THÉORÈME DE GLIVENKO CANTELLI

Nourdin p109 sauf étape 1
Gourdon (pr Dini)

Prérequis:
- Thm de Dini, liminf limsup
- fct° de répartition et ses propriétés.
- LGN faible, cv p-s.
- ~~une fct croissante admet en tt point une limite à g et à d~~ -

16min en allant très vite.
16 min 39 normal
14 min

On rappelle le 2° thm de Dini:

⚠ ne sont pas supposées continues contrairement au Gourdon

Thm: Soient $a, b \in \mathbb{R}$ avec $a<b$ et (f_n) une suite de fonctions croissantes de $[a,b]$ dans $\mathbb{R}$ qui converge simplement vers f continue. Alors (f_n) converge uniformément vers f sur $[a,b]$

Dém: (Ne pas faire le jour J) Gourdon analyse.

Soit $\varepsilon>0$. f est continue sur $[a,b]$ donc d'après le thm de Heine, elle y est unif C^0:
$\exists \eta>0 \quad \forall x,y\in[a,b] \quad |x-y|<\eta \Rightarrow |f(x)-f(y)|<\varepsilon.$ $(*)$

On définit une subdivision $a=x_0<x_1<\dots<x_p=b$ de $[a,b]$ de pas $<\eta$.

- $f_n(x_i)\to f(x_i)$: $\exists N_i \ \forall n\geq N_i \quad |f_n(x_i)-f(x_i)|\leq\varepsilon.$ $\qquad i\in[\![0,p]\!]$. $(**)$

On pose $N=\max N_i$: $\forall n\geq N \ \forall 1\leq i\leq p \quad |f_n(x_i)-f(x_i)|\leq\varepsilon.$

- Soit $x\in[a,b]$. soit $n\geq N$. il existe $i\in[\![1,p]\!]$ tq $x\in[x_i,x_{i+1}]$

alors $|f(x)-f(x_i)|<\varepsilon$ par $(*)$

$$\text{donc } |f(x)-f_n(x)| \leq \underbrace{|f(x)-f(x_i)|}_{*} + \underbrace{|f(x_i)-f_n(x_i)|}_{**} + |f_n(x_i)-f_n(x)|$$
$$\leq 2\varepsilon + (f_n(x)-f_n(x_i)) \quad \text{car } f_n \text{ est croissante}$$
$$\leq 2\varepsilon + f_n(x_{i+1})-f_n(x_i) \quad \text{car } f_n \text{ est croissante}$$
$$\leq 2\varepsilon + \underbrace{|f_n(x_{i+1})-f(x_{i+1})|}_{**} + \underbrace{|f(x_{i+1})-f(x_i)|}_{*} + \underbrace{|f(x_i)-f_n(x_i)|}_{**}$$
$$\leq 5\varepsilon.$$

On a donc la convergence uniforme

Soit $(\Omega,\mathcal{F},\mathbb{P})$ un espace probabilisé.

Thm de Glivenko-Cantelli: Soit (X_n) une suite de v.a. iid de fonction de répartition commune F. Pour tout $t\in\mathbb{R}$ et $n\in\mathbb{N}$ on pose

$$F_n(t)=\frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{]-\infty,t]}(X_k) \quad \text{(C'est une v.a. appelée fonction de répartition empirique)}$$

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

i.e. $\exists A\in\mathcal{F} \ \mathbb{P}(A)=1 \ \forall\omega\in A \ \sup_{t\in\mathbb{R}}|F_n(t)(\omega)-F(t)| \xrightarrow[n\to+\infty]{} 0.$

2 min

Dém: Étape 1: $\sup_{t\in\mathbb{R}}|F_n(t)-F(t)|$ est bien défini et est une v.a.

- Soit $\omega\in\Omega$. On pose $T=\sup_{t\in\mathbb{R}}|F_n(t)(\omega)-F(t)|$. Pour tout $x\in\mathbb{R}$, $F_n(x)(\omega)$ et $F(x)$ sont des réels de $[0,1]$ donc T est fini. Donc $\sup_{t\in\mathbb{R}}|F_n(t)-F(t)|$ est bien défini.

* Soit $\omega \in \Omega$ $\sup\limits_{t\in\mathbb{Q}} |F_n(t)(\omega) - F(t)| \leq \sup\limits_{t\in\mathbb{R}} |F_n(t)(\omega) - F(t)|$ car $\mathbb{Q} \subset \mathbb{R}$

* On va montrer l'autre inégalité : Soit $\omega \in \Omega$, Soit $\varepsilon > 0$, par définition du suprémum, il existe

$x_\varepsilon \in \mathbb{R}$ tq $T - \varepsilon < |F_n(x_\varepsilon)(\omega) - F(x_\varepsilon)| \leq T$ (1)

voir fin

Comme $x \mapsto F_n(x)(\omega)$ et F sont continues à droite en tout point, en particulier elles le sont en x_ε et la valeur absolue de leur différence aussi :

$$\exists \delta > 0 \ \forall t \in]x_\varepsilon, x_\varepsilon + \delta[\quad \big| |F_n(t)(\omega) - F(t)| - |F_n(x_\varepsilon)(\omega) - F(x_\varepsilon)| \big| \leq \varepsilon$$

d'où $-\varepsilon \leq |F_n(t)(\omega) - F(t)| - |F_n(x_\varepsilon)(\omega) - F(x_\varepsilon)| \leq \varepsilon$

d'où en particulier $|F_n(t)(\omega) - F(t)| \geq |F_n(x_\varepsilon)(\omega) - F(x_\varepsilon)| - \varepsilon$

Ceci est vrai en particulier pour tout $t \in]x_\varepsilon, x_\varepsilon + \delta[\cap \mathbb{Q} \neq \emptyset$ par densité de $\mathbb{Q}$ ds $\mathbb{R}$

d'où $\sup\limits_{t\in\mathbb{Q}} |F_n(t)(\omega) - F(t)| \geq |F_n(x_\varepsilon)(\omega) - F(x_\varepsilon)| - \varepsilon$

$\geq T - 2\varepsilon$ par (1)

Ceci est vrai pour tout $\varepsilon > 0$ donc $\sup\limits_{t\in\mathbb{Q}} |F_n(t)(\omega) - F(t)| \geq T$

donc $\forall \omega \in \Omega$ $\sup\limits_{t\in\mathbb{Q}} |F_n(t)(\omega) - F(t)| = \sup\limits_{t\in\mathbb{R}} |F_n(t)(\omega) - F(t)|$

* Or pour $t \in \mathbb{Q}$, $(x_1, .., x_n) \mapsto \left| \frac{1}{n} \sum\limits_{k=1}^{n} 1_{x_i \leq t} - F(t) \right|$ est mesurable donc

$\psi : (x_1, .., x_n) \mapsto \sup\limits_{t\in\mathbb{Q}} \left| \frac{1}{n} \sum\limits_{k=1}^{n} 1_{x_i \leq t} - F(t) \right|$ est mesurable comme sup dénombrable de fonctions mesurables. Donc $\sup\limits_{t\in\mathbb{Q}} |F_n(t) - F(t)| = \psi(X_1, .., X_n)$ est une v.a.

Etape 2 : Problèmes pour appliquer Dini

D'après la LGN forte, $\forall t \in \mathbb{R}$ $\exists A_t \in \mathcal{F}$ $\mathbb{P}(A_t) = 1$ $\forall \omega \in A_t$ $F_n(t)(\omega) \xrightarrow[n\to+\infty]{} F(t)$

De plus pour tout $\omega \in \Omega$, $t \mapsto F_n(t)(\omega)$ est croissante, $n \in \mathbb{N}$.

On a plusieurs problèmes :

* F n'est pas a priori continue
* la convergence n'a pas a priori lieu sur un segment $t \in \mathbb{R}$
* On doit trouver A de mesure pleine tq $\forall \omega \in A$ $\underbrace{\forall t \in \mathbb{R} \ F_n(t)(\omega) \xrightarrow[n\to+\infty]{} F(t)}_{\text{CS du thm de Dini}}$.

Etape 3 : On se ramène à des lois uniforme.

On introduit l'inverse généralisé de F par $\forall u \in [0,1]$ $F^*(u) := \inf\{x \in \mathbb{R} \ F(x) \geq u\}$

$F^*(u)$ est ps à valeurs dans $\mathbb{R}$ car F admet 0 comme limite en $-\infty$ et 1 en $+\infty$.

alors on a $\forall x \in \mathbb{R}$ $\forall u \in [0,1]$ $F^*(u) \leq x \iff u \leq F(x)$. (2) (BL p50)

* Si $F^*(u) \leq x$ alors $\forall s > x$ $\exists t < s$ tq $F(t) \geq u$ (s n'est pas la borne inf) donc par croissance de F, $F(s) \geq F(t) \geq u$ donc $\forall s > x$ $F(s) \geq u$, donc par continuité à droite de F, $F(x) \geq u$.
* Reciproqmt si $u \leq F(x)$ alors $x \in \{y \in \mathbb{R}, F(y) \geq u\}$ donc $x \geq F^*(u) = \inf\{\dots\}$

* F^* est croissante donc mesurable donc $F^*(U_n)$ est une v.a.

Soit (U_n) une suite de v.a. iid de loi $\mathcal{U}([0,1])$ alors $F^*(U_n)$ et X_n ont m loi pr tt $n \in \mathbb{N}$

$\mathbb{P}(F^*(U_n) \leq t) = \mathbb{P}(U_n \leq F(t)) = F(t)$ car U_n est uniforme

donc $\left| \frac{1}{n} \sum\limits_{k=1}^{n} 1_{X_n \leq t} - F(t) \right| \sim \left| \frac{1}{n} \sum\limits_{k=1}^{n} 1_{F^*(U_n) \leq t} - F(t) \right|$

donc $\sup_{t\in\mathbb{Q}} |F_n(t)-F(t)| = \varphi(X_1,\dots,X_n) \sim \varphi(F^*(U_1),\dots,F^*(U_n)) = \sup_{t\in\mathbb{Q}} \left|\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{F^*(U_k)\le t} - F(t)\right|$

D'après l'étape 1 le sup sur $\mathbb{Q}$ ou sur $\mathbb{R}$ c'est pareil.

$$\sup_{t\in\mathbb{R}} \left|\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{F^*(U_k)\le t} - F(t)\right| \overset{(2)}{=} \sup_{t\in\mathbb{R}} \left|\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{U_k\le F(t)} - F(t)\right|$$

$$\overset{s=F(t)}{=} \sup_{s\in F(\mathbb{R})} \left|\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{U_k\le s} - s\right| \le \sup_{s\in[0,1]} \left|\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{U_k\le s} - s\right| \quad \text{car } F(\mathbb{R})\subset[0,1].$$

On pose pour $s\in[0,1]$ $G_n(s) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{U_k\le s}$ et $G(s)=s$.

G est la fonction de répartition de $\mathcal{U}([0,1])$. G_n est la fonction de répartition empirique.

On est donc ramené à étudier des lois $\mathcal{U}([0,1])$ du fait de l'inégalité obtenue.
↳ G est continue et on est sur un segment.

Etape 4 : Etude pour $\mathcal{U}([0,1])$.

• D'après la LGN faible, $\forall s\in[0,1]$ $\exists A_s\in\mathcal{F}$ $\mathbb{P}(A_s)=1$ $\forall\omega\in A_s$ $G_n(s)(\omega) \xrightarrow[n\to+\infty]{} G(s)$

On pose $A = \bigcap_{s\in[0,1]\cap\mathbb{Q}} A_s$ c'est une intersection dénombrable d'ensembles de mesure pleine
donc $\mathbb{P}(A)=1$.
donc $\forall\omega\in A$ $\forall s\in[0,1]\cap\mathbb{Q}$ $G_n(s)(\omega) \xrightarrow[n\to+\infty]{} G(s)$ (3)

* On voudrait que ce soit vrai pour $s\in[0,1]$: Soit $\omega\in A$. Soit $s\in[0,1]$. Soit $\varepsilon>0$.

Comme $\mathbb{Q}\cap[0,1]$ est dense dans $[0,1]$, il existe $p,q\in\mathbb{Q}\cap[0,1]$ tq

$$s-\varepsilon \le p \le s \le q \le s+\varepsilon$$

Or $s\mapsto G_n(s)(\omega)$ est croissante donc $G_n(p)(\omega) \le G_n(s)(\omega) \le G_n(q)(\omega)$.
d'où
$$s-\varepsilon \le p \overset{(3)}{\underset{p\in\mathbb{Q}\cap[0,1]}{=}} \lim G_n(p)(\omega) \le \underline{\lim}\, G_n(s)(\omega) \le \overline{\lim}\, G_n(s)(\omega) \le \lim G_n(p)(\omega) \overset{(3)\ q\in\mathbb{Q}\cap[0,1]}{=} q \le s+\varepsilon$$

C'est vrai pour tout $\varepsilon>0$. donc $\underline{\lim}\, G_n(s)(\omega) = \overline{\lim}\, G_n(s)(\omega) = s = G(s)$

donc $\forall\omega\in A$ $\underbrace{\forall s\in[0,1]\ G_n(s)(\omega) \xrightarrow[n\to+\infty]{} G(s)}_{\text{CS dans le thm de Dini}}$

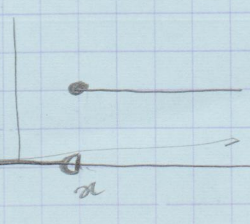
donc d'après le thm de Dini, $\forall\omega\in A$ $\sup_{s\in[0,1]} |G_n(s)(\omega)-G(s)| \xrightarrow[n\to+\infty]{} 0$

Donc d'après l'étape 3), $\forall\omega\in A$ $\sup_{t\in\mathbb{R}} |F_n(t)(\omega)-F(t)| \xrightarrow[n\to+\infty]{} 0$.

Soit $\omega\in\Omega$. $x\mapsto F_n(x)(\omega)$ est continue à droite : ~~Si $t_n\to t_0$ alors~~
l'indicatrice $\mathbb{1}_{X\le t}$ est continue à droite.

Si $t_n\to t$ en décroissant $\mathbb{1}_{X\le t} = \mathbb{1}_{\forall n\, X\le t_n} = \lim_{n\to+\infty} \mathbb{1}_{X\le t_n}$

ou $\mathbb{1}_{x\le t} = \mathbb{1}_{[x,+\infty[}(t)$



c'est bien continue à droite!

donc $\sup_{t\in\mathbb{Q}} |F_n(t) - F(t)| = \psi(X_1,..,X_n) \sim \psi(F^*(U_1),..,F^*(U_n)) = \sup_{t\in\mathbb{Q}} \left| \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{F^*(U_k)\leq t} - F(t)\right|$

D'après l'étape 1 le sup sur $\mathbb{Q}$ ou sur $\mathbb{R}$ c'est pareil.

$$\sup_{t\in\mathbb{R}} \left|\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{F^*(U_k)\leq t} - F(t)\right| \overset{(2)}{=} \sup_{t\in\mathbb{R}} \left|\frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{U_k\leq F(t)} - F(t)\right|$$

$s = F(t)$

$$= \sup_{s\in F(\mathbb{R})} \left|\frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{U_k\leq s} - s\right| \leq \sup_{s\in[0,1]} \left|\frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{U_k\leq s} - s\right| \quad \text{car } F(\mathbb{R})\subset[0,1].$$

On pose pour $s\in[0,1]$ $G_n(s) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{U_k\leq s}$ et $G(s) = s$.

G est la fonction de répartition de $\mathcal{U}([0,1])$. G_n est la fonction de répartition empirique.

On est donc ramené à étudier des lois $\mathcal{U}([0,1])$ du fait de l'inégalité obtenue.
↳ G est continue et on est sur un segment.

Etape 4 : Etude pour $\mathcal{U}([0,1])$.

* D'après la LGN faible, $\forall s\in[0,1]$ $\exists A_s\in\mathcal{F}$ $\mathbb{P}(A_s) = 1$ $\forall \omega\in A_s$ $G_n(s)(\omega) \xrightarrow[n\to+\infty]{} G(s)$

On pose $A = \bigcap_{s\in[0,1]\cap\mathbb{Q}} A_s$ c'est une intersection dénombrable d'ensembles de mesure pleine

donc $\mathbb{P}(A) = 1$.

donc $\forall\omega\in A$ $\forall s\in[0,1]\cap\mathbb{Q}$ $G_n(s)(\omega) \xrightarrow[n\to+\infty]{} G(s)$ (3)

* On voudrait que ce soit vrai pour $s\in[0,1]$: Soit $\omega\in A$. Soit $s\in[0,1]$. Soit $\varepsilon>0$.

Comme $\mathbb{Q}\cap[0,1]$ est dense dans $[0,1]$, il existe $p, q\in\mathbb{Q}\cap[0,1]$ tq

$s-\varepsilon\leq p\leq s\leq q\leq s+\varepsilon$

Or $s\mapsto G_n(s)(\omega)$ est croissante donc $G_n(p)(\omega)\leq G_n(s)(\omega)\leq G_n(q)(\omega)$.

d'où

$$s-\varepsilon\leq p \overset{(3)}{=} \lim G_n(p)(\omega) \leq \underline{\lim}\, G_n(s)(\omega) \leq \overline{\lim}\, G_n(s)(\omega) \leq \lim G_n(q)(\omega) \overset{(3)}{=} q \leq s+\varepsilon$$

$p\in\mathbb{Q}\cap[0,1]$; $q\in\mathbb{Q}\cap[0,1]$.

C'est vrai pour tout $\varepsilon>0$. donc $\underline{\lim}\, G_n(s)(\omega) = \overline{\lim}\, G_n(s)(\omega) = s = G(s)$

donc $\underbrace{\forall\omega\in A\ \forall s\in[0,1]\ G_n(s)(\omega) \xrightarrow[n\to+\infty]{} G(s)}_{\text{CS dans le thm de Dini}}$

Donc d'après le thm de Dini, $\forall\omega\in A$ $\sup_{s\in[0,1]} |G_n(s)(\omega) - G(s)| \xrightarrow[n\to+\infty]{} 0$

Donc d'après l'étape 3), $\forall\omega\in A$

Soit $\omega\in\Omega$. $x\mapsto F_n(x)(\omega)$ est contin
the indicatrice $\mathbb{1}_x$
Si $t_n\to t$ on déc
ou $\mathbb{1}_{x<t} = \mathbb{1}_{[x,}$

* Pour tout n, si $(X_k)_{1\leq k\leq n}$ est un échantillon de taille n, la fonction F_n est appelée la fonction de répartition empirique de l'échantillon $(X_k)_{1\leq k\leq n}$.
* Le théorème de Glivenko-Cantelli est parfois appelé théorème fondamental de la statistique, car il exprime en quoi une loi de probabilité peut être révélée par la connaissance d'un échantillon suffisamment grand de ladite loi de probabilité.
* Le théorème de Glivenko-Cantelli est parfois considéré comme une généralisation du deuxième théorème de Dini, car il ne suppose pas en particulier la continuité de F, ni le fait d'être défini sur un compact.
* Le théorème de Kolmogorov-Smirnov précise l'énoncé du théorème de Glivenko-Cantelli dans le cas où F est continue : il donne une estimation de la vitesse de convergence en $\frac{1}{\sqrt{n}}$:

Théorème 4 (*de Kolmogorov-Smirnov*)

Soit $(X_1, ..., X_n)$ un échantillon de loi μ sur $\mathbb{R}$ de fonction de répartition F. Si F est continue, alors

$$K_n = \sqrt{n}\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \longrightarrow^{\mathcal{L}} \mu_{KS}$$

où μ_{KS} est une loi universelle ne dépendant pas de F. Elle est portée par $\mathbb{R}^+$ et a pour fonction de répartition pour $t\geq 0$:

$$F_{KS}(t) = 1 + 2\sum_{k=1}^{+\infty}(-1)^k e^{-2k^2t^2}$$

* Le théorème de Kolmogorov-Smirnov est à la base du test d'adéquation à une loi de