P. Maurer

ENS Rennes

Recasages : *229, 261, 262*

Référence : Gourdon, Analyse pour le théorème de Dini.

Le reste était inspiré d'un doc sur internet (auteur à retrouver).

Théorème de Glivenko-Cantelli

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

Lemme 1. *(Second théorème de Dini).*

Soit $[a,b]$ un intervalle fermé de $\mathbb{R}$, $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions $f_n\colon [a,b]\to\mathbb{R}$ croissantes. On suppose que $(f_n)_{n\in\mathbb{N}}$ converge simplement vers une fonction $f\colon [a,b]\to\mathbb{R}$ continue. Alors $(f_n)_{n\in\mathbb{N}}$ converge uniformément vers f.

Démonstration. Soit $\varepsilon>0$. Comme f est continue sur le compact $[a,b]$, elle y est uniformément continue par théorème de Heine. Il existe donc $\delta>0$ tel que pour tout $x,y\in[a,b]$, si $|x-y|\leq\delta$, alors $|f(x)-f(y)|\leq\varepsilon$.

On se donne une subdivision $(x_i)_{1\leq i\leq m}$ avec $a=x_1<\cdots<x_m=b$ de pas $\eta=\delta/2$, et $x\in[a,b]$. Il existe donc $i\in[\![1,m-1]\!]$ tel que $x_i\leq x\leq x_{i+1}$. Pour $n\in\mathbb{N}$, une inégalité triangulaire donne

$$|f_n(x)-f(x)| \;\leq\; |f_n(x)-f_n(x_i)|+|f_n(x_i)-f(x_i)|+|f(x_i)-f(x)|.$$

En notant $N\in\mathbb{N}$ un entier tel que $|f_n(x_i)-f(x_i)|\leq\varepsilon$ dès que $n\geq N$, on a donc

$$\begin{aligned}\forall n\geq N\quad |f_n(x)-f(x)| &\leq f_n(x)-f_n(x_i)+2\varepsilon\\ &\leq f_n(x_{i+1})-f_n(x_i)+2\varepsilon\\ &= f_n(x_{i+1})-f(x_{i+1})+f(x_{i+1})-f(x_i)+f(x_i)-f_n(x_i)+2\varepsilon.\end{aligned}$$

Reste à choisir un $N_1\geq N$ tel que $|f_n(x_{i+1})-f(x_{i+1})|\leq\varepsilon$ quand $n\geq N_1$, et on obtient

$$\forall n\geq N_1\quad |f_n(x)-f(x)| \;\leq\; 5\varepsilon.$$

Donc pour $n\geq N_1$, on a $\sup\limits_{x\in[a,b]}|f_n(x)-f(x)|\leq 5\varepsilon,$ ce qui conclut la preuve. □

Théorème 2. *(Glivenko-Cantelli).*

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables réelles aléatoires indépendantes, de même loi. On note F la fonction de répartition commune des X_n et, pour $n\in\mathbb{N}^*$, F_n la fonction de répartition empirique de X_n, c'est-à-dire :

$$\forall x\in\mathbb{R}\quad\forall\omega\in\Omega\quad F_n(x)(\omega)\colon=\frac{1}{n}\sum_{k=1}^{n}\mathbf{1}_{\{X_k(\omega)\leq x\}}.$$

Alors, $\mathbb{P}$-presque sûrement, F_n converge uniformément vers F, c'est-à-dire :

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \underset{n\to+\infty}{\overset{\mathbb{P}\text{-}p.s}{\to}} 0.$$

Démonstration.

L'idée générale est d'utiliser le second théorème de Dini et la croissance des F_n pour conclure. Il y a trois problème à régler :

- Les F_n sont définies sur $\mathbb{R}$ et non pas sur un compact.
- La fonction F, a priori, n'est pas continue.
- Il faut intégrer les questions de théorie de la mesure pour avoir une convergence presque sûre.

Etape 1 : On va utiliser la fonction quantile pour résoudre les deux premiers problèmes.

Si X est une variable aléatoire réelle, on rappelle que sa fonction quantile est définie sur $[0,1]$ par $F_X^{\leftarrow}(u) = \inf\{x \in \mathbb{R} : F_X(x) \geq u\}$. On va démontrer que :

1. On a $F_X^{\leftarrow}(u) \leq x \Longleftrightarrow u \leq F_X(x)$.
2. Si $U \sim \mathcal{U}([0,1])$, alors la variable aléatoire $F_X^{\leftarrow}(U)$ suit la même loi que X.

1. $\boxed{\Longrightarrow}$ Si $F_X^{\leftarrow}(u) \leq x$, il existe $y \leq x$ tel que $F_X(y) \geq u$. Par croissance de F_X, On en déduit $F_X(x) \geq F_X(y) \geq u$.

$\boxed{\Longleftarrow}$ Si $u \leq F_X(x)$, alors $x \in \{x \in \mathbb{R} : F_X(x) \geq u\}$ donc $x \leq F_X^{\leftarrow}(u)$ par définition de la borne inférieure.

2. Soit $U \sim \mathcal{U}([0,1])$. D'après le point 1, pour tout $x \in \mathbb{R}$ on a $F_X^{\leftarrow}(U) \leq x \Longleftrightarrow U \leq F_X(x)$. On en déduit que $\mathbb{P}(F_X^{\leftarrow}(U) \leq x) = \mathbb{P}(U \leq F_X(x)) = F_X(x)$, donc $F_X^{\leftarrow}(U)$ et X ont la même loi.

Etape 2 : On se ramène à montrer le théorème sur des variables uniformes sur $[0,1]$.

Supposons le théorème prouvé pour une suite $(U_n)_{n\in\mathbb{N}}$ de variables i.i.d de loi $\mathcal{U}([0,1])$. On se donne F une fonction de répartition quelconque, et pour tout $n \in \mathbb{N}$, on pose $X_n = F^{\leftarrow}(U_n)$.

Alors les variables $(X_n)_{n\in\mathbb{N}}$ ont pour fonction de répartition commune F d'après ce qui précède, et de plus, on a, pour $n \geq 1$:

$$\begin{aligned}
\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| &= \sup_{x\in\mathbb{R}} \left| \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{F^{\leftarrow}(U_n) \leq x\}} - F(x) \right| \\
&= \sup_{x\in\mathbb{R}} \left| \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{U_n \leq F(x)\}} - F(x) \right| \\
&\leq \sup_{s\in[0,1]} \left| \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{U_n \leq s\}} - s \right|.
\end{aligned}$$

Comme $s \mapsto s$ est la fonction de répartition de la loi uniforme sur $[0,1]$, on en déduit alors que $\mathbb{P}$-presque sûrement, on a $\sup_{s\in[0,1]} \left| \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{U_n \leq s\}} - s \right| \underset{n\to+\infty}{\to} 0$, et donc $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \underset{n\to+\infty}{\to} 0$.

Etape 3 : On prouve le théorème pour une suite de variables uniformes sur $[0,1]$.

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires i.i.d de loi uniforme sur $[0,1]$. D'après la loi forte des grands nombres, pour tout $s\in[0,1]$, il existe $N_s\in\mathcal{F}$ tel que $\mathbb{P}(N_s)=0$ et pour tout $\omega\in X\setminus N_s$,

$$\begin{aligned} \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{U_n(\omega)\leq s\}} & \underset{n\to+\infty}{\to} \mathbb{P}(U_n(\omega)\leq s) \\ & = s. \end{aligned}$$

Cela est en particulier vrai pour $s\in[0,1]\cap\mathbb{Q}$, et on pose ainsi $\mathcal{N}=\bigcup_{s\in[0,1]\cap\mathbb{Q}} N_s$. Cette union étant dénombrable, $\mathcal{N}$ vérifie $\mathbb{P}(\mathcal{N})\leq \sum_{s\in[0,1]\cap\mathbb{Q}} \mathbb{P}(N_s)\leq 0$, donc $\mathbb{P}(\mathcal{N})=0$. Par ailleurs, pour tout $\omega\in X\setminus\mathcal{N}$, on a $\omega\in\bigcap_{s\in[0,1]\cap\mathbb{Q}} X\setminus N_s$ donc $\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{U_n(\omega)\leq s\}} \underset{n\to+\infty}{\to} s$.

Donnons nous un réel $s\in]0,1[$ et $\varepsilon>0$. Par densité de $]0,1[\cap\mathbb{Q}$ dans $]0,1[$, on peut trouver deux rationnels $p,q\in]0,1[\cap\mathbb{Q}$ tels que $s-\varepsilon\leq p\leq s\leq q\leq s+\varepsilon$.

Par croissance de $F_n=\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{U_n\leq s\}}$, on a $F_n(p)\leq F_n(s)\leq F_n(q)$. Aussi, pour $\omega\in X\setminus\mathcal{N}$, il vient

$$F_n(p)(\omega)\leq F_n(s)(\omega)\leq F_n(q)(\omega).$$

Donc $s-\varepsilon\leq p\leq \limsup_n F_n(s)(\omega)$ et $\liminf_n F_n(s)(\omega)\leq q\leq s+\varepsilon$. Ceci étant vrai pour tout $\varepsilon>0$, on en déduit que $\limsup_n F_n(s)(\omega)=s$, et ce pour tout $\omega\in X\setminus\mathcal{N}$.

Fixons un $\omega\in X\setminus\mathcal{N}$. La suite de fonctions $(F_n(\cdot)(\omega))_n$ converge simplement vers $s\mapsto s$, qui est continue sur le compact $[0,1]$. D'après le second théorème de Dini, on en déduit que $(F_n(\cdot)(\omega))_n$ converge uniformément vers $s\mapsto s$, donc $\sup_{s\in[0,1]} |F_n(s)(\omega)-s| \underset{n\to+\infty}{\to} 0$. Ceci étant vrai pour tout $\omega\in X\setminus\mathcal{N}$, on en déduit que $\sup_{s\in[0,1]} |F_n(s)-s|$ converge $\mathbb{P}$-presque sûrement vers zéro, ce qui termine la preuve du théorème. $\square$