathbb{E}\left((\hat{\theta}_0-\bar{X}_n)^2\right)=0 \Rightarrow \boxed{\hat{\theta}_0 = \bar{X}_n}$: $\bar{X}_n$ est le seul estimateur sans biais qui soit préférable à tous les autres estimateurs sans biais.

<u>Exercice</u> Modèle de Poisson : $X_1, \dots, X_n$ iid de loi $\mathcal{P}(\theta)$, $\theta>0$.

On sait que $\mathbb{E}(X_1)=\theta$. Comme, d'après la LGN : $\bar{X}_n \approx \mathbb{E}(X_1)=\theta$, $\bar{X}_n$ est un estimateur "naturel" de θ. Alors, $\bar{X}_n$ est sans biais, d'où

$$R_{\bar{X}_n}(\theta) = var(\bar{X}_n) = var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{var(X_1)}{n}, \text{ car } X_1, \dots, X_n \text{ iid}$$
$$= \frac{\theta}{n}.$$

Calculons maintenant $I(\theta) = var\left(\frac{\partial \ln L(\theta; X_1, \dots, X_n)}{\partial \theta}\right)$.

On a $\mathbb{P}(X_i = y) = e^{-\theta}\frac{\theta^y}{y!}$, $y \in \mathbb{N}$. car $X_i \sim \mathcal{P}(\theta)$.

Ainsi, $\forall y_1, \dots, y_n \in \mathbb{N}$:

$$L(\theta; y_1, \dots, y_n) \overset{\text{déf}}{=} \mathbb{P}(X_1 = y_1, \dots, X_n = y_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = y_i) \text{ par indép. de } X_1, \dots, X_n$$

$$= \prod_{i=1}^{n} e^{-\theta} \frac{\theta^{y_i}}{y_i!} = \frac{e^{-n\theta}}{\prod_{i=1}^{n} y_i!} \underbrace{\theta^{n\overline{y_n}}}_{= e^{n\overline{y_n}\ln\theta}}, \quad \overline{y_n} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Puis,

$$\ln L(\theta; y_1, \dots, y_n) = -n\theta + n\overline{y_n}\ln(\theta) - \ln\left(\prod_{i=1}^{n} y_i!\right),$$

et, de ce fait :

$$\frac{\partial \ln L(\theta; y_1, \dots, y_n)}{\partial \theta} = -n + \frac{n\overline{y_n}}{\theta}$$

Par suite,

$$I(\theta) = \operatorname{var}\left(\frac{\partial \ln L(\theta; X_1, \dots, X_n)}{\partial \theta}\right) = \operatorname{var}\left(-n + \frac{n}{\theta}\overline{X}_n\right) = \frac{n^2}{\theta^2} \underbrace{\operatorname{var}(\overline{X}_n)}_{= \frac{\theta}{n}, \text{ cf + haut}}$$

$$= \frac{n^2}{\theta^2}\frac{\theta}{n} = \frac{n}{\theta}$$

D'après le théo. de Cramer-Rao, pour tout estimateur sans biais $\hat{\theta}$ de θ,

$$R_{\hat{\theta}}(\theta) \geq \frac{1}{I(\theta)} = \frac{\theta}{n} = R_{\bar{X}_n}(\theta)$$

Donc $\bar{X}_n$ est préférable à tous les estimateurs sans biais de θ.