Suite du CT n°5

(3) $X_1, \dots, X_n$ iid de loi $\mathcal{G}(\theta)$, $\theta \in]0,1[$.

Le paramètre d'intérêt est θ. Un estimateur naturel est $\hat{\theta} = \frac{1}{\bar{X}_n}$ et on veut le risque quadratique de cet estimateur $\hat{\theta}$.

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

Or, $\forall i \geq 1$, $X_i \geq 1$ (i.e. $\forall \omega \in \Omega$, $X_i(\omega) \geq 1$) donc $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \geq \frac{1}{n} \sum_{i=1}^{n} 1 = 1$.

Donc

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

Or, $\mathbb{E}(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(X_i) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\theta}$ car $\mathbb{E}(X_i) = \frac{1}{\theta}$ $\forall i = 1, \dots, n$ d'où

$$R_{\hat{\theta}}(\theta) \leq \theta^2 \, \mathbb{E}\left(\left(\mathbb{E}(\bar{X}_n) - \bar{X}_n\right)^2\right) = \theta^2 \, \mathrm{var}(\bar{X}_n) = \theta^2 \, \mathrm{var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \theta^2 \frac{1}{n^2} \mathrm{var}\left(\sum_{i=1}^{n} X_i\right)$$

$$= \frac{\theta^2}{n^2} \sum_{i=1}^{n} \mathrm{Var}(X_i) \text{ car } X_1, \dots, X_n \text{ sont indépendantes}$$

$$= \frac{\theta^2}{n^2} \sum_{i=1}^{n} \frac{1-\theta}{\theta^2}$$ car $X_i \sim \mathcal{G}(\theta)$ donc $\operatorname{var}(X_i) = \frac{1-\theta}{\theta^2}$ (cf feuille TD1).

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

Théorème (Décomposition Biais-Variance). Soient $\theta \in \Theta$ et $X_1, \dots, X_n$ un échantillon iid de la loi $\mathcal{L}_\theta$. Le risque quadratique de l'estimateur $\hat{g} = \hat{g}(X_1, \dots, X_n)$ se décompose comme suit :

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

Relation fondamentale : elle montre que, pour un risque donné, abaisser le biais de l'estimateur augmente sa variance, et réciproquement.

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

$$L(\theta; y_1, \dots, y_n) = \mathbb{P}(X_1 = y_1, \dots, X_n = y_n) \text{ avec } y_1, \dots, y_n \in \mathcal{D} \ (\subset \mathbb{Z}).$$

L est la vraisemblance du modèle (likelihood). On suppose que pour tous $y_1, \dots, y_n \in \mathcal{D}$, la fonction $\theta \mapsto L(\theta; y_1, \dots, y_n)$ est dérivable en θ. De plus, on note

$$\ell(\theta; y_1, \dots, y_n) = \ln L(\theta; y_1, \dots, y_n) \text{ la "log-vraisemblance"}$$

et on suppose que $\theta \mapsto \ell(\theta; y_1, \dots, y_n)$ est dérivable. On note alors $I(\theta)$ l'information de Fisher en θ :

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

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

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

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

Preuve On a ici $g(\theta) = \theta$ donc $g'(\theta) = 1$. Donc pour tout estimateur sans biais $\hat{g}$, on a d'après l'inégalité de Cramer-Rao :

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

Calculons $I(\theta)$. On a $\forall y_1, \dots, y_n \in \{0,1\}$:

$$L(\theta; y_1, \dots, y_n) = \mathbb{P}(X_1 = y_1, \dots, X_n = y_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = y_i)$$ car $X_1, \dots, X_n$ indépendantes

Or,

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

On a donc $L(\theta; y_1, \dots, y_n) = \prod_{i=1}^{n} \theta^{y_i} (1-\theta)^{1-y_i} = \theta^{\sum_{i=1}^{n} y_i} (1-\theta)^{n - \sum_{i=1}^{n} y_i}$

$$L(\theta; y_1, \dots, y_n) = \theta^{n\bar{y}_n} (1-\theta)^{n(1-\bar{y}_n)}$$, si $\bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i \iff n\bar{y}_n = \sum_{i=1}^{n} y_i$

Puis, la log-vraisemblance vaut :

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

Alors, sa dérivée en θ vaut :

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

Enfin, avec $X_1, \dots, X_n$ iid de loi $\mathcal{B}(\theta)$:

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

$$= \text{var}\left(\left(\frac{n}{\theta} + \frac{n}{1-\theta}\right)\bar{X}_n - \frac{n}{1-\theta}\right) = \text{var}\left(\left(\frac{n}{\theta} + \frac{n}{1-\theta}\right)\bar{X}_n\right)$$ car $\text{var}(Z+a) = \text{var}(Z)$ $\forall a \in \mathbb{R}$

$$= \left(\frac{n}{\theta} + \frac{n}{1-\theta}\right)^2 \underbrace{\text{var}(\bar{X}_n)}_{= \frac{\text{var}(X_1)}{n}}$$ car $\text{var}(aZ) = a^2\text{var}(Z)$ $\forall a \in \mathbb{R}$.

car $X_1, \dots, X_n$ iid.

Donc $I(\theta) = \left(\frac{n}{\theta} + \frac{n}{1-\theta}\right)^2 \frac{var(X_1)}{n} = \left(\frac{n}{\theta} + \frac{n}{1-\theta}\right)^2 \frac{\theta(1-\theta)}{n}$

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

On a vu que $R_{\bar{X}_n}(\theta) = \frac{\theta(1-\theta)}{n}$ donc $I(\theta) = \frac{1}{R_{\bar{X}_n}(\theta)}$. Par suite :

$$R_{\hat{g}}(\theta) \geq \frac{1}{I(\theta)} = R_{\bar{X}_n}(\theta) \qquad \forall \hat{g} \text{ estimateur sans biais.}$$

Donc $\bar{X}_n$ est préférable à tout autre estimateur sans biais.

<u>Exercice 5, feuille TD3</u> $X_1, \dots, X_n$ iid de loi $\mathcal{U}(\{1, \dots, k\})$ $(k \in \mathbb{N}^*)$ avec le paramètre d'intérêt k. On note $\hat{k}_1 = 2\bar{X}_n - 1$

① Biais et risque quadratique de $\hat{k}_1$.

$$\mathbb{E}(\hat{k}_1) = \mathbb{E}\left(2\bar{X}_n - 1\right) = 2\,\mathbb{E}(\bar{X}_n) - 1 = \frac{2}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) - 1$$

$$= \frac{2}{n}\sum_{i=1}^{n}\mathbb{E}(X_1) - 1 = 2\,\mathbb{E}(X_1) - 1$$

Or, $\mathbb{E}(X_1) = \sum_{i=1}^{k} i\,\mathbb{P}(X_1 = i) = \sum_{i=1}^{k} i\,\frac{1}{k} = \frac{1}{k}\sum_{i=1}^{k} i = \frac{1}{\cancel{k}}\,\frac{\cancel{k}(k+1)}{2} = \frac{k+1}{2}$

Donc $\mathbb{E}(\hat{k}_1) = 2\,\frac{k+1}{2} - 1 = k+1-1 = k$: $\hat{k}_1$ est sans biais.

Puis, $R_{\hat{k}_1}(k) = \mathbb{E}\left(\left(\hat{k}_1 - k\right)^2\right) = \mathbb{E}\left(\left(\hat{k}_1 - \mathbb{E}(\hat{k}_1)\right)^2\right) = \mathrm{var}(\hat{k}_1) = \mathrm{var}\left(2\bar{X}_n - 1\right)$

$$= 4\,\mathrm{var}(\bar{X}_n) = \frac{4}{n}\,\mathrm{var}(X_1) \text{ car } X_1, \dots, X_n \text{ iid}$$

Pour calculer $\mathrm{var}(X_1)$:

$$\mathbb{E}(X_1^2) = \sum_{i=1}^{k} i^2\,\mathbb{P}(X_1 = i) = \sum_{i=1}^{k} i^2\,\frac{1}{k} = \frac{1}{k}\sum_{i=1}^{k} i^2 = \frac{1}{\cancel{k}}\,\frac{\cancel{k}(k+1)(2k+1)}{6}$$

$$= \frac{(k+1)(2k+1)}{6}$$

Donc $\mathrm{var}(X_1) = \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2 = \frac{(k+1)(2k+1)}{6} - \frac{(k+1)^2}{4} = (k+1)\left(\frac{2k+1}{6} - \frac{k+1}{4}\right) = \frac{k^2-1}{12}$

Donc $R_{\hat{k}_1}(k) = \frac{4}{n}\,\frac{k^2-1}{12} = \frac{k^2-1}{3n}$

② Calculer un IC par excès pour k au niveau de confiance 95%.

$\forall a > 0$:

$$\mathbb{P}(|\hat{k}_1 - k| \geq a) = \mathbb{P}(|\hat{k}_1 - \overbrace{\mathbb{E}(\hat{k}_1)}^{\overset{k \text{ par ①}}{=}}| \geq a) \overset{\text{BT}}{\leq} \frac{\mathrm{Var}(\hat{k}_1)}{a^2} \overset{\text{par ①}}{=} \frac{k^2-1}{3na^2} \leq \frac{k^2}{3na^2}$$

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

On choisit a de sorte que $1 - \frac{k^2}{3na^2} = 0{,}95 \Leftrightarrow \frac{k^2}{3na^2} = 0{,}05 \Leftrightarrow a^2 = \frac{k^2}{3n \times 0{,}05}$

$\Leftrightarrow a = \frac{k}{\sqrt{0{,}15n}}$. Ainsi,

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

Or, $|\hat{k}_1 - k| < \frac{k}{\sqrt{0{,}15n}} \Leftrightarrow -\frac{k}{\sqrt{0{,}15n}} < \hat{k}_1 - k < \frac{k}{\sqrt{0{,}15n}} \Leftrightarrow \begin{cases} k\left(1 - \frac{1}{\sqrt{0{,}15n}}\right) < \hat{k}_1 \\ k\left(1 + \frac{1}{\sqrt{0{,}15n}}\right) > \hat{k}_1 \end{cases}$

$$\iff k \in \underbrace{\left] \frac{\hat{k}_1}{1+\frac{1}{\sqrt{0,15n}}} \,;\, \frac{\hat{k}_1}{1-\frac{1}{\sqrt{0,15n}}} \right[}_{\text{intervalle de confiance pour } k \text{, par excès au niveau } 95\%.}$$