# Probabilité III

STEP, MINES ParisTech*

10 septembre 2020 (#7589d72)

## Table des matières



*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). 

# Probabilités — cadre général

## Interprétation

Les éléments de théorie de la mesure donnés en calcul intégral permettent une relecture des premiers chapitres de probabilités. Le principal avantage est que les différents cas de figures déjà évoqués : lois de probabilités discrètes, à densité, mixtes vont pouvoir être traités dans un cadre unifié. L'intégrale que nous considérons est l'intégrale de Lebesgue. On peut déjà s'apercevoir qu'une probabilité $\mathbb{P}$ définie sur un espace probabilisable (mesurable) $(\Omega, \mathcal{A})$ est une mesure positive *finie* au sens où $\mathbb{P}(\Omega) = 1$ et hérite ainsi de ses propriétés, on parle aussi de *mesure de probabilité*.

**Remarque – Remarque**

— Dans le cas discret ($\Omega$ au plus dénombrable), la mesure de probabilité est une somme pondérée de mesures de Dirac :

$$\mathbb{P} = \sum_{\omega \in \Omega} p_\omega \delta_\omega,$$

où $\sum_{\omega \in \Omega} p_\omega = 1$

— Dans le cas à densité ($\Omega = \mathbb{R}^n$, $n \in \mathbb{N}^*$), la mesure de probabilité s'écrit :

$$\mathbb{P} = f\mu,$$

où $f$ est une densité et $\mu$ la mesure de (Borel-)Lebesgue sur $\mathbb{R}^n$. On dit que $\mathbb{P}$ admet une densité par rapport à la mesure de (Borel-)Lebesgue.

Une variable aléatoire réelle (v.a.r), respectivement un vecteur aléatoire, $X$ est une application mesurable de $(\Omega, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, respectivement dans $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, et sa loi $\mathbb{P}_X$ est la mesure image de $\mathbb{P}$ par $X$.

Le fait que la composition d'un vecteur aléatoire réel par une application $\mathcal{B}(\mathbb{R}^n)/\mathcal{B}(\mathbb{R})$ mesurable est une variable aléatoire s'obtient immédiatement par le résultat du chapitre précédent de calcul intégral relatif à la composition de fonctions mesurables.

On peut généraliser la définition des espaces vectoriels $\mathcal{L}^1$ et $\mathcal{L}^2$ de la manière suivante :

**Définition – Définition — Espace $\mathcal{L}^1$**

Soit $X$ une variable aléatoire réelle. $X$ est intégrable et on note $X \in \mathcal{L}^1$, ou $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, si et seulement si $\mathbb{E}(|X|) = \int_{\mathbb{R}} |x| \mathbb{P}_X(dx) = \int_{\Omega} |X|(\omega) \mathbb{P}(d\omega) < +\infty$.

**Définition – Définition — Espace $\mathcal{L}^2$**

Soit $X$ une variable aléatoire réelle. $X$ est de carré intégrable et on note $X \in \mathcal{L}^2$, ou $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$, si et seulement si $\mathbb{E}(X^2) = \int_{\mathbb{R}} x^2 \mathbb{P}_X(dx) = \int_{\Omega} X^2(\omega)\mathbb{P}(d\omega) < +\infty$.

Les propriétés des espaces $\mathcal{L}^1$ et $\mathcal{L}^2$ données au chapitre 2 du cours de probabilités sont vraies en toute généralité. On peut d'ailleurs étendre ces définitions pour un $p \in \mathbb{N}^*$ quelconque.

**Définition – Définition — Espace $\mathcal{L}^p$**

Soit $X$ une variable aléatoire. On note $X \in \mathcal{L}^p$, ou $\mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$, si et seulement si $\mathbb{E}(|X|^p) = \int_{\mathbb{R}} |x|^p \mathbb{P}_X(dx) = \int_{\Omega} |X|^p(\omega)\mathbb{P}(d\omega) < +\infty$.

Si $X \in \mathcal{L}^p$, on dit qu'elle admet un moment d'ordre $p$. Du fait que $\mathbb{P}$ est une mesure finie, on a la stabilité par inclusion suivante :

**Proposition – Proposition**

Soit $p \in \mathbb{N}^*$, on a l'inclusion :

$$\mathcal{L}^{p+1}(\Omega, \mathcal{A}, \mathbb{P}) \subset \mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$$

**Démonstration** Supposons $X \in \mathcal{L}^{p+1}(\Omega, \mathcal{A}, \mathbb{P})$. On a

$$|X|^p \leq \max(1, |X|^{p+1}) = 1_{|X|<1} + 1_{|X|\geq 1}|X|^{p+1}.$$

Le terme de droite est intégrable, en effet :

$$\mathbb{E}\left(1_{|X|<1} + 1_{|X|\geq 1}|X|^{p+1}\right) \leq \int_{\Omega} \mathbb{P}(d\omega) + \int_{\Omega} |X(\omega)|^{p+1}\mathbb{P}(d\omega) = 1 + \mathbb{E}(|X|^{p+1}).$$

donc $|X|^p$ est intégrable. ■

On donne ici dans le cas général les inégalités de Markov et de Bienaymé-Chebyshev, déjà vues en CPGE.

**Théorème – Inégalité de Markov**

Soient $p \in \mathbb{N}^*$ et $a \in \mathbb{R}^*$. Soit $X \in \mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$, on a

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|^p)}{a^p}$$

**Démonstration** On a

$$|X|^p \geq a^p 1_{[a,+\infty[}(|X|)$$

et en prenant l'espérance, on obtient ainsi

$$\mathbb{E}(|X|^p) \geq a^p \mathbb{E}(1_{[a,+\infty[}(|X|)) = a^p \mathbb{P}(|X| \geq a).$$

■

**Corollaire – Inégalité de Bienaymé-Chebyshev**

Soit $X \in \mathcal{L}^2$, on a

$$\mathbb{P}(|X - \mathbb{E}(X)| > a) \leq \frac{\mathbb{V}(X)}{a^2}$$

**Démonstration** C'est une application immédiate de l'inégalité de Markov (p. 4) à $(X - \mathbb{E}(X))$ avec $p = 2$. ■

**Remarque – Remarque**

L'inégalité de Bienaymé-Chebyshev est très utile en pratique. Elle permet de mesurer la probabilité des grands écarts entre $X$ et sa moyenne. Par exemple, avec $a = 10\sigma X$, il en résulte qu'il est improbable qu'une variable aléatoire $X$ dévie de son espérance $\mathbb{E}(X)$ de plus de 10 fois son écart-type (probabilité inférieure à 0.01). Cette inégalité, tout à fait générale, n'est cependant pas très précise, et surestime très souvent en pratique le membre de gauche. On préférera, quand c'est possible, calculer directement ces probabilités à partir de la loi de $X$.

On peut également réécrire la proposition portant sur l'espérance de la composée d'une variable aléatoire et d'une fonction mesurable avec l'intégrale de Lebesgue.

**Proposition – Proposition**

Soit $X$ une variable aléatoire réelle de loi $\mathbb{P}_X$ et $g$ une fonction $\mathcal{B}(\mathbb{R})/\mathcal{B}(\mathbb{R})$-mesurable (borélienne). Alors $g(X)$ est intégrable si et seulement si l'intégrale

$$\int_{\mathbb{R}} |g(x)|\mathbb{P}_X(dx) = \int_{\Omega} |g(X(\omega))|\mathbb{P}(d\omega)$$

est finie et dans ce cas

$$\mathbb{E}(g(X)) = \int_R g(x)\mathbb{P}_X(dx) = \int_{\Omega} g(X(\omega))\mathbb{P}(d\omega).$$

Ce résultat a été démontré en exercice au chapitre IV de calcul intégral (mesure image).

## Indépendance et suites de variables indépendantes

On notera que $\mathbb{P}$ étant finie, elle est nécessairement $\sigma$-finie. On peut ainsi caractériser l'indépendance de deux variables aléatoires quelconques au moyen du théorème de Fubini et du résultat d'unicité de la mesure produit donné au chapitre V de calcul intégral.

**Proposition – Proposition — indépendance d'un couple de variables aléatoires**

Soient $X$ et $Y$ deux variables aléatoires réelles définies toutes deux sur $(\Omega, \mathcal{A}, \mathbb{P})$. Le couple $Z = (X, Y)$ peut-être considéré comme un vecteur aléatoire à valeurs dans $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}))$, et les deux variables aléatoires $X$ et $Y$ sont indépendantes si et seulement si la loi $\mathbb{P}_{X,Y}$ du couple est égale au produit $\mathbb{P}_X \otimes \mathbb{P}_Y$ des lois de $X$ et $Y$.

**Démonstration** Soit $A$ et $B$ deux boréliens de $\mathbb{R}$. On a évidemment $Z^{-1}(A \times B) = X^{-1}(A) \cap Y^{-1}(B) \in \mathcal{A}$ et donc la mesurabilité $(\mathcal{A}/\left(\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})\right))$de $Z$ découle de la définition de la tribu produit de Borel.

L'indépendance de $X$ et $Y$ revient au fait que pour tous boréliens $A$ et $B$ de $\mathbb{R}$, on ait

$$\mathbb{P}((X, Y) \in A \times B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B),$$

ce qui équivaut à

$$\mathbb{P}_{X,Y}(A \times B) = \mathbb{P}_X(A)\mathbb{P}_Y(B),$$

qui est assuré par l'unicité de la mesure (de probabilité) produit. ∎

Ce résultat nous indique que l'on peut effectivement construire des couples (et même des $n$-uplets en itérant) de variables aléatoires réelles indépendantes en considérant l'espace produit, munis des tribus produit et des (mesures de) probabilités produit. Il est malheureusement beaucoup plus délicat, mais indispensable pour les applications (en particulier la loi des grands nombres), de construire une **suite infinie de variables indépendantes** de lois données.

Plus précisément, pour chaque entier $n$ on se donne une v.a.r. $X_n$ définie sur un espace de probabilité $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$, à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et de loi $\mathbb{P}_{X_n}$ (pour construire chaque $X_n$, on peut procéder comme ci-dessus). Ensuite, on pose

$$\begin{aligned} \Omega &= \prod_{n=1}^{\infty} \Omega_n \text{ (produit cartésien dénombrable)}, \\ \mathcal{A} &= \bigotimes_{n=1}^{\infty} \mathcal{A}_n, \end{aligned}$$

où $\bigotimes_{n=1}^{\infty} \mathcal{A}_n$ désigne la plus petite tribu de $\Omega$ à laquelle appartienne tous les ensembles de la forme

$$A_1 \times A_2 \times \ldots \times A_k \times \Omega_{k+1} \times \Omega_{k+2} \times \ldots, \quad A_i \in \mathcal{A}_i, \quad k \in \mathbb{N}^*$$

On a alors le théorème suivant, que l'on admettra, qui constitue un résultat non trivial de la théorie de la mesure et qui généralise le théorème d'unicité de la mesure produit.

**Théorème – Théorème — mesure produit sur un espace de dimension infinie**

Avec les notations ci-dessus, il existe une unique probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{A})$, telle que

$$\mathbb{P}(A_1 \times \ldots \times A_k \times \Omega_{k+1} \times \ldots) = \prod_{i=1}^{k} \mathbb{P}_i(A_i)$$

pour tous $k \in \mathbb{N}^*$ et $A_i \in \mathcal{A}_i$, $i \in \{1, \ldots, k\}$.

Nous sommes maintenant en mesure de définir une suite de variables aléatoires toutes définies sur le même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

**Définition – Définition — suites de variables aléatoires indépendantes**

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

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

**Proposition – Proposition — conséquences**

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 – 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 $\omega$, nous associons son développement dyadique (unique si l'on impose que les $\omega_i$ 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 $X_i : \Omega \to \{0, 1\}$, qui à $\omega$ associe $X_i(\omega) = \omega_i$ est une variable aléatoire sur $\Omega$. En effet, pour $x_i \in \{0, 1\}$, $1 \le i \le n$,

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

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

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

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 $\Omega$ 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 $\frac{1}{2}$ .

On note enfin le résultat très utile suivant portant sur les suites d'événements indépendants.

**Lemme – Lemme de Borel-Cantelli**

Soit $(A_n)_{n\in\mathbb{N}^*}$ une suite d'événements sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

1. Si $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, alors $\mathbb{P}(\limsup_{n\to\infty} A_n) = \mathbb{P}\left(\bigcap_{n\geq 1} \bigcup_{k\geq n} A_n\right) = 0$.
2. Si $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ et si les événements $A_n$ sont mutuellement indépendants, alors on a $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$.

**Démonstration** Exercice. ■

# Convergences et loi des grands nombres

## Convergences des variables aléatoires

Dans ce paragraphe, on va décrire les notions de convergence de variables (on y inclut les vecteurs) aléatoires. On verra que plusieurs notions sont possibles, non équivalentes, ce qui enrichit mais complique aussi la description des comportements asymptotiques.

En calcul intégral, on a beaucoup étudié le cas de suites de fonctions convergeant simplement. Une variable aléatoire étant une fonction, on a donc la même notion qui revient à écrire qu'une suite $(X_n)_{n\in\mathbb{N}^*}$ de variables aléatoires converge simplement vers $X$ si $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ pour tout $\omega \in \Omega$. Cette définition naturelle est malheureusement à peu près inutile en probabilité, comme l'illustre l'exemple suivant.

**Exemple – Exemple** Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de v.a. réelles qui sont indépendantes et de même loi (on notera *i.i.d* pour *indépendantes et identiquement distribuées*), avec $\mathbb{P}(X_n = 1) = p$ et $\mathbb{P}(X_n = 0) = 1 - p$. Ce sont donc des v.a. de loi de Bernoulli de paramètre $p$ qui modélisent par exemple les résultats d'un jeu de pile ou face. Lorsque $n$ est grand, on s'attend à ce que la proportion de faces ($X_n = 1$) soit à peu près égale à $p$ (c'est l'essence de la conception objectiviste des probabilités). Mathématiquement, on voudrait que

$$\lim_{n\to\infty} \frac{X_1(\omega) + \ldots + X_n(\omega)}{n} = p \text{ pour tout } \omega \in \Omega.$$

C'est en fait complètement faux. En effet, si on considère par exemple la suite $\omega_0 = p, p, p, \ldots$ qui ne contient que des piles, on obtient

$$\lim_{n\to\infty} \frac{X_1(\omega_0) + \ldots + X_n(\omega_0)}{n} = 0$$

et plus généralement, on a $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i(\omega) = 0$ pour tout $\omega$ dans l'ensemble $A = \{\omega : \text{ il n'y a qu'un nombre fini de faces}\}$. On peut même trouver des $\omega$ pour lesquels la fréquence converge vers n'importe quel nombre fixé dans $[0,1]$ ! Bien entendu, l'événement $A$ est invraisemblable et on peut montrer qu'il vérifie $\mathbb{P}(A) = 0$ [1]. On verra cependant que la **loi des grands nombres** assure :

$$\mathbb{P}\left(\left\{\omega; \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i(\omega) = p\right\}\right) = 1.$$

Ce type de convergence, pour lequel on n'a pas convergence pour tout $\omega$, mais seulement pour presque tout $\omega$ (autrement dit $\mathbb{P}$-presque partout) est typiquement ce qui arrive pour des variables aléatoires.

On considère une suite $(X_n)_{n\in\mathbb{N}^*}$ de variables 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 valeurs absolue dans $\mathbb{R}$ ou la norme euclidienne dans $\mathbb{R}^d$.

**Définition – Définition — convergences**

1. La suite $(X_n)_{n\in\mathbb{N}^*}$ *converge presque sûrement* vers $X$, ce qui s'écrit $X_n \to X$ p.s., s'il existe un ensemble $N \in \mathcal{A}$ de probabilité nulle tel que
$$X_n(\omega) \xrightarrow[n\to\infty]{} X(\omega), \ \ \forall \omega \notin N.$$
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 $\forall \varepsilon > 0$ on a
$$\mathbb{P}(|X_n - X| \geq \varepsilon) \xrightarrow[n\to\infty]{} 0.$$
3. La suite $(X_n)_{n\in\mathbb{N}^*}$ *converge en moyenne* (ou dans $\mathcal{L}^1$) 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
$$\mathbb{E}(|X_n - X|) \xrightarrow[n\to\infty]{} 0.$$

1. Considérer les événements $A_n = \{\text{on a face au } n\text{-ième tirage}\}$, puis on montre que $\sum_{n=1}^\infty \mathbb{P}(A_n) = \infty$ et on conclut en invoquant le lemme de Borel-Cantelli.

**Remarque – Remarque**

La définition de la convergence dans $\mathcal{L}^1$ se généralise aux ordres supérieurs, pour $p \in \mathbb{N}^*$, on parle de convergence dans $\mathcal{L}^p$ (en moyenne quadratique si $p = 2$) ce qui s'écrit $X_n \xrightarrow{\mathcal{L}^p} X$, si $X_n$ et $X$ sont dans $\mathcal{L}^p$ et si

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

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

**Exemple – Exemples**

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

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

  Pour tout $\varepsilon \in ]0, 1[$, la probabilité $\mathbb{P}(|X_n| \geq \varepsilon) = \frac{1}{n}$ 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 $\mathbb{E}(X_n) = \frac{1}{n}$, elle tend également en moyenne vers 0.
  Mais si on considère maintenant une suite $(Y_n)_{n\in\mathbb{N}^*}$ de variables aléatoires de Bernoulli à valeurs dans $\{0, n^2\}$ telles que

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

  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

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

  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.

On étudie maintenant les liens entre ces différentes convergences.

**Proposition – Proposition — p.s. / $\mathcal{L}^1$ $\Rightarrow$ en proba**

La convergence presque sûre et la convergence en moyenne entraînent la convergence en probabilité.

**Démonstration** Soient $n \in \mathbb{N}^*$, $\varepsilon > 0$ et $A_{n,\varepsilon} = \{|X_n - X| > \varepsilon\}$.

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

$$\mathbb{P}(A_{n,\varepsilon}) \le \mathbb{P}(N^c \cap A_{n,\varepsilon}) + \mathbb{P}(N) = \mathbb{P}(N^c \cap A_{n,\varepsilon}) = \mathbb{E}(Y_{n,\varepsilon}) \xrightarrow[n\to\infty]{} 0.$$

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

$$\mathbb{P}(A_{n,\varepsilon}) \le \frac{1}{\varepsilon}\mathbb{E}(|X_n - X|) \xrightarrow[n\to\infty]{} 0.$$

■

La convergence en probabilité n'entraîne pas la convergence en moyenne, comme nous l'avons vu dans l'exemple précédent, 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 – Proposition — cas borné**

S'il existe une constante $a$ telle que $|X_n| \le a$ 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 (p. 10), dont on reprend les notations, il suffit de montrer que la convergence en probabilité implique la convergence en moyenne lorsque $|X_n| \le a$.

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

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

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

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

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

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

On en déduit que $\limsup_{n\to\infty} \mathbb{E}(|X_n - X|) \le \varepsilon$, et comme $\varepsilon$ 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 plus haut (p. 10), mais "à peine plus", comme le montre le résultat suivant.

**Proposition – Proposition — Existence d'une sous-suite convergente**

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

**Démonstration** On remarque d'abord que l'on peut réécrire la définition de la convergence en probabilité de la manière suivante : $\forall \varepsilon > 0$ et $\delta > 0$, il existe $N = N(\delta, \varepsilon)$ tel que

$$n \geq N \Rightarrow \mathbb{P}(|X_n - X| > \varepsilon) < \delta.$$

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

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

On a alors

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

et en appliquant le lemme de Borel-Cantelli (p. 8) aux ensembles

$$A_j = \left\{|X_{n_j} - X| > \frac{1}{2^j}\right\},$$

on obtient que la suite $(X_{n_j})_{j\in\mathbb{N}^*}$ converge presque-sûrement. ■

**Exemple – 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 $\omega$, et on n'a pas $X_n \to X$ presque-sûrement ; cependant comme la série $\sum_n 1/n^2$ converge, il s'en suit que $X_{n^2} \to X = 0$ presque-sûrement. Nous avons donc la convergence presque-sûre de la sous-suite $(X_{n^2})_{n\in\mathbb{N}^*}$.

**Proposition – Proposition — continuité**

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

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

**Démonstration**

1. Soit $N$ l'ensemble de probabilité nulle en dehors duquel on a $X_n(\omega) \xrightarrow[n\to\infty]{} X(\omega)$. Si $\omega \notin N$, il vient

$$\lim_{n\to\infty} f(X_n(\omega)) = f(\lim_{n\to\infty} X_n(\omega)) = f(X(\omega))$$

par continuité de $f$, d'où le résultat.

2. On remarqe d'abord que si $K > 0$ et $\varepsilon > 0$,

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

La fonction $f$ est uniformément continue sur $\{x : |x| \leq K\}$, donc il existe $\eta > 0$ tel que $|x - y| < \eta$ et $|x| \leq 2K$ et $|y| \leq 2K$ impliquent $|f(x) - f(y)| < \varepsilon$. On a donc

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

d'où

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

Par hypothèse, il vient

$$\lim_{n\to\infty} \mathbb{P}(|f(X_n) - f(X)| \geq \varepsilon) \leq \mathbb{P}(|X| > K).$$

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

■

## La loi des grands nombres

On présente maintenant 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 p.s. vers la probabilité de réalisation de cet événement. Ce résultat, appelé **Loi des grands nombres**, a d'autres portées fondamentales. Il est en particulier 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, en Mathématiques Financières, dans les méthodes de quantification d'incertitudes.

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 :

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

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 $n$).

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

**Théorème – Théorème — loi des grands nombres cas $\mathcal{L}^2$**

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 espérance. Alors la suite $(M_n)_{n\in\mathbb{N}^*}$

définie par

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

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 :

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

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 (p. 5) (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 (p. 8).

### Démonstration

Notons $\sigma^2$ 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

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

d'où la convergence en moyenne quadratique.

Comme $\mathbb{E}(Y)^2 \leq \mathbb{E}(Y^2)$, on en déduit que $\mathbb{E}(|M_n - m|) \to 0$, 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.

La convergence dans $\mathcal{L}^1$ impliquant la convergence en probabilité, on sait qu'on peut extraire de $(M_n)_n$ une sous-suite convergeant p.s. vers 0. Cependant, cela ne suffit pas puisqu'on veut que la suite $(M_n)_n$ elle-même converge p.s. Pour le montrer, on construit d'abord une sous suite-particulière qui converge p.s. puis on traite les termes qui se trouvent entre deux éléments successifs de la sous-suite.

D'après l'inégalité de Bienaymé-Chebyshev (p. 5) et ce qui précède, on a pour $q \in \mathbb{N}^*$

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

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 (p. 8), on obtient que $\mathbb{P}(C_q) = 0$. En conséquence, si on pose $N = \cup_{q\in\mathbb{N}^*} C_q$, on obtient $\mathbb{P}(N) \leq \sum_{q\in\mathbb{N}^*} \mathbb{P}(C_q) = 0$.

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 siginfie 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, $M_{n^2} \to 0$ si $\omega \notin N$, avec $\mathbb{P}(N) = 0$, d'où

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

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 puique les variables aléatoires de la somme sont indépendantes, il vient

$$\begin{aligned}
\mathbb{E}\left(\left(M_n - \frac{p(n)^2}{n} M_{p(n)^2}\right)^2\right) &= \frac{n - p(n)^2}{n^2}\sigma^2 \\
&\leq \frac{2p(n)+1}{n^2}\sigma^2 \\
&\leq \frac{2\sqrt{n}+1}{n^2}\sigma^2 \leq \frac{3}{n^{3/2}}\sigma^2
\end{aligned}$$

où on a utilisé le fait que $p(n) \leq \sqrt{n}$.

En appliquant de nouveau l'inégalité de Bienaymé-Chebyshev (p. 5), 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

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

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

Plus généralement, le théorème suivant donne les hypothèses minimales assurant la validité de la loi des grands nombres, à savoir que les $X_n$ sont dans $\mathcal{L}_1$ (on se référera par exemple à ce document en ligne ou à Jacod and Protter (2003) pour la démonstration).

**Théorème – Théorème — loi des grands nombres cas $\mathcal{L}^1$**

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 espérance. Alors la suite $(M_n)_{n\in\mathbb{N}^*}$ définie par

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

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

# 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 – 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 réelle $f$ **continue bornée** sur $\mathbb{R}^d$,

$$\mathbb{E}(f(X_n)) \xrightarrow[n\to\infty]{} \mathbb{E}(f(X)).$$

**Exemple – 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

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

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. En particulier, le cas de la convergence de la loi binomiale vers la loi de Poisson a été traité en CPGE.

### Remarque – Remarque

Dans la définition (p. 16), les v.a. $X_n$ et $X$ peuvent être définies sur des univers distincts puisque seules leurs lois sont en cause. Il arrive même qu'une suite $X_n$ converge vers une limite $X$ qui ne peut pas exister sur les espaces sur lesquels sont définies les $X_n$, parce que ceux-ci sont trop "petits" : par exemple, si $X_n$ est une variable binomiale à $n$ modalités, convenablement normalisée, et la limite $X$ est gaussienne (on pourra justifier ceci par le théorème central limite (p. 25));

l'espace naturel sur lequel est définie $X_n$ contient $n+1$ points, et sur un tel espace toutes les v.a. sont discrètes. La convergence en loi permet donc une sorte de convergence pour des v.a. pour lesquelles toute autre forme de convergence serait impossible.

**Exemple – 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

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

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 – Proposition — en proba $\Rightarrow$ en loi**

Soient $(X_n)_{n\in\mathbb{N}*}$ et $X$ des v.a., toutes définies sur le même espace de probabilité $(\Omega,\mathcal{A},\mathbb{P})$. Si $X_n \xrightarrow{\mathbb{P}} X$, alors $X_n \xrightarrow{\mathcal{L}} X$.

**Démonstration** Soit $f$ une fonction réelle continue bornée. D'après la proposition de continuité (p. 12), 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 — cas borné (p. 11). Comme $|\mathbb{E}(Y)|\leq\mathbb{E}(|Y|)$ pour toute variable aléatoire réelle $Y$, on en déduit $\mathbb{E}(f(X_n))\to\mathbb{E}(f(X))$. ■

On peut finalement résumer les implications entre les différents modes de convergence à l'aide de la figure suivante :

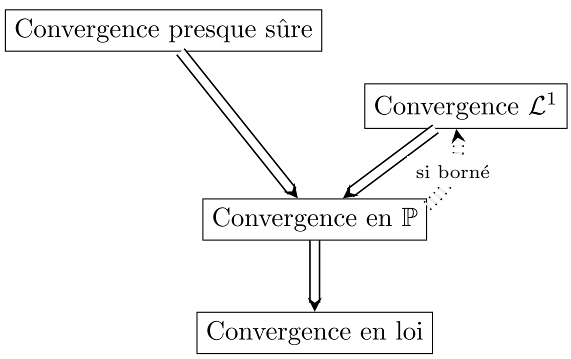


Figure 1 – Relations entre modes de convergence

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

**Proposition – Proposition — convergence en loi et fonction de répartition**

Soient $(X_n)_{n\in\mathbb{N}^*}$ 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 $F_n(x) \underset{n\to\infty}{\longrightarrow} F(x)$ 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 [2] 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, $\mathbb{E}(f_{p,b}(X_n)) \to \mathbb{E}(f_{p,b}(X))$ 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 $\liminf_n F_n(a) \geq F(a)$, puique $F(a-) = F(a)$. Ces deux résultats impliquent que $F_n(a) \underset{n\to\infty}{\longrightarrow} F(a)$.
2. Inversement, supposons $F_n(x) \underset{n\to\infty}{\longrightarrow} F(x)$ pour tout $x \in T$, où $T$ est une partie dense de $\mathbb{R}$. Soit $f$ une fonction continue bornée sur $\mathbb{R}$ et $\varepsilon > 0$. Soient $a, b \in T$ avec $F(a) \leq \varepsilon$ et $F(b) \geq 1-\varepsilon$. Il existe $n_0$ tel que
$$n \geq n_0 \Rightarrow \mathbb{P}(X_n \notin ]a,b]) = 1 - F_n(b) + F_n(a) \leq 3\varepsilon.$$
La fonction $f$ est uniformément continue sur $[a,b]$, donc il existe un nombre fini de points $a_0 = a < a_1 < \ldots < a_k = b$ appartenant tous à $T$, et tels que $|f(x) - f(a_i)| \leq \varepsilon$ 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 $|f-g| \leq \varepsilon$ sur $]a,b]$. Si $M = \sup_x |f(x)|$, il vient alors
$$|\mathbb{E}(f(X_n)) - \mathbb{E}(g(X_n))| \leq M\mathbb{P}(X_n \notin [a,b]) + \varepsilon,$$

2. Soient $\Omega$ un espace topologique et $A$ une partie de $\Omega$. On dit que $A$ est *dense* dans $\Omega$ si l'une des propriétés équivalentes est satisfaite : tout ouvert non vide de $\Omega$ contient des éléments de $A$ ; l'adhérence de $A$ est égale à $\Omega$ ; tout point de $\Omega$ est adhérent à $A$ ; le complémentaire de $A$ est d'intérieur vide.

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

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

Finalement, on a

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

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

■

**Corollaire – 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 $a < b$,

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

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

## Fonctions caractéristiques

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 nécessaire, via le théorème de Lévy, pour démontrer le théorème central limite. L'essentiel de cette section peut être considéré hors programme, dans le sens où, bien que très utile en pratique, sa connaissance ne sera pas évaluée à l'examen.

On notera $< x, y >$ le produit scalaire de deux vecteurs de $\mathbb{R}^n$ . Si $u \in \mathbb{R}^n$ , la fonction (complexe) $x \mapsto e^{i<u,x>}$ est continue, de module 1. Donc si $X$ est un vecteur aléatoire à valeurs dans $\mathbb{R}^n$ , nous pouvons considérer $e^{i<u,X>}$ comme une variable aléatoire à valeurs complexes. Ses parties réelle $Y = \cos(< u, X >)$ et imaginaire $Z = \sin(< u, X >)$ 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 $e^{i<u,x>}$ est

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

**Définition – Définition**

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

$$\phi_X(u) = \mathbb{E}(e^{i<u,X>}) = \int_{\mathbb{R}^n} e^{i<u,x>}\mathbb{P}_X(dx).$$

**Remarque – 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 $G_X(s) = \mathbb{E}(s^X)$, $s \in [0,1]$, vue en CPGE dans le cas discret. Elle vérifie

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

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

**Proposition – Proposition — propriétés de la fonction caractéristique**

$\phi_X$ 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 :

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

Pour montrer la continuité, considérons une suite $u_p \xrightarrow[p\to\infty]{} u$. Il y a convergence simple de $e^{i<u_p,X>}$ vers $e^{i<u,X>}$. Comme ces variables aléatoires sont de module inférieur à 1, le théorème de convergence dominée assure que $\phi_X(u_p) \xrightarrow[p\to\infty]{} \phi_X(u)$. $\phi_X$ est donc continue. ■

**Proposition – Proposition — transformation linéaire**

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

$$\phi_{a+AX}(u) = e^{i<u,a>}\phi_X(A^t u), \forall u \in \mathbb{R}^m$$

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

**Exemple – 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 $\mathcal{U}_{[a,b]}$ :$\phi_X(u) = \frac{e^{iua}-e^{iub}}{iu(b-a)}$.
4. $X$ suit une loi exponentielle $\mathcal{E}(\lambda)$, $\lambda > 0$ : $\phi_X(u) = \frac{\lambda}{\lambda - iu}$.

5. $X$ suit une loi normale $\mathcal{N}(0,1)$ : $\phi_X(u) = e^{-u^2/2}$. (calcul en annexe (p. 28))
6. $X$ suit une loi normale $\mathcal{N}(\mu,\sigma^2)$ : $\phi_X(u) = e^{iu\mu - u^2\sigma^2/2}$. (application directe de la proposition ci-dessus (p. 20))

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 – Théorème — caractérisation**

La fonction caractéristique $\phi_X$ 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 $\sigma > 0$ :

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

On note que $f_\sigma$ est la densité d'un vecteur gaussien $Z$ de dimension $n$, centré et de matrice de covariance $\sigma^2 I_n$, autrement dit dont les composantes sont indépendantes, centrées, de variances $\sigma^2$.

On a

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

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

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

Supposons que $X$ et $Y$ admettent la même fonction caractéristique $\phi_X = \phi_Y$. En utilisant le théorème de Fubini, on a

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

et la même égalité reste vraie pour $\mathbb{P}_Y$. On en déduit que

$$\mathbb{E}(g(X)) = \int g(u)\mathbb{P}_X(du) = \int g(u)\mathbb{P}_Y(du) = \mathbb{E}(g(X'))$$

pour toute fonction $g$ de l'espace vectoriel de fonction engendrées par $u \mapsto f_\sigma(u-v)$.

D'après le théorème de Stone-Weiertrass [3], 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 (la norme est le sup sur $\mathbb{R}^n$).

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 – Corollaire**

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

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

où $\phi_X$ désigne la fonction caractéristique du vecteur aléatoire $X$, et $\phi_{X_j}$ celle de la composante $X_j$ pour chaque $j$.

**Démonstration** On a $< u, X >= \sum_{j=1}^n u_j X_j$. Si les $X_i$ sont indépendantes, et comme $e^{i<u,x>} = \prod_j e^{iu_j x_j}$, nous obtenons directement le résultat en utilisant la proposition (valable dans le cas général) qui montre que l'espérance du produit de fonctions de variables aléatoires indépendantes est égal au produit des espérances.

Supposons inversement qu'on ait $\phi_X(u_1, \ldots, u_n) = \prod_{j=1}^n \phi_{X_j}(u_j)$. 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 $X' = (X'_1, \ldots, X'_n)$, 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

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

d'où l'indépendance. ∎

3. voir par exemple Simmons (2003) p.160.

**Proposition – Proposition — somme**

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 $e^{i<u,X+Y>} = e^{i<u,X>} e^{i<u,Y>}$, il suffit d'appliquer la proposition déjà utilisée dans la preuve du corollaire ci-dessus. ■

**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 ci-dessus (p. 20) point 6. et la proposition ci-dessus (p. 23).
2. Si $X$ et $Y$ suivent des lois de Poisson de paramètres $\theta$ et $\theta'$, alors $Z$ suit une loi de Poisson de paramètre $\theta + \theta'$, d'après l'exemple ci-dessus (p. 20) point 2. et la proposition ci-dessus (p. 23).
3. Si $X$ suit une loi binomiale $\mathcal{B}(n, p)$ et $Y$ la loi biomiale $\mathcal{B}(m, p)$, alors $Z$ suit une loi binomiale $\mathcal{B}(n + m, p)$, d'après l'exemple ci-dessus (p. 20) point 1. et la proposition ci-dessus (p. 23).

**Proposition – Proposition — fonction caractéristique et moments**

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

$$\frac{\partial^m}{\partial u_{i_1} \ldots \partial u_{i_m}} \phi_X(u) = i^m \mathbb{E}(e^{i<u,X>} X_{i_1} \ldots X_{i_m}),$$

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 – Remarque**

En prenant $u = 0$ dans la proposition — somme (p. 23), la formule permet de calculer $\mathbb{E}(X_{i_1} \ldots X_{i_m})$ en fonction des dérivées à l'origine de $\phi_X$, 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), on a

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

On donne ici la définition générale d'un vecteur gaussien, qui nous sera utile pour la démonstration du théorème central limite.

**Définition – Définition**

Un vecteur aléatoire $X = (X_1, \dots, X_n)$ à valeurs dans $\mathbb{R}^n$ est dit gaussien si toute combinaison linéaire de ses composantes, soit $\sum_{j=1}^n a_j X_j$ où $a_j \in \mathbb{R}$, suit une loi normale uni-dimensionnelle (éventuellement dégénérée, par exemple si on prend $a_j = 0$ pour tout j).

**Théorème – Théorème — cas gaussien**

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

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

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

**Démonstration**

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

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

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

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

$$\mathbb{E}(Y) =< a, m >, \mathbb{V}(Y) =< a, Ca >$$

Par hypothèse, $Y$ suit une loi normale donc vu le point 6. de l'exemple plus haut (p. 20), sa fonction caractéristique est

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

Mais $\phi_Y(1) = \phi_{<a,X>}(1) = \mathbb{E}(e^{i<a,X>}) = \phi_X(a)$, 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 – 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 $\phi_{X_n}$ converge simplement vers $\phi_X$.
2. Si les $\phi_{X_n}$ convergent simplement vers une fonction (complexe) $\phi$ 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 $g_u(x) = e^{i<u,x>}$. On applique alors la définition (p. 16).
2. Se reporter à Jacod and Protter (2003).

■

## 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 aléatoire $(X_n)_{n\in\mathbb{N}^*}$ indépendantes, de même loi et de carré intégrable. On note $m$ et $\sigma^2$ l'espérance et la variance commune aux variables $X_n$, et

$$S_n = X_1 + \ldots + X_n$$

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 a 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 amené à étudier la limite éventuelle de la suite $n^\alpha(\frac{S_n}{n} - m)$ pour différentes valeurs de $\alpha$ : si $\alpha$ est "petit" cette suite va encore tendre vers 0, et elle va "exploser" si $\alpha$ est "grand". On peut espérer que pour une (et alors nécessairement une seule) valeur de $\alpha$, 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 $\alpha$. 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).

**Théorème – Théorème central limite**

Si les $X_n$ sont des variables aléatoires réelles, indépendantes et de même loi, de carré intégrable, d'espérance $m$ 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 $\phi$ la fonction caractéristique de $X_n - m$, et $Y_n = \frac{S_n - nm}{\sigma\sqrt{n}}$. Comme les $X_n$ sont indépendantes, la proposition — somme (p. 23) entraîne que la fonction caractéristique de $Y_n$ est

$$\begin{aligned}
\phi_{Y_n}(u) &= \phi_{\frac{1}{\sigma\sqrt{n}}\sum_{j=1}^n (X_j - m)}(u) \\
&= \phi_{\sum_{j=1}^n (X_j - m)}\left(\frac{u}{\sigma\sqrt{n}}\right) \\
&= \prod_{j=1}^n \phi\left(\frac{u}{\sigma\sqrt{n}}\right) \\
&= \phi\left(\frac{u}{\sigma\sqrt{n}}\right)^n
\end{aligned}$$

Comme $\mathbb{E}(X_n - m) = 0$ et $\mathbb{E}((X_n - m)^2) = \sigma^2$, la proposition — fonction caractéristique et moments (p. 23) entraîne que

$$\phi'(0) = 0 \text{ et } \phi''(0) = -\sigma^2$$

Si on fait le développement de Taylor à l'ordre 2 au voisinage de zéro, on obtient

$$\phi(u) = 1 - \frac{u^2\sigma^2}{2} + u^2 h(|u|),$$

où $h(u) \to 0$ quand $u \to 0$. On a ainsi

$$\begin{aligned}
\phi_{Y_n}(u) &= \phi\left(\frac{u}{\sigma\sqrt{n}}\right)^n \\
&= e^{n \log \phi\left(\frac{u}{\sigma\sqrt{n}}\right)} \\
&= e^{n \log(1 - \frac{u^2}{2n} + \frac{u^2}{n\sigma^2} h(\frac{u}{\sigma\sqrt{n}}))}
\end{aligned}$$

où log désigne la valeur principale du logarithme complexe [4] (elle vaut 0 au point 1 et est continue dans le cercle complexe de centre 1 et de rayon 1/2 et admet le même développement limité au voisinage de $z = 1$ que le logarithme réel). Comme $\phi(0) = 1$ et que $\phi$ est continue en 0, on a que pour $u$ fixé et $n$ assez grand,

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

En faisant $n \to \infty$, on obtient

$$\lim_{n\to\infty} \phi_n(u) = \exp -\frac{u^2}{2}$$

Le théorème de Lévy (p. 24) implique alors que $Y_n$ converge en loi vers $Z$ de fonction caractéristique $\phi_Z(u) = e^{-u^2/2}$ où l'on reconnaît la fonction caractéristique de la loi $\mathcal{N}(0, 1)$. ■

**Remarque – 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. $\alpha > 1/2$).

4. voir par exemple https://fr.wikipedia.org/wiki/Logarithme_complexe

**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 $X_1+\ldots+X_n$ 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 $\theta$.

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

$$\frac{S_n}{n}\xrightarrow{\mathbb{P}}p,$$
$$\frac{S_n-np}{\sqrt{np(1-p)}}\xrightarrow{\mathcal{L}}Y\sim\mathcal{N}(0,1)$$

Si on désigne par $\Phi$ 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 $\Phi$ sont tabulées) et on obtient

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

Le théorème (p. 25) 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 espérance $m=E(X_n)$, et une matrice de covariance $C=(c_{ij})_{i,j=1,\ldots,d}$ avec $c_ij=\mathrm{Cov}(X_i,X_j)$. On peut alors énoncer le TCL multi-dimensionnel.

**Théorème – 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. d'espérance nulle), de matrice $C$.

**Remarque – Remarque**

Il est important de noter ici que la vitesse de convergence ne dépend pas de la dimension des vecteurs $X_n$.

# Annexe

## Fonction caractéristique d'une variable gaussienne

Soit $X \sim \mathcal{N}(0,1)$. On a

$$\begin{aligned}\phi_X(u) = \mathbb{E}(e^{iuX}) &= \int_{\mathbb{R}} e^{iux} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= \int_{\mathbb{R}} \frac{\cos(ux)}{\sqrt{2\pi}} e^{-x^2/2} dx + \int_{\mathbb{R}} \frac{i \sin(ux)}{\sqrt{2\pi}} e^{-x^2/2} dx\end{aligned}$$

Comme $x \mapsto \frac{\sin(ux)}{\sqrt{2\pi}} e^{-x^2/2}$ est impaire et intégrable, son intégrale est nulle, d'où

$$\phi_X(u) = \int_{\mathbb{R}} \frac{\cos(ux)}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

D'après la proposition — fonction caractéristique et moments (p. 23), on peut dériver les deux membres par rapport à $u$, et on obtient

$$\phi'_X(u) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} -x \sin(ux) e^{-x^2/2} dx$$

puis par intégration par parties

$$\phi'_X(u) = -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} u \cos(ux) e^{-x^2/2} dx = -u\phi_X(u).$$

Ainsi, $\phi_X$ satisfait à l'équation différentielle $\phi'_X(u) = -u\phi_X(u)$, dont la solution générale est

$$\phi_X(u) = Ce^{-u^2/2}.$$

Comme $\phi_X(0) = 1$, on en déduit finalement que

$$\phi_X(u) = e^{-u^2/2}.$$

# Exercices

## Inégalités de concentration

**Question 1** Soit $(X_n)_n$ une suite de variables aléatoires indépendantes et identiquement distribuées d'espérance $\mathbb{E}(X_i) = m$ et de variance $\mathbb{V}(X_i) = \sigma^2 \leq 1$. Montrer que pour tout $\delta \in ]0,1[$, avec probabilité au moins $1-\delta$ on a

$$\left| \frac{1}{n} \sum_{i=1}^{n} X_i - m \right| \leq \sqrt{\frac{1}{\delta n}}$$

On peut trouver des bornes à décroissance beaucoup plus rapide dans des cas particuliers. (Solution p. 33.)

**Question 2** On suppose désormais que les $X_i$ sont de loi $\mathcal{N}(m, \sigma^2)$.

i) On pose pour $u \in \mathbb{R} : M(u) = \mathbb{E}(e^{u(X_1-m)})$. Calculer $M(u)$.

ii) On pose $S_n = X_1 + \ldots + X_n$ . Montrer que $\forall a \in \mathbb{R}, \forall u \in \mathbb{R}_+$,

$$\mathbb{P}(S_n - nm > a) \leq e^{-ua}(M(u))^n,$$

et que $\forall a \in \mathbb{R}, \forall u \in \mathbb{R}_-$,

$$\mathbb{P}(S_n - nm < -a) \leq e^{ua}(M(u))^n.$$

iii) Soit $Y_n = \frac{S_n}{n}$. Montrer que $\forall \varepsilon > 0$,

$$\mathbb{P}(|Y_n - m| > \varepsilon) \leq 2 \exp\left(\frac{-n\varepsilon^2}{2\sigma^2}\right).$$

Cette inégalité est appelée inégalité de Chernov.

On peut dériver ce type d'inégalités pour différentes lois de probabilité. On voit qu'ici la concentration auprès de la moyenne se fait à vitesse exponentielle. On a le même type de résultats pour la loi de Bernoulli par exemple ce qui est très utile en apprentissage statistique dans les problèmes de classification. (Solution p. 33.)

**Question 3**

iv) On suppose que $m = 1$ et que $\sigma^2 = 10$. Quelle taille d'échantillon doit-on choisir pour obtenir

$$\mathbb{P}(|Y_n - m| < \varepsilon) \geq \alpha,$$

avec $\alpha = 0,95$ et $\varepsilon = 0,05$,

— en utilisant l'inégalité de Bienaymé-Chebyshev ?

— en utilisant l'inégalité de Chernov ?

(Solution p. 33.)

## Lemme de Borel-Cantelli

Soit $A_n$ une suite d'événements sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

**Question 1** On suppose que $\sum_{i=1}^{n} \mathbb{P}(A_n) < \infty$. Montrer que $\mathbb{P}(\limsup_{n\to\infty} A_n) = \mathbb{P}\left(\bigcap_{n\geq 1} \bigcup_{k\geq n} A_n\right) = 0$. (Solution p. 34.)

**Question 2** On suppose maintenant que les événements $A_n$ sont mutuellement indépendants. Montrer que si $\sum_{i=1}^{n} \mathbb{P}(A_n) = \infty$, alors on a $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$. (Solution p. 34.)

**Question 3** Donner un exemple où $\sum_{i=1}^{n} \mathbb{P}(A_n) = \infty$ et $\mathbb{P}(\limsup_{n\to\infty} A_n) < 1$ quand les $A_n$ ne sont pas indépendants. (Solution p. 34.)

**Question 4** On considère le jeu de pile ou face infini de l'exemple en début de cours (p. 9). Montrer que l'événement $A = \{\omega : \text{ il n'y a qu'un nombre fini de faces}\}$ est de probabilité nulle. (Considérer les événements $A_n = \{\text{on a face au } n\text{-ième tirage}\}$, puis montrer que $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$). (Solution p. 34.)

## Convergence vers une constante

Soient $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles, $X$ une autre variable aléatoire réelle et $a \in \mathbb{R}$.

**Convergence $\mathcal{L}^2$** On suppose que $X$ et chaque $X_n$, $n \in \mathbb{N}^*$, sont de carré intégrable.

1. Montrer l'implication

$$X_n \xrightarrow[n\to+\infty]{\mathcal{L}^2} X \Rightarrow \left| \begin{array}{l} \mathbb{E}\left(X_n\right) \xrightarrow[n\to+\infty]{} \mathbb{E}(X), \\ \mathbb{V}\left(X_n\right) \xrightarrow[n\to+\infty]{} \mathbb{V}(X). \end{array} \right.$$

2. Vérifier l'équivalence

$$X_n \xrightarrow[n\to+\infty]{\mathcal{L}^2} a \Leftrightarrow \left| \begin{array}{l} \mathbb{E}\left(X_n\right) \xrightarrow[n\to+\infty]{} a, \\ \mathbb{V}\left(X_n\right) \xrightarrow[n\to+\infty]{} 0. \end{array} \right.$$

(Solution p. 34.)

**Convergences en loi et probabilité** Montrer que si $X_n$ converge en loi vers $a$ quand $n \to +\infty$, alors elle converge aussi en probabilité vers $a$. (Solution p. 36.)

## Loi faible des grands nombres

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. Montrer que la suite $(M_n)_{n\in\mathbb{N}^*}$ définie par

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

converge vers $m$ en probabilité quand $n$ tend vers l'infini.

## Fonction de répartition empirique

Soient $X$ une variable aléatoire réelle de fonction de répartition $F$ et $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes, de même loi que $X$. Pour tout $n \in \mathbb{N}^*$ on définit la fonction de répartition empirique comme suit :

$$F_n : x \in \mathbb{R} \mapsto \frac{1}{n}\sum_{i=1}^{n} 1_{]-\infty,x]}(X_i).$$

Son nom est issu de la modélisation probabiliste en statistique. Lorsque l'on souhaite étudier une variable physique dont on n'est pas à même de prédire parfaitement les valeurs (e.g. la hauteur d'eau de la Seine à la station d'Austerlitz), le statisticien va la considérer comme une variable aléatoire $X$. Pour en retrouver les propriétés, il va observer un certain nombre $n \in \mathbb{N}^*$ de réalisations de cette variable (en réalisant des mesures sur le terrain), qu'il va à nouveau considérer comme des variables aléatoires $X_1, \dots, X_n$ de même loi que $X$. Lorsque cette hypothèse est raisonnable, il va supposer que ces dernières sont même indépendantes. Nous allons voir dans cet exercice que pour tout $x \in \mathbb{R}$, $F_n(x)$, calculée à partir des données de terrain (d'où le nom attribué à $F_n$), approche bien la quantité théorique $F(x) = \mathbb{P}(X \le x)$.

On considère $x \in \mathbb{R}$ fixé.

**Préliminaire** Quelle est la loi de la variable aléatoire $1_{]-\infty,x]}(X)$ ? Expliciter son espérance et sa variance. (Solution p. 36.)

**Biais** Soit $n \in \mathbb{N}^*$. Calculer l'espérance de $F_n(x)$. On dit que $F_n(x)$ est un *estimateur sans biais* de $F(x)$. (Solution p. 36.)

**Erreur quadratique moyenne** Montrer que $F_n(x) \xrightarrow{\mathcal{L}^2} F(x)$ quand $n \to +\infty$. (Solution p. 36.)

**Consistance** Montrer que $F_n(x) \to F(x)$ p.s. quand $n \to +\infty$. On dit que $F_n(x)$ est un *estimateur fortement consistant* de $F(x)$. (Solution p. 37.)

## Théorème de Weierstrass sur $[0, 1]$

Le théorème de Weierstrass est un résultat classique d'analyse dont l'énoncé est le suivant : soient $I$ un segment de $\mathbb{R}$ et $f : I \to \mathbb{R}$ une fonction continue. Alors il existe une suite $(P_n)_{n\in\mathbb{N}^*}$ de polynômes $\mathbb{R} \to \mathbb{R}$ convergeant uniformément vers $f$ sur $I$.

Dans cet exercice, nous allons voir une démonstration constructive de ce théorème dans le cas particulier où $I = [0, 1]$.

**Préliminaires** Nous allons avoir besoin de deux résultats intermédiaires pour établir la preuve du théorème de Weierstrass sur $[0, 1]$.

— **Théorème de convergence dominée.** Soient $(X, \mathcal{A}, \mu)$ un espace mesuré, $(f_n)_{n\in\mathbb{N}^*}$ une suite de fonctions mesurables $X \to [-\infty, +\infty]$ et $g : X \to [-\infty, +\infty]$ une fonction intégrable, telles que pour tout $n \in \mathbb{N}^*$ on a $|f_n| \leq g$ $\mu$-**presque partout**. Supposons qu'il existe $f : X \to [-\infty, +\infty]$ telle que $f_n$ converge simplement vers $f$ $\mu$-**presque partout** quand $n \to +\infty$. Alors $f$ est intégrable et

$$\int_X f_n \mu \xrightarrow[n\to+\infty]{} \int_X f\mu.$$

— **Inégalité de Jensen.** Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, $X : \Omega \to \mathbb{R}$ une variable aléatoire intégrable et $f : \mathbb{R} \to \mathbb{R}$ une fonction mesurable convexe, telle que $f(X) \in \mathcal{L}^1$. Alors

$$f(\mathbb{E}(X)) \leq \mathbb{E}(f(X)).$$

Démontrer ce résultat.

(Solution p. 37.)

**Preuve du théorème.** Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variable aléatoires indépendantes, suivant toutes la même loi de Bernoulli de paramètre $x \in ]0, 1[$. Pour tout $n \in \mathbb{N}^*$ on pose $M_n := \frac{1}{n} \sum_{i=1}^n X_i$. On considère une fonction $f : [0, 1] \to \mathbb{R}$ continue.

1. Soit $n \in \mathbb{N}^*$. Vérifier que $\mathbb{E}(f(M_n))$ peut s'écrire sous forme de polynôme.
2. Montrer que $\mathbb{E}(f(M_n)) \xrightarrow[n\to+\infty]{} f(x)$.
3. En déduire qu'il existe une suite de polynômes $[0, 1] \to \mathbb{R}$ qui convergent simplement vers $f$.
4. Montrer que la suite de polynômes exhibée à la question précédente converge même uniformément vers $f$.

(Solution p. 38.)

## Théorème de Slutsky

**Question 1** Soient $(X_n)_n$ et $X$ des variables aléatoires. Montrer que $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si $\mathbb{E}(f(X_n)) \xrightarrow[n\to\infty]{} \mathbb{E}(f(X))$ pour toute fonction $f$ lipschitzienne bornée. (Solution p. 40.)

**Question 2** Soient $(X_n)_n$ et $(Y_n)_n$ deux suites de variables aléatoires définies sur le même espace de probabilité. On suppose que $(X_n)_n$ converge en loi vers $X$ et que $(X_n - Y_n)_n$ converge vers 0 en probabilité. Montrer que $(Y_n)_n$ converge en loi vers $X$. (Solution p. 40.)

**Question 3** Montrer ce résultat à l'aide du théorème de Lévy (p. 24).

---

(Solution p. 40.)

# Solutions

## Inégalités de concentration

**Question 1** L'inégalité de Chebyshev nous donne pour tout $a > 0$

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - m\right| > a\right) \leq \frac{\mathbb{V}(X_1)}{na^2}$$

Prenant, $\frac{\mathbb{V}(X_1)}{na^2} = \delta$, on obtient $a = \frac{1}{\sqrt{n\delta}}$, d'où le résultat. On retrouve au passage la même vitesse de convergence que celle donnée par le TCL.

**Question 2**

i) $M(u) = \mathbb{E}(e^{u(X_1-m)}) = \phi_{X_1}(-iu) = e^{u^2\sigma^2/2}$, où $\phi_{X_1}$ est la fonction caractéristique de $X_1$.

ii) On a $e^{u(M_n-nm)} \geq e^{ua}1_{S_n-nm\geq a}$, d'où

$$\mathbb{P}(S_n - nm \geq a) \leq \frac{\mathbb{E}(e^{u(M_n-nm)})}{e^{ua}} = e^{-ua}M(u)^n$$

par indépendance des $X_i$.

iii)

$$\begin{aligned}\mathbb{P}(|Y_n - m| \geq \varepsilon) &= \mathbb{P}(S_n - nm \geq n\varepsilon) + P(S_n - nm \leq -n\varepsilon)\\ &\leq e^{-nu\varepsilon}(M(u))^n + e^{-n(-u)(-\varepsilon)}(M(-u))^n = 2e^{-nu\varepsilon}(M(u))^n\\ &\leq 2e^{-nu\varepsilon}e^{nu^2\sigma^2/2}, \ \forall u \geq 0\end{aligned}$$

La meilleure majoration va être obtenue en minimisant l'exposant, c'est-à-dire pour $u = \varepsilon/\sigma^2$. Nous en déduisons l'inégalité de Chernov.

**Question 3** Par l'inégalité de Bienaymé-Chebyshev, $\mathbb{P}(|Y_n - m| > \varepsilon) \leq \frac{\mathbb{V}(Y_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$.

Ainsi, $\mathbb{P}(|Y_n - m| \leq \varepsilon) \leq \alpha$ dès que $1 - \frac{\sigma^2}{n\varepsilon^2}$. Avec $\varepsilon = 0,05$, il vient $n \geq 80000$.

Par l'inégalité de Chernov, nous obtenons $2e^{-n\varepsilon^2/20} \leq 0,05$. Il vient que $n \geq 29512$. Pour avoir une évaluation du même ordre de la probabilité cherchée, nous pouvons donc prendre un échantillon beaucoup plus petit si nous utilisons l'inégalité de Chernov. A taille d'échantillon fixée, nous aurons une meilleure évaluation avec cette inégalité.

## Lemme de Borel-Cantelli

**Question 1** On voit dans un premier temps que $\bigcap_{n\geq 0}\bigcup_{k\geq n} A_n \in \mathcal{A}$ par unions et intersections dénombrables. On a

$$\mathbb{P}(\limsup_n A_n) = \lim_{p\to\infty} \mathbb{P}(\cup_{n\geq p} A_n) \leq \lim_{p\to\infty} \sum_{n\geq p} \mathbb{P}(A_n),$$

où on remarque que les deux suites sont décroissantes.

Si la série $\sum_n \mathbb{P}(A_n)$ est convergente, le reste de cette série tend vers 0 et l'inégalité implique que $\mathbb{P}(\limsup_n An) = 0$.

**Question 2** Supposons maintenant que les $A_n$ soient indépendants et que la série diverge. Soit $m$ un nombre entier. Nous avons

$$\begin{aligned}\mathbb{P}(\cup_{i=p}^m A_i) &= 1 - \mathbb{P}(\cap_{i=p}^m A_i^c) = 1 - \prod_{i=p}^m \mathbb{P}(A_i^c) \text{ par indépendance} \\ &= 1 - \prod_{i=p}^m (1 - \mathbb{P}(A_i)) \geq 1 - e^{-\sum_{i=p}^m \mathbb{P}(A_i)}\end{aligned}$$

du fait de l'inégalité $1 - x \leq e^{-x}$ pour $x \geq 0$. Ainsi,

$$\mathbb{P}(\cup_{i=p}^\infty A_i) \geq 1 - e^{-\sum_{i=p}^\infty \mathbb{P}(A_i)} = 1$$

et l'on conclut finalement que pour tout $p$, $\mathbb{P}(\cup_{i=p}^\infty A_i) = 1$, ce qui implique finalement que $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$.

**Question 3** Prendre tous les $A_n$ égaux à un même événement $A$ de probabilité $\mathbb{P}(A) \in ]0, 1[$.

**Question 4** En suivant l'indication, on a $\mathbb{P}(A_n) = p$, d'où $\sum_{n=1}^\infty \mathbb{P}(A_n) = \infty$. Le lemme de Borel-Cantelli nous indique alors que $\mathbb{P}(\limsup_{n\to\infty} A_n) = \mathbb{P}\left(\cap_{n\geq 1} \cup_{k\geq n} A_n\right) = 1$. Autrement dit, on a presque sûrement un nombre infini de faces. En passant au complémentaire, on en déduit que l'événement $A$ est de probabilité nulle.

## Loi faible des grands nombres

Appliquer l'inégalité de Bienaymé-Chebyshev à la variable aléatoire $M_n$.

## Convergence vers une constante

### Convergence $\mathcal{L}^2$

1. Supposons $X_n \xrightarrow{\mathcal{L}^2} X$ quand $n \to +\infty$, i.e. $\mathbb{E}\left((X_n - X)^2\right) \xrightarrow[n\to+\infty]{} 0$.

Par positivité de la variance, on a tout d'abord

$$\mathbb{E}\left((X_n - X)^2\right) \geq (\mathbb{E}(X_n - X))^2 \geq 0,$$

qui garantit que $|\mathbb{E}(X_n - X)| = |\mathbb{E}(X_n) - \mathbb{E}(X)| \to 0$ quand $n \to +\infty$.

Ensuite, par inégalité triangulaire on a

$$\begin{aligned}|\mathbb{V}(X_n) - \mathbb{V}(X)| &= \left|\mathbb{E}\left(X_n^2\right) - \mathbb{E}(X_n)^2 - \mathbb{E}\left(X^2\right) + \mathbb{E}(X)^2\right| \\ &\leq \left|\mathbb{E}\left(X_n^2 - X^2\right)\right| + \left|\mathbb{E}(X_n)^2 - \mathbb{E}(X)^2\right|.\end{aligned}$$

Or, comme nous avons vu que $\mathbb{E}(X_n) \xrightarrow[n\to+\infty]{} \mathbb{E}(X)$, on a directement que $\left|\mathbb{E}(X_n)^2 - \mathbb{E}(X)^2\right| \xrightarrow[n\to+\infty]{} 0$. Par ailleurs,

$$\begin{aligned}\left|\mathbb{E}\left(X_n^2 - X^2\right)\right| &= \left|\mathbb{E}\left((X_n - X)^2\right) - 2\mathbb{E}\left(X^2\right) + 2\mathbb{E}(X\, X_n)\right| \\ &= \left|\mathbb{E}\left((X_n - X)^2\right) + 2\mathbb{E}(X(X_n - X))\right| \\ &\leq \mathbb{E}\left((X_n - X)^2\right) + 2\left|\mathbb{E}(X(X_n - X))\right|.\end{aligned}$$

Par l'inégalité de Cauchy-Schwartz on a

$$|\mathbb{E}(X(X_n - X))|^2 \leq \mathbb{E}\left(X^2\right)\, \mathbb{E}\left((X_n - X)^2\right),$$

donc

$$\left|\mathbb{E}\left(X_n^2 - X^2\right)\right| \leq \mathbb{E}\left((X_n - X)^2\right) + 2\sqrt{\mathbb{E}\left(X^2\right)}\sqrt{\mathbb{E}\left((X_n - X)^2\right)},$$

qui tend bien vers 0 quand $n \to +\infty$ par hypothèse. On en conclut que $|\mathbb{V}(X_n) - \mathbb{V}(X)| \xrightarrow[n\to+\infty]{} 0$.

2. Par définition, $X_n \xrightarrow{\mathcal{L}^2} a$ quand $n \to +\infty$ ssi $\mathbb{E}\left((X_n - a)^2\right) \xrightarrow[n\to+\infty]{} 0$.

Le premier sens de l'équivalence est immédiatement obtenu par la question précédente, en prenant $X = a$ presque-sûrement.

Réciproquement, pour tout $n \in \mathbb{N}^*$, par linéarité de l'espérance on a

$$\begin{aligned}\mathbb{E}\left((X_n - a)^2\right) &= \mathbb{E}\left(X_n^2 + a^2 - 2aX_n\right) = \mathbb{E}\left(X_n^2\right) + a^2 - 2a\mathbb{E}(X_n) \\ &= \mathbb{V}(X_n) + a^2 - 2a\mathbb{E}(X_n) + \mathbb{E}(X_n)^2 \\ &= \mathbb{V}(X_n) + (\mathbb{E}(X_n) - a)^2.\end{aligned}$$

On en déduit immédiatement le deuxième sens de l'équivalence.

**Convergences en loi et probabilité** Supposons que $X_n \xrightarrow{\mathcal{L}} a$ quand $n \to +\infty$. Cela revient à dire que $X_n$ tend en loi vers la variable aléatoire $X = a$ p.s. quand $n \to +\infty$. Dans ce cas, $X$ a pour fonction de répartition $x \in \mathbb{R} \mapsto 1_{[a,+\infty[}(x)$ et pour tout $x \in \mathbb{R}\backslash\{a\}$, on a

$$\mathbb{P}(X_n \leq x) \xrightarrow[n\to+\infty]{} 1_{[a,+\infty[}(x).$$

Soit maintenant $\varepsilon > 0$. D'après la croissance de la fonction de répartition et l'hypothèse de convergence en loi, on a

$$\begin{aligned}\mathbb{P}(|X_n - a| \geq \varepsilon) &= \mathbb{P}(X_n \geq a + \varepsilon) + \mathbb{P}(X_n \leq a - \varepsilon)\\ &= 1 - \mathbb{P}(X_n < a + \varepsilon) + \mathbb{P}(X_n \leq a - \varepsilon)\\ &\leq 1 - \mathbb{P}\left(X_n \leq a + \frac{\varepsilon}{2}\right) + \mathbb{P}(X_n \leq a - \varepsilon)\\ &\xrightarrow[n\to+\infty]{} 1 - 1_{[a,+\infty[}\left(a + \frac{\varepsilon}{2}\right) + 1_{[a,+\infty[}(a - \varepsilon) = 0.\end{aligned}$$

Ainsi, on a bien $X_n \xrightarrow{\mathbb{P}} a$ quand $n \to +\infty$.

## Fonction de répartition empirique

**Préliminaire** La variable aléatoire $1_{]-\infty,x]}(X)$ est à valeurs dans $\{0,1\}$ et

$$\mathbb{P}\left(1_{]-\infty,x]}(X) = 1\right) = F(x) = 1 - \mathbb{P}\left(1_{]-\infty,x]}(X) = 0\right).$$

On reconnaît une loi de Bernoulli de paramètre $F(x)$. On en déduit directement que $\mathbb{E}(1_{]-\infty,x]}(X)) = F(x)$ et $\mathbb{V}(1_{]-\infty,x]}(X)) = F(x)\,(1 - F(x))$.

**Biais** Sous les hypothèses de l'exercice, on a

$$\begin{aligned}\mathbb{E}(F_n(x)) &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(1_{]-\infty,x]}(X_i)\right) \quad \text{par linéarité de l'espérance,}\\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(1_{]-\infty,x]}(X)\right) \quad \text{car } X_1,\ldots,X_n \text{ ont même loi que } X,\\ &= \frac{1}{n}\sum_{i=1}^{n}F(x) \quad \text{d'après la première question,}\\ &= F(x).\end{aligned}$$

**Erreur quadratique moyenne** Soit $n \in \mathbb{N}^*$. D'après la question précédente, on a

$$\mathbb{E}\left((F_n(x) - F(x))^2\right) = \mathbb{E}\left((F_n(x) - \mathbb{E}(F_n(x)))^2\right) = \mathbb{V}(F_n(x)).$$

Or

$$\begin{aligned}
\mathbb{V}\left(F_n(x)\right) &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}\left(1_{]-\infty,x]}(X_i)\right) \quad \text{par indépendance des variables,}\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}\left(1_{]-\infty,x]}(X)\right) \quad \text{car } X_1,\dots,X_n \text{ ont même loi que } X,\\
&= \frac{1}{n^2}\sum_{i=1}^{n}F(x)\left(1-F(x)\right) \quad \text{d'après la première question,}\\
&= \frac{1}{n}F(x)\left(1-F(x)\right) \xrightarrow[n\to+\infty]{} 0.
\end{aligned}$$

On en conclut que $F_n(x) \xrightarrow{\mathcal{L}^2} F(x)$ quand $n \to +\infty$.

**Consistance** On remarque que pour tout $n \in \mathbb{N}^*$, $F_n(x)$ n'est autre que la moyenne de $n$ variables aléatoires indépendantes de même loi de Bernoulli de paramètre $F(x)$. La loi forte des grands nombres nous assure donc qu'elle converge presque-sûrement vers l'espérance de cette loi, qui n'est autre que $F(x)$.

**Remarque**

On peut en fait aller plus loin et montrer que l'on a la convergence presque sûre uniformément sur $\mathbb{R}$ voir par exemple ce document, c'est le théorème de Glivenko-Cantelli très utile en statistiques.

## Théorème de Weierstrass sur $[0,1]$

**Préliminaires** **Théorème de convergence dominée presque partout** — Puisque $f_n$ converge vers $f$ $\mu$-presque partout, il existe un ensemble $E$ mesurable et négligeable tel que

$$f_n(x) \to f(x), \;\; \forall x \in X \setminus E$$

De même, puisque $|f_n(x)| \leq g(x)$ $\mu$-presque partout, les ensembles $F_n = \{x \in X; |f(x)| > g(x)\}$ sont mesurables et négligeables pour tout $n \in \mathbb{N}^*$. Alors $N = E \cup_{n\in\mathbb{N}^*} F_n$ est mesurable et négligeable.

Soit $\tilde{f}_n = 1_{N^c} f_n$ et $\tilde{f} = 1_{N^c} f$, les restrictions à $N^c$ des $f_n$ et $f$. Alors le théorème de convergence dominée s'applique à la suite des $(\tilde{f}_n)_{n\in\mathbb{N}^*}$. On a ainsi que $\tilde{f}$ est intégrable et

$$\int_X \tilde{f}\mu = \lim_{n\to\infty}\int_X \tilde{f}_n\mu.$$

Puisque $f = \tilde{f}$ $\mu$-p.p. et $f_n = \tilde{f}_n$ $\mu$-p.p., on a $\int_X f\mu = \int_X \tilde{f}\mu$ et $\int_X f_n\mu = \int_X \tilde{f}_n\mu$. Alors $f$ est intégrable et

$$\int_X f_n\mu \xrightarrow[n\to+\infty]{} \int_X f\mu.$$

**Inégalité de Jensen** — Puisque $f$ est convexe, pour tout $a \in \mathbb{R}$ il existe $\lambda_a \in \mathbb{R}$ tel que pour tout $x \in \mathbb{R}$ on a

$$f(x) \geq f(a) + \lambda_a (x - a).$$

C'est une conséquence directe de la caractérisation de la convexité par les inégalités des pentes. C'est vrai en particulier pour $x = X(\omega)$, $\omega \in \Omega$, et $a = \mathbb{E}(X)$ : pour tout $\omega \in \Omega$,

$$f(X(\omega)) \geq f(\mathbb{E}(X)) + \lambda_{\mathbb{E}(X)} (X(\omega) - \mathbb{E}(X)).$$

En intégrant de chaque côté de l'inégalité, on obtient bien

$$\begin{aligned}\int_\Omega f(X(\omega))\, \mathbb{P}(d\omega) = \mathbb{E}(f(X)) &\geq f(\mathbb{E}(X)) + \lambda_{\mathbb{E}(X)} \left( \int_\Omega X(\omega)\, \mathbb{P}(d\omega) - \mathbb{E}(X) \right) \\ &= f(\mathbb{E}(X)) + \lambda_{\mathbb{E}(X)} (\mathbb{E}(X) - \mathbb{E}(X)) \\ &= f(\mathbb{E}(X)).\end{aligned}$$

**Preuve du théorème.**

1. Comme $X_1, \ldots, X_n$ sont indépendantes de même loi de Bernoulli de paramètre $x \in ]0,1[$, $M_n$ est à valeurs dans $\left\{ \frac{k}{n} : k \in \{0, \ldots, n\} \right\}$ et pour tout $k \in \{0, \ldots, n\}$ on a

   $$\mathbb{P}\left( M_n = \frac{k}{n} \right) = \binom{k}{n} x^k (1-x)^{n-k}.$$

   Ainsi,

   $$\mathbb{E}(f(M_n)) = \sum_{k=0}^{n} \binom{k}{n} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

Les polynômes de la forme $B_{n,k} : u \in [0,1] \mapsto \binom{n}{k} u^k (1-u)^{n-k}$ sont appelés *polynômes de Bernstein.*

2. Comme une variable de Bernoulli admet une espérance égale à son paramètre, la loi (forte) des grands nombres nous assure que $M_n \to x$ p.s. quand $n \to +\infty$. Puisque $f$ est continue, on a de même $f(M_n) \to f(x)$ p.s. quand $n \to +\infty$.

Par ailleurs, la continuité de $f$ sur $[0,1]$ nous assure que $f$ est bornée (l'image d'un compact par une fonction continue $\mathbb{R} \to \mathbb{R}$ est un compact). Ainsi, pour tout $n \in \mathbb{N}^*$, la variable aléatoire $f(M_n)$ l'est également. Cela nous permet d'appliquer le théorème de convergence dominée de la question 1 : en notant $(\Omega, \mathcal{A}, \mathbb{P})$ l'espace probabilisé sur lequel sont définies nos variables aléatoires,

$$\mathbb{E}(f(M_n)) = \int_\Omega f(M_n(\omega))\, \mathbb{P}(d\omega) \xrightarrow[n \to +\infty]{} \int_\Omega f(x)\, \mathbb{P}(d\omega) = f(x).$$

3. En combinant les résultats des deux questions précédentes, on obtient que la suite $(P_n)_{n \in \mathbb{N}^*}$ des polynômes définis pour tout $n \in \mathbb{N}^*$ par

   $$P_n : x \in ]0,1[ \mapsto \sum_{k=0}^{n} \binom{k}{n} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}$$

converge simplement vers $f(x)$. On étend simplement ce résultat à $[0,1]$ en remarquant que pour tout $n \in \mathbb{N}^*$ on a directement l'égalité

$$P_n(0) = f\left(\frac{0}{n}\right) = f(0) \text{ et } P_n(1) = f\left(\frac{n}{n}\right) = f(1).$$

On a même toujours l'égalité $P_n(x) = \mathbb{E}(f(M_n))$ pour $x \in [0,1]$ si l'on remarque qu'une loi de Bernoulli de paramètre $a \in \{0,1\}$ n'est autre d'une Dirac en $\{a\}$, ce qui signifie que $M_n = a$ p.s. et donc que $\mathbb{E}(f(M_n)) = f(a) = P_n(a)$.

4. On souhaite montrer que

$$\forall \varepsilon > 0 \; \exists N_\varepsilon \in \mathbb{N}^* : (n \geq N_\varepsilon) \Rightarrow (\forall x \in [0,1] : |P_n(x) - f(x)| \leq \varepsilon).$$

Remarquons tout d'abord que la fonction $f$ étant bornée, elle admet un maximum sur $[0,1]$, que l'on note $K$. Puisqu'elle est continue sur un segment réel, elle est aussi uniformément continue :

$$\forall \varepsilon > 0 \quad \exists \delta_\varepsilon > 0 : \forall (x,y) \in ]0,1[^2, \quad |x-y| < \delta_\varepsilon \Rightarrow |f(x) - f(y)| < \varepsilon.$$

Soient $\varepsilon > 0$, puis $\delta = \delta_{\varepsilon/2} > 0$ pour lequel l'implication précédente est vraie avec $\frac{\varepsilon}{2}$. Alors pour tous $n \in \mathbb{N}^*$ et $x \in ]0,1[$ on a

$$\begin{aligned}
|\mathbb{E}(f(M_n)) - f(x)| &\leq \mathbb{E}(|f(M_n) - f(x)|) \quad \text{par l'inégalité de Jensen,} \\
&= \mathbb{E}\left(|f(M_n) - f(x)| \; 1_{[0,\delta[}(|M_n - x|)\right) \\
&\quad + \mathbb{E}\left(|f(M_n) - f(x)| \; 1_{[\delta,+\infty[}(|M_n - x|)\right) \\
&\leq \frac{\varepsilon}{2} + 2|K| \, \mathbb{P}(|M_n - x| \geq \delta) \quad \text{par inégalité triangulaire,} \\
&\leq \frac{\varepsilon}{2} + 2|K| \, \frac{\mathbb{V}(M_n)}{\delta^2} \quad \text{par Bienaymé-Chebyshev.}
\end{aligned}$$

Or pour tout $n \in \mathbb{N}^*$

$$\begin{aligned}
\mathbb{V}(M_n) &= \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}(X_i) \quad \text{par indépendance de } X_1, \ldots, X_n, \\
&= \frac{1}{n^2}\sum_{i=1}^n x(1-x) \quad \text{car } X_1, \ldots, X_n \text{ suivent la même loi de Bernoulli,} \\
&= \frac{x(1-x)}{n} < \frac{1}{4n} \quad \text{car } x \in ]0,1[.
\end{aligned}$$

Par conséquent, pour tout $n \in \mathbb{N}^*$,

$$|\mathbb{E}(f(M_n)) - f(x)| < \frac{\varepsilon}{2} + \frac{|K|}{2n\delta^2}.$$

En prenant $N_\varepsilon$ le plus petit entier naturel supérieur ou égal à $\frac{|K|}{\varepsilon\,\delta^2}$, on obtient alors que pour tout $n \geq N_\varepsilon$, $|\mathbb{E}(f(M_n)) - f(x)| < \varepsilon$. Comme $\varepsilon$ et $N_\varepsilon$ sont les mêmes quel que soit $x \in [0,1]$, on obtient bien la convergence uniforme désirée.

### Théorème de Slutsky

**Question 1** Il suffit de montrer que dans la preuve de la proposition sur la convergence des f.d.r. (p. 18), on peut remplacer les fonctions continues bornées $f_{p,b}$ approchant $1_{]-\infty,b]}$ par des fonctions lipschitziennes bornées, ce qui est immédiat.

**Question 2** Du fait du résultat de la question 1, il suffit de montrer que $\lim_{n\to\infty} \mathbb{E}(f(Y_n)) = \mathbb{E}(f(X))$, pour toute fonction lipschitzienne bornée. Soit $f$ une telle fonction. On a alors $|f(x)| \leq k$ et $|f(x) - f(y)| \leq C|x-y|$, pour des constantes $k$ et $C$.

Soit $\varepsilon > 0$ donné. On a

$$\begin{aligned}|\mathbb{E}(f(Y_n)) - \mathbb{E}(f(X_n))| &\leq \mathbb{E}(|f(Y_n) - f(X_n)|) \\ &\leq \mathbb{E}(|f(Y_n) - f(X_n)|(1_{|Y_n - X_n| \leq \varepsilon} + 1_{|Y_n - X_n| > \varepsilon})) \\ &\leq C\varepsilon + 2kP(|X_n - Y_n| > \varepsilon)\end{aligned}$$

Le deuxième terme du membre de droite tend vers 0 quand $n$ tend vers l'infini, car $(X_n - Y_n)_n$ converge en probabilité vers 0. Comme $\varepsilon$ est arbitrairement petit, nous en déduisons que $\lim_{n\to\infty} |\mathbb{E}(f(Y_n)) - \mathbb{E}(f(X_n))| = 0$, d'où le résultat.

**Question 3** Soit $u \in R^d$ . On a :

$$|\phi_{Y_n}(u) - \phi_X(u)| \leq |\phi_{Y_n}(u) - \phi_{X_n}(u)| + |\phi_{X_n}(u) - \phi_X(u)|$$

D'une part, le théorème de Lévy (p. 24) partie 1. montre que $|\phi_{X_n}(u) - \phi_X(u)|$ tend vers 0 quand $n \to \infty$. D'autre part,

$$|\phi_{Y_n}(u) - \phi_{X_n}(u)| = |\mathbb{E}(e^{i<u,Y_n>} - e^{i<u,X_n>})| = |\mathbb{E}(e^{i<u,X_n>}(e^{i<u,Y_n-X_n>} - 1))| \leq \mathbb{E}(|e^{i<u,Y_n-X_n>} - 1|)$$

tend vers 0 quand $n \to \infty$ d'après la proposition — cas borné (p. 11) appliquée aux variables aléatoires $e^{i<u,Y_n-X_n>} - 1$ dont la convergence en proba vers 0 est assurée par la propriété de continuité (p. 12).

# Références

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

Simmons, George F. 2003. *Introduction to Topology and Modern Analysis*. Reprint edition. Malabar, Fla: Krieger Publishing Company.