

Probabilité IV

STEP, Mines Paristech*

3 septembre 2019 (#2bf3c07)

Table des matières

Suites de variables aléatoires indépendantes	2
Définition	3
Proposition	3
Exemple	3
Inégalités de concentrations -> demi amphi théorie de la mesure (en lien avec l'espérance)	4
Convergences et loi des grands nombres	4
Convergences de variables aléatoires	4
Définition	4
Remarque	4
Exemples	5
Proposition	5
Proposition	6
Proposition	6
Exemple	7
Proposition	7
La loi des grands nombres	8
Théorème	8
Démonstration	8
Théorème	10
Exemple TODO cf garnier	10
Convergence en loi — fonction caractéristique — théorème central limite	10
Convergence en loi	10
Définition	10
Exemple	10
Exemple	11

*Ce document est un des produits du projet  **boisgera/CDIS**, initié par la collaboration de (S)ébastien Boisgérault (CAOR), (T)homas Romary et (E)milie Chautru (GEOSCIENCES), (P)auline Bernard (CAS), avec la contribution de Gabriel Stoltz (Ecole des Ponts ParisTech, CERMICS). Il est mis à disposition selon les termes de la licence Creative Commons “attribution – pas d'utilisation commerciale – partage dans les mêmes conditions” 4.0 internationale.

Proposition	11
Proposition	11
Corollaire	12
Fonction caractéristique	13
Définition	13
Remarque	13
Proposition	13
Proposition	14
Exemples	14
Théorème	14
Corollaire	15
Proposition	16
Exemples :	16
Proposition	16
Remarque	17
Théorème	17
Théorème de Lévy	17
Théorème central limite	18
Théorème central limite	18
Remarque	19
Exemple : convergence des lois binomiales	19
Théorème central limite multi-dimensionnel	20

Nous allons présenter dans ce chapitre l'un des résultats essentiels de la théorie des probabilités. Ce résultat montre rigoureusement que, quand le nombre de répétitions de l'expérience tend vers l'infini, la fréquence de réalisation d'un événement converge vers la probabilité de réalisation de cet événement. Ce résultat, appelé **Loi des grands nombres**, a d'autres portées fondamentales. Philosophique tout d'abord, puisqu'il permet de voir le monde déterministe comme la limite macroscopique d'une accumulation de phénomènes élémentaires microscopiques aléatoires. Portée numérique aussi, car nous verrons que ce théorème est à l'origine de méthodes de calcul numérique appelées Méthodes de Monte-Carlo, qui sont extrêmement puissantes et robustes. Elles sont par exemple très utilisées en Physique ou en Mathématiques Financières. Plus généralement, nous allons étudier les suites de variables aléatoires indépendantes et les notions de convergence de variables aléatoires. Nous verrons que plusieurs notions sont possibles, non équivalentes, ce qui enrichit mais complique aussi la description des comportements asymptotiques.

Nous nous plaçons dans le cadre général des variables aléatoires réelles qui n'admettent pas nécessairement de densité.

Suites de variables aléatoires indépendantes

On considère une suite **infinie** $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires définies sur l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

Définition

La suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires est dite *indépendante* si pour tout n , la famille finie X_1, \dots, X_n est indépendante.

Il est facile de vérifier que l'indépendance est préservée par certaines transformations.

Proposition

L'indépendance de la suite $(X_n)_{n \in \mathbb{N}^*}$ entraîne celle de

1. toute sous-suite $(X_{i_k})_{k \in \mathbb{N}^*}$,
2. toute suite de vecteurs issus de X_n ,
3. toute suite de la forme $(f_n(X_n))_{n \in \mathbb{N}^*}$, où les fonctions f_n sont des fonctions mesurables.

Exemple

Nous considérons l'ensemble $\Omega = [0, 1[$ muni de la tribu borélienne restreinte à cet ensemble, et de la mesure de Lebesgue. A chaque réel ω , nous associons son développement dyadique (unique si l'on impose que les ω_i ne soient pas tous égaux à 1 à partir d'un certain rang) :

$$\omega = \sum_{i \in \mathbb{N}^*} \frac{\omega_i}{2^i}, \omega_i \in \{0, 1\}$$

. L'application $X_i : \Omega \rightarrow \{0, 1\}$, qui à ω associe $X_i(\omega) = \omega_i$ est une variable aléatoire sur Ω . En effet, pour $x_i \in \{0, 1\}$, $1 \leq i \leq n$,

$$\{X_i = x_i\} = \bigcup_{x_1, \dots, x_{i-1} \in \{0, 1\}} \left[\sum_{j=1}^i \frac{x_j}{2^j}, \sum_{j=1}^i \frac{x_j}{2^j} + \frac{1}{2^i} \right[,$$

qui est bien un élément de la tribu borélienne de $\Omega = [0, 1[$, et

$$\mathbb{P}(\{X_i = x_i\}) = \frac{1}{2^i} \sum_{x_1, \dots, x_{i-1} \in \{0, 1\}} 1 = \frac{1}{2}.$$

Montrons l'indépendance des variables aléatoires $(X_i)_{1 \leq i \leq n}$. Nous avons

$$\bigcap_{1 \leq i \leq n} \{X_i = x_i\} = \left[\sum_{i=1}^n \frac{x_i}{2^i}, \sum_{i=1}^n \frac{x_i}{2^i} + \frac{1}{2^n} \right[$$

si bien que

$$\mathbb{P} \left(\bigcap_{1 \leq i \leq n} \{X_i = x_i\} \right) = \frac{1}{2^n} = \prod_{1 \leq i \leq n} \mathbb{P}(\{X_i = x_i\}).$$

Cela démontre que les variables aléatoires X_i sont indépendantes et de loi de Bernoulli de paramètre $\frac{1}{2}$. Ainsi, nous venons de construire sur Ω une suite de variables aléatoires telle que toute sous-suite finie est constituée de variables aléatoires indépendantes. C'est donc une suite de variables aléatoires indépendantes de loi de Bernoulli de paramètre $\frac{1}{2}$.

Inégalités de concentrations -> demi amphi théorie de la mesure (en lien avec l'espérance)

Convergences et loi des grands nombres

Convergences de variables aléatoires

Dans ce paragraphe, nous allons étudier des modes de convergence impliquant la proximité des variables aléatoires elles-mêmes, contrairement au cas de la convergence en loi qui sera étudiée ultérieurement.

On considère une suite $(X_n)_{n \in \mathbb{N}^*}$ de vecteurs aléatoires, définis sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, et à valeurs dans \mathbb{R}^d . On considère également sur le même espace un vecteur "limite" X . On notera $|\cdot|$ la valeur absolue dans \mathbb{R} ou la norme dans \mathbb{R}^d .

Définition

1. La suite $(X_n)_{n \in \mathbb{N}^*}$ converge *presque sûrement* vers X , ce qui s'écrit $X_n \rightarrow X$ p.s., s'il existe un ensemble $N \in \mathcal{A}$ de probabilité nulle tel que

$$X_n(\omega) \xrightarrow[n \rightarrow \infty]{} X(\omega) \forall \omega \notin N.$$

2. La suite $(X_n)_{n \in \mathbb{N}^*}$ converge *en probabilité* vers X , ce qui s'écrit $X_n \xrightarrow{\mathbb{P}} X$, si $\forall \epsilon > 0$ on a

$$\mathbb{P}(|X_n - X| \geq \epsilon) \xrightarrow[n \rightarrow \infty]{} 0.$$

3. La suite $(X_n)_{n \in \mathbb{N}^*}$ converge *en moyenne* vers X , ce qui s'écrit $X_n \xrightarrow{\mathcal{L}^1} X$, si X_n et X sont dans \mathcal{L}^1 et si

$$\mathbb{E}(|X_n - X|) \xrightarrow[n \rightarrow \infty]{} 0.$$

Remarque

La convergence presque sûre est la plus proche de la convergence simple des fonctions. Mais ici, nous permettons à certains ω de ne pas vérifier $X_n(\omega) \rightarrow X(\omega)$, si toutefois la probabilité de réalisation de l'ensemble de ces ω est nulle.

Ces convergences ne sont pas équivalentes comme le montre les exemples suivants.

Exemples

- Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires de Bernoulli à valeurs dans $\{0, 1\}$ telles que

$$\mathbb{P}(X_n = 1) = \frac{1}{n}; \mathbb{P}(X_n = 0) = 1 - \frac{1}{n}.$$

Pour tout $\epsilon \in]0, 1[$, la probabilité $\mathbb{P}(|X_n| \geq \epsilon) = \frac{1}{n}$ tend vers 0 quand n tend vers l'infini. Ainsi, la suite $(X_n)_{n \in \mathbb{N}^*}$ tend vers $X = 0$ en probabilité. Comme $\mathbb{E}(\frac{1}{n}) = 0$, elle tend également en moyenne vers 0. Considérons maintenant une suite $(Y_n)_{n \in \mathbb{N}^*}$ de variables aléatoires de Bernoulli à valeurs dans $\{0, n^2\}$ telles que

$$\mathbb{P}(Y_n = n^2) = \frac{1}{n}; \mathbb{P}(Y_n = 0) = 1 - \frac{1}{n}.$$

Par le même argument que ci-dessus, nous voyons que la suite $(Y_n)_{n \in \mathbb{N}^*}$ converge en probabilité vers 0, mais comme $\mathbb{E}(Y_n) = n$, la suite ne converge pas en moyenne vers 0 (ni vers aucune autre limite finie).

- Soit U une variable aléatoire uniforme sur $[0, 1]$. Posons $Z_n = 1_{\{U \leq \frac{1}{n}\}}$. Alors

$$\mathbb{P}(Z_n = 1) = \frac{1}{n}; \mathbb{P}(Z_n = 0) = 1 - \frac{1}{n}$$

Si $\omega \in \{U > 0\}$ est fixé, alors $\exists n_0 \in \mathbb{N}$ tel que $U(\omega) > \frac{1}{n_0}$, et donc tel que $Z_n(\omega) = 0$ pour tout $n \geq n_0$. Comme $\mathbb{P}(U > 0) = 1$, ceci montre que la suite $(Z_n)_{n \in \mathbb{N}^*}$ converge presque sûrement vers 0.

Étudions maintenant les liens entre ces différentes convergences.

Proposition

La convergence p.s. et la convergence en moyenne entraînent la convergence en probabilité.

Démonstration Soit $A_{n,\epsilon} = \{|X_n - X| > \epsilon\}$.

- Supposons que $X_n \rightarrow X$ p.s. et soit N l'ensemble de probabilité nulle en dehors duquel on a $X_n(\omega) \rightarrow X(\omega)$. Si $\omega \notin N$, on a $\omega \notin A_{n,\epsilon}$ pour tout $n \geq n_0$, où n_0 dépend de ω et de ϵ , ce qui implique que les variables aléatoires $Y_{n,\epsilon} = 1_{N^c \cap A_{n,\epsilon}}$ tendent simplement vers 0 lorsque $n \rightarrow \infty$. Comme on a aussi $0 \leq Y_{n,\epsilon} \leq 1$ le théorème de convergence dominée implique que $\mathbb{E}(Y_{n,\epsilon}) \xrightarrow{n \rightarrow \infty} 0$. Mais

$$\mathbb{P}(A_{n,\epsilon}) \leq \mathbb{P}(N^c \cap A_{n,\epsilon}) + \mathbb{P}(N) = \mathbb{P}(N^c \cap A_{n,\epsilon}) = \mathbb{E}(Y_{n,\epsilon}) \xrightarrow{n \rightarrow \infty} 0.$$

- Supposons que $X_n \xrightarrow{\mathcal{L}^1} X$. Pour $\epsilon > 0$, on a $1_{A_{n,\epsilon}} \leq \frac{1}{\epsilon} |X_n - X|$, donc

$$\mathbb{P}(A_{n,\epsilon}) \leq \frac{1}{\epsilon} \mathbb{E}(|X_n - X|) \rightarrow 0.$$

La convergence en probabilité n'entraîne pas la convergence en moyenne, comme nous l'avons vu dans l'exemple ci-dessus, ne serait-ce que parce qu'elle n'implique pas l'appartenance de X_n et X à \mathcal{L}^1 . Si les X_n ne sont “pas trop grands”, il y a cependant équivalence entre les deux modes de convergence. En voici un exemple : ■

Proposition

S'il existe une constante a telle que $|X_n| \leq a$ presque sûrement, il y a équivalence entre $X_n \xrightarrow{\mathbb{P}} X$ et $X_n \xrightarrow{\mathcal{L}^1} X$.

Démonstration Etant donnée la proposition précédente, dont on reprend les notations, il suffit de montrer que la convergence en probabilité implique la convergence en moyenne lorsque $|X_n| \leq a$.

Comme $|X_n| \leq a$, on a $\{|X| > a + \epsilon\} \subset A_{n,\epsilon}$, et donc $\mathbb{P}(|X| > a + \epsilon) \leq \mathbb{P}(A_{n,\epsilon})$. En faisant $n \rightarrow \infty$, on en déduit que $\mathbb{P}(|X| > a + \epsilon) = 0$. Ceci est vrai pour tout ϵ et donc

$$\mathbb{P}(|X| > a) = 0.$$

Comme $|X_n| \leq a$, on a aussi

$$|X_n - X| \leq \epsilon + (X_n + X)1_{A_{n,\epsilon}} \leq \epsilon + 2a1_{A_{n,\epsilon}}$$

sur l'ensemble $\{|X| \leq a\}$ qui est de probabilité 1. On a ainsi

$$\mathbb{E}(|X_n - X|) \leq \epsilon + 2a\mathbb{P}(A_{n,\epsilon})$$

On en déduit que $\limsup_n \mathbb{E}(|X_n - X|) \leq \epsilon$, et comme ϵ est arbitrairement proche de 0, on a le résultat souhaité. ■

Les rapports entre convergence presque-sûre et convergence en probabilité sont plus subtils. La première de ces deux convergences est plus forte que la seconde d'après la proposition, mais “à peine plus”, comme le montre le résultat suivant.

Proposition

Si $X_n \xrightarrow{\mathbb{P}} X$, il existe une sous-suite (n_k) telle que $X_{n_k} \rightarrow X$ p.s. quand $k \rightarrow \infty$.

Démonstration Comme la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en probabilité vers X , on peut définir une sous-suite de la manière suivante : posons $n_1 = 1$, et soit

$$n_j = \inf \left\{ n > n_{j-1}; \mathbb{P} \left(|X_r - X_s| > \frac{1}{2^j} \right) < \frac{1}{3^j}, \text{ pour } r, s \geq n \right\}.$$

Il résulte alors de

$$\sum_j \mathbb{P} \left(|X_{n_{j+1}} - X_{n_j}| > \frac{1}{2^j} \right) < \sum_j \frac{1}{3^j} < \infty$$

que la suite $(X_{n_j})_{j \in \mathbb{N}^*}$ converge presque-sûrement. C'est en effet une conséquence du lemme de Borel-Cantelli appliqué aux ensembles

$$A_j = \left\{ |X_{n_{j+1}} - X_{n_j}| > \frac{1}{2^j} \right\}.$$

■

Exemple

Soient $\Omega = \mathbb{R}$ muni de sa tribu borélienne et \mathbb{P} la probabilité uniforme sur $[0, 1]$. Soit $X_n = 1_{A_n}$, où A_n est un intervalle de $[0, 1]$ de longueur $\frac{1}{n}$. Ainsi, $\mathbb{E}(X_n) = \frac{1}{n}$, et la suite X_n tend vers $X = 0$ en moyenne, et donc en probabilité. Supposons que les A_n soient placés bout-à-bout, en recommençant en 0 chaque fois qu'on arrive au point 1. Il est clair que l'on parcourt indéfiniment l'intervalle $[0, 1]$ (car la série de terme général $1/n$ diverge). Ainsi la suite numérique $X_n(\omega)$ ne converge pour aucun ω , et on n'a pas $X_n \rightarrow X$ presque-sûrement ; cependant comme la série $\sum_n 1/n^2$ converge, il s'en suit que $X_{n^2} \rightarrow X = 0$ presque-sûrement. Nous avons donc la convergence presque-sûre de la sous-suite $(X_{n^2})_{n \in \mathbb{N}^*}$.

Proposition

Soit f une fonction continue de \mathbb{R}^d dans \mathbb{R} .

1. Si $X_n \rightarrow X$ presque-sûrement, alors $f(X_n) \rightarrow f(X)$ presque-sûrement.
2. Si $X_n \xrightarrow{\mathbb{P}} X$, alors $f(X_n) \xrightarrow{\mathbb{P}} f(X)$.

Démonstration Le point 1. est évident. Pour 2., remarquons d'abord que si $K > 0$ et $\epsilon > 0$,

$$\{|f(X_n) - f(X)| \geq \epsilon\} \subset \{|X| > K\} \cup \{|X| \leq K, |f(X_n) - f(X)| \geq \epsilon\}.$$

La fonction f est uniformément continue sur $\{x : |x| \leq K\}$, donc il existe $\eta > 0$ tel que $|x - y| < \epsilon$ et $|x| \leq K$ et $|y| \leq K$ impliquent $|f(x) - f(y)| < \epsilon$. On a donc

$$\{|f(X_n) - f(X)| \geq \epsilon\} \subset \{|X| > K\} \cup \{|X_n - X| \geq \eta\}$$

d'où

$$\mathbb{P}(|f(X_n) - f(X)| \geq \epsilon) \leq \mathbb{P}(|X| > K) + \mathbb{P}(|X_n - X| \geq \eta).$$

Par hypothèse, il vient

$$\limsup_n \mathbb{P}(|f(X_n) - f(X)| \geq \epsilon) \leq \mathbb{P}(|X| > K).$$

Enfin, $\lim_{K \rightarrow +\infty} \mathbb{P}(|X| > K) = 0$ (par convergence dominée) et donc la \limsup ci-dessus est nulle. On a ainsi le résultat. ■

La loi des grands nombres

Dans ce paragraphe, on considère une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires **indépendantes et de même loi** (ou indépendantes et identiquement distribuées, i.i.d. en abrégé). On considère la “moyenne” des n premières variables aléatoires :

$$M_n = \frac{X_1 + \dots + X_n}{n},$$

et notre objectif est de montrer que M_n converge vers l’espérance des X_n lorsque celle-ci existe (comme les X_n ont même loi, cette espérance est la même pour tout n). Il s’agit là d’un des résultats essentiels de toute la théorie des probabilités, connu sous le nom de **loi des grands nombres**.

Nous allons démontrer dans un premier temps la loi des grands nombres pour des variables aléatoires de carré intégrable.

Théorème

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes, de même loi et de **carré intégrable**, et $m = \mathbb{E}(X_n)$ leur moyenne. Alors la suite $(M_n)_{n \in \mathbb{N}^*}$ définie par

$$M_n = \frac{X_1 + \dots + X_n}{n}$$

converge vers m , **presque sûrement et en moyenne**, quand n tend vers l’infini. Elle converge donc aussi en probabilité. On a même convergence en *moyenne quadratique*, à savoir que :

$$\mathbb{E}((M_n - m)^2) \xrightarrow{n \rightarrow \infty} 0.$$

Le résultat sur la convergence en probabilité est appelé *loi faible des grands nombres*. Sa preuve est presque immédiate. Elle résulte de l’inégalité de Bienaymé-Chebyshev (exercice). Le résultat est peu informatif et permet d’obtenir certains contrôles d’erreurs. Le résultat prouvant la convergence presque-sûre est appelé *loi forte des grands nombres*. Sa preuve est plus délicate et utilise le lemme de Borel-Cantelli.

Démonstration

Notons σ^2 la variance des variables X_n , bien définie puisqu’on les a supposées de carré intégrable. En vertu de la linéarité de l’espérance, on a

$$\mathbb{E}(M_n) = m, \mathbb{E}((M_n - m)^2) = \mathbb{V}(M_n) = \frac{\sigma^2}{n},$$

d’où la convergence en moyenne quadratique.

Comme $\mathbb{E}(Y)^2 \leq \mathbb{E}(Y^2)$, on en déduit que $\mathbb{E}(|M_n - m|) \rightarrow 0$, donc on a aussi la convergence en moyenne.

La preuve de la convergence presque-sûre est plus délicate.

Quitte à remplacer X_n par $X_n - m$ (et donc M_n par $M_n - m$), nous pouvons supposer que $m = 0$.

Montrons tout d'abord que la sous-suite $(M_{n^2})_{n \in \mathbb{N}^*}$ converge presque-sûrement vers 0.

D'après l'inégalité de Bienaymé-Chebyshev et ce qui précède, on a pour $q \in \mathbb{N}^*$

$$\mathbb{P}(|M_n^2| \geq \frac{1}{q}) \leq \frac{\sigma^2 q^2}{n^2}$$

Donc si $A_{n,q} = \{|M_n^2| \geq \frac{1}{q}\}$, nous obtenons que $\sum_{n \geq 1} \mathbb{P}(A_{n,q}) < \infty$. Posons ensuite $B_{n,q} = \cup_{m \geq n} A_{m,q}$ et $C_q = \cap_{n \geq 1} B_{n,q} = \limsup_n A_{n,q}$. En appliquant le lemme de Borel-Cantelli, on obtient que $\mathbb{P}(C_q) = 0$. En conséquence, si on pose $N = \cup_{q \in \mathbb{N}^*} C_q$, on obtient $\mathbb{P}(N) \leq \sum_{q \in \mathbb{N}^*} 0 = 0$.

Si $\omega \notin N$, alors $\omega \in \cap_{q \in \mathbb{N}^*} (C_q)^c$. Ainsi, $\omega \notin C_q$ pour tout $q \geq 1$, et donc $\omega \notin B_{n,q}$ pour n assez grand (car $B_{n,q}$ est décroissant en n). Cela signifie que pour tout $\omega \notin N$, pour tout $q \geq 1$, il existe un n assez grand tel que $M_{k^2} \leq \frac{1}{q}$ dès que $k \geq n$. Autrement dit, $M_{n^2} \rightarrow 0$ si $\omega \notin N$, avec $\mathbb{P}(N) = 0$, d'où

$$M_{n^2} \xrightarrow[n \rightarrow \infty]{} 0 \text{ p.s.}$$

Montrons maintenant que la suite $(M_n)_{n \in \mathbb{N}^*}$ tend presque-sûrement vers 0.

Pour tout entier n , notons $p(n)$, l'entier tel que $p(n)^2 \leq n \leq (p(n) + 1)^2$. Alors,

$$M_n - \frac{p(n)^2}{n} M_{p(n)^2} = \frac{1}{n} \sum_{p=p(n)^2+1}^n X_p,$$

et puisque les variables aléatoires de la somme sont indépendantes, il vient

$$\begin{aligned} \mathbb{E} \left(\left(M_n - \frac{p(n)^2}{n} M_{p(n)^2} \right)^2 \right) &= \frac{n - p(n)^2}{n^2} \sigma^2 \\ &\leq \frac{2p(n) + 1}{n^2} \sigma^2 \leq \frac{2\sqrt{n} + 1}{n^2} \sigma^2 \end{aligned}$$

En appliquant de nouveau l'inégalité de Bienaymé-Chebyshev, on obtient

$$\mathbb{P} \left(\left| M_n - \frac{p(n)^2}{n} M_{p(n)^2} \right| > a \right) \leq \frac{2\sqrt{n} + 1}{n^2} \frac{\sigma^2}{a^2}$$

Comme la série $\sum_n \frac{2\sqrt{n}+1}{n^2}$ converge, le même raisonnement que pour M_{n^2} décrit ci-dessus, montre que

$$M_n - \frac{p(n)^2}{n} M_{p(n)^2} \rightarrow 0 \text{ p.s.}$$

Par ailleurs, on a déjà montré que $M_{p(n)^2} \rightarrow 0$ p.s. et $\frac{p(n)^2}{n} \rightarrow 1$. On en déduit que $M_n \rightarrow 0$ p.s.

Plus généralement, on a le résultat suivant (se référer par exemple à Jacod and Protter (2003) pour la démonstration)

Théorème

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes, de même loi et **intégrables**, et $m = \mathbb{E}(X_n)$ leur moyenne. Alors la suite $(M_n)_{n \in \mathbb{N}^*}$ définie par

$$M_n = \frac{X_1 + \dots + X_n}{n}$$

converge vers m , **presque sûrement et en moyenne**, quand n tend vers l'infini.

Exemple TODO cf garnier

Convergence en loi — fonction caractéristique — théorème central limite

Nous allons introduire maintenant une nouvelle notion de convergence de suites de variables aléatoires. La convergence en loi définie dans ce paragraphe va concerner les lois des variables aléatoires. Elle signifiera que les lois sont asymptotiquement “proches”, sans que les variables aléatoires elles-mêmes le soient nécessairement.

Convergence en loi

On considère des vecteurs aléatoires X_n et X , tous à valeurs dans le même espace \mathbb{R}^d , mais pouvant éventuellement être définis sur des espaces de probabilité différents.

Définition

On dit que la suite $(X_n)_{n \in \mathbb{N}^*}$ *converge en loi* vers X et on écrit $X_n \xrightarrow{\mathcal{L}} X$, si pour toute fonction f continue bornée sur \mathbb{R}^d ,

$$\mathbb{E}(f(X_n)) \xrightarrow{\mathcal{L}} \mathbb{E}(f(X)).$$

Exemple

Un cas très simple est celui où toutes les variables aléatoires X_n prennent un nombre fini de valeurs $\{x_i, 1 \leq i \leq N\}$. Alors, la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en loi vers X si et seulement si

$$\lim_{n \rightarrow +\infty} \mathbb{P}(X_n = x_i) = \mathbb{P}(X = x_i), \forall 1 \leq i \leq N$$

Il suffit d'écrire pour f continue bornée

$$\mathbb{E}(f(X_n)) = \sum_{i=1}^N f(x_i) P(X_n = x_i)$$

Dans l'exemple ci-dessus, N est fini et fixé. Mais nous avons un résultat analogue (en faisant tendre N vers l'infini) si les variables aléatoires ont un nombre dénombrable de valeurs. Le cas de la convergence de la loi binomiale vers la loi de Poisson est notamment traité en CPGE.

Exemple

Soit $(X_n)_{n \in \mathbb{N}^*}$ et X des variables aléatoires de lois respectives $\mathcal{N}(0, \sigma_n^2)$ et $\mathcal{N}(0, \sigma^2)$. On suppose que la suite de réels positifs $(\sigma_n)_{n \in \mathbb{N}^*}$ converge vers $\sigma > 0$ quand n tend vers l'infini. Alors la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en loi vers X . En effet, soit f une fonction continue bornée sur \mathbb{R} . On a

$$\begin{aligned}\mathbb{E}(f(X_n)) &= \int_{\mathbb{R}} f(y) \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{y^2}{2\sigma_n^2}\right) dy \\ &\rightarrow \int_{\mathbb{R}} f(y) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy\end{aligned}$$

où l'on a utilisée le théorème de convergence dominée.

La convergence en loi est plus faible que la convergence en probabilité et donc aussi que les convergences presque-sûre et en moyenne.

Proposition

Si $X_n \xrightarrow{\mathbb{P}} X$, alors $X_n \xrightarrow{\mathcal{L}} X$.

Démonstration Soit f une fonction continue bornée. D'après la proposition, on a $f(X_n) \xrightarrow{\mathbb{P}} f(X)$ et donc $f(X_n)$ converge aussi en moyenne vers $f(X)$ par la proposition. Comme $|\mathbb{E}(Y)| \leq \mathbb{E}(|Y|)$ pour toute variable aléatoire réelle Y , on en déduit $\mathbb{E}(f(X_n)) \rightarrow \mathbb{E}(f(X))$. ■

Un moyen efficace de caractériser la convergence en loi passe par l'étude de la suite des fonctions de répartition.

Proposition

Soient X_n et X des variables aléatoires réelles de fonctions de répartition respectives F_n et F . Pour que $X_n \xrightarrow{\mathcal{L}} X$, il faut et il suffit que $F_n(x) \xrightarrow[n \rightarrow \infty]{} F(x)$ pour tout x en lequel F est continue.

Notons que puisque la fonction F est continue à droite et croissante, l'ensemble des points où F est continue est l'ensemble $D = \{x : F(x-) = F(x)\}$, et son complémentaire est au plus dénombrable. Ainsi, D est dense dans \mathbb{R} .

Démonstration

1. Supposons d'abord que $X_n \xrightarrow{\mathcal{L}} X$. Soit $a \in \mathbb{R}$ tel que $F(a-) = F(a)$. Pour tout $p \in \mathbb{N}^*$ et tout $b \in \mathbb{R}$, il existe une fonction $f_{p,b}$ continue bornée sur \mathbb{R} telle que

$$1_{]-\infty, b]} \leq f_{p,b} \leq 1_{]-\infty, b + \frac{1}{p}]}$$

Alors, par définition, $\mathbb{E}(f_{p,b}(X_n)) \rightarrow \mathbb{E}(f_{p,b}(X))$ quand n tend vers l'infini. L'inégalité ci-dessus implique que $F_n(a) = \mathbb{P}(X_n \leq a) \leq \mathbb{E}(f_{p,a}(X_n))$ et $\mathbb{E}(f_{p,a}(X)) \leq F(a + 1/p)$; donc $\limsup_n F_n(a) \leq F(a + 1/p)$ pour tout p et donc on a aussi $\limsup_n F_n(a) \leq F(a)$. On a également que $F_n(a) \geq \mathbb{E}(f_{p,a-1/p}(X_n))$ et $\mathbb{E}(f_{p,a-1/p}(X)) \geq F(a - 1/p)$; donc $\liminf_n F_n(a) \geq F(a - 1/p)$ pour tout p et donc aussi $\liminf_n F_n(a) \geq F(a)$, puisque $F(a-) = F(a)$. Ces deux résultats impliquent que $F_n(a) \xrightarrow{n \rightarrow \infty} F(a)$.

2. Inversement, supposons $F_n(x) \xrightarrow{n \rightarrow \infty} F(x)$ pour tout $x \in T$, où T est une partie dense de \mathbb{R} . Soit f une fonction continue bornée sur \mathbb{R} et $\epsilon > 0$. Soient $a, b \in T$ avec $F(a) \leq \epsilon$ et $F(b) \geq 1 - \epsilon$. Il existe n_0 tel que

$$n \geq n_0 \Rightarrow \mathbb{P}(X_n \notin]a, b]) = 1 - F_n(b) + F_n(a) \leq 3\epsilon.$$

La fonction f est uniformément continue sur $[a, b]$, donc il existe un nombre fini de points $a_0 = a < a_1 < \dots < a_k = b$ appartenant tous à T , et tels que $|f(x) - f(a_i)| \leq \epsilon$ si $a_{i-1} \leq x \leq a_i$. Donc

$$g(x) = \sum_{i=1}^k f(a_i) 1_{]a_{i-1}, a_i]}(x)$$

vérifie $|f - g| \leq \epsilon$ sur $]a, b]$. Si $M = \sup_x |f(x)|$, il vient alors

$$|\mathbb{E}(f(X_n)) - \mathbb{E}(g(X_n))| \leq M \mathbb{P}(X_n \notin [a, b]) + \epsilon,$$

de même pour X . Enfin, $\mathbb{E}(g(X_n)) = \sum_{i=1}^k f(a_i)(F_n(a_i) - F_n(a_{i-1}))$, et de même pour X par définition de g . Comme $(F_n(a_i))_{n \in \mathbb{N}^*}$ converge vers $F(a_i)$ pour tout i , on en déduit l'existence de n_1 tel que

$$n \geq n_1 \Rightarrow |\mathbb{E}(g(X_n)) - \mathbb{E}(g(X))| \leq \epsilon.$$

Finalement, on a

$$n \geq \sup(n_0, n_1) \Rightarrow |\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| \leq 3\epsilon + 5M\epsilon.$$

Vu l'arbitraire sur ϵ , on en déduit que $\mathbb{E}(f(X_n))$ converge vers $\mathbb{E}(f(X))$, d'où le résultat.

■

Corollaire

Si la suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires réelles converge en loi vers X , et si la loi de X admet une densité, alors pour tous $a < b$,

$$\mathbb{P}(X_n \in]a, b]) \xrightarrow{n \rightarrow \infty} \mathbb{P}(X \in]a, b]).$$

Démonstration La fonction de répartition de X est alors continue en tout point. (Mais pas nécessairement celles des variables aléatoires X_n .)

Slutsky ???? voir avec CAA -> stat/ML ■

Fonction caractéristique

Dans ce paragraphe, nous introduisons un outil important en calcul des probabilités : il s'agit de ce que l'on appelle *la fonction caractéristique* d'une variable aléatoire, et qui dans d'autres branches des mathématiques s'appelle aussi *la transformée de Fourier*. Elle nous sera notamment très utile pour démontrer le théorème central limite.

On notera $\langle x, y \rangle$ le produit scalaire de deux vecteurs de \mathbb{R}^n . Si $u \in \mathbb{R}^n$, la fonction (complexe) $x \mapsto e^{i\langle u, x \rangle}$ est continue, de module 1. Donc si X est un vecteur aléatoire à valeurs dans \mathbb{R}^n , nous pouvons considérer $e^{i\langle u, X \rangle}$ comme une variable aléatoire à valeurs complexes. Ses parties réelle $Y = \cos(\langle u, X \rangle)$ et imaginaire $Z = \sin(\langle u, X \rangle)$ sont des variables aléatoires réelles. Ces variables aléatoires réelles sont de plus bornées par 1, donc elles admettent une espérance. Il est alors naturel d'écrire que l'espérance de $e^{i\langle u, X \rangle}$ est

$$\mathbb{E}(e^{i\langle u, X \rangle}) = \mathbb{E}(Y) + i\mathbb{E}(Z) = \mathbb{E}(\cos \langle u, X \rangle) + i\mathbb{E}(\sin \langle u, X \rangle)$$

Définition

Si X est un vecteur aléatoire à valeurs dans \mathbb{R}^n , sa *fonction caractéristique* est la fonction ϕ_X de \mathbb{R}^n dans \mathbb{C} définie par

$$\phi_X(u) = \mathbb{E}(e^{i\langle u, X \rangle}).$$

Remarque

La fonction caractéristique ne dépend en fait **que de la loi \mathbb{P}_X de X** : c'est la "transformée de Fourier" de la loi \mathbb{P}_X .

Nous verrons que cette fonction porte bien son nom, au sens où elle caractérise la loi \mathbb{P}_X . C'est une notion qui, de ce point de vue, généralise la fonction génératrice G qui a été vue en CPGE. Elle vérifie

$$\phi_X(u) = G_X(e^{iu}) = \mathbb{E}(e^{iuX})$$

pour une variable X à valeurs dans \mathbb{N} .

Proposition

ϕ_X est de module inférieur à 1, continue, avec

$$\phi(0) = 1; \phi_X(-u) = \overline{\phi_X(u)}.$$

Démonstration $|z|$ désigne le module d'un nombre complexe z .

Comme $\mathbb{E}(Y)^2 \leq \mathbb{E}(Y^2)$ pour toute variable réelle Y , on a :

$$|\phi_X(u)|^2 = \mathbb{E}(\cos \langle u, X \rangle)^2 + \mathbb{E}(\sin \langle u, X \rangle)^2 \leq \mathbb{E}(\cos^2 \langle u, X \rangle + \sin^2 \langle u, X \rangle) = 1.$$

Pour montrer la continuité, considérons une suite $u_p \xrightarrow{p \rightarrow \infty} u$. Il y a convergence simple de $e^{i\langle u_p, X \rangle}$ vers $e^{i\langle u, X \rangle}$. Comme ces variables aléatoires sont de module borné par 1, le théorème de convergence dominée assure que $\phi_X(u_p) \xrightarrow{p \rightarrow \infty} \phi_X(u)$. ϕ_X est donc continue. ■

Proposition

Si X est un vecteur aléatoire à valeurs dans \mathbb{R}^n , si $a \in \mathbb{R}^m$ et A est une matrice réelle de taille $m \times n$, alors

$$\phi_a + AX(u) = e^{i\langle u, a \rangle} \phi_X(A^t u), \forall u \in \mathbb{R}^m$$

Démonstration Nous avons $e^{i\langle u, a+AX \rangle} = e^{i\langle u, a \rangle} e^{i\langle A^t u, X \rangle}$. En effet, $\langle u, AX \rangle = \langle A^t u, X \rangle$. On prend ensuite les espérances pour obtenir le résultat. ■

Exemples

1. X suit une loi binomiale $\mathcal{B}(n, p)$: $\phi_X(u) = (pe^{iu} + 1 - p)^n$.
2. X suit une loi de Poisson $\mathcal{P}(\theta)$: $\phi_X(u) = e^{\theta(e^{iu} - 1)}$.
3. X suit une loi uniforme $\mathcal{U}_{[a, b]}$: $\phi_X(u) = \frac{e^{iub} - e^{iua}}{iu(b-a)}$.
4. X suit une loi exponentielle $\mathcal{E}(\lambda)$, $\lambda > 0$: $\phi_X(u) = \frac{\lambda}{\lambda - iu}$.
5. X suit une loi normale $\mathcal{N}(0, 1)$: $\phi_X(u) = e^{-u^2/2}$.
6. X suit une loi normale $\mathcal{N}(\mu, \sigma^2)$: $\phi_X(u) = e^{iu\mu - u^2\sigma^2/2}$.

L'intérêt majeur de la fonction caractéristique réside dans le fait qu'elle caractérise la loi de la variable aléatoire (d'où son nom).

Théorème

La fonction caractéristique ϕ_X caractérise la loi du vecteur aléatoire X . Ainsi, si deux vecteurs aléatoires X et Y ont même fonction caractéristique, ils ont même loi.

Démonstration Soient les fonctions suivantes avec $\sigma > 0$:

$$f_\sigma(x) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{|x|^2}{2\sigma^2}\right) \text{ et } \hat{f}_\sigma(u) = \exp\left(-\frac{|u|^2}{2\sigma^2}\right).$$

On a

$$\begin{aligned}\int f_\sigma(x) e^{i\langle u, x \rangle} dx &= \int \mathbb{R}^n \prod_{j=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_j^2}{2\sigma^2} + iu_j x_j\right) dx_1 \dots dx_n \\ &= \prod_{j=1}^n \int \mathbb{R} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2} + iu_j t\right) dt = \widehat{f}_\sigma(u)\end{aligned}$$

d'après l'exemple 5 ci-dessus et en utilisant le théorème de Fubini. On remarque ainsi que

$$f_\sigma(u-v) = \frac{1}{(2\pi\sigma^2)^{n/2}} \widehat{f}_\sigma\left(\frac{u-v}{\sigma^2}\right) = \frac{1}{(2\pi\sigma^2)^{n/2}} \int f_\sigma(x) e^{i\langle u-v, x \rangle / \sigma^2} dx.$$

Supposons que X et X' admettent la même fonction caractéristique $\phi_X = \phi_{X'}$. En utilisant le théorème de Fubini, on a

$$\begin{aligned}\int f_\sigma(u-v) \mathbb{P}_X(du) &= \int_{\mathbb{R}^n} \frac{1}{(2\pi\sigma^2)^{n/2}} \left(\int_{\mathbb{R}^n} f_\sigma(x) e^{i\langle u-v, x \rangle / \sigma^2} dx \right) \mathbb{P}_X(du) \\ &= \int_{\mathbb{R}^n} f_\sigma(x) \frac{1}{(2\pi\sigma^2)^{n/2}} \phi_X\left(\frac{x}{\sigma^2}\right) e^{-i\langle v, x \rangle / \sigma^2} dx,\end{aligned}$$

et la même égalité reste vraie pour $\mathbb{P}_{X'}$. Par suite $\mathbb{E}(g(X)) = \mathbb{E}(g(X'))$ pour toute fonction g de l'espace vectoriel de fonction engendrées par $u \mapsto f_\sigma(u-v)$. D'après le théorème de Stone-Weierstrass, cet espace est dense dans l'ensemble C_0 des fonctions continues sur \mathbb{R}^n et ayant une limite nulle à l'infini, pour la topologie de la convergence uniforme. Par suite, $\mathbb{E}(g(X)) = \mathbb{E}(g(X'))$ pour toute fonction $g \in C_0$. Comme l'indicatrice de tout ouvert est limite croissante de fonctions de C_0 , on en déduit que $\mathbb{P}_X(A) = \mathbb{E}(1_A(X))$ est égal à $\mathbb{P}_{X'}(A) = \mathbb{E}(1_A(X'))$ pour tout ouvert A , ce qui implique $P_X = P_{X'}$. ■

La fonction caractéristique offre également un moyen commode de caractériser l'indépendance.

Corollaire

Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire à valeurs dans \mathbb{R}^n . Ses composantes X_i sont indépendantes si et seulement si pour tous $u_1, \dots, u_n \in \mathbb{R}$,

$$\phi_X(u_1, \dots, u_n) = \prod_{j=1}^n \phi_{X_j}(u_j)$$

où ϕ_X désigne la fonction caractéristique du vecteur aléatoire X , et ϕ_{X_j} celle de la composante X_j pour chaque j .

Démonstration On a $\langle u, X \rangle = \sum_{j=1}^n u_j X_j$. Si les X_i sont indépendantes, et comme $e^{i\langle u, x \rangle} = \prod_j e^{iu_j x_j}$, nous obtenons directement le résultat en utilisant la proposition.

Supposons inversement qu'on ait $\phi_X(u_1, \dots, u_n) = \prod_{j=1}^n \phi_{X_j}(u_j)$. On peut alors construire des variables aléatoires X'_j indépendantes, telles que X'_j et X_j aient même loi pour tout j et donc telles que $\phi_{X'_j} = \phi_{X_j}$. Si $X' = (X'_1, \dots, X'_n)$, on a donc $\phi_{X'} = \phi_X$. Donc X et X' ont même loi, ce qui entraîne que pour tous boréliens A_j , on ait

$$\mathbb{P}\left(\bigcap_j \{X_j \in A_j\}\right) = \mathbb{P}\left(\bigcap_j \{X'_j \in A_j\}\right) = \prod_j \mathbb{P}(\{X'_j \in A_j\}) = \prod_j \mathbb{P}(\{X_j \in A_j\})$$

d'où l'indépendance. ■

Proposition

Si X et Y sont deux vecteurs aléatoires indépendants à valeurs dans \mathbb{R}^n , la fonction caractéristique de la somme $X + Y$ est donnée par

$$\phi_{X+Y} = \phi_X \phi_Y$$

Démonstration Comme $e^{i\langle u, X+Y \rangle} = e^{i\langle u, X \rangle} e^{i\langle u, Y \rangle}$, il suffit d'appliquer la proposition. ■

Exemples :

Soient X et Y deux variables aléatoires réelles indépendantes et $Z = X + Y$:

1. Si X suit la loi normale $\mathcal{N}(m, \sigma^2)$ et Y suit $\mathcal{N}(m', \sigma'^2)$, alors Z suit une loi $\mathcal{N}(m + m', \sigma^2 + \sigma'^2)$, d'après l'exemple point 6. et la proposition.
2. Si X et Y suivent des lois de Poisson de paramètres θ et θ' , alors Z suit une loi de Poisson de paramètre $\theta + \theta'$, d'après l'exemple point 2. et la proposition.
3. Si X suit une loi binomiale $\mathcal{B}(n, p)$ et Y la loi binomiale $\mathcal{B}(m, p)$, alors Z suit une loi binomiale $\mathcal{B}(n + m, p)$, d'après l'exemple point 1. et la proposition.

Proposition

Soit X un vecteur aléatoire de \mathbb{R}^n . Si la variable $|X|^m$ (ou $|\cdot|$ désigne la norme euclidienne) est intégrable pour un entier m , la fonction ϕ_X est m fois continûment différentiable sur \mathbb{R}^n et pour tout choix des indices i_1, \dots, i_m ,

$$\frac{\partial^m}{\partial u_{i_1} \dots \partial u_{i_m}} \phi_X(u) = i^m \mathbb{E}(e^{i\langle u, X \rangle} X_{i_1} \dots X_{i_m}),$$

où les X_j sont les composantes de X .

Démonstration Le résultat se démontre par application itérée du théorème de dérivation sous le signe somme. ■

Remarque

En prenant $u = 0$ dans la proposition, la formule permet de calculer $\mathbb{E}(X_{i_1} \dots X_{i_m})$ en fonction des dérivées à l'origine de ϕ_X , autrement dit de calculer tous les moments du vecteur X , s'ils existent. Par exemple, si X est à valeurs réelles et est intégrable (respectivement de carré intégrable), nous avons

$$\mathbb{E}(X) = i\phi'_X(0), (\text{resp. } \mathbb{E}(X^2) = \phi''_X(0))$$

Théorème

X est un vecteur gaussien si et seulement si sa fonction caractéristique s'écrit

$$\phi_X(u) = e^{i\langle u, m \rangle - \frac{1}{2}\langle u, Cu \rangle}$$

où $m = \mathbb{E}(X) \in \mathbb{R}^n$ et C est la matrice de covariance de X .

Démonstration

1. Supposons $\phi_X(u) = e^{i\langle u, m \rangle - \frac{1}{2}\langle u, Cu \rangle}$. Pour toute combinaison linéaire $Y = \sum_j a_j X_j = \langle a, X \rangle$, et pour tout $v \in \mathbb{R}$, on a

$$\phi_Y(v) = \phi_X(va) = e^{iv\langle a, m \rangle - \frac{v^2}{2}\langle a, Ca \rangle}$$

donc Y suit la loi $\mathcal{N}(\langle a, m \rangle, \langle a, Ca \rangle)$.

2. Inversement, soit C la matrice de covariance de X et m son vecteur moyenne. Si $Y = \langle a, X \rangle = \sum_{j=1}^n a_j X_j$ avec $a \in \mathbb{R}^n$, un calcul simple montre

$$\mathbb{E}(Y) = \langle a, m \rangle, \mathbb{V}(Y) = \langle a, Ca \rangle$$

Par hypothèse, Y suit une loi normale donc vu le point 6. de l'exemple plus haut, sa fonction caractéristique est

$$\phi_Y(v) = e^{iv\langle a, m \rangle - \frac{v^2}{2}\langle a, Ca \rangle}$$

Mais $\phi_Y(1) = \phi_{\langle a, X \rangle}(1) = \mathbb{E}(e^{i\langle a, X \rangle}) = \phi_X(a)$, d'où le résultat.

Le théorème suivant caractérise la convergence en loi à l'aide des fonctions caractéristiques. C'est un critère extrêmement utile dans la pratique. ■

Théorème de Lévy

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de vecteurs aléatoires à valeurs dans \mathbb{R}^d .

1. Si la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en loi vers X , alors ϕ_{X_n} converge simplement vers ϕ_X .
2. Si les ϕ_{X_n} convergent simplement vers une fonction (complexe) ϕ sur \mathbb{R}^d et si cette fonction est **continue** en 0, alors c'est la fonction caractéristique d'une variable aléatoire X et $X_n \xrightarrow{\mathcal{L}} X$.

Démonstration

1. On remarque que $\phi_{X_n}(u) = \mathbb{E}(g_u(X_n))$ et $\phi_X(u) = \mathbb{E}(g_u(X))$ où g_u est la fonction continue bornée $g_u(x) = e^{i\langle u, x \rangle}$. On applique alors la définition.
2. Admis.

■

Théorème central limite

Ce théorème est aussi connu sous le nom de théorème de la limite centrale. Plus simplement, il apparaît souvent sous l'abréviation TCL.

On considère une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}^*}$ indépendantes, de même loi et de carré intégrable. On note m et σ^2 la moyenne et la variance commune aux variables X_n , et

$$S_n = X_1 + \dots + X_n$$

ainsi ($S_n = nM_n$). On a vu que la loi des grands nombres assure que M_n converge vers m presque-sûrement et en moyenne. On va s'intéresser à la vitesse à laquelle cette convergence a lieu.

Pour évaluer cette vitesse, c'est-à-dire trouver un équivalent de $\frac{S_n}{n} - m$, on est amenés à étudier la limite éventuelle de la suite $n^\alpha(\frac{S_n}{n} - m)$ pour différentes valeurs de α : si α est "petit" cette suite va encore tendre vers 0, et elle va "exploser" si α est "grand". On peut espérer que pour une (et alors nécessairement une seule) valeur de α , cette suite converge vers une limite qui n'est ni infinie ni nulle.

Il se trouve que la réponse à cette question a un aspect "négatif" : la suite $n^\alpha(\frac{S_n}{n} - m)$ ne converge au sens presque-sûr, ou même en probabilité, pour aucune valeur de α . Elle a aussi un aspect "positif" : cette suite converge, au sens de la convergence en loi, pour la même valeur $\alpha = 1/2$ quelle que soit la loi des X_n , et toujours vers une loi normale.

Ce résultat, qui peut sembler miraculeux, a été énoncé par Laplace (1749-1827) et démontré beaucoup plus tard par Lyapounov (1901). Il montre le caractère universel de la loi normale en probabilités (d'où son nom). Il fait l'objet du théorème suivant, appelé théorème central limite (TCL), ou de la limite centrale.

Théorème central limite

Si les X_n sont des variables aléatoires réelles, indépendantes et de même loi, de carré intégrable, de moyenne m et de variance $\sigma^2 > 0$, alors les variables

$$\frac{S_n - nm}{\sigma\sqrt{n}}$$

convergent en loi vers une variable aléatoire de loi $\mathcal{N}(0, 1)$.

En d'autres termes, $\sqrt{n}(M_n - m)$ converge vers une variable normale de loi $\mathcal{N}(0, \sigma^2)$.

Démonstration Soit ϕ la fonction caractéristique de $X_n - m$, et $Y_n = \frac{S_n - nm}{\sigma\sqrt{n}}$. D'après la proposition et la proposition, la fonction caractéristique de Y_n est

$$\phi_n(u) = \phi\left(\frac{u}{\sigma\sqrt{n}}\right)^n.$$

Comme $\mathbb{E}(X_n - m) = 0$ et $\mathbb{E}((X_n - m)^2) = \sigma^2$, la proposition entraîne

$$\phi(u) = 1 - \frac{u^2\sigma^2}{2} + u^2 o(|u|) \text{ quand } u \rightarrow 0.$$

Comme $\phi(0) = 1$ et que ϕ est continue en 0, on a que pour u fixé et n assez grand,

$$\left| \phi\left(\frac{u}{\sigma\sqrt{n}}\right) - 1 \right| \leq 1/2.$$

Il est possible de généraliser la notion de logarithme aux complexes z tels que $|z - 1| \leq 1/2$. La fonction $\log z$ définie sur le disque $\{z \in \mathbb{C}; |z - 1| \leq 1/2\}$ admet le même développement limité au voisinage de $z = 1$ que le logarithme réel. Ainsi

$$\phi_n(u) = \exp\left(n \log\left(1 - \frac{u^2}{2n} + \frac{1}{n}\epsilon_n(u)\right)\right),$$

où $\epsilon_n(u)$ tend vers 0 quand n tend vers l'infini, et on en déduit que $\phi_n(u)$ tend vers $\exp(-u^2/2)$. Le résultat découle alors du théorème de Lévy. ■

Remarque

On peut déduire de ce résultat que $n^\alpha(\frac{S_n}{n} - m)$ converge vers 0 (resp. $+\infty$) en probabilité lorsque $\alpha < 1/2$ (resp. $\alpha > 1/2$).

Exemple : convergence des lois binomiales

Supposons que S_n suive une loi binomiale $\mathcal{B}(p, n)$. Cela revient à dire que S_n a la même loi qu'une somme $X_1 + \dots + X_n$ de n variables aléatoires X_i indépendantes de loi $\mathcal{B}(p, 1)$, i.e. $\mathbb{P}(X_i = 1) = p$ et $\mathbb{P}(X_i = 0) = 1 - p$. On a alors $m = p$ et $\sigma^2 = p(1 - p)$.

On veut calculer $\mathbb{P}(S_n \leq x)$ pour x fixé et n grand.

Si p est petit de sorte que $\theta = np$ ne soit pas trop grand (en pratique, $\theta \leq 5$ convient), on peut utiliser l'approximation par une loi de Poisson, vue en CPGE. Si p est très proche de 1, de sorte que $\theta = n(1 - p)$ soit comme ci-dessus, alors $n - S_n$ suit à son tour une loi proche de la loi de Poisson de paramètre θ .

Dans les autres cas, on utilise la loi des grands nombres et le théorème central limite :

$$\begin{aligned} \frac{S_n}{n} &\xrightarrow{\mathbb{P}} p, \\ \frac{S_n - np}{\sqrt{np(1-p)}} &\xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \end{aligned}$$

Si on désigne par Φ la fonction de répartition de la loi $\mathcal{N}(0, 1)$, il vient

$$\mathbb{P}(S_n \leq x) \approx \Phi\left(\frac{x - np}{\sqrt{np(1-p)}}\right)$$

Imaginons que l'on lance 1000 fois une pièce (non truquée). On cherche la probabilité d'obtenir plus de 545 fois le côté Face. Le calcul exact utilisant les lois binomiales est extrêmement lourd. Le résultat ci-dessus nous donne une très bonne approximation. On a

$$\mathbb{P}(S_{1000} > 545) = \mathbb{P}\left(\frac{S_{1000} - 500}{\sqrt{250}} > \frac{45}{\sqrt{250}}\right) \approx \int_{\frac{45}{\sqrt{250}}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Cette dernière intégrale se calcule numériquement (on trouve encore des abaques où les valeurs de Φ sont tabulées) et on obtient

$$\mathbb{P}(S_{1000} > 545) \approx 1 - \Phi(2, 84) \approx 0,0023.$$

Le théorème admet une version multidimensionnelle, de preuve similaire. On considère des vecteurs aléatoires X_n à valeurs dans \mathbb{R}^d , indépendants et de même loi, dont les composantes sont de carré intégrable. On a un vecteur moyenne $m = E(X_n)$, et une matrice de covariance $C = (c_{ij})_{i,j=1,\dots,d}$ avec $c_{ij} = \text{Cov}(X_i, X_j)$. On peut alors énoncer le TCL multi-dimensionnel.

Théorème central limite multi-dimensionnel

Les vecteurs aléatoires $\frac{S_n - nm}{\sqrt{n}}$ convergent en loi vers un vecteur aléatoire gaussien centré (i.e. de moyenne nulle), de matrice C .

Jacod, J., and P. Protter. 2003. *L'essentiel En Théorie Des Probabilités*. Cassini. <https://hal.archives-ouvertes.fr/hal-00104956>.