Leçon 250 : Loi des grands nombres. Théorème de la limite centrale. Applications.

1 Loi des grands nombres

1.1 Loi faible des grands nombres

Proposition 1. Inégalité de Tchebychev
Soit X une variable aléatoire réelle admettant un moment d'ordre 2. Pour tout réel $\epsilon > 0$, on a

$$P(|X - EX| > \epsilon) \leq \frac{\sigma_X}{\epsilon^2}$$

Théorème 1. Loi faible des grands nombres
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires définies sur l'espace probabilisé $(\Omega, \mathcal{A}, P)$, indépendantes et admettant un moment d'ordre 2. On suppose que :

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} EX_j = m \text{ et } \lim_{n\to\infty} \frac{1}{n^2}\sum_{j=1}^{n} \sigma_{X_j}^2 = 0.$$

Alors la suite des variables aléatoires $\bar{X}_n = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} X_j$ converge en probabilité vers m.

Théorème 2. Théorème de Bernoulli
Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements indépendants de même probabilité p. La suite des variables aléatoires $\frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{A_j}$ converge en probabilité vers p.

Application 1. Théorème de Bernstein
Soient $f : [0,1] \to \mathbb{C}$ une fonction continue, $\omega(h) = sup\{|f(u) - f(v)|; |u - v| \leq h\}$. Pour $n \geq 1$, on considère le polynôme $B_n(f,x) = B_n(x) = \displaystyle\sum_{k=0}^{n} \binom{n}{k} x^k(1-x)^{n-k} f\left(\frac{k}{n}\right)$, le n-ième polynôme de Bernstein de f. Alors :

a)$(B_n)_{n\in\mathbb{N}}$ converge vers f uniformément sur $[0,1]$.

b)Plus précisément, on a $\|f - B_n\|_\infty \leq C\omega\left(\dfrac{1}{\sqrt{n}}\right)$.

c)L'estimation du b) est optimale : il existe une fonction lipschitzienne f pour laquelle $\|f - B_n\|_\infty \geq \dfrac{\delta}{\sqrt{n}}$, où δ est une constante numérique.

Théorème 3. Théorème de Khintchine
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires définies sur l'espace probabilisé $(\Omega, \mathcal{A}, P)$, deux à deux indépendantes de même loi μ et admettant une moyenne. Alors, la suite des variables aléatoires $\bar{X}_n = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} X_j$ converge en probabilité vers la moyenne commune EX_1.

Théorème 4. Estimation des grands écarts
Soit $p \in [0,1]$. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes, identiquement distribuées, de loi de Bernoulli, de paramètre p. On pose $S_n = \sum_{j=1}^{n} X_j$, de loi $\mathscr{B}(n,p)$. Pour $\epsilon > 0$, on pose

$$h_+(\epsilon) = (p+\epsilon)\ln\left(\frac{p+\epsilon}{\epsilon}\right) + (1-p-\epsilon)\ln\left(\frac{1-p-\epsilon}{1-p}\right).$$

Alors

a) $h_+ > 0$

b) $\forall n \geq 1, P\left(\dfrac{S_n}{n} \geq p + \epsilon\right) \leq \exp(nh_+(\epsilon))$

c) Optimalité : $\dfrac{1}{n}\ln\left(P\left(\dfrac{S_n}{n} \geq p+\epsilon\right)\right) \underset{n\to+\infty}{\longrightarrow} -h_+(\epsilon)$

1.2 Loi forte des grands nombres

Théorème 5. Loi forte des grands nombres
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires définies sur l'espace probabilisé $(\Omega, \mathcal{A}, P)$, indépendantes et admettant un moment d'ordre 2. On suppose que :

$$\lim_{n\to\infty} EX_n = m \quad \text{et} \quad \sum_{j=1}^{\infty} \frac{1}{j^2}\sigma_{X_j}^2 < +\infty.$$

Alors la suite des variables aléatoires $\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$ converge P-p.s. et dans L^2 vers m.

Exemple 1. La suite de variables aléatoires indépendantes $(X_n)_{n\geq 2}$ de lois données par

$$P_{X_n} = \frac{1}{2n\ln(n)}(\delta_n + \delta_{-n}) + \left(1 - \frac{1}{2n\ln(n)}\right)\delta_0$$

suit la loi faible des grands nombres mais pas la loi forte.

Théorème 6. Théorème de Kolmogorov-Khintchine
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires définies sur l'espace probabilisé $(\Omega, \mathcal{A}, P)$, indépendantes, de même loi. Les assertions suivantes sont équivalentes :

a)Il existe un réel c tel que la suite des variables aléatoires $\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$ converge P-p.s. vers c.

b)$X_1 \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$.
Si l'assertion a) est vraie, on a $c = EX_1$.

Application

Soit X une variable aléatoire de loi μ. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles, indépendantes, de même loi μ et de fonction de répartition F.

Définition 1. Le vecteur aléatoire $(X_1, X_2, \ldots, X_n)$ est appelé échantillon de taille n de X. La fonction F_n de $\mathbb{R}\times\Omega$ dans $[0,1]$ définie par

$$\forall(x,\omega)\in\mathbb{R}\times\Omega \quad F_n(x,\omega) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{(X_j\leq x)}(\omega)$$

est appelée fonction de répartition empirique (associée à X) basée sur l'échantillon $(X_1, X_2, \ldots, X_n)$.

Théorème 7. Théorème de Glivenko-Cantelli
Pour P-presque tout ω, la suite des fonctions de répartition $F_n(\cdot,\omega)$ converge uniformément vers F, autrement dit, on a

$$\text{P-p.s.} \lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x,\cdot) - F(x)| = 0.$$

F continue

Remarque 1. Le théorème de Kolmogorov-Smirnov affirme que si F_n est une fonction de répartition empirique, $D_n = \sqrt{n}\|F_n - F\|$ converge en loi vers une mesure dite de Kolmogorov-Smirnov, caractérisée par exemple par sa fonction de répartition :

$$\forall t > 0, \quad F_{KS}(t) = \sum_{k=-\infty}^{\infty} (-1)^k \exp(-2k^2t^2)$$

Cette convergence est à la base du test de Kolmogorov-Smirnov qui teste la fonction de répartition d'un échantillon, contre une autre.

Proposition 2. Méthode de Monte-Carlo pour le calcul d'intégrales
Soient D un domaine de $\mathbb{R}^d$ et f une fonction réelle définie sur $\mathbb{R}^d$ mesurable, tels que $\mathbf{1}_D \cdot f$ soit Lebesgue-intégrable. Soit $(U_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes, de loi uniforme sur $[0,1]$. On définit pour tout n la variable aléatoire $\underline{\underline{U_n}}$ à valeurs dans $\mathbb{R}^d$ par

$$\underline{\underline{U_n}} = (U_{nd+1}, U_{nd+2}, \ldots, U_{(n+1)d})$$

et la variable aléatoire réelle $X_n = (\mathbf{1}_D \cdot f) \circ \underline{\underline{U_n}}$..
Alors la suite de terme général $S_n = \frac{1}{n}\sum_{j=1}^n X_j$ converge P-p.s. vers l'intégrale $I = \displaystyle\int_{D\cap[0,1]^d} f(x)dx$ et si f est bornée par $c > 0$, on a, pour tout $\epsilon > 0$,

$$P(|S_n - I| > \epsilon) \leq \frac{c^2}{n\epsilon^2}.$$

Nombres normaux

Soient x un irrationnel dans $[0,1]$ et un entier $r \geq 2$. x se développe de façon unique en base r, c'est-à-dire :

$$x = \sum_{n=1}^{+\infty} \frac{\epsilon_n(x)}{r^n}, \quad \text{où } \epsilon_n(x) \in \{0, 1, \ldots, r-1\} = A.$$

Soit $b = (b_1, \ldots, b_k) \in A^k$. On note n_b le nombre des $i \leq n-k+1$ tels que $\epsilon_i(x) = b_1, \epsilon_{i+1}(x) = b_2, \ldots, \epsilon_{i+k-1}(x) = b_k$.

Définition 2. x est normal en base r si, pour tout $k \geq 1$ et tout $b \in A^k$, on a $\dfrac{n_b}{n} \underset{n\to+\infty}{\longrightarrow} \dfrac{1}{r^k}$.
x est normal s'il est normal en toute base $r \geq 2$.

Théorème 8. Presque tout irrationnel est normal.

2 Théorème central limite

Définition 3. Soit pour $n \in \mathbb{N}$ une variable aléatoire X_n,définie sur un espace probabilisé $(\Omega, \mathcal{A}, P^n)$, à valeurs dans $\mathbb{R}^d$, et soit μ une probabilité sur $\mathbb{R}^d$. La suite des variables aléatoires $(X_n)_{n\in\mathbb{N}}$ converge en loi vers μ si la suite $(P^n_{X_n})_{n\in\mathbb{N}}$ des lois des X_n converge étroitement vers la loi μ.

Théorème 9. Théorème de Lévy
Soit pour $n \in \mathbb{N}$ une variable aléatoire X_n,définie sur un espace probabilisé $(\Omega_n, \mathcal{A}_n, P^n)$, à valeurs dans $\mathbb{R}^d$, de fonction caractéristique ϕ_{X_n}.
a) Si la suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X, où X est une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, à valeurs dans $\mathbb{R}^d$, alors la suite $(\phi_{X_n})_{n\in\mathbb{N}}$ des fonctions caractéristiques converge simplement (et même uniformément sur tout compact de $\mathbb{R}^d$) vers la fonction caractéristique ϕ_X de X.
b) Inversement, si la suite $(\phi_{X_n})_{n\in\mathbb{N}}$ des fonctions caractéristiques converge simplement vers une fonction ϕ continue en 0, alors ϕ est la transformée de Fourier d'une probabilité μ sur $\mathbb{R}^d$, et la suite des variables aléatoires $(X_n)_{n\in\mathbb{N}}$ converge en loi vers μ.
De plus il existe une variable aléatoire (non unique) X définie sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, à valeurs dans $\mathbb{R}^d$, telle que la suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X.

Lemme 1. Si la variable aléatoire réelle X admet un moment d'ordre 2, sa fonction caractéristique ϕ_X admet un développement limité d'ordre 2 en 0 donnée par, pour tout réel t,

$$\phi_X(t) = 1 + itEX - \frac{t^2}{2}EX^2 + o(t^2)$$

Théorème 10. Théorème central limite
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{A}, P)$,à valeurs dans $\mathbb{R}^d$,indépendantes, de même loi et admettant un moment d'ordre 2. La suite de terme général Y_n défini pour tout $n \in \mathbb{N}^*$ par

$$Y_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (X_j - EX_j)$$

converge en loi vers la loi gaussienne $\mathcal{N}_{\mathbb{R}^d}(0, C_{X_1})$, où C_{X_1} est la matrice des covariances des X_j.

Théorème 11. Théorème des événements rares de Poisson
Soit pour tout $n \in \mathbb{N}^*$, une famille finie $\{A_{n,j} | 1 \leq j \leq M_n\}$ d'événements indépendants définis sur un espace probabilisé $(\Omega, \mathcal{A}, P)$. On pose $P(A_{n,j}) = p_{n,j}$ et on note

$$S_n = \sum_{j=1}^{M_n} \mathbf{1}_{A_{n,j}}.$$

On suppose que la suite de terme général M_n tend en croissant vers $+\infty$, que

$$\max_{1 \leq j \leq M_n} p_{n,j} \underset{n \to +\infty}{\longrightarrow} 0 \text{ et que } \sum_{j=1}^{M_n} p_{n,j} \underset{n \to +\infty}{\longrightarrow} \lambda,$$

où $\lambda > 0$. Alors la suite $(S_n)_{n\in\mathbb{N}^*}$ converge en loi vers la loi de Poisson $\mathscr{P}(\lambda)$.

Théorème 12. Théorème de Karl-Pearson
Soit $k \in \mathbb{N}^*$ fixé. Pour tout $n \in \mathbb{N}^*$, on considère une partition $(A_j^n)_{1 \leq j \leq k}$ de Ω par des ensembles $\mathcal{A}$-mesurables. On suppose que ces partitions sont indépendantes, c'est-à-dire que les familles indexées sur n, constituées par les éléments de ces partitions sont indépendantes. On suppose de plus que,

$$\forall n \in \mathbb{N}^*, \quad \forall j = 1, \dots, k, \quad P(A_j^n) = p_j,$$

où $p_j > 0$ et $\sum_{j=1}^{n} p_j = 1$. On définit, pour tout $j = 1, \dots, k$, les variables aléatoires réelles

$$N_j^n = \sum_{l=1}^{n} \mathbf{1}_{A_j^l},$$

puis la variable aléatoire

$$\chi_{k,n}^2 = \sum_{j=1}^{k} \frac{(N_j^n - np_j)^2}{np_j} = \sum_{j=1}^{k} \frac{(\frac{1}{n} N_j^n - p_j)^2}{p_j}.$$

Alors la suite des lois $P_{\chi_{k,n}^2}$ converge étroitement vers la loi χ_{k-1}^2 du χ-deux à $k-1$ degrés de liberté ; autrement dit, la suite des variables aléatoires $\chi_{k,n}^2$ converge en loi vers la loi du χ-deux à $k-1$ degrés de liberté.