VM
24/04/15
1

Théorème de Glivenko - Cantelli.

v.a à densité : 249, 262, 263
v.a quelconques : 229

Théorème :

Soit (X_k) une suite de v.a iid. Soit $F : t \mapsto \mathbb{P}(X_1 \leq t)$ la fonction de répartition associée. Soit $F_n : t \mapsto \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{X_k \leq t}$.

Alors $\|F_n - F\|_\infty \xrightarrow[n\to+\infty]{ps} 0$.

dém : Cas où la loi de X_1 est à densité, i.e F continue.

On démontre tout d'abord le théorème suivant :

Théorème (loi forte des grands nombres, variables de Bernoulli)

Soit $(X_i)_i$ une suite iid de loi de Bernoulli de paramètre p.
Soit $S_n = \sum_{i=1}^{n} X_i$. Alors $\frac{S_n}{n} \xrightarrow[n\to+\infty]{ps} p$.

dém : On pose $Y_i = X_i - p$. On a $|Y_i| \leq 1 \quad \forall i$.

Soit $\varepsilon > 0$.

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) = \mathbb{P}\left(\left|\frac{S_n}{n} - p\right|^4 > \varepsilon^4\right)$$

$$\underset{\text{inégalité de Markov}}{\leq} \frac{1}{\varepsilon^4}\, \mathbb{E}\left[\left(\frac{S_n}{n} - p\right)^4\right].$$

On $\left(\frac{S_n}{n} - p\right)^4 = \frac{1}{n^4}\left(\sum_{i=1}^{n} Y_i\right)^4 = \frac{1}{n^4}\sum_{1\leq i_1, i_2, i_3, i_4 \leq n} Y_{i_1} Y_{i_2} Y_{i_3} Y_{i_4}$

Si $i_1 \notin \{i_2, i_3, i_4\}$, par indépendance des Y_i, on a

$$\mathbb{E}[Y_{i_1} Y_{i_2} Y_{i_3} Y_{i_4}] = \mathbb{E}[Y_{i_1}]\, \mathbb{E}[Y_{i_2} Y_{i_3} Y_{i_4}] = 0.$$

Ainsi, $\mathbb{E}\left[\left(\frac{S_n}{n} - p\right)^4\right] = \frac{1}{n^4}\left[\sum_{i=1}^{n} \underbrace{\mathbb{E}[Y_i^4]}_{\leq 1} + \binom{4}{2}\sum_{i<j} \underbrace{\mathbb{E}[Y_i^2 Y_j^2]}_{\leq 1}\right]$

$$\leq \frac{1}{n^4}\left[n + 6\,\frac{n(n-1)}{2}\right] \leq \frac{K}{n^2} \text{ pour un certain } K.$$

Donc $\sum_{n\geq 1} \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \leq \frac{K}{\varepsilon^4}\sum_{n\geq 1}\frac{1}{n^2} < +\infty$.

D'après le lemme de Borel-Cantelli, $\mathbb{P}\left(\exists N, \forall n \geq N, \left|\frac{S_n}{n} - p\right| \leq \varepsilon\right) = 1$.

On par intersection dénombrable,

$$\mathbb{P}\left(\frac{S_n}{n} \to p\right) = \mathbb{P}\left(\forall k, \exists N, \forall n \geq N, \left|\frac{S_n}{n} - p\right| \leq \frac{1}{k}\right) = 1.$$

$\square$

démo (Glivenko-Cantelli) :

F_n et F sont continues à droite, donc

$$\|F_n - F\|_\infty = \sup_{x \in \mathbb{Q}} |F_n(x) - F(x)|$$

donc c'est bien une variable aléatoire.

Soit $t \in \mathbb{R}$. Alors $(\mathbb{1}_{X_j \leq t})_{j \in \mathbb{N}}$ est une suite de v.a de Bernoulli de paramètre $\mathbb{E}[\mathbb{1}_{X_j \leq t}] = \mathbb{P}(X_1 \leq t) = F(t)$.

Donc par la loi forte des grands nombres,

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{X_i \leq t} \xrightarrow[n \to +\infty]{ps} F(t).$$

Soit $k \in \mathbb{N} \setminus \{0\}$.

On définit $x_{j,k} = \inf \left\{ x \in \mathbb{R} \mid F(x) = \frac{j}{k} \right\}$ pour tout $1 \leq j \leq k-1$

$x_{0,k} = -\infty$; $x_{k,k} = +\infty$.

Soit $\Delta_n(k) = \max_{1 \leq j \leq k-1} \left\{ |F_n(x_{j,k}) - F(x_{j,k})| \right\}$

Comme $\forall j \in \{1, \dots, k-1\}$, $F_n(x_{j,k}) \xrightarrow[n \to +\infty]{ps} F(x_{j,k})$,

on a $\Delta_n(k) \xrightarrow[n \to +\infty]{ps} 0$.

Soit $x \in \mathbb{R}$. S'il existe $j \in \{1, \dots, k-1\}$ tel que $x \in [x_{j,k}, x_{j+1,k}]$,

alors $F_n(x_{j,k}) \leq F_n(x) \leq F_n(x_{j+1,k}) \quad \forall n$

$F(x_{j,k}) \leq F(x) \leq F(x_{j+1,k})$ par croissance de F_n et F

et $F(x_{j+1,k}) - F(x_{j,k}) = \frac{1}{k}$

donc $F_n(x) - F(x) \leq F_n(x_{j+1,k}) - F(x_{j,k})$

$\leq F_n(x_{j+1,k}) - F(x_{j+1,k}) + F(x_{j+1,k}) - F(x_{j,k}) \leq \Delta_n(k) + \frac{1}{k}.$

et $\quad F(x) - F_n(x) \leq F(x_{j+1,k}) - F(x_{j,k}) + F(x_{j,k}) - F_n(x_{j,k})$
$\leq \frac{1}{k} + \Delta_n(k).$

De même, si $x < x_{1,k}$ ou si $x > x_{k-1,k}$, on trouve la même chose en considérant $x_{0,k} = -\infty$, $x_{k,k} = +\infty$.

Donc $\quad \|F_n - F\|_\infty \leq \Delta_n(k) + \frac{1}{k}$.

$\forall k$, ps, $\limsup_n \|F_n - F\|_\infty \leq \frac{1}{k}$.

par dénombrabilité

Donc ps, $\limsup_{n\to+\infty} \|F_n - F\|_\infty \leq 0$

donc $\|F_n - F\|_\infty \xrightarrow[n\to+\infty]{} 0$.

$\square$

<u>Cas où F n'est pas continue</u> :

· Les choses à rajouter dans la preuve :

Soit $t \in \mathbb{R}$.

$F(t-) := \lim_{\substack{u\to t\\ u<t}} F(u) = \mathbb{P}(X_1 < t)$.

$F_n(t-) := \lim_{\substack{u\to t\\ u<t}} F_n(u) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{X_i < t}$.

Donc par la loi des grands nombres,

$F_n(t-) \longrightarrow F(t-)$.

On définit <u>désormais</u> $x_{j,k} = \inf\{x \in \mathbb{R} \mid F(x) \geq \frac{j}{k}\}$

et $\quad \Delta_n^1(k) = \max_{1\leq j\leq k-1} |F_n(x_{j,k}) - F(x_{j,k})|$

$\Delta_n^2(k) = \max_{1\leq j\leq k-1} |F_n(x_{j,k}-) - F(x_{j,k}-)|$.

Comme avant ; $\Delta_n^1(k) \to 0$

$\Delta_n^2(k) \to 0$.

S'il existe $j \in \{1, .., k-1\}$ tel que $x \in]x_{j,k}, x_{j+1,k}[$

$$F(x_{j,k}) \leq F(x) \leq F(x_{j+1,k}-)$$

$$F_n(x_{j,k}) \leq F_n(x) \leq F_n(x_{j+1,k}-)$$

$$\text{et} \quad 0 \leq F(x_{j+1,k}-) - F(x_{j,k}) \leq \frac{1}{k}.$$

et on fait de même qu'avant :

$$\|F_n - F\|_\infty \leq \max(\Delta_n^1(k), \Delta_n^2(k)) + \frac{1}{k}$$

...

ps $\|F_n - F\|_\infty \xrightarrow[n \to +\infty]{} 0$.