M2RI Asymptotic Statistics Exercices

Exercice n°1

Let P a measure on $\mathbb{R}$. F_1, F_2 two sets of functions from $\mathbb{R} \to \mathbb{R}$ s.t. $\forall f \in F_1 \cup F_2$, $\int |f| dP < \infty$.

Show that $\forall \varepsilon > 0$, $N_{[\,]}(F_1 + F_2, L^1(P), 2\varepsilon)$

$$\leq N_{[\,]}(F_1, L^1(P), \varepsilon) N_{[\,]}(F_2, L^1(P), \varepsilon).$$

Solution Let $f = f_1 + f_2 \in F_1 + F_2$. There exists $[\ell_1, u_1]$ (resp $[\ell_2, u_2]$) a bracket containing f_1 (resp f_2) st $\int (u_1 - \ell_1) dP$ (resp. $\int (u_2 - \ell_2) dP$) $< \varepsilon$.

Then $[\ell_1 + \ell_2, u_1 + u_2]$ is a bracket for $f_1 + f_2$.

Furthermore, $\int (u_1 + u_2) - (\ell_1 + \ell_2) \, dP \leq \int u_1 - \ell_1 \, dP + \int u_2 - \ell_2 \, dP$

$$\leq 2\varepsilon.$$

Exercice n°2

$X_1, \dots, X_n \overset{iid}{\sim} \mathrm{Unif}([-1,1])$.

S.t. $\prod_{i=1}^{n} (1 + X_i/\sqrt{n})$ converges in distribution and determine the limit.

Hint: we can use that $|\log(1 + X_i/\sqrt{n}) - X_i/\sqrt{n} + \frac{1}{2} X_i^2/n| \leq a_i / n^{3/2}$

where $a_i \geq 0$ and $\max_{1 \leq i \leq n} a_i = O(1)$. (To show if used).

Solution First, let's apply Taylor-Lagrange to $f(x) = \ln(1+x)$

in 0. $f'(x) = \frac{1}{\dots}$; $f''(x) = \frac{-1}{(1+x)^2}$; $f'''(x) = \frac{2}{(1+x)^3}$

$$\forall i,\ \exists u \in \left(0, \frac{X_i}{\sqrt{n}}\right) \text{ s.t. } \log\left(1+\frac{X_i}{\sqrt{n}}\right) = \frac{X_i}{\sqrt{n}} - \frac{1}{2}\frac{X_i^2}{n} + \frac{1}{6}\frac{2}{(1+u)^2}\frac{X_i^3}{n^{3/2}}$$

Furthermore, since a.s. $|X_i| < 1$,

$$\text{a.s. } \left|\frac{1}{6}\frac{2}{(1+u)^2} X_i^3\right| \leq \frac{1}{3}\frac{1}{\left(1-\frac{1}{\sqrt{n}}\right)^2} = O(1).$$

Then, $$\log\left(\prod_{i=1}^{n}\left(1+\frac{X_i}{\sqrt{n}}\right)\right) = \sum_{i=1}^{n}\log\left(1+\frac{X_i}{\sqrt{n}}\right)$$
$$= \sum_{i=1}^{n} \frac{X_i}{\sqrt{n}} - \frac{1}{2}\frac{X_i^2}{n} + R_i$$
with $|R_i| \leq \frac{C}{n^{3/2}}$ a.s.

CLT: $$\sum_{i=1}^{n}\frac{X_i}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma^2) \quad \sigma^2 = Var(X_i) = \frac{1}{3}$$

LLN: $$-\sum_{i=1}^{n}\frac{X_i^2}{n} \xrightarrow{\mathbb{P}} -\mathbb{E}(X_1^2) = -\frac{1}{3}$$
$$\sum_{i=1}^{n} R_i \xrightarrow{a.s.} 0$$

So Slutsky + CMT $\Rightarrow$ $$\log\left(\prod_{i=1}^{n}\left(1+\frac{X_i}{\sqrt{n}}\right)\right) \longrightarrow \mathcal{N}\left(-\frac{1}{6}, \sigma^2\right).$$
$$= \mathcal{N}\left(-\frac{1}{6}, \frac{1}{3}\right).$$

Exercice n°3:

$(X_1,Y_1), \dots, (X_n,Y_n)$ iid s.t. $X_1 \sim \mathcal{L}$ on $\mathbb{R}$ with density $f > 0$ continuous and $Y_1 = X_1 + Z_1$, $Z_1 \perp X_1$ and $Z_1 \sim \mathcal{N}(0,1)$.

Assumption: $\mathbb{E}(|X_1|) < \infty$.

$$\hat{\theta} \in \operatorname*{argmin}_{\theta \in [\frac{1}{2}, 2]} \sum_{i=1}^{n} |Y_i - \theta X_i|$$

<u>1)</u> Show that $\theta \mapsto \mathbb{E}(|Y_1 - \theta X_1|)$ is continuous on $[\frac{1}{2}, 2]$ and admits a unique minimiser at $\theta_0 = 1$.

<u>Hint:</u> We can use (without proof) that $\mathbb{E}(|W+a|) > \mathbb{E}(|W|)$ whenever $a \neq 0$ and $W \sim \mathcal{N}(0,1)$.

<u>2)</u> Show that

$$\sup_{\theta \in [\frac{1}{2}, 2]} \left| \frac{1}{n} \sum_{i=1}^{n} |Y_i - \theta X_i| - \mathbb{E}(|Y_1 - \theta X_1|) \right| \xrightarrow{\mathbb{P}} 0 .$$

<u>3)</u> Conclude that $\hat{\theta}_n \xrightarrow{\mathbb{P}} 1$.

<u>Solution:</u>

<u>1)</u> $|Y_1 - \theta X_1| \leq |Y_1| + |\theta| |X_1| \leq |Y_1| + 2|X_1|$

and $\mathbb{E}(|Y_1| + 2|X_1|) < \infty$.

So by continuity under the integral, $\theta \mapsto \mathbb{E}(|Y_1 - \theta X_1|)$ is continuous.

Furthermore, if $\delta \in [-\frac{1}{2}, 1]$,

$$\mathbb{E}(|Y_1 - (1+\delta)X_1|) = \mathbb{E}(|X_1 + Z_1 - (1+\delta)X_1|)$$

$$= \mathbb{E}(|Z_1 - \delta X_1|) = \mathbb{E}\Big(\underbrace{\mathbb{E}(|Z_1 - \delta X_1| \mid X_1)}_{> \mathbb{E}(|Z_1|) \text{ if } \delta \neq 0}\Big)$$

$$> \mathbb{E}(|Z_1|) \text{ if } \delta \neq 0$$

So it is minimal for $\delta = 0$ and this minimiser is unique.

2) $\mathcal{F} = \{ f_\theta(x,y) = |y - \theta x| : \theta \in [\frac{1}{2}, 2] \}$.

$$G(x,y) = |y| + 2|x|$$

$\forall f \in \mathcal{F}$, $\int |f| \, d\mathbb{P} \leq \int |G| \, d\mathbb{P} < \infty$.

Lipschitz property $|f_\theta(x,y) - f_{\theta'}(x,y)|$

$$= \big| |y - \theta x| - |y - \theta' x| \big|$$

$$\leq |\theta - \theta'| |x|$$

Brackets Construction: Let $\varepsilon > 0$, $\Delta := \dfrac{\varepsilon}{4 \mathbb{E}(|X_1|)}$

$\theta_k = \frac{1}{2} + k\Delta$, $k = 0, 1, \dots, K$ $\quad K \simeq \dfrac{2 - \frac{1}{2}}{\Delta}$

$$[l_k, u_k] := [f_{\theta_k} - \Delta |x|, f_{\theta_k} + \Delta |x|].$$

- If $|\theta - \theta_k| \leq \Delta$, $f_\theta \in [l_k, u_k]$.
- $\int |u_k - l_k| \, d\mathbb{P} = \mathbb{E}(2\Delta |X|) = \varepsilon / 2 < \varepsilon$

So, $\mathcal{N}_{[\,]}(\mathcal{F}, L^1, \varepsilon) \leq K + 1 < +\infty$.

So $\mathcal{F}$ is Glivenko Cantelli.

3) The consistency theorem applies and $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0 = 1$.