Chapitre 8

Lois des grands nombres

Soit $(X_k)_{k\geq 1}$ une suite de variables aléatoires réelles définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$. On étudie dans ce chapitre le comportement asymptotique de la suite $(M_n)_{n\geq 1}$ des moyennes arithmétiques :

$$M_n := \frac{1}{n} S_n = \frac{1}{n} \sum_{k=1}^{n} X_k.$$

(Re)lecture conseillée : [ICP], chapitre 6.

8.1 Convergences de suites de variables aléatoires

Commençons par faire le point sur les divers modes de convergences pour une suite $(Y_n)_{n\geq 1}$ de variables aléatoires réelles. Nous connaissons déjà la convergence presque sûre qui est l'autre nom de la convergence $\mathbf{P}$-presque partout sur Ω et les convergences L^p $(1 \leq p < +\infty)$:

$$Y_n \xrightarrow[n\to+\infty]{\mathrm{L}^p} Y \quad \Leftrightarrow \quad \|Y_n - Y\|_p \xrightarrow[n\to+\infty]{} 0 \quad \Leftrightarrow \quad \mathbf{E}|Y_n - Y|^p \xrightarrow[n\to+\infty]{} 0.$$

Il est temps d'y ajouter la convergence *en probabilité.*

Définition 8.1. *La suite Y_n converge en probabilité vers Y (notation $Y_n \xrightarrow[n\to+\infty]{\mathrm{Pr}} Y$) si*

$$\forall \varepsilon > 0, \quad \mathbf{P}\big(|Y_n - Y| \geq \varepsilon\big) \xrightarrow[n\to+\infty]{} 0.$$

Grâce à l'inégalité de Markov, on voit immédiatement que cette nouvelle convergence est impliquée par n'importe quelle convergence L^p $(1 \leq p < +\infty)$:

$$\forall \varepsilon > 0, \quad \mathbf{P}\big(|Y_n - Y| \geq \varepsilon\big) \leq \frac{1}{\varepsilon^p} \mathbf{E}|Y_n - Y|^p.$$

Comme la convergence L^∞ implique toutes les convergences L^p (cf. théorème 6.18), elle implique aussi la convergence en probabilité.

Proposition 8.2. *La convergence presque-sûre implique la convergence en probabilité.*

Démonstration. Fixons $\varepsilon > 0$. L'hypothèse de convergence presque sûre de Y_n vers Y signifie que l'événement

$$\Omega' := \{\omega \in \Omega;\ \lim_{n\to+\infty} Y_n(\omega) = Y(\omega)\}$$

a pour probabilité 1. Définissons

$$\Omega'_\varepsilon := \{\omega \in \Omega;\ \exists k_0 = k_0(\omega),\ \forall n \geq k_0,\ |Y_n(\omega) - Y(\omega)| < \varepsilon\}.$$

C'est bien un évènement (i.e. $\Omega'_\varepsilon \in \mathcal{F}$) puisqu'il s'écrit

$$\Omega'_\varepsilon = \bigcup_{k\in\mathbb{N}^*} \bigcap_{n\geq k} \{|Y_n - Y| < \varepsilon\}.$$

De plus Ω'_ε contient Ω', donc $\mathbf{P}(\Omega'_\varepsilon) = 1$. Pour tout $k \geq 1$, notons

$$A_k := \{\omega \in \Omega;\ \forall n \geq k,\ |Y_n(\omega) - Y(\omega)| < \varepsilon\} = \bigcap_{n\geq k} \{|Y_n - Y| < \varepsilon\}.$$

La suite $(A_k)_{k\geq 1}$ est clairement croissante pour l'inclusion et sa réunion est Ω'_ε. Par continuité séquentielle croissante de $\mathbf{P}$, on a donc $\mathbf{P}(A_k) \uparrow \mathbf{P}(\Omega'_\varepsilon) = 1$ $(k \to +\infty)$. Par conséquent,

$$\forall \delta > 0,\ \exists k_1, \quad P(A_{k_1}) > 1 - \delta.$$

Pour tout $n \geq k_1$, l'évènement $\{|Y_n - Y| < \varepsilon\}$ contient A_{k_1}, d'où

$$\forall n \geq k_1, \quad \mathbf{P}(|Y_n - Y| < \varepsilon) > 1 - \delta$$

et en passant à l'évènement complémentaire

$$\forall n \geq k_1, \quad \mathbf{P}(|Y_n - Y| \geq \varepsilon) < \delta.$$

Ceci établit la convergence vers 0 de $\mathbf{P}(|Y_n - Y| \geq \varepsilon)$. Comme ε était quelconque, on a bien convergence en probabilité de Y_n vers Y. □

Remarque 8.3. La convergence en probabilité n'implique pas la convergence presque-sûre. Voici un contre-exemple. On prend comme espace probabilisé $(]0,1], \mathrm{Bor}(]0,1]), \lambda)$. On définit les Y_n comme suit :

$$\begin{aligned}
&Y_1 = \mathbf{1}_{]0,1]},\\
&Y_2 = \mathbf{1}_{]0,1/2]}, Y_3 = \mathbf{1}_{]1/2,1]},\\
&Y_4 = \mathbf{1}_{]0,1/4]}, Y_5 = \mathbf{1}_{]1/4,1/2]}, Y_6 = \mathbf{1}_{]1/2,3/4]}, Y_7 = \mathbf{1}_{]3/4,1]},\\
&Y_8 = \mathbf{1}_{]0,1/8]}, Y_9 = \mathbf{1}_{]1/8,1/4]}, \dots\dots\dots\dots\dots\dots\dots\dots, Y_{15} = \mathbf{1}_{]7/8,1]},\\
&Y_{16} = \mathbf{1}_{]0,1/16]}, \dots\dots
\end{aligned}$$

Le lecteur qui ne se satisferait pas de cette définition informelle peut toujours s'exercer à trouver une formule explicite pour Y_n. Sans entrer dans ces détails techniques, on peut facilement se convaincre de deux choses :

1. Pour tout $\omega \in]0,1]$, la suite de « bits » $(Y_n(\omega))_{n\geq 1}$ est formée d'une infinité de 0 et d'une infinité de 1. Elle ne peut donc converger (sa limite inférieure vaut 0 et sa limite supérieure 1). Ainsi non seulement on n'a pas de convergence presque sûre de Y_n, mais en plus $Y_n(\omega)$ ne converge pour *aucun* $\omega \in \Omega$.
2. Pour $0 < \varepsilon < 1$, $\mathbf{P}(|Y_n - 0| > \varepsilon) = \mathbf{P}(Y_n = 1) = \lambda(I_n)$, en notant I_n l'intervalle dyadique dont Y_n est l'indicatrice. La longueur de cet intervalle tend vers zéro quand n tend vers l'infini (à la même vitesse que l'inverse du logarithme en base deux de n). Donc Y_n converge vers 0 en probabilité.

Il est toutefois possible d'obtenir la convergence presque sûre à partir de la convergence en probabilité, à condition d'avoir une *bonne vitesse* de convergence en probabilité. C'est l'objet du résultat suivant.

Proposition 8.4. *Soient Y et $(Y_n)_{n\geq 1}$ des variables aléatoires réelles définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$ et vérifiant*

$$\forall \varepsilon > 0, \quad \sum_{n=1}^{+\infty} \mathbf{P}(|Y_n - Y| \geq \varepsilon) < +\infty. \tag{8.1}$$

Alors Y_n converge presque sûrement vers Y.

La condition (8.1) implique évidemment la convergence en probabilité puisque le terme général de la série doit tendre vers 0. On dit d'une suite $(Y_n)_{n\geq 1}$ vérifiant (8.1) qu'elle converge *presque complètement* vers Y.

Démonstration. Posons $B_{n,\varepsilon} := \{|Y_n - Y| > \varepsilon\}$ et $A_\varepsilon := \limsup B_{n,\varepsilon}$. Par le premier lemme de Borel-Cantelli, on déduit de (8.1) que $\mathbf{P}(A_\varepsilon) = 0$, d'où $\mathbf{P}(A_\varepsilon^c) = 1$. Notons que

$$A_\varepsilon^c = \big\{\omega \in \Omega;\ \exists n_0 = n_0(\omega, \varepsilon),\ \forall n \geq n_0,\ |Y_n(\omega) - Y(\omega)| < \varepsilon\big\}.$$

Discrétisons le ε en posant par exemple $\varepsilon_i = 2^{-i}$ et posons

$$\Omega' := \bigcap_{i\geq 1} A_{\varepsilon_i}^c.$$

Alors Ω' est un évènement de probabilité 1 comme intersection dénombrable d'évènements de probabilité 1. De plus,

$$\forall \omega \in \Omega', \quad \forall i \geq 1, \exists n_0 = n_0(\omega, \varepsilon_i),\ \forall n \geq n_0,\ |Y_n(\omega) - Y(\omega)| < \varepsilon_i\big\}.$$

On en déduit immédiatement que si $\omega \in \Omega'$, $Y_n(\omega)$ converge vers $Y(\omega)$. Comme $\mathbf{P}(\Omega') = 1$, ceci entraîne la convergence presque sûre de Y_n vers Y. $\square$

Corollaire 8.5. *Si la suite $(Y_n)_{n\geq 1}$ converge en probabilité vers Y, on peut en extraire une sous-suite $(Y_{n_i})_{i\geq 1}$ qui converge presque sûrement vers Y.*

Démonstration. Notons encore $\varepsilon_i = 2^{-i}$. La convergence en probabilité de Y_n vers Y implique pour tout $i \geq 1$, la convergence de $\mathbf{P}(|Y_n - Y| > \varepsilon_i)$ vers 0 quand n tend vers l'infini. On en déduit l'existence d'une suite strictement croissante d'indices n_i telle que

$$\forall i \geq 1, \quad \mathbf{P}\big(|Y_{n_i} - Y| \geq \varepsilon_i\big) \leq \frac{1}{i^2}.$$

Vérifions maintenant que

$$\forall \varepsilon > 0, \quad \sum_{i=1}^{+\infty} \mathbf{P}(|Y_{n_i} - Y| \geq \varepsilon) < +\infty.$$

En effet la convergence vers 0 de ε_i nous assure de l'existence d'un $i_0 = i_0(\varepsilon)$ tel que pour tout $i \geq i_0$, $\varepsilon_i < \varepsilon$. Pour $i \geq i_0$, on a donc $\{|Y_{n_i} - Y| \geq \varepsilon\} \subset \{|Y_{n_i} - Y| \geq \varepsilon_i\}$ et cette inclusion d'évènements nous permet de majorer le terme général de la série ci-dessus par i^{-2} à partir du rang i_0. Ainsi la suite (Y_{n_i}) converge presque complètement vers Y donc aussi presque sûrement. $\square$

Le diagramme de la figure 8.1 résume les relations entre les différents modes de convergence. Les flèches en trait plein représentent des implications (si la suite converge selon le mode de la case de départ, alors elle converge aussi selon celui de la case d'arrivée). Les flèches en tirets signifient l'existence d'une sous-suite convergente selon le mode de la case d'arrivée. La convergence en loi sera étudiée ultérieurement.

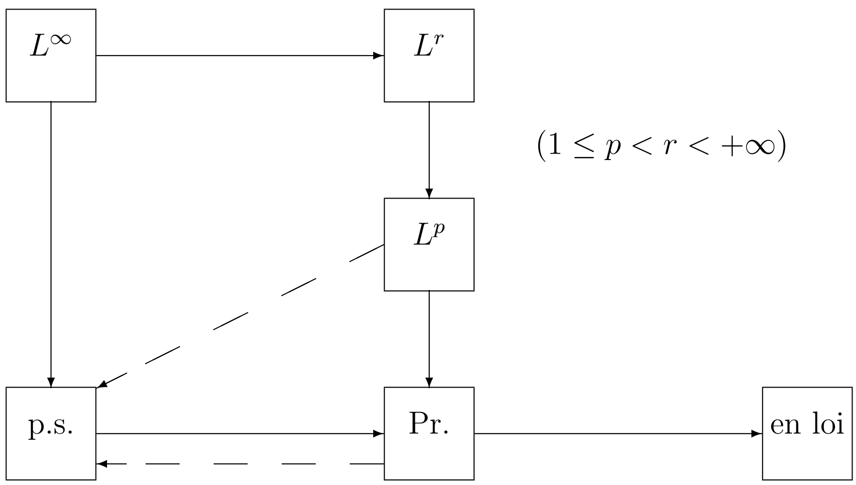

Fig. 8.1 – Diagramme des convergences des suites de v.a.

8.2 Loi faible des grands nombres

Rappelons que si X est de carré intégrable ($\mathbf{E}X^2 < +\infty$), sa variance est définie par

$$\operatorname{Var} X := \mathbf{E}(X - \mathbf{E}X)^2.$$

Si les X_k sont deux à deux *non corrélées* (ce qui a lieu en particulier lorsqu'elles sont indépendantes), on a l'égalité

$$\operatorname{Var} S_n = \sum_{k=1}^{n} \operatorname{Var} X_k. \tag{8.2}$$

Proposition 8.6 (Inégalité de Bienaymé-Tchebycheff). *Si les X_k sont de carré intégrable et deux à deux non corrélées,*

$$\forall t > 0, \quad \mathbf{P}\Big(\Big| \sum_{k=1}^{n} (X_k - \mathbf{E}X_k) \Big| \geq t \Big) \leq \frac{1}{t^2} \sum_{k=1}^{n} \operatorname{Var} X_k. \tag{8.3}$$

Preuve. Il suffit d'écrire :

$$\begin{aligned} \mathbf{P}\big(|S_n - \mathbf{E}S_n| \geq t\big) = \mathbf{P}\big(|S_n - \mathbf{E}S_n|^2 \geq t^2\big) \quad &\leq \quad \frac{1}{t^2}\mathbf{E}(S_n - \mathbf{E}S_n)^2 & (8.4) \\ &= \quad \frac{1}{t^2} \operatorname{Var} S_n, \end{aligned}$$

où l'inégalité dans (8.4) est l'inégalité de Markov appliquée à la variable aléatoire positive $|S_n - \mathbf{E}S_n|^2$. On conclut avec (8.2). □

Théorème 8.7 (Loi faible des G.N.). *Si les X_k sont de même loi, de carré intégrable et deux à deux non corrélées, on a la convergence en probabilité :*

$$\frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow[n \to +\infty]{\text{Pr}} \mathbf{E}X_1. \tag{8.5}$$

Preuve. Comme les X_k ont même loi, on a pour tout k les égalités $\mathbf{E}X_k = \mathbf{E}X_1$ et $\operatorname{Var} X_k = \operatorname{Var} X_1$. Comme elles sont aussi deux à deux non-corrélées, (8.2) nous donne $\operatorname{Var} S_n = n \operatorname{Var} X_1$. Par linéarité de l'espérance on a aussi $\mathbf{E}S_n = n\mathbf{E}X_1$. L'inégalité de Bienaymé-Tchebycheff nous dit alors que :

$$\forall t > 0, \quad \mathbf{P}\big(|S_n - \mathbf{E}S_n| \geq t\big) = \mathbf{P}\big(|S_n - n\mathbf{E}X_1| \geq t\big) \leq \frac{n \operatorname{Var} X_1}{t^2}.$$

Posant $t = n\varepsilon$, on en déduit :

$$\forall \varepsilon > 0,\ \forall n \in \mathbb{N}^*, \quad \mathbf{P}\big(|S_n - n\mathbf{E}X_1| \geq n\varepsilon\big) = \mathbf{P}\Big(\Big| \frac{S_n}{n} - \mathbf{E}X_1 \Big| \geq \varepsilon \Big) \leq \frac{n \operatorname{Var} X_1}{n^2\varepsilon^2} = \frac{\operatorname{Var} X_1}{n\varepsilon^2}.$$

Pour tout $\varepsilon > 0$ fixé, on a ainsi

$$\mathbf{P}\Big(\Big| \frac{S_n}{n} - \mathbf{E}X_1 \Big| \geq \varepsilon \Big) \leq \frac{\operatorname{Var} X_1}{n\varepsilon^2} \xrightarrow[n \to +\infty]{} 0,$$

ce qui traduit exactement la convergence en probabilité de la suite de variables aléatoires S_n/n vers la variable aléatoire constante $\mathbf{E}X_1$. □

8.3 Lois fortes des grands nombres

Théorème 8.8 (Inégalité de Kolmogorov). *Soient $X_1, \dots, X_n$ des variables aléatoires réelles indépendantes, de carré intégrable et d'espérance nulle et $S_k := \sum_{i=1}^k X_i$. Alors*

$$\forall t > 0, \quad \mathbf{P}\Big(\max_{1\le k\le n} |S_k| \ge t\Big) \le \frac{1}{t^2} \operatorname{Var} S_n. \tag{8.6}$$

Remarque 8.9. Si les X_k ne sont pas centrées ($\mathbf{E}X_k \neq 0$), en posant $X'_k := X_k - \mathbf{E}X_k$, en appliquant (8.6) aux X'_k et en notant que $\operatorname{Var} X_k = \operatorname{Var} X'_k$, on obtient la version plus générale de l'inégalité de Kolmogorov :

$$\forall t > 0, \quad \mathbf{P}\Big(\max_{1\le k\le n} |S_k - \mathbf{E}S_k| \ge t\Big) \le \frac{1}{t^2} \operatorname{Var} S_n. \tag{8.7}$$

Remarque 8.10. L'inégalité de Kolmogorov ressemble formellement à celle de Bienaymé-Tchebycheff. Elle est cependant beaucoup plus puissante, puisqu'elle permet de contrôler en probabilité les déviations de *toute la suite* finie $(S_k - \mathbf{E}S_k)_{1\le k\le n}$ au lieu de seulement son dernier terme $S_n - \mathbf{E}S_n$ pour Bienaymé-Tchebycheff. L'hypothèse est aussi plus restrictive (indépendance mutuelle au lieu de non-corrélation deux à deux).

Preuve du théorème 8.8. Définissons les évènements

$$A_1 := \{|S_1| \ge t\} = \{|X_1| \ge t\} \quad \text{et} \quad A_k := \{|S_k| \ge t\} \cap \Big(\bigcap_{j=1}^{k-1} \{|S_j| < t\}\Big) \quad (2 \le k \le n).$$

Autrement dit $\omega \in A_k$ (l'évènement élémentaire ω réalise l'évènement A_k) si et seulement si la suite finie $\big(|S_j(\omega)|\big)_{1\le j\le n}$ atteint ou dépasse t la première fois pour $j = k$. Avec cette interprétation, il est clair que les A_k sont deux à deux disjoints et que :

$$A := \bigcup_{k=1}^n A_k = \big\{\max_{1\le k\le n} |S_k| \ge t\big\}. \tag{8.8}$$

On en déduit par croissance et additivité de l'intégrale (ou de la mesure de densité S_n^2 par rapport à $\mathbf{P}$) la minoration :

$$\mathbf{E}S_n^2 = \int_\Omega S_n^2 \,\mathrm{d}\mathbf{P} \ge \int_A S_n^2 \,\mathrm{d}\mathbf{P} = \sum_{k=1}^n \int_{A_k} S_n^2 \,\mathrm{d}\mathbf{P}. \tag{8.9}$$

En intégrant sur A_k l'inégalité élémentaire

$$S_n^2 = \big[S_k + (S_n - S_k)\big]^2 \ge S_k^2 + 2S_k(S_n - S_k),$$

on obtient

$$\int_{A_k} S_n^2 \,\mathrm{d}\mathbf{P} \ge \int_{A_k} S_k^2 \,\mathrm{d}\mathbf{P} + 2\int_{A_k} S_k(S_n - S_k) \,\mathrm{d}\mathbf{P}. \tag{8.10}$$

On remarque alors que $\int_{A_k} S_k(S_n - S_k) \,\mathrm{d}\mathbf{P}$ peut s'écrire $\mathbf{E}(YZ)$ avec $Y := \mathbf{1}_{A_k} S_k$ et $Z := S_n - S_k$. Ces deux variables aléatoires sont de carré intégrable comme combinaisons linéaires de variables aléatoires de carré intégrable. La première Y est une fonction mesurable du seul vecteur aléatoire $(X_1, \dots, X_k)$ tandis que Z est fonction mesurable du seul vecteur aléatoire $(X_{k+1}, \dots, X_n)$. Par conséquent Y et Z sont indépendantes (corollaire 5.41) et $\mathbf{E}(YZ) = \mathbf{E}Y\mathbf{E}Z$ (proposition 5.43). Or

$$\mathbf{E}Z = \mathbf{E}(S_n - S_k) = \mathbf{E} \sum_{j=k+1}^{n} X_j = \sum_{j=k+1}^{n} \mathbf{E}X_j = 0,$$

d'où $\int_{A_k} S_k(S_n - S_k)\,\mathrm{d}\mathbf{P} = 0$ (résultat évident directement dans le cas particulier où $k = n$). En reportant ceci dans (8.10), on en déduit compte tenu de (8.9) :

$$\operatorname{Var} S_n = \mathbf{E}S_n^2 \geq \sum_{k=1}^{n} \int_{A_k} S_k^2\,\mathrm{d}\mathbf{P}. \tag{8.11}$$

Vu la définition de A_k, on a sur cet évènement $|S_k| \geq t$, d'où $\int_{A_k} S_k^2\,\mathrm{d}\mathbf{P} \geq t^2\mathbf{P}(A_k)$. En injectant cette minoration dans (8.11), en utilisant (8.8) et en rappelant que les A_k sont deux à deux disjoints, on aboutit à :

$$\operatorname{Var} S_n \geq \sum_{k=1}^{n} t^2\mathbf{P}(A_k) = t^2\mathbf{P}\Big(\bigcup_{k=1}^{n} A_k\Big) = t^2\mathbf{P}(A).$$

L'inégalité $\mathbf{P}(A) \leq t^{-2} \operatorname{Var} S_n$ ainsi établie est exactement (8.6). □

Théorème 8.11 (Loi forte des grands nombres de Kolmogorov). *On suppose que $(X_k)_{k\geq 1}$ est une suite de variables aléatoires réelles indépendantes, de carré intégrable, d'espérance nulle et que*

$$\sum_{j=1}^{+\infty} \frac{\operatorname{Var} X_j}{j^2} < +\infty. \tag{8.12}$$

Alors

$$\frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow[n\to+\infty]{\text{p.s.}} 0. \tag{8.13}$$

Remarque 8.12. Si les X_k ne sont pas centrées, en appliquant (8.13) aux $X'_k := X_k - \mathbf{E}X_k$, on obtient

$$\frac{1}{n}S_n - \frac{1}{n}\sum_{k=1}^{n} \mathbf{E}X_k \xrightarrow[n\to+\infty]{\text{p.s.}} 0. \tag{8.14}$$

Ce résultat est intéressant notamment dans le cas où la suite déterministe $(\mathbf{E}X_n)_{n\geq 1}$ a une limite finie ℓ quand n tend vers l'infini. En effet le théorème de Césaro nous donne alors la convergence vers ℓ de $n^{-1}\sum_{k=1}^{n} \mathbf{E}X_k$, d'où la convergence presque sûre de $n^{-1}S_n$ vers ℓ. En particulier si les X_k ont même espérance, $n^{-1}S_n$ converge p.s. vers $\mathbf{E}X_1$.

Démonstration du théorème 8.11. Une des difficultés lorsque l'on veut établir une convergence presque sûre est la gestion du « $\forall\varepsilon > 0$ » qui induit une intersection non dénombrable. Nous allons procéder en trois étapes, la première étant consacrée à la résolution de ce problème.

Première étape : réduction à un $\varepsilon > 0$ fixé. Supposons déjà prouvé que pour tout $\varepsilon > 0$,

$$\mathbf{P}\Big(\limsup_{n\to+\infty}\Big\{\Big|\frac{S_n}{n}\Big| \geq \varepsilon\Big\}\Big) = 0, \tag{8.15}$$

autrement dit qu'il y a une probabilité nulle que la suite (S_n/n) ait une infinité de termes hors de l'intervalle $]-\varepsilon, +\varepsilon[$. L'évènement contraire Ω_ε de cette limite supérieure s'énonce « à partir d'un certain rang (aléatoire) fini, plus aucun terme de la suite (S_n/n) n'est hors de $]-\varepsilon, +\varepsilon[$ », ce qui s'écrit encore :

$$\Omega_\varepsilon := \Big(\limsup_{n\to+\infty}\Big\{\Big|\frac{S_n}{n}\Big| \geq \varepsilon\Big\}\Big)^c = \Big\{\omega \in \Omega;\ \exists n_0 = n_0(\omega,\varepsilon),\ \forall n \geq n_0,\ \Big|\frac{S_n(\omega)}{n}\Big| < \varepsilon\Big\}.$$

D'après (8.15), on a $\mathbf{P}(\Omega_\varepsilon) = 1$. Discrétisons alors le ε en le remplaçant par une suite $(\varepsilon_i) \downarrow 0$, disons $\varepsilon_i := 2^{-i}$. Posons

$$\Omega' := \bigcap_{i\in\mathbb{N}} \Omega_{\varepsilon_i}.$$

Cet évènement est de probabilité 1 puisque :

$$\mathbf{P}(\Omega'^c) = \mathbf{P}\Big(\bigcup_{i\in\mathbb{N}} \Omega_{\varepsilon_i}^c\Big) \leq \sum_{i\in\mathbb{N}} \mathbf{P}(\Omega_{\varepsilon_i}^c) = 0.$$

Or par définition de Ω', tout $\omega \in \Omega'$ appartient à *chacun* des Ω_{ε_i}, autrement dit,

$$\forall \omega \in \Omega',\ \forall i \in \mathbb{N},\ \exists n_0 = n_0(\omega,\varepsilon_i),\ \forall n \geq n_0,\quad \Big|\frac{S_n(\omega)}{n}\Big| < \varepsilon_i.$$

On en déduit que pour tout $\omega \in \Omega'$, $n^{-1}S_n(\omega)$ tend vers zéro quand n tend vers $+\infty$. Donc Ω' est inclus dans l'évènement $\{n^{-1}S_n \to 0\}$ et comme $\mathbf{P}(\Omega') = 1$, $\mathbf{P}(n^{-1}S_n \to 0) = 1$, ce qui signifie la convergence presque sûre de $n^{-1}S_n$ vers zéro[1]. Ainsi le théorème sera démontré si on prouve que pour un $\varepsilon > 0$ quelconque fixé, on a l'égalité (8.15). □

Deuxième étape : réduction de (8.15) à une condition de Borel-Cantelli. On travaille désormais avec un $\varepsilon > 0$ fixé. On va se ramener au premier lemme de Borel-Cantelli en utilisant une partition de $\mathbb{N}^*$ en « octaves » $[2^k, 2^{k+1}[$. On remarque en effet que l'inégalité $n^{-1}|S_n(\omega)| \geq \varepsilon$ se réalise pour une infinité d'indices n si et seulement si ω appartient pour une infinité d'indices k, à l'évènement B_k défini par

$$B_k := \big\{\omega \in \Omega;\ \exists n = n(\omega) \in [2^k, 2^{k+1}[,\ |S_n(\omega)| \geq n\varepsilon\big\},\quad k \in \mathbb{N}.$$

Cette équivalence se traduit par l'égalité d'évènements :

$$\limsup_{n\to+\infty}\Big\{\Big|\frac{S_n}{n}\Big| \geq \varepsilon\Big\} = \limsup_{k\to+\infty} B_k.$$

1. En écrivant $\mathbf{P}(n^{-1}S_n \to 0)$, on a admis implicitement que l'ensemble $\Omega'' := \{\omega \in \Omega;\ n^{-1}S_n(\omega) \to 0\}$ est bien un *évènement*, c'est à dire appartient à la tribu $\mathcal{F}$. Une façon rapide de le justifier est d'utiliser la mesurabilité de la variable aléatoire $Y := \limsup_{n\to+\infty} |n^{-1}S_n|$ et l'équivalence entre $n^{-1}S_n \to 0$ et $Y = 0$, d'où $\Omega'' = Y^{-1}(\{0\})$.

Cette égalité et le premier lemme de Borel-Cantelli montrent que pour établir (8.15), il suffit de prouver la convergence de série :

$$\sum_{k=0}^{+\infty} \mathbf{P}(B_k) < +\infty. \tag{8.16}$$

$\square$

Troisième étape : preuve de (8.16) par l'inégalité de Kolmogorov. De l'écriture de l'évènement B_k sous la forme

$$B_k = \bigcup_{2^k \le n < 2^{k+1}} \{|S_n| \ge n\varepsilon\},$$

on déduit immédiatement les inclusions :

$$B_k \subset \bigcup_{2^k \le n < 2^{k+1}} \{|S_n| \ge 2^k\varepsilon\} = \Big\{ \max_{2^k \le n < 2^{k+1}} |S_n| \ge 2^k\varepsilon \Big\} \subset \Big\{ \max_{1 \le j \le 2^{k+1}} |S_j| \ge 2^k\varepsilon \Big\}.$$

Ces inclusions nous permettent de majorer $\mathbf{P}(B_k)$ grâce à l'inégalité de Kolmogorov (c'est seulement maintenant que nous avons besoin de l'indépendance des X_j et de leur appartenance à $\mathrm{L}^2(\mathbf{P})$) :

$$\mathbf{P}(B_k) \le \mathbf{P}\Big(\max_{1 \le j \le 2^{k+1}} |S_j| \ge 2^k\varepsilon \Big) \le \frac{1}{\varepsilon^2 2^{2k}} \operatorname{Var}\big(S_{2^{k+1}}\big) = \frac{1}{\varepsilon^2 2^{2k}} \sum_{j=1}^{2^{k+1}} \operatorname{Var} X_j.$$

Cette inégalité nous conduit à la sommation triangulaire suivante pour contrôler la somme de la série des $\mathbf{P}(B_k)$.

$$\begin{aligned}
\sum_{k=0}^{+\infty} \mathbf{P}(B_k) &\le \frac{1}{\varepsilon^2} \sum_{k=0}^{+\infty} \frac{1}{2^{2k}} \sum_{j=1}^{2^{k+1}} \operatorname{Var} X_j \\
&= \frac{1}{\varepsilon^2} \sum_{k=0}^{+\infty} \sum_{j=1}^{+\infty} \frac{1}{2^{2k}} \operatorname{Var} X_j \, \mathbf{1}_{\{1 \le j \le 2^{k+1}\}} \\
&= \frac{1}{\varepsilon^2} \sum_{j=1}^{+\infty} \operatorname{Var} X_j \sum_{k=0}^{+\infty} \frac{1}{2^{2k}} \mathbf{1}_{\{2^{k+1} \ge j\}} \\
&= \frac{1}{\varepsilon^2} \sum_{j=1}^{+\infty} \operatorname{Var} X_j \sum_{k=m_j}^{+\infty} \frac{1}{2^{2k}},
\end{aligned}$$

où l'on a posé $m_j := \min\big\{k \in \mathbb{N};\ 2^{k+1} \ge j\big\}$. Un simple calcul de série géométrique nous donne :

$$\sum_{k=m_j}^{+\infty} \frac{1}{2^{2k}} = \sum_{k=m_j}^{+\infty} \frac{1}{4^k} = \frac{1}{4^{m_j}} \sum_{l=0}^{+\infty} \frac{1}{4^l} = \frac{1}{4^{m_j}} \frac{1}{1 - \frac{1}{4}} = \frac{4}{3 \times 4^{m_j}}.$$

De la définition de m_j, on tire les inégalités $2^{m_j} \geq j/2$, puis $4^{m_j} \geq j^2/4$, d'où

$$\frac{4}{3 \times 4^{m_j}} \leq \frac{16}{3j^2}.$$

Finalement,

$$\sum_{k=0}^{+\infty} \mathbf{P}(B_k) \leq \frac{16}{3\varepsilon^2} \sum_{j=1}^{+\infty} \frac{\operatorname{Var} X_j}{j^2},$$

ce qui établit (8.16) grâce à l'hypothèse (8.12). □

Par la deuxième étape, on en déduit (8.15) et comme ε était quelconque, ceci entraîne la convergence presque sûre vers 0 de S_n/n, d'après la première étape. Le théorème est maintenant complètement démontré. □

Théorème 8.13 (Loi forte des grands nombres de Khintchine). *On suppose les X_k indépendantes,* de même loi *et* $\mathbf{E}|X_1| < +\infty$. *Alors*

$$\frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow[n \to +\infty]{\text{p.s.}} \mathbf{E}X_1. \tag{8.17}$$

Démonstration. Quitte à remplacer X_k par $X'_k := X_k - \mathbf{E}X_k = X_k - \mathbf{E}X_1$, il suffit de faire la preuve dans le cas où $\mathbf{E}X_1 = 0$. On utilise une méthode de *troncature* permettant de remplacer les X_k par des variables aléatoires *bornées* Y_k (les bornes dépendant de k), donc ayant une variance et relevant ainsi de la loi forte des grands nombres de Kolmogorov. On pose pour cela

$$Y_k(\omega) := \begin{cases} X_k(\omega) & \text{si } |X_k(\omega)| \leq k, \\ 0 & \text{si } |X_k(\omega)| > k. \end{cases}$$

Autrement dit,

$$Y_k = X_k \mathbf{1}_{\{|X_k| \leq k\}}.$$

Bien que les X_k aient même loi, il n'en va pas de même des Y_k, car la troncature dépend de k. Remarquons cependant que

$$m_k := \mathbf{E}\big(X_k \mathbf{1}_{\{|X_k| \leq k\}}\big) = \mathbf{E}\big(X_1 \mathbf{1}_{\{|X_1| \leq k\}}\big) \xrightarrow[k \to +\infty]{} \mathbf{E}X_1, \tag{8.18}$$

par convergence dominée, grâce à l'hypothèse $\mathbf{E}|X_1| < +\infty$.

Première étape : réduction à la l.f.g.n. pour (Y_k). Nous commençons par montrer que si $\mathbf{E}X_1 = 0$, (8.17) est impliquée par :

$$\frac{1}{n} \sum_{k=1}^{n} Y_k \xrightarrow[n \to +\infty]{\text{p.s.}} 0. \tag{8.19}$$

L'idée est d'utiliser le premier lemme de Borel-Cantelli pour prouver que les suites $(X_k)_{k \geq 1}$ et $(Y_k)_{k \geq 1}$ ont presque sûrement tous leurs termes égaux à partir d'un certain rang (aléatoire). Les moyennes arithmétiques auront alors p.s. le même comportement asymptotique. Ceci nous amène à étudier la convergence de la série $\sum_{k=1}^{+\infty} \mathbf{P}(X_k \neq Y_k)$. On note d'abord que

$$\sum_{k=1}^{+\infty} \mathbf{P}(X_k \neq Y_k) = \sum_{k=1}^{+\infty} \mathbf{P}(|X_k| > k) = \sum_{k=1}^{+\infty} \mathbf{P}(|X_1| > k),$$

la deuxième égalité reposant sur l'équidistribution [2] des X_k. Pour comparer cette dernière série à une intégrale, posons $g(t) := \mathbf{P}(|X_1| > t)$, ce qui définit une fonction décroissante $g : \mathbb{R}_+ \to [0, 1]$. Cette fonction est clairement λ-intégrable sur tout intervalle borné de $\mathbb{R}_+$ et on a :

$$\forall k \in \mathbb{N}^*, \quad \int_{[k-1,k[} g(t)\,\mathrm{d}\lambda(t) \geq \int_{[k-1,k[} g(k)\,\mathrm{d}\lambda = g(k)\lambda([k-1,k[) = g(k).$$

On en déduit la majoration :

$$\sum_{k=1}^{+\infty} \mathbf{P}(X_k \neq Y_k) = \sum_{k=1}^{+\infty} g(k) \leq \sum_{k=1}^{+\infty} \int_{[k-1,k[} g(t)\,\mathrm{d}\lambda(t) = \int_{[0,+\infty[} \mathbf{P}(|X_1| > t)\,\mathrm{d}\lambda(t).$$

Or d'après la proposition 5.10 appliquée à la variable aléatoire positive $|X_1|$,

$$\int_{[0,+\infty[} \mathbf{P}(|X_1| > t)\,\mathrm{d}\lambda(t) = \mathbf{E}|X_1|.$$

Ainsi la convergence

$$\sum_{k=1}^{+\infty} \mathbf{P}(X_k \neq Y_k) < +\infty \tag{8.20}$$

résulte de l'hypothèse $\mathbf{E}|X_1| < +\infty$.

On sait maintenant que l'évènement

$$A := \big\{\omega \in \Omega;\ \exists k_0 = k_0(\omega),\ \forall k \geq k_0,\ X_k(\omega) = Y_k(\omega)\big\}$$

a pour probabilité 1, puisque (8.20) et le premier lemme de Borel-Cantelli nous donnent :

$$\mathbf{P}(A^c) = \mathbf{P}\big(\{\omega \in \Omega; X_k(\omega) \neq Y_k(\omega) \text{ pour une infinité d'indices } k\}\big) = 0.$$

Si $\omega \in A$, pour tout $n \geq k_0(\omega)$,

$$\frac{1}{n}\sum_{k=1}^{n} X_k(\omega) - \frac{1}{n}\sum_{k=1}^{n} Y_k(\omega) = \frac{1}{n}\underbrace{\left[\sum_{k=1}^{k_0(\omega)} X_k(\omega) - \sum_{k=1}^{k_0(\omega)} Y_k(\omega)\right]}_{\text{ne dépend pas de } n},$$

2. L'équidistribution d'une suite de variables aléatoires X_k est la propriété : « les X_k ont même loi ».

d'où la convergence vers 0 de cette différence des moyennes arithmétiques. Ainsi on a l'inclusion d'évènements :

$$A \subset \left\{ \frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} Y_k \xrightarrow[n\to+\infty]{} 0 \right\}$$

et comme $P(A) = 1$, on en déduit la convergence presque sûre :

$$\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} Y_k \xrightarrow[n\to+\infty]{\text{p.s.}} 0.$$

Il est maintenant clair que (8.17) est impliquée par (8.19). □

Deuxième étape : réduction de la preuve de (8.19). Centrons les variables Y_k en les remplaçant par $Y'_k := Y_k - \mathbf{E}Y_k = Y_k - m_k$. Il suffit pour établir (8.19) de prouver que

$$\frac{1}{n}\sum_{k=1}^{n} Y'_k \xrightarrow[n\to+\infty]{\text{p.s.}} 0. \tag{8.21}$$

En effet $n^{-1}\sum_{k=1}^{n} Y_k = n^{-1}\sum_{k=1}^{n} Y'_k + n^{-1}\sum_{k=1}^{n} m_k$ et par (8.18) et le théorème de Césaro,

$$\frac{1}{n}\sum_{k=1}^{n} m_k \xrightarrow[n\to+\infty]{} \mathbf{E}X_1 = 0.$$

□

Troisième étape : preuve de (8.21). Nous allons établir (8.21) en appliquant aux Y'_k la loi forte des grands nombres de Kolmogorov. Pour cela, il nous faut vérifier que

$$\sum_{k=1}^{+\infty} \frac{\operatorname{Var} Y'_k}{k^2} < +\infty. \tag{8.22}$$

Comme $\operatorname{Var} Y'_k = \operatorname{Var} Y_k$ et $\operatorname{Var} Y_k = \mathbf{E}Y_k^2 - m_k^2$, on voit que

$$\operatorname{Var} Y'_k \leq \mathbf{E}Y_k^2 = \int_{\{|X_k|\leq k\}} X_k^2 \,\mathrm{d}\mathbf{P}.$$

Comme X_k et X_1 ont même loi, on a par transferts :

$$\int_{\{|X_k|\leq k\}} X_k^2 \,\mathrm{d}\mathbf{P} = \int_{[-k,k]} x^2 \,\mathrm{d}P_{X_k}(x) = \int_{[-k,k]} x^2 \,\mathrm{d}P_{X_1}(x) = \int_{\{|X_1|\leq k\}} X_1^2 \,\mathrm{d}\mathbf{P}.$$

Nous pouvons maintenant écrire les majorations suivantes.

$$\begin{aligned}
\sum_{k=1}^{+\infty} \frac{\operatorname{Var} Y'_k}{k^2} &\leq \sum_{k=1}^{+\infty} \frac{1}{k^2} \int_{\{|X_1| \leq k\}} X_1^2 \,\mathrm{d}\mathbf{P} \\
&= \sum_{k=1}^{+\infty} \frac{1}{k^2} \sum_{i=1}^{k} \int_{\{i-1<|X_1|\leq i\}} X_1^2 \,\mathrm{d}\mathbf{P} \quad (\text{sur } \{i-1<|X_1|\leq i\},\ X_1^2 \leq i|X_1|) \\
&\leq \sum_{k=1}^{+\infty} \frac{1}{k^2} \sum_{i=1}^{k} \int_{\{i-1<|X_1|\leq i\}} i|X_1| \,\mathrm{d}\mathbf{P} \\
&= \sum_{k=1}^{+\infty} \sum_{i=1}^{+\infty} \frac{i}{k^2} \mathbf{1}_{\{1\leq i\leq k\}} \int_{\{i-1<|X_1|\leq i\}} |X_1| \,\mathrm{d}\mathbf{P} \\
&= \sum_{i=1}^{+\infty} i \left(\sum_{k=i}^{+\infty} \frac{1}{k^2} \right) \int_{\{i-1<|X_1|\leq i\}} |X_1| \,\mathrm{d}\mathbf{P}.
\end{aligned}$$

Pour achever la majoration, une classique comparaison série-intégrale nous donne

$$\forall i \geq 1, \quad \sum_{k=i}^{+\infty} \frac{1}{k^2} = \frac{1}{i^2} + \sum_{k=i+1}^{+\infty} \frac{1}{k^2} \leq \frac{1}{i^2} + \int_i^{+\infty} \frac{\mathrm{d}x}{x^2} = \frac{1}{i^2} + \frac{1}{i} \leq \frac{2}{i}.$$

Finalement,

$$\sum_{k=1}^{+\infty} \frac{\operatorname{Var} Y'_k}{k^2} \leq 2 \sum_{i=1}^{+\infty} \int_{\{i-1<|X_1|\leq i\}} |X_1| \,\mathrm{d}\mathbf{P} = 2\mathbf{E}|X_1| < +\infty,$$

ce qui établit (8.22) et donc aussi (8.21) par la loi forte des grands nombres de Kolmogorov. □

La loi forte des grands nombres de Khintchine est maintenant complètement démontrée. □

L'hypothèse d'intégrabilité de X_1 dans la loi forte des grands nombres de Khintchine est optimale. Plus précisément, dans le cas d'une suite (X_k) i.i.d., on a équivalence entre la convergence presque sûre de S_n/n et l'intégrabilité de X_1. Cette équivalence résulte du théorème 8.13 et du théorème suivant.

Théorème 8.14. *Soit $(X_k)_{k\geq 1}$ une suite de variables aléatoires indépendantes et de même loi telle que S_n/n converge presque sûrement. Alors $\mathbf{E}|X_1| < +\infty$ et la limite* p.s. *de S_n/n est la constante $\mathbf{E}X_1$.*

Démonstration. Nous exposons la preuve sous l'hypothèse supplémentaire que la limite p.s. de S_n/n est une *constante* c. La loi du zéro-un de Kolmogorov (qui figure au programme de Maîtrise) montre que si S_n/n converge p.s., sa limite est nécessairement une constante. Alors

$$\frac{X_n}{n} = \frac{S_n - S_{n-1}}{n} = \frac{S_n}{n} - \frac{n-1}{n} \times \frac{S_{n-1}}{n-1} \xrightarrow[n\to+\infty]{\text{p.s.}} c - c = 0.$$

En fixant $\varepsilon > 0$, on en déduit

$$\mathbf{P}\Big\{\omega \in \Omega; \exists n_0 = n_0(\omega), \forall n \geq n_0, \frac{|X_n(\omega)|}{n} < \varepsilon\Big\} = 1,$$

soit en passant au complémentaire

$$\mathbf{P}\Big\{\frac{|X_n|}{n} \geq \varepsilon \text{ une infinité de fois}\Big\} = 0.$$

Par le second lemme de Borel-Cantelli (lemme 5.37), on a alors

$$\sum_{n\geq 1} \mathbf{P}\Big\{\frac{|X_n|}{n} \geq \varepsilon\Big\} < +\infty.$$

Les X_i ayant même loi, ceci s'écrit

$$\sum_{n=1}^{+\infty} \mathbf{P}(|X_1| \geq n\varepsilon) < +\infty. \tag{8.23}$$

Pour finir la preuve, on observe que

$$\begin{aligned}
\mathbf{E}|X_1| &\leq \mathbf{E}\Big(\sum_{n=0}^{+\infty}(n+1)\varepsilon \mathbf{1}_{\{n\varepsilon \leq |X_1| < (n+1)\varepsilon\}}\Big) \\
&= \sum_{n=0}^{+\infty}(n+1)\varepsilon \mathbf{P}(n\varepsilon \leq |X_1| < (n+1)\varepsilon) \\
&= \varepsilon \sum_{n=0}^{+\infty} \mathbf{P}(n\varepsilon \leq |X_1|) < +\infty,
\end{aligned}$$

la dernière ligne s'obtenant par sommation triangulaire à partir des décompositions en unions disjointes

$$\{|X_1| \geq n\varepsilon\} = \bigcup_{k\geq n}\{k\varepsilon \leq |X_1| < (k+1)\varepsilon\}.$$

□

8.4 Quelques applications des lois des grands nombres

Nous présentons dans cette section quelques applications de la loi forte des grands nombres qui concernent essentiellement la statistique. Nous commençons par le cas facile des variables aléatoires indépendantes, de même loi et bornées. Elles sont donc automatiquement intégrables et vérifient la loi forte des grands nombres de Khintchine.

8.4.1 Applications de la LFGN pour des v.a. bornées

L'application la plus simple et aussi une des plus importantes du théorème 8.13 est la convergence des fréquences de succès dans une suite d'épreuves répétées de Bernoulli indépendantes. Ce résultat explique *a posteriori* l'approche fréquentiste dans la définition d'une probabilité. En effet si $(X_k)_{k\geq 1}$ est une suite de variables aléatoires de Bernoulli indépendantes et de même paramètre p, le théorème 8.13 nous donne la convergence presque sûre de $n^{-1}S_n$ vers $\mathbf{E}X_1 = p$. La figure 8.2 montre une simulation réalisée en Scilab avec $n = 1000$ et $p = 0.6$. Soit maintenant $A \in \mathfrak{F}$ un évènement et $(A_k)_{k\geq 1}$ une suite d'évènements indépendants de même probabilité que A. En prenant $X_k = \mathbf{1}_{A_k}$ et en notant que $\mathbf{E}\mathbf{1}_{A_k} = \mathbf{P}(A_k) = \mathbf{P}(A)$, on obtient

$$\frac{1}{n}\sum_{k=1}^{n}\mathbf{1}_{A_k} \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbf{P}(A).$$

Par exemple si A est l'évènement « obtention d'un double lors du lancer d'une paire de dés équilibrés », ceci nous dit que la *fréquence* d'obtention d'un double en n lancers converge presque sûrement vers $1/6$ lorsque n tend vers l'infini. À titre d'exemple historique, on peut mentionner le problème de l'aiguille de Buffon (cf. D.M. n° 4, Annales IFP 2002–03).

Le théorème 8.13 a une traduction statistique fondamentale : il permet de justifier la convergence de la fonction de répartition empirique. Considérons une suite (Y_k) de variables aléatoires indépendantes et de même loi de fonction de répartition F. On définit la *fonction de répartition empirique* F_n construite sur l'*échantillon* $Y_1, \dots, Y_n$ par

$$F_n(x) := \frac{1}{n}\sum_{k=1}^{n}\mathbf{1}_{\{Y_k\leq x\}}, \quad x \in \mathbb{R}. \tag{8.24}$$

Le théorème 8.13 appliqué aux variables aléatoires bornées $X_k = \mathbf{1}_{\{Y_k\leq x\}}$ nous donne immédiatement pour tout $x \in \mathbb{R}$ la convergence presque sûre de $F_n(x)$ vers $F(x)$, en remarquant que $\mathbf{E}X_1 = P(Y_1 \leq x) = F(x)$. Ainsi une *loi inconnue* peut être reconstituée approximativement à partir de l'observation d'un échantillon de grande taille. En fait, on peut obtenir mieux que la convergence simple presque sûre de F_n vers F.

Théorème 8.15 (Glivenko-Cantelli). *Soit (Y_k) une suite de variables aléatoires indépendantes, de même loi et (F_n) la suite de fonctions de répartition empiriques associées. Alors*

$$\|F_n - F\|_\infty := \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \xrightarrow[n\to+\infty]{\text{p.s.}} 0. \tag{8.25}$$

Ce théorème sera démontré en cours de Maîtrise.

La LFGN pour des variables aléatoires bornées donne aussi immédiatement la convergence presque sûre des fonctions caractéristiques empiriques.

Proposition 8.16. *Soit (Y_k) une suite de vecteurs aléatoires dans $\mathbb{R}^d$, indépendants et de même loi de fonction caractéristique φ définie par $\varphi(u) := \mathbf{E}\exp(i\langle u, Y_1\rangle)$, $u \in \mathbb{R}^d$.*

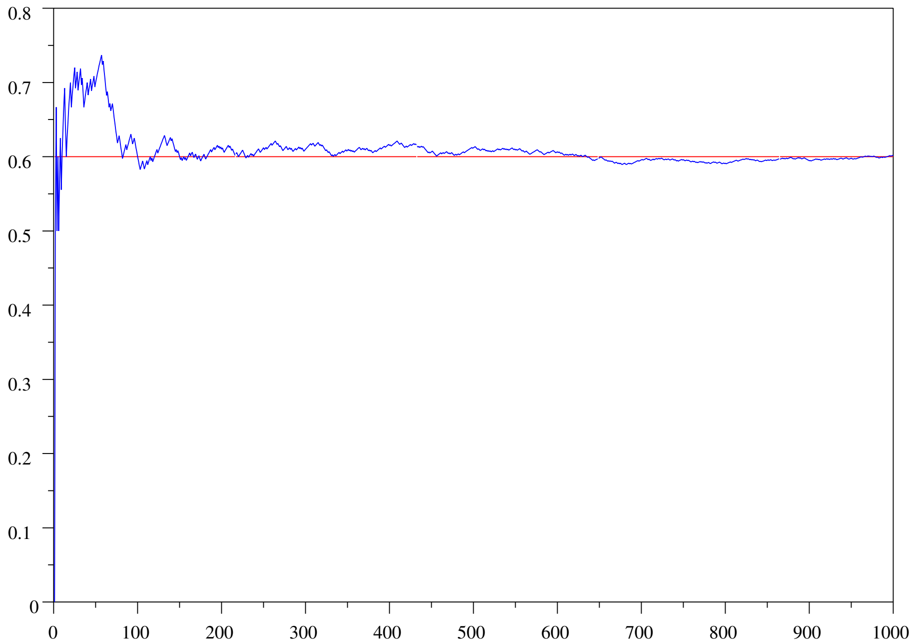

FIG. 8.2 – Ligne polygonale de sommets $(k, S_k/k)$, $1 \leq k \leq 1000$, $X_k \sim \text{Bern}(0.6)$.

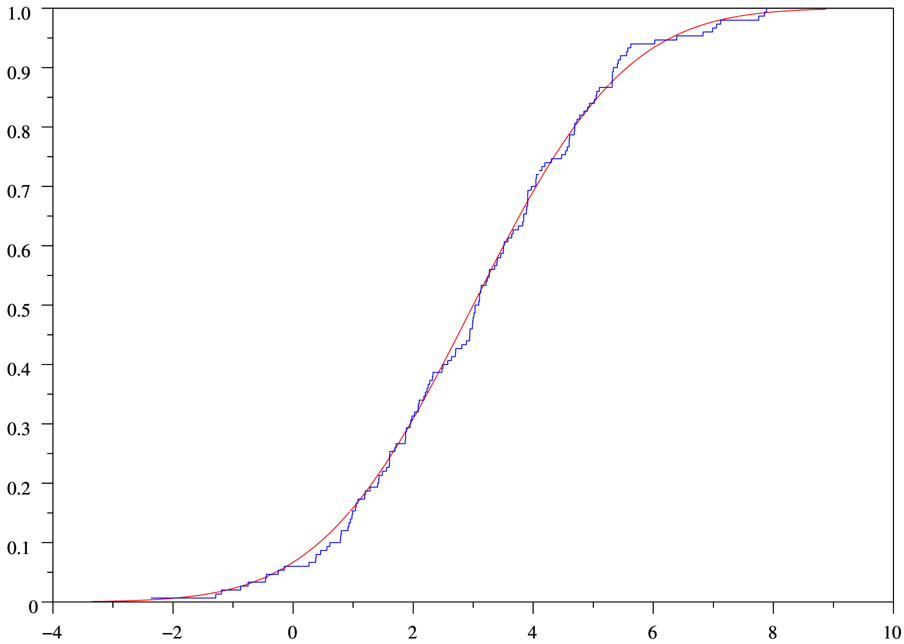

FIG. 8.3 – F_{150} et F pour un échantillon de la loi $\mathfrak{N}(3, 2)$.

Alors la fonction caractérisitique empirique

$$\varphi_n(u) := \frac{1}{n}\sum_{k=1}^{n} \exp(i\langle u, Y_k\rangle)$$

converge ponctuellement presque sûrement sur $\mathbb{R}^d$ *vers* φ.

Vérification. Il suffit d'appliquer le théorème 8.13 aux variables aléatoires réelles $X'_k = \cos(\langle u, Y_k\rangle)$ et $X''_k = \sin(\langle u, Y_k\rangle)$. □

8.4.2 La méthode de Monte Carlo

La loi des grands nombres fournit une méthode de calcul approché d'intégrales, intéressante lorsque la fonction à intégrer est très irrégulière ou lorsque la dimension de l'espace est élevée. Supposons que l'on veuille effectuer un calcul approché de

$$I := \int_{[0,1]^d} f(x)\,\mathrm{d}x,$$

où f est Lebesgue intégrable sur $[0,1]^d$. Soit $(U_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes de même loi uniforme sur $[0,1]$. On déduit facilement de la LFGN de Kolmogorov-Khintchine que :

$$\frac{1}{n}\sum_{k=1}^{n} f\big(U_{(k-1)d+1}, U_{(k-1)d+2}, \dots, U_{kd}\big) \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbf{E} f(U_1,\dots,U_d) = \int_{[0,1]^d} f(x)\,\mathrm{d}x.$$

Le théorème limite central permet ensuite d'obtenir un *intervalle de confiance* pour I si l'on a des hypothèses supplémentaires permettant de contrôler la variance de $f(U_1,\dots,U_d)$, par exemple f bornée...

8.4.3 Estimation de paramètres

La LFGN permet de définir des estimateurs convergents de paramètres d'une loi inconnue μ (ou partiellement inconnue, par exemple on sait qu'il s'agit d'une loi de Poisson de paramètre α dont on ignore la valeur). Pour cela on utilise une suite d'observations indépendantes $X_1(\omega),\dots,X_n(\omega)$ où les X_i sont i.i.d. de même loi μ. On souhaite estimer un paramètre θ de la forme $\theta = \int_{\mathbb{R}} H\,\mathrm{d}\mu$. L'idée est de remplacer la mesure déterministe mais inconnue μ par la mesure *aléatoire* μ_n calculable à partir des observations :

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}.$$

Cette mesure est appelée *mesure empirique*. La fonction de répartition empirique déjà vue en (8.24) est simplement sa fonction de répartition : $F_n(x) = \mu_n(]-\infty, x])$. On propose d'estimer θ par

$$\widehat{\theta}_n := \int_{\mathbb{R}} H\,\mathrm{d}\mu_n = \frac{1}{n}\sum_{i=1}^{n} H(X_i).$$

La définition de θ suppose implicitement que H est μ intégrable. Cette intégrabilité s'écrit encore $\mathbf{E}|H(X_1)| < +\infty$. Ainsi par la loi forte des grands nombres,

$$\widehat{\theta}_n \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbf{E}H(X_1) = \int_{\mathbb{R}} H \, \mathrm{d}\mu = \theta.$$

On dit que $\widehat{\theta}_n$ est une estimateur *fortement consistant* de θ. Il est aussi *sans biais* puisque $\mathbf{E}\widehat{\theta}_n = \theta$.

Cette méthode permet notamment d'estimer les moments de μ : en prenant $H(x) = x^r$, $\theta = \mathbf{E}X_1^r = \int_{\mathbb{R}} x^r \mu(\mathrm{d}x)$. Le cas $r = 1$ revêt une importance particulière. L'estimateur $\widehat{\theta}_n$ est alors simplement la *moyenne empirique*

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i.$$

On peut ainsi estimer notamment
- le paramètre p d'une loi de Bernoulli car $\mathbf{E}X_1 = p$;
- le paramètre α d'une loi de Poisson car $\mathbf{E}X_1 = \alpha$;
- le paramètre m d'une loi $\mathfrak{N}(m, \sigma^2)$ car $\mathbf{E}X_1 = m$;
- le paramètre θ d'une loi uniforme[3] sur $[0, \theta]$ car $\theta = 2\mathbf{E}X_1$.

Dans le même ordre d'idées, on peut estimer le paramètre a d'une loi exponentielle de densité $f(t) = a\exp(-at)\mathbf{1}_{\mathbb{R}^+}(t)$ par $\tilde{\theta}_n = 1/\bar{X}_n$. En effet, $\mathbf{E}X_1 = 1/a$. On garde un estimateur fortement consistant, mais il n'est plus sans biais car $\mathbf{E}(1/\bar{X}_n) \neq 1/\mathbf{E}X_1$.

On peut de même estimer la variance σ^2 d'une loi μ d'espérance *connue* m. Il suffit de prendre $H(x) = (x-m)^2$ et on obtient l'estimateur fortement consistant et sans biais

$$\widehat{\theta}_n = \frac{1}{n}\sum_{i=1}^n (X_i - m)^2.$$

Quand m est inconnu, on l'estime par $\bar{X}_n$ et la variance σ^2 est estimée par la *variance empirique*

$$V_n = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2,$$

la dernière égalité résultant simplement de la formule de Koenig pour la variance de la loi de probabilité $\mu_n(\omega)$ qui est exactement $V_n(\omega)$. On a toujours un estimateur fortement consistant par la LFGN, par contre il n'est plus qu'*asymptotiquement* sans biais puisque :

$$\mathbf{E}V_n = \mathbf{E}X_1^2 - \mathbf{E}\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2 = \cdots = \frac{n-1}{n}\sigma^2.$$

Ceci explique pourquoi pour les petites valeurs de n on préfère l'estimateur sans biais $\frac{n}{n-1}V_n$ noté souvent σ_{n-1}^2 par un de ces abus d'écriture qui font le charme si particulier de la littérature statistique...

3. En fait dans ce cas, un meilleur estimateur est $\tilde{\theta}_n = \max_{1\leq i\leq n} X_i$, affaire à suivre...