Leçons: 241, 261, 262, 266

Théorème de Glivenko - Cantelli

Réf:
- Probabilité 2, J.-Y. Ouvrard
- Agrégation de mathématiques, Y. Nourdin

Théorème (Dini)

Soient a, b dans $\mathbb{R}$ tels que $a < b$.
Soit (f_n) une suite de fonctions croissantes de $[a,b]$ dans $\mathbb{R}$ convergeant simplement vers une fonction continue f.
Alors (f_n) converge uniformément vers f.

Preuve:

Soit $\varepsilon \in \mathbb{R}_+^*$. Comme f est continue sur le compact $[a,b]$ le théorème de Heine assure que f est uniformément continue sur $[a,b]$.

Soit $\delta \in \mathbb{R}_+^*$ tel que

$$\forall (x,y) \in [a,b]^2 \quad |x-y| < \delta \Rightarrow |f(x)-f(y)| < \varepsilon$$

Soit $(a_i)_{1\leq i\leq k}$ une subdivision de $[a,b]$ de pas strictement inférieur à δ.

Soit $n_0 \in \mathbb{N}$ tel que pour tout $n \in \mathbb{N}$

$$n \geq n_0 \Rightarrow \forall i \in [\![1,k]\!] \quad |f_n(a_i) - f(a_i)| < \varepsilon$$

Soit $x \in [a,b]$. Il existe $i \in [\![1,k]\!]$ tel que $x \in [a_i, a_{i+1}]$.
Alors on a, pour tout $n \in \mathbb{N}$ tel que $n \geq n_0$

$$\begin{aligned} |f(x) - f_n(x)| &\leq |f(x) - f(a_i)| + |f(a_i) - f_n(a_i)| + |f_n(a_i) - f_n(x)| \\ &\leq \varepsilon + \varepsilon + f_n(x) - f_n(a_i) \\ &\leq 2\varepsilon + f_n(a_{i+1}) - f_n(a_i) \leq 5\varepsilon. \end{aligned}$$

Cela conclut.

Théorème (Glivenko-Cantelli)

Soient (X_n) des vaiid de fonction de répartition F.
Pour $t \in \mathbb{R}$ et $n \in \mathbb{N}^*$, on note $F_n(t)$ la somme

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{]-\infty,t]}(X_k)$$

Alors on a presque sûrement

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

Preuve :

Etape 1 : Pour tout $n \in \mathbb{N}^*$, $\sup_{t\in\mathbb{R}} |F_n(t)-F(t)|$ est une variable aléatoire.

Soit $n \in \mathbb{N}^*$. Soit $\omega \in \Omega$ tel que pour tout $i \in [\![1,n]\!]$ $X_i(\omega)$ soit défini.
Notons τ la quantité $\sup_{t\in\mathbb{R}} |F_n(t)(\omega) - F(t)|$. Comme les fonctions F_n et F sont bornées par 1, τ est un réel (positif).

- Montrons que $\tau = \sup_{t\in\mathbb{Q}} |F_n(t) - F(t)|$ et cela conclura le premier point. (supremum dénombrable de fonctions mesurables)

  ↳ Soit $\varepsilon \in \mathbb{R}_+^*$. Par définition / propriété du supremum, il existe $x_\varepsilon \in \mathbb{R}$ tel que $\tau - \varepsilon < |F_n(x_\varepsilon)(\omega) - F(x_\varepsilon)|$
  Comme les applications $t \longmapsto F_n(t)(\omega)$ et F sont continues à droite, l'application $g : t \longmapsto |F_n(t)(\omega) - F(t)|$ l'est aussi
  Soit $\delta \in \mathbb{R}_+^*$ tel que $\forall t \in]x_\varepsilon, x_\varepsilon+\delta[$ $|g(t) - g(x_\varepsilon)| < \varepsilon$.
  On a alors

$$\forall t \in]x_\varepsilon, x_\varepsilon+\delta[\qquad |F_n(t)(\omega) - F(t)| \geqslant |F_n(x_\varepsilon)(\omega) - F(x_\varepsilon)| - \varepsilon$$

  Par densité de $\mathbb{Q}$ dans $\mathbb{R}$, on en déduit
  $\sup_{t\in\mathbb{Q}} |F_n(t)(\omega) - F(t)| \geqslant |F_n(x_\varepsilon)(\omega) - F(x_\varepsilon)| - \varepsilon \geqslant \tau - 2\varepsilon$.

D'où l'égalité annoncée, et la fin de l'étape 1.

Etape 2 : La loi des grands nombres assure qu'on a

$$\forall t \in \mathbb{R} \qquad F_n(t) \xrightarrow[n \to +\infty]{p.s.} F(t)$$

et de plus, pour presque tout $\omega \in \Omega$, les applications

$$\left(t \longmapsto F_n(t) \right)_{n \in \mathbb{N}^*}$$

sont croissantes.

On voudrait appliquer le théorème de Dini énoncé précédem--ment, mais il y a trois "problèmes" :

* la régularité de F ;
* la convergence ne se fait pas sur un segment ;
* Il faut " p.s. $\forall t \in \mathbb{R} \quad \lim\limits_{n \to +\infty} F_n(t) = F(t)$ ".

Etape 3 : On se ramène à des lois uniformes.

Rappel : Inverse généralisé

↳ on note $\forall u \in [0,1] \quad F^*(u) = \inf \left\{ x \in \mathbb{R} \,/\, F(x) \geqslant u \right\}$

L'application F^* vérifie

$$\forall x \in \mathbb{R} \; \forall u \in [0,1] \quad F^*(u) \leqslant x \iff u \leqslant F(x)$$

et si $U \sim \mathcal{U}[0,1]$ alors $F^*(U)$ a pour fonction de réparti--tion F.

Alors $\Psi(X_1, \dots, X_n) := \sup\limits_{t \in \mathbb{Q}} |F_n(t) - F(t)|$ suit la même loi que $\Psi(F^*(U_1), \dots, F^*(U_n))$ où (U_n) sont des v.a.i.i.d. de loi $\mathcal{U}[0,1]$.

On a :

$$\Psi(F^*(U_1), \dots, F^*(U_n)) \overset{\text{étape 1}}{=} \sup_{t \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{F^*(U_i) \leqslant t\}} - F(t) \right|$$

$$" = \sup_{s\in F(\mathbb{R})} \left| \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i \leq s\}} - s \right|$$

$$" \leq \sup_{s\in [0,1]} \left| \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i \leq s\}} - s \right|$$

Notons $G_n(s) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i \leq s\}}$ et $G(s) := s$.

Alors G est la fonction de répartition de la loi $\mathcal{U}[0,1]$ et G_n est sa fonction de répartition empirique

Par l'inégalité précédente nous ramène donc à montrer le théorème dans le cas $\mathcal{U}([0,1])$ (qui élimine 2 problèmes)

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

Par la loi des grands nombres, pour tout $s \in [0,1]$, il existe un ensemble mesurable A_s tel que $\mathbb{P}(A_s) = 1$ et tel que

$$\forall \omega \in A_s \quad \lim_{n\to+\infty} G_n(s)(\omega) = G(s)$$

Notons A l'ensemble $\bigcap_{t\in[0,1]\cap\mathbb{Q}} A_t$, qui est de proba 1.

Par caractère dénombrable, on a

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

Montrons que cela reste vrai pour $s \in [0,1]$.

↳ Soit $s \in [0,1]$ et $\varepsilon \in \mathbb{R}_+^*$. Par densité de $\mathbb{Q}\cap[0,1]$ dans $[0,1]$ il existe p, q dans $[0,1]\cap\mathbb{Q}$ tels que $s-\varepsilon \leq p \leq s \leq q \leq s+\varepsilon$.

Par croissance de $s \longmapsto \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i(\omega) \leq s\}}$, on obtient

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i(\omega) \leq p\}} \leq \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i(\omega) \leq s\}} \leq \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i(\omega) \leq q\}}$$

pour tout $\omega \in A$.

On en déduit

$$p = \liminf \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i(\omega)\leq p\}} \leq \liminf \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i(\omega)\leq s\}}$$

$$q = \limsup \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i(\omega)\leq q\}} \geq \limsup \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i(\omega)\leq s\}}$$

et donc $s-\varepsilon \leq \liminf \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i(\omega)\leq s\}} \leq \limsup \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i(\omega)\leq s\}} \leq s+\varepsilon.$

(pour tout $\omega \in A$)

D'où

$$\forall \omega \in A \quad \forall s \in [0,1] \qquad \lim_{n\to+\infty} G_n(s)(\omega) = s$$

Le théorème de Dini énoncé conclut. ■

<u>Rmq</u>:

(1) on utilise le fait que si (X_n), (Y_n) sont deux suites de v.a. telles que

sur le même espace de proba

$$\forall n \in \mathbb{N} \quad X_n \sim Y_n$$

et $X_n \xrightarrow[n\to+\infty]{p.s} 0$, alors $Y_n \xrightarrow[n\to+\infty]{p.s} 0$.

le 0 est important !

(2) On pourrait montrer que l'on a

$$\| F_n - F \|_\infty = \max\left(\max_{1\leq k\leq n} \left| \frac{k}{n} - F(X_{(k)})\right|, \max_{1\leq k\leq n} \left| \frac{k-1}{n} - F(X_{(k)})\right| \right)$$

(3) 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

Soit (X_n) une suite de v.a.i.i.d de fonction de répartition F continue.

Alors on a

$$\sqrt{n}\, \| F_n - F \|_\infty \xrightarrow[n \to +\infty]{\text{Loi}} \mu_{KS}$$

où μ_{KS} est la loi de fonction de répartition

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

ne dépend pas de F

pour $t \in \mathbb{R}_+^*$.