# Développement: Théorème de Dini
# et application au théorème de Glivenko Cantelli

Adrien Fontaine

1er juillet 2014

*Référence :* Gourdon, Analyse
Nourdin, Agrégation de mathématiques, épreuve orale

**Théorème 1** (***Deuxième théorème de Dini***)

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

**Démonstration :** Soit $\varepsilon > 0$. La fonction $f$ étant continue sur le compact $[a, b]$, elle est uniformément continue sur $[a, b]$ par le théorème de Heine. Il existe donc $\eta > 0$ tel que

$$\forall x, y \in [a, b], |x - y| < \eta \Rightarrow |f(x) - f(y)| < \varepsilon$$

Soit alors $S = (a = a_0 < a_1 < ... < a_{n-1} < a_n = b)$ une subdivision de $[a, b]$ de pas $< \eta$. Puisque $(f_n)$ converge simplement vers $f$, et que les $a_i$ sont en nombre fini, il existe un entier $n_0 \in \mathbb{N}$, tel que $\forall n \geq n_0$, et $\forall 1 \leq i \leq n$, on ait $|f(a_i) - f_n(a_i)| < \varepsilon$.
Soit $x \in [a, b]$ et $n \geq n_0$. Il existe $i \in \{1, ..., n\}$ tel que $x \in [a_i, a_{i+1}]$. Alors,

$$|f(x) - f(a_i)| < \varepsilon$$

Donc,

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

La convergence est donc uniforme. ∎

**Application :**

**Théorème 2 (*de Glivenko-Cantelli*)**

*Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a.i.i.d. Notons $F$ la fonction de répartition commune des $X_n$ et posons si $t \in \mathbb{R}$ et $n \in \mathbb{N}^*$, la variable aléatoire :*

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

*Alors, presque surement, on a :*

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

**Démonstration : Idée de la preuve :** La loi forte des grands nombres nous donne facilement $F_n(t) \longrightarrow^{p.s}_{n\to+\infty} F(t)$ pour tout $t \in \mathbb{R}$. Par ailleurs, chaque fonction $F_n$ est croissante. On a donc envie d'appliquer le deuxième théorème de Dini. Plusieurs problèmes se posent alors :
— la fonction $F$ n'est pas, a priori, continue.
— la convergence n'a pas, a priori, lieu sur un segment.
— la loi forte des grands nombres nous donne

$$\forall t \in \mathbb{R}, \exists A_t/\mathbb{P}(A_t) = 1 \text{ et } \forall \omega \in A_t, F_n(t)(\omega) \underset{n\to+\infty}{\longrightarrow} F(t)(\omega)$$

Pour appliquer le théorème de Dini, il nous faut trouver un ensemble $A$ de mesure pleine uniforme pour tous les $t \in \mathbb{R}$. Autrement dit, on cherche $A$ tel que $\mathbb{P}(A) = 1$ et

$$\forall \omega \in A, \forall t \in \mathbb{R}, F_n(t)(\omega) \underset{n\to+\infty}{\longrightarrow} F(t)(\omega)$$

Les deux premiers points seront résolus grâce à l'inverse généralisé de la fonction de répartition, le troisième grâce à la séparabilité de $\mathbb{R}$.
Commençons par deux résultats fondamentaux sur l'inverse généralisé de la fonction $F$.

**Lemme 1**

*On introduit l'inverse généralisé de la fonction $F$ :*

$$\forall u \in [0,1], F^{\leftarrow}(u) := \inf\{x \in \mathbb{R}/F(x) \geq u\}$$

*Alors, on a l'équivalence suivante, pour tout $x \in \mathbb{R}$ et $u \in [0,1]$,*

$$F^{\leftarrow}(u) \leq x \Longleftrightarrow u \leq F(x)$$

**Démonstration :** Si $F^{\leftarrow}(u) \leq x$, alors il existe $y \leq x$ tel que $F(y) \geq u$. Mais par croissance de $F$, on a $F(y) \leq F(x)$, et donc

$$u \leq F(y) \leq F(x)$$

Réciproquement, si $u \leq F(x)$, alors $x \in \{y \in \mathbb{R}/F(y) \geq u\}$, et donc$x \geq \inf\{y \in \mathbb{R}/F(y) \geq u\}$, i.e

$$F^{\leftarrow}(u) \leq x$$

■

**Corollaire 1**

*Si $Y$ est une v.a réelle de fonction de répartition $F$ et $U \sim \mathcal{U}([0,1])$, alors $F^{\leftarrow}(U) \sim Y$.*

**Démonstration :** Il suffit d'écrire

$$\mathbb{P}(F^{\leftarrow}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x)$$

car la restriction à $[0,1]$ de la fonction de répartition d'une loi uniforme est l'identité. ■

On va maintenant pouvoir se ramener, dans la preuve du théorème de Glivenko-Cantelli, au cas de v.a qui suivent une loi uniforme sur $[0,1]$.
En effet, soit $(U_n)_{n\in\mathbb{N}}$ une suite de v.a réelles indépendantes de même loi $\mathcal{U}([0,1])$. Alors, on a :

$$\sup_{t\in\mathbb{R}} |\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k \leq t} - F(t)| \sim \sup_{t\in\mathbb{R}} |\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{F^{\leftarrow}(U_k) \leq t} - F(t)| \sim \sup_{t\in\mathbb{R}} |\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{U_k \leq F(t)} - F(t)|$$

Par ailleurs, si on pose $s = F(t)$, il vient :

$$\sup_{t\in\mathbb{R}} |\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{U_k \leq F(t)} - F(t)| = \sup_{s\in F(\mathbb{R})} |\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{U_k \leq s} - s| \leq \sup_{s\in[0,1]} |\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{U_k \leq s} - s|$$

Ainsi, il suffit de montrer que le théorème de Glivenko-Cantelli est vrai dans le cas particulier de v.a qui suivent des lois $\mathcal{U}([0,1])$, et où $s \in [0,1]$.
Grâce à loi forte des grands nombres, on sait que pour tout $s \in [0,1]$, il existe $A_s$ de mesure pleine tel que

$$\forall \omega \in A_s, \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{U_k(\omega) \leq s} \to s$$

On va maintenant essayer de trouver un ensemble $A$ de mesure pleine, qui soit uniforme pour tous les $s \in [0,1]$.
$\mathbb{Q}$ étant dénombrable, et une intersection dénombrable d'ensembles de mesure pleine étant de mesure pleine, on en déduit l'existence d'un ensemble $A$ de mesure pleine tel que

$$\forall \omega \in A, \forall s \in [0,1] \cap \mathbb{Q}, \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{U_k(\omega) \leq s} \to s$$

Montrons que la propriété ci-dessus est vraie pour tout $s \in [0,1]$ et pas seulement pour tout $s \in [0,1] \cap \mathbb{Q}$.
Fixons $s \in [0,1]$ et $\varepsilon > 0$. $\mathbb{Q} \cap [0,1]$ étant dense dans $[0,1]$, il existe $p$ et $q$ deux éléments de $\mathbb{Q} \cap [0,1]$ tels que $s - \varepsilon \leq p \leq s \leq q \leq s + \varepsilon$.
Alors, par croissance de $s \mapsto \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{U_k(\omega) \leq s}$, on a

$$\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{U_k(\omega) \leq p} \leq \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{U_k(\omega) \leq s} \leq \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{U_k(\omega) \leq q}$$

pour tout $\omega \in A$.
D'où, en passant à la $\limsup$ à droite et à la $\liminf$ à gauche, on en déduit (la propriété étant vérifiée pour $p$ et $q$) :

$$s - \varepsilon \leq \liminf_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{U_k(\omega) \leq s} \leq \limsup_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{U_k(\omega) \leq s} \leq s + \varepsilon$$

pour tout $\omega \in A$.
Pour chaque $\omega \in A$, la suite de fonctions croissantes $\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{U_k(\omega) \leq s}$ converge donc simplement vers $s$ sur $[0,1]$. Le deuxième théorème de Dini, nous assure alors que la convergence est uniforme sur $[0,1]$. Ce qui achève la démonstration. ■

**Remarques :**

- Pour justifier que l'application définie sur $\Omega$ par $\sup_{t\in\mathbb{R}} |F_n(t) - F(t)|$ est bien une variable aléatoire (ce qui justifie par la même occasion la suite d'équivalences en loi de la preuve du théorème de Glivenko-Cantelli), on a besoin du lemme suivant :

**Lemme 2**

*La fonction de répartition empirique $F_n$ bâtie sur les variables aléatoires $X_1, ..., X_n$ de fonction de répartition $F$ vérifie :*

$$\forall \omega \in \Omega, \sup_{x\in\mathbb{R}} |F_n(x)(\omega) - F(x)| = \sup_{x\in\mathbb{Q}} |F_n(\omega)(x) - F(x)|$$

*En conséquence, $\|F_n - F\|_\infty$ est une variable aléatoire réelle.*

**Démonstration :** Fixons $\omega$ quelconque dans $\Omega$ et notons pour alléger

$$\tau := \|F_n(.)(\omega) - F\|_\infty = \sup_{x\in\mathbb{R}} |F_n(x)(\omega) - F(x)|$$

Comme $F_n(x)(\omega)$ et $F(x)$ sont toujours deux réels de $[0, 1]$, ce suprémum $\tau$ est fini. Alors, pour tout $\varepsilon > 0$, il existe un $x_\varepsilon$ tel que

$$\tau - \varepsilon < |F_n(x_\varepsilon)(\omega) - F(x_\varepsilon)| \leq \tau$$

Les fonctions de répartitions $F_n(.)(\omega)$ et $F$ étant continues à droite au point $x_\varepsilon$, la valeur absolue de leur différence l'est aussi. Il existe donc un $\delta > 0$ tel que

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

Par ailleurs, $\mathbb{Q}$ étant dense dans $\mathbb{R}$, il y a au moins un rationnel $t$ dans l'intervalle $]x_\varepsilon, x_\varepsilon + \delta[$. Ce rationnel vérifiant l'inégalité ci-dessus, on en déduit

$$\sup_{r\in\mathbb{Q}} |F_n(r)(\omega) - F(r)| > \tau - 2\varepsilon$$

Le premier membre ne dépendant pas de $\varepsilon$ et $\varepsilon > 0$ étant arbitraire, on en déduit

$$\sup_{r\in\mathbb{Q}} |F_n(r)(\omega) - F(r)| \geq \tau := \sup_{r\in\mathbb{R}} |F_n(r)(\omega) - F(r)|$$

Puisque $\omega$ était quelconque, ceci vaut pour tout $\omega \in \Omega$. L'inégalité dans l'autre sens est évidente, donc l'égalité du lemme est démontrée pour tout $\omega \in \Omega$. ■

- Il existe bien sur un premier théorème de Dini dont l'énoncé est le suivant :

**Théorème 3 (*Premier théorème de Dini*)**

*Soient $a, b \in \mathbb{R}$ avec $a < b$, et $(f_n)_{n\in\mathbb{N}}$ une suite croissante de fonctions continues de $[a, b]$ dans $\mathbb{R}$, qui converge simplement vers une fonction continue $f$. Alors, $(f_n)_{n\in\mathbb{N}}$ converge uniformément vers $f$ sur $[a, b]$.*

Prendre garde au fait que dans ce théorème, on a besoin de la continuité des $f_n$, contrairement au deuxième théorème de Dini. Dans *Gourdon*, les fonctions $f_n$ sont supposées continues dans les deux théorèmes de Dini, mais comme on le voit ici, cette hypothèse n'est pas nécessaire pour le deuxième. La démonstration du théorème de Glivenko-Cantelli nécessite d'ailleurs cet énoncé plus précis, puisqu'a priori, il n'y a aucune raison pour que les fonctions $F_n$ soient continues.

- 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 $\mu$ 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 $\mu_{KS}$ 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 Kolomogorov-Smirnov.