

Exercices, 6

SOLUTION 1 (Exercice 7). —

1. Pour une variable Z on note φ_Z sa fonction caractéristique. Soit $t \in \mathbb{R}$, on a

$$\begin{aligned}\varphi_{\hat{m}_{naive}}(t) &= \varphi_{n^{-1} \sum_{i=1}^n X_i}(t) = \varphi_{\sum_{i=1}^n X_i}(t/n) \\ &= \varphi_{X_1}(t/n)^n \quad (\text{les } X_i \text{ sont i.i.d.}) \\ &= \left(e^{im(t/n) - |t|/n} \right)^n = e^{imt - |t|} = \varphi_{X_1}(t).\end{aligned}$$

On en déduit que $\hat{m}_{naive} \sim \mathcal{C}(1)$, et donc qu'il ne peut pas être consistant: par exemple

$$\begin{aligned}P_m(|\hat{m}_{naive} - m| \geq 1) &= \mathbb{P}(|\mathcal{C}(1)| \geq 1) \\ &= \frac{2}{\pi} \int_1^{+\infty} \frac{1}{1+x^2} dx \\ &= \frac{2}{\pi} [\arctan(x)]_1^{+\infty} = 1/2 \not\rightarrow_{n \rightarrow +\infty} 0.\end{aligned}$$

2. Le modèle de cette expérience s'écrit

$$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (m + \mathcal{C}(1))_{m \in \mathbb{R}}^{\otimes n}).$$

Pour $x_1, \dots, x_n \in \mathbb{R}$, la vraisemblance s'écrit

$$V_{x_{1:n}}(m) = \pi^{-n} \prod_{i=1}^n (1 + (x_i - m)^2)^{-1}.$$

Comme $V_{x_{1:n}} > 0$ pour tout $x_{1:n}$, $\hat{m}_{EMV}$ maximise nécessairement

$$\begin{aligned}\ell_{x_{1:n}}(m) &= \log(V_{x_{1:n}}(m)) \\ &= -n \log(\pi) - \sum_{i=1}^n \log(1 + (x_i - m)^2).\end{aligned}$$

Comme $\ell_{x_{1:n}}$ tend vers $-\infty$ lorsque $m \rightarrow \pm\infty$, $\hat{m}_{EMV}$ est nécessairement un extremum local de $\ell_{x_{1:n}}$, et est donc solution de l'équation

$$\begin{aligned}\ell'_{x_{1:n}}(m) = 0 &\Leftrightarrow 2 \sum_{i=1}^n \frac{(m - x_i)}{1 + (m - x_i)^2} = 0 \\ &\Leftrightarrow \sum_{i=1}^n (m - x_i) \prod_{j \neq i} (1 + (m - x_j)^2) = 0.\end{aligned}$$

$\hat{m}_{EMV}$ est donc racine du polynôme ci-dessus (de degré $2n - 1$).

3. Comme $\frac{1}{2n+1} |\{X_i \mid X_i \geq X_{(n+1)}\}| = \frac{1}{2n+1} |\{X_i \mid X_i \leq X_{(n+1)}\}| = (n+1)/(2n+1) \geq 1/2$, on déduit que $\hat{m} = \hat{X}_{(n+1)}$.

4. Soit $\varepsilon > 0$. On a

$$\begin{aligned} P_m(\hat{m} \leq m - \varepsilon) &= P_m\left(\frac{1}{2n+1} \sum_{i=1}^{2n+1} \mathbb{1}_{X_i \leq m - \varepsilon} \geq 1/2\right) \\ &= \mathbb{P}\left(\frac{1}{2n+1} \sum_{i=1}^{2n+1} \mathbb{1}_{Z_i \leq -\varepsilon} \geq 1/2\right), \end{aligned}$$

où $Z_1, \dots, Z_n$ sont i.i.d. de loi $\mathcal{C}(1)$. Notons

$$p(x) = \mathbb{P}(Z_1 \leq x) = \frac{1}{\pi} \left(\arctan(x) + \frac{\pi}{2} \right) := 1/2 + \delta(x).$$

On a alors

$$\begin{aligned} P_m(\hat{m} \leq m - \varepsilon) &= \mathbb{P}\left(\frac{1}{2n+1} \left(\sum_{i=1}^{2n+1} \mathbb{1}_{Z_i \leq -\varepsilon} - p(-\varepsilon) \right) \geq -\delta(-\varepsilon)\right) \\ &= \mathbb{P}\left(\frac{1}{2n+1} \left(\sum_{i=1}^{2n+1} \mathbb{1}_{Z_i \leq -\varepsilon} - p(-\varepsilon) \right) \geq \delta(\varepsilon)\right) \end{aligned}$$

Comme $\mathbb{1}_{Z_1 \leq -\varepsilon}$ est intégrable, la loi des grands nombres donne

$$P_m(\hat{m} \leq m - \varepsilon) \xrightarrow{n \rightarrow +\infty} 0.$$

5. Soit $\alpha \in]0, 1[$, $\varepsilon > 0$, et notons B_ε la variable $\sum_{i=1}^{2n+1} \mathbb{1}_{Z_i \leq \varepsilon}$. De la question précédente, en utilisant l'inégalité de Hoeffding, on déduit

$$\begin{aligned} P_m(\hat{m} \leq m - \varepsilon) &= \mathbb{P}\left(\frac{1}{2n+1} (B_{-\varepsilon} - \mathbb{E}(B_{-\varepsilon})) \geq \delta(\varepsilon)\right) \\ &\leq \exp(-2n\delta(\varepsilon)^2). \end{aligned}$$

Dans l'autre sens, on a

$$\begin{aligned} P_m(\hat{m} \geq m + \varepsilon) &= P_m\left(\frac{1}{2n+1} \sum_{i=1}^{2n+1} \mathbb{1}_{X_i \leq m + \varepsilon} \leq \frac{1}{2}\right) \\ &= \mathbb{P}\left(\frac{1}{2n+1} (B_\varepsilon - \mathbb{E}(B_\varepsilon)) \leq \delta(\varepsilon)\right) \\ &\leq \exp(-2n\delta(\varepsilon)^2). \end{aligned}$$

En posant $\varepsilon_\alpha = \tan\left(\pi \sqrt{\frac{\log(2/\alpha)}{2n}}\right)$, on a

$$P_m(|\hat{m} - m| \geq \varepsilon_\alpha) \leq \alpha,$$

et donc $[\hat{m} - \varepsilon_\alpha, \hat{m} + \varepsilon_\alpha]$ est un intervalle de niveau de confiance $1 - \alpha$.

6. On pose $\varepsilon_n = \sqrt{\frac{10 \log(n)}{2n+1}}$, et $A_n = \{|\hat{m} - m| \geq \varepsilon_n\}$. D'après la question 5-, on a

$$\begin{aligned} P_m(A_n) &\leq 2 \exp\left(-2n/\pi^2 \arctan^2\left(\sqrt{\frac{10 \log(n)}{2n+1}}\right)\right) \\ &\leq 2 \exp\left(-\frac{10}{\pi^2} \log(n) + O(\log(n)^2/n)\right) \\ &= O\left(n^{-10/\pi^2}\right). \end{aligned}$$

On en déduit que $\sum_{n \geq 1} \mathbb{P}(A_n) < +\infty$, et donc, en utilisant le Lemme de Borel Cantelli, que

$$P_m(\limsup A_n) = 0.$$

On en déduit alors que $P_m(\liminf A_n^c) = 1$, et donc que $\hat{m} \xrightarrow[n \rightarrow +\infty]{p.s.} m$.

7. Soit $x \in \mathbb{R}$. On a

$$\begin{aligned} P_m\left(\frac{2}{\pi}\sqrt{2n+1}(\hat{m} - m) \leq x\right) &= P_m\left(\hat{m} \leq m + \frac{\pi x}{2\sqrt{2n+1}}\right) \\ &= \mathbb{P}\left(\frac{1}{2n+1}(B_{\varepsilon(x)} - \mathbb{E}(B_{\varepsilon(x)})) \geq -\delta(\varepsilon(x))\right), \end{aligned}$$

avec $\varepsilon(x) = \frac{\pi x}{2\sqrt{2n+1}}$. Notons $p_{2n+1} = 1/2 + \delta(\varepsilon(x))$. On a $(2n+1)p_{2n+1}(1-p_{2n+1}) \xrightarrow[n \rightarrow +\infty]{} +\infty$, on en déduit alors que

$$\frac{1}{\sqrt{2n+1}p_{2n+1}(1-p_{2n+1})}(B_{\varepsilon(x)} - \mathbb{E}(B_{\varepsilon(x)})) \rightsquigarrow \mathcal{N}(0, 1).$$

On a alors

$$\begin{aligned} P_m\left(\frac{2}{\pi}\sqrt{2n+1}(\hat{m} - m) \leq x\right) \\ = \mathbb{P}\left(\frac{1}{\sqrt{(2n+1)p_{2n+1}(1-p_{2n+1})}}(B_{\varepsilon(x)} - \mathbb{E}(B_{\varepsilon(x)})) \geq -\sqrt{\frac{2n+1}{p_{2n+1}(1-p_{2n+1})}}\delta(\varepsilon(x))\right). \end{aligned}$$

Or,

$$\sqrt{\frac{2n+1}{p_{2n+1}(1-p_{2n+1})}}\delta(\varepsilon(x)) \sim_{n \rightarrow +\infty} 2\sqrt{2n+1}\frac{\varepsilon(x)}{\pi} \xrightarrow[n \rightarrow +\infty]{} x.$$

On en déduit que

$$P_m\left(\frac{2}{\pi}\sqrt{2n+1}(\hat{m} - m) \leq x\right) \xrightarrow[n \rightarrow +\infty]{} \mathbb{P}(\mathcal{N}(0, 1) \geq -x) = \mathbb{P}(\mathcal{N}(0, 1) \leq x),$$

et donc que $\frac{2}{\pi}\sqrt{2n+1}(\hat{m} - m) \rightsquigarrow \mathcal{N}(0, 1)$.