Autour du théorème de Glivenko-Cantelli

Def : $F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \le t}$ est la fonction de répartition empirique associée à l'échantillon $(X_1, \dots, X_n)$

Rq : $nF_n(t) = \sum_{i=1}^{n} \mathbb{1}_{X_i \le t} \sim \sum_{i=1}^{n} \mathcal{B}(F_X(t)) \sim \mathcal{B}(n, F_X(t))$. $X_i \sim X$

En triant les X_i en $X_{(i)}$ on a aussi : $F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_{(i)} \le t}$.

Prop : F_n est une fonction en escalier, croissante, continue à droite et admettant une limite à gauche. Elle est discontinue aux points $(X_{(i)})_{1\le i\le n}$ et constante sur $[X_{(i)}, X_{(i+1)}[$.

Prop : $F_n(t)$ est un estimateur sans biais et fortement consistant de $F(t)$.
De plus, il est asymptotiquement normal : $\sqrt{n}\,(F_n(t) - F(t)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, F(t)(1-F(t)))$.

Dem : Sans biais : trivial. Fortement consistant : LfGN.
Asymptotiquement normal : TCL

Thm : (Glivenko-Cantelli) On a plus que la cvps on a de la cvu :

$$\sup_{t\in\mathbb{R}} |F_n(t) - F(t)| \xrightarrow[n\to+\infty]{} 0 \text{ p.s.}$$

Lm : (2e thm de Dini) Soit (f_n) une suite de fonctions réelles croissantes définies sur $[a,b]$. Si (f_n) cvs vers $f \in C^0([a,b])$ alors il y a cvu.

Dem du Lm : $f \in C^0([a,b]) \xRightarrow{\text{Heine}}$ pour $\varepsilon > 0$, $\exists \eta > 0$ tq $\forall x, y \in [a,b]$ $|x-y| < \eta \Rightarrow |f(x) - f(y)| < \varepsilon$
On se donne $a = x_0 < \dots < x_p = b$ une subdivision de pas $< \eta$
Puisque $f_n(x_i) \to f(x_i)$ on a $\exists N$ tq $\forall n \ge N$ $|f(x_i) - f_n(x_i)| < \varepsilon$.
Soit $x \in [a,b]$ et i tq $x \in [x_i, x_{i+1}]$. Par ↗ de f_n :

$$\begin{aligned}
|f(x) - f_n(x)| &\le |f(x) - f(x_i)| + |f(x_i) - f_n(x_i)| + |f_n(x_i) - f_n(x)| \\
&\le 2\varepsilon + f_n(x) - f_n(x_i) \le 2\varepsilon + f_n(x_{i+1}) - f_n(x_i) \\
&\le 2\varepsilon + |f_n(x_{i+1}) - f(x_{i+1})| + |f(x_{i+1}) - f(x_i)| + |f(x_i) - f_n(x_i)| \\
&\le 2\varepsilon + \varepsilon + \varepsilon + \varepsilon = 5\varepsilon.
\end{aligned}$$

Donc $\forall \varepsilon > 0\ \exists N \in \mathbb{N}$ tq $\forall n \ge N$ $\sup_{x\in[a,b]} |f(x) - f_n(x)| < \varepsilon$. □

Dem : (Glivenko - Cantelli)

On utilise F^{-1} l'inverse généralisée : $F^{-1}(u) \leq x$ ssi $u \leq F(x)$ $\forall x \in \mathbb{R}$ $\forall u \in [0,1]$

$F^{-1}(u) = \inf\{x \in \mathbb{R}, F(x) \geq u\}$ et si $\begin{cases} Y \text{ a pour fdr } G \\ U \sim \mathcal{U}([0,1]) \end{cases}$ alors $G^{-1}(U) \sim Y$

Donc $\sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \sim \sup_{t \in \mathbb{R}} \left| \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{F^{-1}(U_k) \leq t} - F(t) \right| = \sup_{t \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{U_k \leq F(t)\}} - F(t) \right|$

$$\leq \sup_{s \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{U_k \leq s\}} - s \right| \quad (*)$$

Par LfGN, $\frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{U_k \leq s} \xrightarrow{ps} s$ $\forall s$ ($\forall s \, \exists N_s$ tq $P(N_s) = 0$ ... et on intervertit pour $s \in \mathbb{Q}$ et on encadre)

On prend $s \in \mathbb{Q}$ et $w \in N^c$ (où $P(N) = 0$) on a : $\frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{U_k \leq s} \xrightarrow[n \to +\infty]{} s$

Et on revient à $s \in \mathbb{R}$ en encadrant la somme avec $p \leq s \leq q \in \mathbb{Q}$, $p \in \mathbb{Q}$

Or $s \mapsto \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{U_i \leq s}$ est $\nearrow$

D'où : $\forall w \in N^c$: $\cdot$ $s \mapsto \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{U_k(w) \leq s\}}$ est $\nearrow$ $\forall n \geq 1$

$\cdot$ $\frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{U_k(w) \leq s\}} \to s$

$\cdot$ $s \mapsto s \in C^0([0,1])$

Par Dini, il y a cvu. (*) conclut. $\square$

Application (Kolmogorov-Smirnov)

Si $F \in C^0$ alors $\sqrt{n} \|F_n - F\|_\infty \xrightarrow{\mathcal{L}} \mu_{KS}$.

Et on a $F_n(X_{(j)}) = j/n$ d'où $\|F_n - F\|_\infty = \max_{1 \leq j \leq n} \max \left(\left| \frac{j}{n} - F(X_{(j)}) \right|, \left| F(X_{(j)}) - \frac{j-1}{n} \right| \right)$ où $\frac{j}{n} = F_n(X_{(j)}^+)$, $\frac{j-1}{n} = F_n(X_{(j)}^-)$

On calcule $D(F_n, F_0) = \|F_n - F_0\|_\infty$ où H_0 "$F = F_0$"

H_1 = "$F = F_1$"

Et on cherche $q_{1-\alpha}$ le quantile de la loi de K-S : on ne rejette pas H_0 si $\sqrt{n}\|F_n - F_0\|_\infty < q_{1-\alpha}$