

Examen partiel de statistique

Exercice 1 : Questions de cours :

1. (2 points) Donner la définition du biais et de l'erreur quadratique moyenne d'un estimateur.
2. (2 points) Soit $\theta \in \mathbb{R}$ et $\hat{\theta}_n$ un estimateur admettant un moment d'ordre 2. Démontrer la décomposition biais variance suivante :

$$\text{EQM}(\hat{\theta}_n, \theta) = B(\hat{\theta}_n, \theta)^2 + \mathbb{V}[\hat{\theta}_n].$$

Correction : Cours.

Exercice 2 : On considère des X_i i.i.d suivant une loi de Rademacher de paramètre $\theta \in (0, 1)$, i.e on a

$$\mathbb{P}[X_1 = 1] = \theta \quad \text{et} \quad \mathbb{P}[X_1 = -1] = 1 - \theta.$$

1. (2 points) Montrer que

$$\mathbb{E}[X_1] = 2\theta - 1 \quad \text{et} \quad \mathbb{V}[X_1] = 4\theta(1 - \theta).$$

2. (1 point) Proposer un estimateur à l'aide de la méthode de moments.
3. (2 points) Montrer qu'il est consistant.
4. (2 points) Montrer qu'il est asymptotiquement normal.
5. (3 points) En déduire un intervalle de confiance asymptotique de niveau $1 - \alpha$ avec $\alpha \in (0, 1)$.
6. (1 point) Montrer que la densité $f_\theta(\cdot)$ de X_1 peut s'écrire pour tout $x \in \{-1, 1\}$ de la forme

$$f_\theta(x) = \theta^{\frac{x+a}{2}} (1 - \theta)^{\frac{a-x}{2}}$$

pour un choix approprié de a .

7. (2 points) Donner l'estimateur du maximum de vraisemblance.

Correction :

1. **Option 1 :** On peut remarquer que X a la même loi que $2Y - 1$ ou $Y \sim \mathcal{B}(\theta)$ et on a directement

$$\mathbb{E}[X] = 2\mathbb{E}[Y] - 1 = 2\theta - 1 \quad \text{et} \quad \mathbb{V}[X] = 4\mathbb{V}[Y] = 4\theta(1 - \theta).$$

Option 2 : On peut directement calculer

$$\mathbb{E}[X] = (-1) \times \mathbb{P}[X = -1] + 1 \times \mathbb{P}[X = 1] = -(1 - \theta) + \theta = 2\theta - 1$$

et

$$\mathbb{E}[X^2] = (-1)^2 \times \mathbb{P}[X = -1] + 1^2 \times \mathbb{P}[X = 1] = (1 - \theta) + \theta = 1$$

et donc

$$\mathbb{V}[X] = 1 - (2\theta - 1)^2 = 1 - 4\theta^2 + 4\theta - 1 = 4\theta(1 - \theta).$$

2. On a $\theta = \frac{\mathbb{E}[X]+1}{2}$ et on prend donc l'estimateur

$$\hat{\theta}_n = \frac{\bar{X}_n + 1}{2}.$$

3. Par la loi forte des grands nombres, on a \bar{X}_n qui converge presque sûrement vers $\mathbb{E}[X] = 2\theta - 1$ et donc, comme la fonction $g : x \mapsto (x + 1)/2$ est continue,

$$\hat{\theta}_n = g(\bar{X}_n) \xrightarrow[n \rightarrow +\infty]{p.s.} g(2\theta - 1) = \theta.$$

A noter que l'on aurait pu avoir le résultat en remarquant que la transformation est affine.

4. Le TLC nous donne

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 4\theta(1 - \theta))$$

Option 1 : Comme $g'(x) = 1/2$, et en particulier $g'(2\theta - 1) = 1$, on a par la méthode delta

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \theta(1 - \theta)).$$

Option 2 : On peut juste "diviser par deux le TCL" et on a

$$\sqrt{n} \left(\frac{\bar{X}_n}{2} - (\theta - 1/2) \right) = \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \theta(1 - \theta)).$$

5. On peut réécrire la normalité asymptotique comme

$$\sqrt{n} \frac{\hat{\theta}_n - \theta}{\sqrt{\theta(1 - \theta)}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

De plus, on a

$$\sqrt{n} \frac{\hat{\theta}_n - \theta}{\sqrt{\hat{\theta}_n(1 - \hat{\theta}_n)}} = \frac{\sqrt{\theta(1 - \theta)}}{\sqrt{\hat{\theta}_n(1 - \hat{\theta}_n)}} \sqrt{n} \frac{\hat{\theta}_n - \theta}{\sqrt{\theta(1 - \theta)}}$$

Comme la fonction

$$x \mapsto \frac{\sqrt{\theta(1 - \theta)}}{\sqrt{x(1 - x)}}$$

est continue en $\theta \in (0, 1)$, on a

$$\frac{\sqrt{\theta(1 - \theta)}}{\sqrt{\hat{\theta}_n(1 - \hat{\theta}_n)}} \xrightarrow[n \rightarrow +\infty]{p.s.} 1.$$

A l'aide du théorème de Slutsky, on obtient donc

$$\sqrt{n} \frac{\hat{\theta}_n - \theta}{\sqrt{(\hat{\theta}_n(1 - \hat{\theta}_n))}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Soit $q_{1-\alpha/2}$ le quantile d'ordre $1 - \alpha/2$ de la loi normale centrée réduite, on a donc

$$\mathbb{P} \left[-q_{1-\alpha/2} \leq \sqrt{n} \frac{\hat{\theta}_n - \theta}{\sqrt{(\hat{\theta}_n(1 - \hat{\theta}_n))}} \leq q_{1-\alpha/2} \right] \xrightarrow[n \rightarrow +\infty]{} 1 - \alpha$$

et on obtient donc l'intervalle de confiance

$$IC_{1-\alpha}(\theta) = \left[\bar{X}_n \pm q_{1-\alpha/2} \frac{\sqrt{\hat{\theta}_n(1 - \hat{\theta}_n)}}{\sqrt{n}} \right].$$

6. En prenant $a = 1$, on retrouve

$$f_\theta(1) = \theta \quad \text{et} \quad f_\theta(-1) = 1 - \theta$$

7. On a la vraisemblance

$$L_{\mathbf{X}}(\theta) = \prod_{i=1}^n \theta^{\frac{X_i+1}{2}} (1 - \theta)^{\frac{1-X_i}{2}} = \theta^{\sum_{i=1}^n \frac{X_i+1}{2}} (1 - \theta)^{\sum_{i=1}^n \frac{1-X_i}{2}} = \theta^{\frac{1}{2}(n\bar{X}_n+n)} (1 - \theta)^{\frac{1}{2}(n-n\bar{X}_n)}$$

On a donc la log vraisemblance

$$\ell_{\mathbf{X}}(\theta) = \frac{1}{2}(n\bar{X}_n + n) \log(\theta) + \frac{1}{2}(n - n\bar{X}_n) \log(1 - \theta).$$

En dérivant et en mettant au même dénominateur, on obtient

$$\ell'_X(\theta) = \frac{n\bar{X}_n + n}{2\theta} - \frac{n - n\bar{X}_n}{2(1-\theta)} = \frac{n\bar{X}_n + n - 2n\theta}{2\theta(1-\theta)} \geq 0 \iff \theta \leq \frac{n\bar{X}_n + n}{2n} = \frac{\bar{X}_n + 1}{2}$$

et on obtient donc l'estimateur du maximum de vraisemblance $\hat{\theta}_n^{MV} = \frac{\bar{X}_n + 1}{2}$.

Exercice 3 : On considère des variables aléatoires X_i indépendantes et $\theta > 0$ tels que pour tout $i \geq 1$,

$$\mathbb{E}[X_i] = \frac{i+1}{i}\theta \quad \text{et} \quad \mathbb{V}[X_i] = \left(\frac{i+1}{i}\right)^2 \theta^2$$

L'objectif est d'estimer θ . Dans ce qui suit, on considère l'estimateur $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

1. Calculer son biais.
2. Calculer son erreur quadratique moyenne.
3. Montrer qu'il converge en probabilité. On admettra

$$\sum_{i=1}^n \frac{1}{i} \leq 1 + \log(n) \quad \text{et} \quad \sum_{i=1}^n \frac{1}{i^2} \leq 2.$$

4. On considère maintenant l'estimateur $\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n \frac{i}{i+1} X_i$. Calculer son biais et son erreur quadratique moyenne.
5. Quel estimateur choisiriez vous ?

Correction :

1. (2 points) On a par linéarité

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n} \sum_{i=1}^n \theta \frac{i+1}{i} = \theta + \frac{\theta}{n} \sum_{i=1}^n \frac{1}{i}.$$

On obtient donc pour tout $\theta > 0$

$$B(\bar{X}_n, \theta) = \frac{\theta}{n} \sum_{i=1}^n \frac{1}{i}$$

2. (2 points) On a par indépendance

$$\begin{aligned} \mathbb{V}[\bar{X}_n] &= \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}[X_i] = \frac{1}{n^2} \sum_{i=1}^n \left(\frac{i+1}{i}\right)^2 \theta^2 \\ &= \frac{\theta^2}{n^2} \sum_{i=1}^n \left(1 + \frac{2}{i} + \frac{1}{i^2}\right) \\ &= \frac{\theta^2}{n} + \frac{2\theta^2}{n^2} \sum_{i=1}^n \frac{1}{i} + \frac{\theta^2}{n^2} \sum_{i=1}^n \frac{1}{i^2} \end{aligned}$$

Pour tout $\theta > 0$, à l'aide de la décomposition biais variance, on obtient l'erreur quadratique moyenne

$$EQM(\bar{X}_n, \theta) = \frac{\theta^2}{n} + \frac{2\theta^2}{n^2} \sum_{i=1}^n \frac{1}{i} + \frac{\theta^2}{n^2} \sum_{i=1}^n \frac{1}{i^2} + \frac{\theta^2}{n^2} \left(\sum_{i=1}^n \frac{1}{i} \right)^2$$

3. (2 points) Grâce aux inégalités données dans l'énoncé, on a

$$EQM(\bar{X}_n, \theta) \leq \frac{\theta^2}{n} + \frac{2\theta^2}{n^2} (1 + \log(n)) + \frac{2\theta^2}{n^2} + \frac{\theta^2}{n^2} (1 + \log(n))^2 \xrightarrow{n \rightarrow +\infty} 0.$$

L'estimateur converge donc en moyenne quadratique et donc en probabilité.

4. (3 points) On a par linéarité

$$\mathbb{E}[\hat{\theta}_n] = \frac{1}{n} \sum_{i=1}^n \frac{i}{i+1} \mathbb{E}[X_1] = \frac{1}{n} \sum_{i=1}^n \frac{i}{i+1} \frac{i+1}{i} \theta = \theta.$$

L'estimateur est donc sans biais. De plus, à l'aide de la décomposition biais variance et par indépendance

$$\begin{aligned} EQM(\hat{\theta}_n, \theta) &= \mathbb{V}[\hat{\theta}_n] = \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}\left[\frac{i}{i+1} X_1\right] \\ &= \frac{1}{n^2} \sum_{i=1}^n \left(\frac{i}{i+1}\right)^2 \mathbb{V}[X_1] \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{i}{i+1}\right)^2 \left(\frac{i+1}{i}\right)^2 \theta^2 \\ &= \frac{\theta^2}{n} \end{aligned}$$

5. (1 point) On a pour tout $\theta > 0$

$$EQM(\bar{X}_n, \theta) - EQM(\hat{\theta}_n, \theta) = \frac{2\theta^2}{n^2} \sum_{i=1}^n \frac{1}{i} + \frac{\theta^2}{n^2} \sum_{i=1}^n \frac{1}{i^2} + \frac{\theta^2}{n^2} \left(\sum_{i=1}^n \frac{1}{i} \right)^2 > 0$$

et donc $\hat{\theta}_n$ est un meilleur estimateur.