Théorème de Glivenko-Cantelli.

Rémi Lajugie

Définition 1 *Soit $(X_i)_{i\in\mathbb{N}}$, une suite de variables aléatoires indépendantes réelles et identiquement distribuées (i.i.d), on appelle fonction de répartition empirique d'ordre n la fonction définie $\forall t \in \mathbb{R}, F_n(t) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{X_i \le t}$.*

Théorème 1 *Soit $(X_i)_{i\in\mathbb{N}}$ une suite de variables aléatoires i.i.d de loi μ. On note F_n la fonction de répartition empirique et F la fonction de répartition de μ. Alors $\|F - F_n\|_\infty \to 0$ et ce, presque sûrement.*

Preuve :

Etape 1 : $\|F - F_n\|_\infty$ est une variable aléatoire mesurable

Soit $n \in \mathbb{N}$. On note $(X^{(i)})_{i\in\{1,\dots,n\}}$ la statistique d'ordre (on ordonne les X_i dans l'ordre croissant). F est croissante sur $\mathbb{R}$, F_n est constante et vaut i/n sur les intervalles $[X^{(i)}, X^{(i+1)}[$ $F - F_n$ est monotone sur les intervalles on a $\|F - F_n\|_\infty = \max_{i=1\dots n}(|\frac{i}{n} - F(X^i)\|, |\frac{i}{n} - F^-(X^{i+1})\|)$, où F^- est la limite à gauche de la fonction F.

Ainsi, comme le max de variables aléatoires est une variable aléatoire, $\|F - F_n\|_\infty$ est mesurable.

Etape 2 : $\mathbb{P}(\|F - F_n\|_\infty \to 0)$ ne dépend que de la loi μ.

En effet $\|F - F_n\|_\infty = \psi_n(X_1, \dots, X_n)$ où $\psi_n : (x_1, \dots, x_n) \mapsto sup_{t\in\mathbb{R}}|F(t) - \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{X_i \le t}|$ est une fonction mesurable de $\mathbb{R}^n$ dans $\mathbb{R}$, via le raisonnement qui précède. La question de la convergence presque certaine revient donc à étudier la probabilité

$$\mathbb{P}(\|F_n - F\| \to 0) = \mathbb{P}_\mu(\psi_n(\pi_1, \dots, \pi_n) \to 0).$$

Ainsi la convergence presque sûre est une propriété de la mesure image de μ.

Etape 3 : on peut choisir l'espace probabilisé qui nous arrange.

On sait que si $(U_i)_{i\in\mathbb{N}}$ est une suite de variables aléatoires i.i.d, suivant une loi uniforme sur $[0, 1]$, alors les variables aléatoires $Q(U_i)$ où Q est la pseudo inverse ($Q(y) = \inf_{t\in\mathbb{R}} F(x) \ge u$) suit la loi μ.

Il vient alors que $\|F_n - F\|_\infty = sup_{t\in\mathbb{R}}|F(t) - \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{Q(U_i)\le t}| \le sup_{t\in\mathbb{R}}|y - \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{U_i \le y}|$.

Ainsi il suffit de montrer le théorème dans le cas réel.

Etape 4 : On conclut avec le théorème de Dini.

On considère une énumération des rationnels, par la loi des grands nombres, on a convergence ponctuelle de $F_n(q)$ pour tout rationnel vers F sauf sur un ensemble A_q de mesure nulle. Soit A l'intersection des A_q. Comme F_n est croissante, par on a en fait convergence ponctuelle pour tout $\omega \in A_q$. On peut alors appliquer le théorème de Dini car on a une suite de fonction croissantes sur un compact (le support de la loi uniforme est $[0, 1]$). Cela permet de conclure.

Références

— Garet-Kurtzmann.