cours 24, le jeudi 6 mai 2010

Lois des grands nombres

On peut penser que le fait de considérer la moyenne

$$\mathrm{Z}_n(\omega) = \frac{\mathrm{X}_1(\omega) + \cdots + \mathrm{X}_n(\omega)}{n}$$

des résultats indépendants successifs $(\mathrm{X}_n(\omega))$ d'une même expérience aléatoire, pour n assez grand, a des chances de stabiliser le phénomène, en réduisant l'influence de l'aléa. Ça n'est pas toujours le cas, comme l'indique l'exercice qui suit.

Exercice. Si les (X_j) sont des v.a. de Cauchy indépendantes, de loi

$$\frac{\mathrm{d}x}{\pi(1+x^2)},$$

la variable aléatoire Z_n définie ci-dessus a toujours la même loi de Cauchy !

Solution. — On a vu que la fonction caractéristique de ces variables de Cauchy est donnée par

$$\forall t \in \mathbb{R}, \quad \varphi_{\mathrm{X}_j}(t) = \mathrm{e}^{-|t|}\,;$$

par conséquent,

$$\varphi_{\mathrm{Z}_n}(t) = \mathrm{E}\,\mathrm{e}^{\mathrm{i}t\mathrm{Z}_n} = \mathrm{E}\Bigl(\prod_{j=1}^{n} \mathrm{e}^{\mathrm{i}\,(t/n)\mathrm{X}_j}\Bigr) = \prod_{j=1}^{n} \varphi_{\mathrm{X}_j}(t/n) = \bigl(\mathrm{e}^{-|t|/n}\bigr)^n = \mathrm{e}^{-|t|}\,.$$

Loi faible des grands nombres

L'effet régularisant escompté aura lieu pour les lois qui ont un moment absolu d'ordre 1, autrement dit pour les variables aléatoires intégrables. On va commencer par un résultat assez simple qui est à notre portée.

Théorème (loi faible des grands nombres). *Si les variables aléatoires* $(\mathrm{X}_n)_{n\geqslant 1}$ *sont indépendantes, de même loi et intégrables, la variable aléatoire*

$$\frac{\mathrm{X}_1 + \cdots + \mathrm{X}_n}{n}$$

converge en loi vers la constante $\mathrm{E}\,\mathrm{X}_1$.

Preuve. — On recentre en posant

$$\mathrm{Y}_j = \mathrm{X}_j - \mathrm{E}\,\mathrm{X}_j,$$

et il s'agit de montrer que $\mathrm{V}_n = (\mathrm{Y}_1 + \cdots + \mathrm{Y}_n)/n$ tend en loi vers la variable aléatoire constante $\mathbf{0}$, partout égale à 0 ; d'après le cours précédent, on sait que la convergence étroite des lois $\mathrm{P}_{\mathrm{V}_n}$ des v.a. V_n vers la mesure de Dirac δ_0 au point 0 équivaut à la convergence simple sur $\mathbb{R}$ des fonctions caractéristiques. Or

$$\varphi_{\mathbf{0}}(t) = \mathrm{E}\,\mathrm{e}^{\mathrm{i}t\cdot 0} = 1,$$

et

$$\varphi_{\mathrm{Y}}(t) = 1 + t\varphi'_{\mathrm{Y}}(t) + t\varepsilon(t) = 1 + \mathrm{i}t\,\mathrm{E}\,\mathrm{Y} + t\varepsilon(t) = 1 + t\varepsilon(t)\,;$$

il en résulte que

$$\varphi_{\mathrm{V}_n}(t) = \mathrm{E}\,\mathrm{e}^{\mathrm{i}t\mathrm{V}_n} = \mathrm{E}\Bigl(\prod_{j=1}^{n} \mathrm{e}^{\mathrm{i}(t/n)\mathrm{Y}_j}\Bigr) = \prod_{j=1}^{n} \mathrm{E}\,\mathrm{e}^{\mathrm{i}(t/n)\mathrm{Y}_j} =$$

$$= \prod_{j=1}^{n} \varphi_{\mathrm{Y}_j}(t/n) = \bigl(\varphi_{\mathrm{Y}}(t/n)\bigr)^n = \Bigl(1 + \frac{t}{n}\,\varepsilon(t/n)\Bigr)^n \underset{n}{\longrightarrow} 1.$$

On vérifie l'affirmation précédente en prenant le logarithme,

$$\ln \varphi_{\mathrm{V}_n}(t) = n \ln\Bigl(1 + \frac{t}{n}\,\varepsilon(t/n)\Bigr) \sim n\,\Bigl(\frac{t}{n}\,\varepsilon(t/n)\Bigr) = t\varepsilon(t/n) \underset{n}{\longrightarrow} 0.$$

Le cas L^2

Si les Y_i sont des v.a. réelles dans L^2, indépendantes et centrées, elles sont orthogonales ; en effet, pour $i \neq j$,

$$\langle \mathrm{Y}_i, \mathrm{Y}_j \rangle = \int_\Omega \mathrm{Y}_i(\omega)\overline{\mathrm{Y}_j(\omega)}\,\mathrm{dP}(\omega) = \mathrm{E}\,\mathrm{Y}_i\mathrm{Y}_j = \mathrm{E}\,\mathrm{Y}_i\,\mathrm{E}\,\mathrm{Y}_j = 0.$$

Pour des v.a. $\mathrm{X}_1, \mathrm{X}_2$ non nécessairement centrées, on introduit la *covariance*, définie par

$$\mathrm{Cov}(\mathrm{X}_1, \mathrm{X}_2) = \mathrm{E}\bigl((\mathrm{X}_1 - \mathrm{E}\,\mathrm{X}_1)(\mathrm{X}_2 - \mathrm{E}\,\mathrm{X}_2)\bigr).$$

Si X_1 et X_2 sont indépendantes, leur covariance est nulle d'après ce qui précède. On note que $\mathrm{Cov}(\mathrm{X}, \mathrm{X}) = \mathrm{Var}(\mathrm{X})$.

Quand $\mathrm{Cov}(\mathrm{X}_1, \mathrm{X}_2) = 0$, on dit que X_1 et X_2 sont *non corrélées*. Si X_1 et X_2 sont indépendantes, elles sont non corrélées. Il faut savoir que la réciproque est vastement fausse : la non corrélation de deux v.a., ou bien l'orthogonalité de deux variables aléatoires centrées est une propriété beaucoup plus faible que l'indépendance.

Voici un exemple à l'appui de la phrase précédente. Si G est gaussienne (par exemple) et si X est une variable de Bernoulli centrée, prenant les valeurs ± 1, indépendante de G, alors G et GX sont orthogonales :

$$\mathrm{E}(\mathrm{G})(\mathrm{GX}) = \mathrm{E}\,\mathrm{G}^2\mathrm{X} = \mathrm{E}\,\mathrm{G}^2\,\mathrm{E}\,\mathrm{X} = 0,$$

mais elles ne sont pas indépendantes, sinon les fonctions $|\mathrm{G}|$ et $|\mathrm{GX}| = |\mathrm{G}|$ de ces variables seraient indépendantes aussi, or $|\mathrm{G}|$ n'est pas indépendante d'elle même ; en effet, on devrait avoir

$$\mathrm{P}(|\mathrm{G}| < 1) = \mathrm{P}\bigl((|\mathrm{G}| < 1)\,\&\,(\mathrm{G} < 1)\bigr) = (\mathrm{P}(\mathrm{G} < 1))^2,$$

qui n'est possible que si $\mathrm{P}(|\mathrm{G}| < 1) = 0$ ou 1, et ceci est faux pour une gaussienne. En modifiant cet argument, on pourra voir que les seules v.a. indépendantes d'elles mêmes sont les constantes.

Il résulte de l'orthogonalité (« théorème de Pythagore » dans l'espace $\mathrm{L}^2(\Omega, \mathcal{F}, \mathrm{P})$) que pour des v.a. de carré intégrable centrées et indépendantes (Y_j), on a

$$\mathrm{E}\Big(\sum_{j=1}^{n} \mathrm{Y}_j\Big)^2 = \Big\|\sum_{j=1}^{n} \mathrm{Y}_j\Big\|_2^2 = \sum_{j=1}^{n} \|\mathrm{Y}_j\|_2^2 = \sum_{j=1}^{n} \mathrm{E}\,\mathrm{Y}_j^2.$$

Proposition 1. *Si les variables aléatoires* X_i *sont indépendantes et sont dans l'espace* $\mathrm{L}^2(\Omega, \mathcal{F}, \mathrm{P})$, *on a*

$$\mathrm{Var}\Big(\sum_{i=1}^{n} \mathrm{X}_i\Big) = \sum_{i=1}^{n} \mathrm{Var}(\mathrm{X}_i).$$

Preuve. — Posons $\mathrm{Y}_i = \mathrm{X}_i - \mathrm{E}\,\mathrm{X}_i$; les (Y_j) sont indépendantes et centrées, ce qui ramène au cas précédent : on a

$$\mathrm{Var}\Big(\sum_{i=1}^{n} \mathrm{X}_i\Big) = \mathrm{E}\Big(\sum_{i=1}^{n} \mathrm{X}_i - \sum_{i=1}^{n} \mathrm{E}\,\mathrm{X}_i\Big)^2 = \mathrm{E}\Big(\sum_{i=1}^{n} \mathrm{Y}_i\Big)^2 = \sum_{i=1}^{n} \mathrm{E}\,\mathrm{Y}_i^2 = \sum_{i=1}^{n} \mathrm{Var}(\mathrm{X}_i).$$

Avec l'inégalité de Tchebychev, on obtient pour tout $\delta > 0$ et tout $n \geq 1$, lorsque les (X_j) sont indépendantes de même loi, de carré intégrable

$$\mathrm{P}\Big(\Big|\frac{\sum_{i=1}^{n} \mathrm{X}_i}{n} - \mathrm{E}\,\mathrm{X}\Big| \geq \delta\Big) = \mathrm{P}\Big(\Big|\sum_{i=1}^{n} \mathrm{X}_i - \mathrm{E}\Big(\sum_{i=1}^{n} \mathrm{X}_i\Big)\Big| \geq n\delta\Big) \leq \frac{\mathrm{Var}(\sum_{i=1}^{n} \mathrm{X}_i)}{n^2\delta^2} = \frac{\mathrm{Var}(\mathrm{X})}{\delta^2 n}.$$

On retrouve la loi faible, mais avec une estimation de la vitesse de convergence, alors qu'il n'y avait aucune information de vitesse dans le théorème précédent.

Remarque. Si on arrivait à gagner un $\varepsilon > 0$, sous la forme

$$\mathrm{P}\Big(\Big|\frac{\sum_{i=1}^{n} \mathrm{X}_i}{n} - \mathrm{E}\,\mathrm{X}\Big| \geq \delta\Big) \leq \frac{\mathrm{C}(\mathrm{X}, \delta)}{n^{1+\varepsilon}},$$

on aurait une série convergente et on pourrait déduire la loi forte des grands nombres par Borel-Cantelli (voir la preuve donnée à la fin de ce cours).

La loi forte implique la faible : de façon générale, si une suite (Z_n) de v.a. converge presque sûrement vers une v.a. Z, il en résulte que la loi de Z_n converge étroitement vers la loi de Z. En effet, pour tout t réel, on a la convergence

$$\mathrm{e}^{\mathrm{i}t\mathrm{Z}_n} \xrightarrow[n]{} \mathrm{e}^{\mathrm{i}t\mathrm{Z}}$$

presque partout, avec un module égal à la fonction constante 1 qui est P-intégrable. Par convergence dominée,

$$\varphi_{\mathrm{Z}_n}(t) = \mathrm{E}\,\mathrm{e}^{\mathrm{i}t\mathrm{Z}_n} \xrightarrow[n]{} \mathrm{E}\,\mathrm{e}^{\mathrm{i}t\mathrm{Z}} = \varphi_{\mathrm{Z}}(t).$$

Weierstrass par Bernstein

Sur l'ensemble à deux points $\{0,1\}$ considérons pour tout $x \in [0,1]$ la probabilité μ_x définie par

$$\mu_x(\{1\}) = x, \quad \mu_x(\{0\}) = 1 - x.$$

Sur $\Omega_n = \{0,1\}^n$ on regarde les fonctions coordonnées : pour $\omega = (\omega_1, \ldots, \omega_n) \in \Omega_n$ et $i = 1, \ldots, n$ on pose

$$\mathrm{X}_i(\omega) = \omega_i \in \{0,1\} \subset \mathbb{R}.$$

On introduit sur Ω_n les probabilités $(\mathrm{P}_x)_{x\in[0,1]}$ produit,

$$\mathrm{P}_x = \mu_x \otimes \cdots \otimes \mu_x = \mu_x^{\otimes n}$$

produit tensoriel de n facteurs égaux à μ_x. Comme P_x est une mesure produit, on a quand $\mathrm{A}_j \subset \{0,1\}$, $j = 1, \ldots, n$,

$$(*) \qquad \mathrm{P}_x(\mathrm{A}_1 \times \cdots \times \mathrm{A}_n) = \prod_{j=1}^{n} \mu_x(\mathrm{A}_j).$$

Pour la proba P_x, on a

$$\mathrm{P}_x(\mathrm{X}_i = 1) = x,$$

résultat obtenu en prenant $\mathrm{A}_i = \{1\}$ et les A_j, $j \neq i$, égaux à $\{0,1\}$,

$$\mathrm{P}_x(\mathrm{X}_i = 1) = \mu_x(\{1\}) \prod_{j \neq i} \mu_x(\{0,1\}) = \mu_x(\{1\}) = x.$$

Sous la loi P_x, les (X_i) sont donc de même loi μ_x. L'égalité $(*)$ peut se récrire sous la forme

$$\mathrm{P}_x\big((\mathrm{X}_1 \in \mathrm{A}_1) \,\&\, \ldots \,\&\, (\mathrm{X}_n \in \mathrm{A}_n)\big) = \prod_{j=1}^{n} \mathrm{P}_x(\mathrm{X}_j \in \mathrm{A}_j),$$

qui montre que sous la loi P_x, les variables (X_i) sont indépendantes.

Ces proba P_x sont « polynomiales par rapport au paramètre $x \in [0,1]$ », puisque pour tout singleton $\{(t_1, \ldots, t_n)\}$ contenu dans Ω_n on a

$$\mathrm{P}_x(\{(t_1, \ldots, t_n)\}) = x^k (1-x)^{n-k},$$

où k est le nombre de 1 dans la suite $(t_1, \ldots, t_n) \in \Omega_n$; le résultat précédent est obtenu en prenant $\mathrm{A}_j = \{t_j\}$, $j = 1, \ldots, n$. On a par ailleurs

$$\mathrm{E}_x\, \mathrm{X}_i = 1.\mathrm{P}_x(\mathrm{X}_i = 1) + 0.\mathrm{P}_x(\mathrm{X}_i = 0) = x,$$

et

$$\mathrm{Var}_x\, \mathrm{X}_i = \mathrm{E}_x(\mathrm{X}_i - x)^2 = (1-x)^2 x + x^2(1-x) = x(1-x) \leq 1/4.$$

Posons

$$\mathrm{S}_n = \sum_{j=1}^{n} \mathrm{X}_j, \quad \mathrm{M}_n = \frac{\mathrm{S}_n}{n}.$$

On vérifie que

$$\mathrm{E}_x\,\mathrm{M}_n = x.$$

Par la proposition 1, on a $\mathrm{Var}_x\,\mathrm{S}_n = n\,\mathrm{Var}_x\,\mathrm{X}_1 = nx(1-x)$, donc

$$\mathrm{Var}_x(\mathrm{M}_n) = \mathrm{Var}_x(\mathrm{S}_n/n) = \mathrm{Var}_x(\mathrm{S}_n)/n^2 = nx(1-x)/n^2 \le 1/(4n).$$

Donc par Tchebychev,

$$\mathrm{P}_x(|\mathrm{M}_n - x| \ge \delta) \le \frac{1}{4\delta^2 n}.$$

On retrouve le principe de la loi des grands nombres : il y a de grandes chances, pour la probabilité P_x, que la valeur de M_n soit proche de x.

Si f est continue sur $[0,1]$, il y aura aussi de grandes chances que la valeur de $f(\mathrm{M}_n)$ soit proche de $f(x)$, et en particulier que l'espérance $\mathrm{E}_x\,f(\mathrm{M}_n)$ soit proche de $f(x)$. Cela étant vrai pour tout x, on aura approché $x \to f(x)$ par $x \to \mathrm{E}_x\,f(\mathrm{M}_n)$, qui se trouve être polynomiale. On aura ainsi une preuve du théorème d'approximation polynomiale de Weierstrass.

Il faut préciser les choses ; si $\varepsilon > 0$ est donné, il existe $\delta > 0$ tel que $|f(y) - f(x)| < \varepsilon$ dès que $|x - y| < \delta$. On va voir que $x \to \mathrm{E}_x\,f(\mathrm{M}_n)$ est polynomiale, et uniformément proche de $x \to f(x)$, ce qui démontre le théorème de Weierstrass dans le cas de $[0,1]$.

Déterminons la loi de S_n ; il est clair que S_n prend des valeurs entières qui peuvent varier de $k = 0$ à $k = n$; pour ces valeurs de k, on a

$$\mathrm{P}_x(\mathrm{S}_n = k) = \binom{n}{k} x^k (1-x)^{n-k}.$$

C'est assez clair directement (pour que la somme de valeurs 0 ou 1 soit égale à k, il faut qu'il y ait exactement k valeurs égales 1, chacune obtenue avec probabilité x, et indépendamment, ce qui donne le facteur x^k, mais les k places avec des 1 sont n'importe où parmi n, ce qui amène le coefficient du binôme), mais on va le confirmer par récurrence. Le résultat est clair quand $n = 1$: pour $k = 0, 1$,

$$\mathrm{P}_x(\mathrm{S}_1 = k) = \mathrm{P}_x(\mathrm{X}_1 = k) = x^k(1-x)^{1-k} = \binom{1}{k} x^k (1-x)^{1-k}.$$

Passons de $n \ge 1$ à $n+1$. On a

$$\mathrm{P}_x(\mathrm{S}_{n+1} = k) = \mathrm{P}_x\big((\mathrm{S}_n = k)\,\&\,(\mathrm{X}_{n+1} = 0)\big) + \mathrm{P}_x\big((\mathrm{S}_n = k-1)\,\&\,(\mathrm{X}_{n+1} = 1)\big) =$$

$$= \mathrm{P}_x(\mathrm{S}_n = k)\,(1-x) + \mathrm{P}_x(\mathrm{S}_n = k-1)\,x =$$

$$= \binom{n}{k} x^k (1-x)^{n-k}(1-x) + \binom{n}{k-1} x^{k-1}(1-x)^{n-k+1} x =$$

$$= \binom{n+1}{k} x^k (1-x)^{n+1-k}.$$

On vérifie ainsi que

$$\mathrm{E}_x\,f(\mathrm{M}_n) = \sum_{k=0}^{n} f\Big(\frac{k}{n}\Big)\,\mathrm{P}(\mathrm{S}_n = k) = \sum_{k=0}^{n} f\Big(\frac{k}{n}\Big)\binom{n}{k} x^k (1-x)^{n-k},$$

polynôme de Bernstein $x \to \mathrm{P}_n(f,x)$ de degré n. D'un autre côté, en découpant suivant que $|\mathrm{M}_n - x| < \delta$ ou non, on obtient

$$\Bigl|\mathrm{E}_x\, f(\mathrm{M}_n) - f(x)\Bigr| \le \mathrm{E}_x\bigl|f(\mathrm{M}_n) - f(x)\bigr| \le$$

$$\le \varepsilon \mathrm{P}(|\mathrm{M}_n - x| < \delta) + 2\|f\|_\infty \mathrm{P}(|\mathrm{M}_n - x| \ge \delta) \le \varepsilon + \frac{\|f\|_\infty}{2\delta^2 n}.$$

Cela est valable pour tout $x \in [0,1]$, donc

$$(**) \qquad \|\mathrm{P}_n(f) - f\|_{\mathrm{C}([0,1])} \le \varepsilon + \frac{\|f\|_\infty}{2\delta^2 n}.$$

Si on donne $\varepsilon > 0$, on lui associe $\delta > 0$ (dépendant de f), on peut ensuite choisir n_0 tel que

$$\frac{\|f\|_\infty}{2\delta^2 n_0} \le \varepsilon.$$

Pour tout $n \ge n_0$, on aura alors

$$\|\mathrm{P}_n(f) - f\|_{\mathrm{C}([0,1])} \le 2\varepsilon.$$

Remarque. Si f est 1-lipschitzienne on peut prendre $\delta = \varepsilon$ et la preuve indique, en supposant aussi $\|f\|_\infty \le 1$, qu'il faut associer ε et n de façon que

$$\frac{1}{\varepsilon^2 n} \sim \varepsilon\,;$$

si on donne $n \ge 1$ et si on pose $\varepsilon = n^{-1/3}$, on trouve d'après $(**)$

$$\|\mathrm{P}_n(f) - f\|_{\mathrm{C}([0,1])} \le 2n^{-1/3}.$$

Mais en fait la preuve se simplifie dans ce cas Lipschitz. On peut écrire directement

$$\bigl|\mathrm{E}_x\, f(\mathrm{M}_n) - f(x)\bigr| \le \mathrm{E}_x\,|\mathrm{M}_n - x| \le \bigl(\mathrm{E}_x(\mathrm{M}_n - x)^2\bigr)^{1/2} = \sqrt{\frac{x(1-x)}{n}} \le \frac{1}{2\sqrt{n}}.$$

On ne peut pas faire mieux que $n^{-1/2}$ comme vitesse de convergence dans le cas lipschitzien, comme le montre l'exemple $f(x) = |x - 1/2|$ et une application du théorème de la limite centrale. Dans le cas où f est de classe C^2, on peut écrire avec Taylor-Lagrange

$$\bigl|f(y) - f(x) - (y-x)f'(x)\bigr| \le \|f''\|_\infty \frac{(y-x)^2}{2}$$

qui implique, comme $\mathrm{E}_x\,\mathrm{M}_n = x$, que

$$\mathrm{E}_x(\mathrm{M}_n - x)f'(x) = 0,$$

donc

$$\bigl|\mathrm{E}_x\, f(\mathrm{M}_n) - f(x)\bigr| = \bigl|\mathrm{E}_x\bigl(f(\mathrm{M}_n) - f(x) - (\mathrm{M}_n - x)f'(x)\bigr)\bigr| \le$$

$$\le \|f''\|_\infty\, \mathrm{E}_x(\mathrm{M}_n - x)^2/2 = \|f''\|_\infty \frac{x(1-x)}{2n} \le \frac{\|f''\|_\infty}{8n}.$$

On ne peut pas faire mieux que la vitesse $1/n$, comme montre l'exemple $f(x) = x^2$. En effet,

$$\mathrm{E}_x\, \mathrm{M}_n^2 = \mathrm{E}_x\big((\mathrm{M}_n - x) + x\big)^2 = \mathrm{E}_x(\mathrm{M}_n - x)^2 + x^2,$$

donc l'écart entre la valeur $\mathrm{E}_x\, f(\mathrm{M}_n) = \mathrm{E}_x\, \mathrm{M}_n^2$ donnée par Bernstein et la vraie valeur $f(x) = x^2$ est

$$\mathrm{E}_x\, \mathrm{M}_n^2 - x^2 = \mathrm{E}_x(\mathrm{M}_n - x)^2 = \frac{x(1-x)}{n},$$

de l'ordre de $1/n$; précisément, on a pour le polynôme d'approximation $\mathrm{P}_n(f)$, dans le cas présent où $f(x) = x^2$,

$$\|\mathrm{P}_n(f) - f\|_{\mathrm{C}([0,1])} = \frac{1}{4n}.$$

Le vrai théorème : la loi forte

Théorème (loi forte des grands nombres). *Si les $(\mathrm{X}_n)_{n\geqslant 1}$ sont indépendantes, de même loi et intégrables, on a presque sûrement*

$$\frac{\mathrm{X}_1 + \cdots + \mathrm{X}_n}{n} \xrightarrow[n]{} \mathrm{E}\,\mathrm{X}.$$

On ne montrera qu'un cas particulier simple, celui où $\mathrm{E}\,\mathrm{X}^4 < +\infty$.

Proposition. *On suppose que les $(\mathrm{Y}_i)_{i\geqslant 1}$ sont indépendantes, centrées et*

$$b^4 = \sup_i \mathrm{E}\,\mathrm{Y}_i^4 < +\infty.$$

Alors pour tout $\delta > 0$ et tout $n \geq 1$, on a

$$\mathrm{P}\Big(\Big|\frac{\mathrm{Y}_1 + \cdots + \mathrm{Y}_n}{n}\Big| \geq \delta\Big) \leq \frac{3b^4}{\delta^4 n^2}.$$

Preuve. — On pose $\mathrm{S}_n = \sum_{j=1}^n \mathrm{Y}_j$ et on développe sauvagement

$$\mathrm{E}\,\mathrm{S}_n^4 = \mathrm{E}\prod_{k=1}^4 \Big(\sum_{j_k=1}^n \mathrm{Y}_{j_k}\Big) = \sum_{(j_1,j_2,j_3,j_4)} \mathrm{E}\big(\mathrm{Y}_{j_1}\mathrm{Y}_{j_2}\mathrm{Y}_{j_3}\mathrm{Y}_{j_4}\big).$$

Par l'indépendance et le centrage, l'espérance du produit de quatre est nulle dès qu'un terme Y_{j_k} n'apparaît qu'une seule fois, comme dans (j,i,j,j), avec $i \neq j$. Il ne reste que les termes de la forme (i,i,i,i) ou bien les (i,i,j,j), (i,j,i,j), (i,j,j,i), où le deuxième indice i a trois positions possibles. Il en résulte, en notant que $(\mathrm{E}\,\mathrm{Y}^2)^2 \leq \mathrm{E}\,\mathrm{Y}^4$, que

$$\mathrm{E}\,\mathrm{S}_n^4 = \sum_{i=1}^n \mathrm{E}\,\mathrm{Y}_i^4 + 3\sum_{i\neq j} \mathrm{E}\,\mathrm{Y}_i^2\, \mathrm{E}\,\mathrm{Y}_j^2 \leq nb^4 + 3(n^2 - n)b^4 = (3n^2 - 2n)b^4 \leq 3n^2 b^4.$$

On conclut avec Markov,

$$\mathrm{P}\big(|\mathrm{S}_n|/n \geq \delta\big) = \mathrm{P}(\mathrm{S}_n^4 \geq \delta^4 n^4) \leq \frac{\mathrm{E}\,\mathrm{S}_n^4}{\delta^4 n^4} \leq \frac{3b^4}{\delta^4 n^2}.$$

Proposition. *On suppose que les* $(\mathrm{X}_i)_{i\geqslant 1}$ *sont indépendantes de même loi et que*

$$a^4 = \sup_i \mathrm{E}\,\mathrm{X}_i^4 < +\infty.$$

Il en résulte que presque sûrement

$$\frac{\mathrm{X}_1 + \cdots + \mathrm{X}_n}{n} \xrightarrow[n]{} \mathrm{E}\,\mathrm{X}_1.$$

Preuve. — On a pour les intégrales par rapport à une probabilité

$$\mathrm{E}\,|\mathrm{X}| \leq (\mathrm{E}\,\mathrm{X}^2)^{1/2} \leq (\mathrm{E}\,\mathrm{X}^4)^{1/4}$$

donc

$$|\,\mathrm{E}\,\mathrm{X}_j| \leq a, \quad \mathrm{E}\,\mathrm{X}_j^2 \leq a^2,$$

puisque par hypothèse $\mathrm{E}\,\mathrm{X}_j^4 \leq a^4$. Les variables $\mathrm{Y}_i = \mathrm{X}_i - \mathrm{E}\,\mathrm{X}_i$ sont centrées indépendantes et $(\mathrm{E}\,\mathrm{Y}_i^4)^{1/4} \leq 2a$ (inégalité triangulaire dans L^4). Avec $b = 2a$, on trouve

$$\mathrm{P}\Big(\Big|\frac{\mathrm{X}_1 + \cdots + \mathrm{X}_n}{n} - \mathrm{E}\,\mathrm{X}\Big| \geq \delta\Big) = \mathrm{P}\Big(\Big|\frac{\mathrm{Y}_1 + \cdots + \mathrm{Y}_n}{n}\Big| \geq \delta\Big) \leq \frac{3b^4}{\delta^4 n^2}.$$

On a donc que la suite des événements

$$\mathrm{A}_n = \mathrm{A}_n(\delta) = \Big\{\Big|\frac{\mathrm{Y}_1 + \cdots + \mathrm{Y}_n}{n}\Big| \geq \delta\Big\}$$

vérifie $\sum \mathrm{P}(\mathrm{A}_n) < +\infty$, autrement dit

$$\int_\Omega \Big(\sum_{n=1}^{+\infty} \mathbf{1}_{\mathrm{A}_n}(\omega)\Big)\,\mathrm{dP}(\omega) = \sum_{n=1}^{+\infty} \int_\Omega \mathbf{1}_{\mathrm{A}_n}(\omega)\,\mathrm{dP}(\omega) = \sum_{n=1}^{+\infty} \mathrm{P}(\mathrm{A}_n) < +\infty,$$

donc la fonction sous l'intégrale est finie presque sûrement. Cela veut dire qu'il existe un ensemble négligeable $\mathrm{N}(\delta) \in \mathcal{F}$ tel que pour tout $\omega \notin \mathrm{N}(\delta)$, il n'existe qu'un nombre fini de valeurs de n telles que $\omega \in \mathrm{A}_n(\delta)$. On peut donc trouver un entier $n_0(\omega)$ tel que

$$n \geq n_0(\omega) \;\Rightarrow\; \Big|\frac{\mathrm{Y}_1(\omega) + \cdots + \mathrm{Y}_n(\omega)}{n}\Big| \leq \delta.$$

Pour chaque $\delta = 2^{-k}$, $k \in \mathbb{N}$, on dispose d'un ensemble négligeable $\mathrm{N}(2^{-k})$. Par réunion dénombrable en $k \in \mathbb{N}$ de négligeables, on obtient $\mathrm{N} \in \mathcal{F}$ tel que $\mathrm{P}(\mathrm{N}) = 0$ et tel que pour tout $k \geq 0$, tout $\omega \notin \mathrm{N}$, il existe un entier $n_0(k,\omega)$ tel que pour $n \geq n_0(k,\omega)$, on ait $\omega \notin \mathrm{A}_n(2^{-k})$. C'est la convergence presque sûre ; pour tout $\omega \notin \mathrm{N}$ on a le résultat suivant : pour tout entier k, il existe un entier $n_0 = n_0(k,\omega)$ tel que pour tout $n \geq n_0$, on ait

$$\Big|\frac{\mathrm{Y}_1(\omega) + \cdots + \mathrm{Y}_n(\omega)}{n}\Big| < 2^{-k}.$$

Autrement dit, pour tout $\omega \notin \mathrm{N}$, c'est-à-dire pour presque tout $\omega \in \Omega$, la suite

$$\frac{\mathrm{Y}_1(\omega) + \cdots + \mathrm{Y}_n(\omega)}{n}$$

tend vers $0 = \mathrm{E}\,\mathrm{Y}$ quand n tend vers l'infini.