

Synthèse du cours de probabilités

Thibaut Lemoine

Ces notes sont un résumé du cours, inspiré majoritairement par le polycopié de Zhan Shi qui fait office de référence pour ce semestre, mais aussi empreint de remarques plus personnelles. Cette synthèse n'est surtout pas à prendre comme une substitution du cours – la plupart des résultats sont énoncés sans démonstration – mais plutôt comme un guide de lecture. Son but est de donner une vue d'ensemble, et quelques pistes pour aborder les problèmes les plus fréquents, par exemple comment trouver la loi d'une variable aléatoire ou comment montrer la convergence presque-sûre d'une suite de variables aléatoires. Les éventuelles coquilles/imprécisions/erreurs ne sont pas à exclure, et, s'il y en a, elles sont entièrement imputables à l'auteur.

Espaces de probabilité, variables aléatoires

Définition 1. *Un espace de probabilité est un triplet $(\Omega, \mathcal{F}, \mathbb{P})$, où (Ω, \mathcal{F}) est un espace mesurable et \mathbb{P} une mesure de probabilité sur cet espace. Les éléments de \mathcal{F} sont appelés événements.*

Note : hormis quelques cas spécifiques, l'espace Ω sera souvent dépourvu de toute structure pertinente, et peut être vu comme un espace abstrait, "inaccessible" : on n'émet aucune hypothèse sur ses éléments. En revanche, on dispose en général d'informations sur les événements – tout du moins certains d'entre eux. Et ces informations se traduisent justement par leur probabilité.

Définition 2. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Une variable aléatoire à valeurs dans un espace mesurable (E, \mathcal{E}) est une fonction mesurable $X : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$. Sa loi est la mesure image¹ de \mathbb{P} par X , autrement dit $\mathcal{P}_X = X_*\mathbb{P} = \mathbb{P} \circ X^{-1}$.*

On dit qu'une variable aléatoire admet un *moment d'ordre p* si elle est dans $L^p(\Omega, \mathbb{P})$, c'est-à-dire si $\int_{\Omega} |X(\omega)|^p d\mathbb{P}(\omega) < \infty$. Si elle admet un moment d'ordre 1, on définit son *espérance* comme $\mathbb{E}(X) = \int_{\Omega} X(\omega) d\mathbb{P}(\omega)$. En particulier, tout le cours sur l'intégrale de Lebesgue par rapport à une mesure finie s'applique aux variables aléatoires et leur espérance.

Définition 3. *Soit $X : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$ une variable aléatoire. Sa tribu engendrée $\sigma(X)$ est la sous-tribu de \mathcal{F} définie par*

$$\sigma(X) = \sigma(\{X^{-1}(A), A \in \mathcal{E}\}). \quad (1)$$

On rencontre essentiellement deux types de variables aléatoires dans ce cours : celles à valeurs dans \mathbb{N} et celles à valeurs dans \mathbb{R} . Nous verrons que les résultats généraux prennent des formes plus simples dans chacun de ces deux cas. Le premier de ces résultats est la formule de transfert, qui permet de *transférer* l'intégrale sur Ω en une intégrale sur l'espace image, lorsqu'on cherche à calculer des espérances.

Théorème 1 (Formule de transfert). *Soit $X : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$ une variable aléatoire. Pour toute fonction $f : E \rightarrow \mathbb{R}$ mesurable bornée (ou positive), on a*

$$\mathbb{E}[f(X)] = \int_E f(x) d\mathcal{P}_X(x). \quad (2)$$

En particulier :

1. C'est exactement la même notion que celle, que vous verrez peut-être plus tard en géométrie différentielle, de *puissé-en-avant* ou *push-forward*.

1. Si $E = \mathbb{N}$, alors

$$\mathbb{E}[f(X)] = \sum_{n \in \mathbb{N}} f(n) \mathbb{P}(X = n). \quad (3)$$

2. Si $E = \mathbb{R}$ et si X a pour densité f_X par rapport à la mesure de Lebesgue sur \mathbb{R} , alors

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x) f_X(x) dx. \quad (4)$$

Comme on le constate, une fois la formule de transfert appliquée, le calcul d'espérance se ramène à une intégrale réelle ou à une somme sur \mathbb{N} . En d'autres termes, l'essentiel de ce qui se passe au-delà de la formule de transfert repose sur des cours précédents : celui sur l'intégration de Lebesgue et celui sur les séries numériques.

On voit également, par la formule de transfert et par le lemme de classe monotone, que la loi d'une variable aléatoire entière (resp. réelle) est entièrement caractérisée par $(\mathbb{P}(X = n), n \in \mathbb{N})$ (resp. $(\mathbb{P}(X \leq x), x \in \mathbb{R})$).

Définition 4. Soit X une variable aléatoire réelle. Sa fonction de répartition est la fonction $F_X : \mathbb{R} \rightarrow [0, 1]$ définie par $F_X(x) = \mathbb{P}(X \leq x)$.

La fonction de répartition F_X d'une v.a. X vérifie les propriétés suivantes :

1. Elle est croissante² sur \mathbb{R} ;
2. Elle est continue à droite ;
3. Ses limites en $-\infty$ et $+\infty$ sont respectivement 0 et 1 ;
4. Si F_X est continue et presque-partout dérivable, telle que F_X soit une fonction primitive de sa dérivée notée f , alors X admet f pour densité. Réciproquement, si X a pour densité f_X , alors

$$F_X(x) = \int_{-\infty}^x f_X(t) dt.$$

Pour résumer :

- Lorsqu'on veut déterminer la loi d'une variable aléatoire à valeurs dans un espace (E, \mathcal{E}) général³, une méthode qui fonctionne systématiquement est celle de la *fonction muette* (ou *fonction test*) : si l'on sait calculer $\mathbb{E}[f(X)]$ pour toute fonction $f : E \rightarrow \mathbb{R}$ mesurable positive ou bornée, alors cela détermine la loi de X .
- Dans le cas où X est à valeurs dans \mathbb{N} on peut se contenter de calculer $\mathbb{P}(X = n)$ pour tout $n \in \mathbb{N}$
- Dans le cas où X est à valeurs réelles on peut se contenter de calculer sa fonction de répartition et/ou sa densité.

Notons qu'il n'est pas anodin que l'on mette l'accent sur les v.a. à densité et celles à valeurs entières/discrètes : il existe en effet un théorème (hors programme) qui dit, à peu de chose près, que toute mesure de Borel sur \mathbb{R} se décompose en une somme de mesures à densité et de mesures discrètes. C'est pourquoi il n'est pas étonnant que l'on rencontre dans ce cours essentiellement des mesures de ce type.

On termine cette section avec un commentaire sur les lois marginales. Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire réel et $X' = (X_{i_1}, \dots, X_{i_k})$ un sous-vecteur. La *loi marginale* de X' est simplement la loi du sous-vecteur et se calcule en intégrant la loi de X par rapport aux composantes de X qui ne font pas partie de X' . Cela se voit simplement : soit π la projection définie par

$$\pi : \begin{cases} \mathbb{R}^n & \rightarrow \mathbb{R}^k \\ (x_1, \dots, x_n) & \mapsto (x_{i_1}, \dots, x_{i_k}) \end{cases} ,$$

2. En particulier cela implique qu'elle admet au plus un nombre dénombrable de points de discontinuité.

3. Par exemple, si l'on considère un couple de variables aléatoires dont l'une est réelle et l'autre est entière, le couple peut être vu comme une variable aléatoire à valeurs dans $(\mathbb{R} \times \mathbb{N}, \mathcal{B}(\mathbb{R}) \otimes \mathcal{P}(\mathbb{N}))$.

alors on peut déterminer la loi marginale en calculant, pour toute $f : \mathbb{R}^k \rightarrow \mathbb{R}$ mesurable bornée, $\mathbb{E}[f(X')] = \mathbb{E}[f \circ \pi(X)]$. Ainsi, cela donne

$$\mathbb{E}[f(X')] = \int_{\mathbb{R}^k} f(x_{i_1}, \dots, x_{i_k}) \int_{\mathbb{R}^{n-k}} d\mathcal{P}_X(x_1, \dots, x_n),$$

et on en déduit que $d\mathcal{P}_{X'}(x_{i_1}, \dots, x_{i_k}) = \int_{\mathbb{R}^{n-k}} d\mathcal{P}_X(x_1, \dots, x_n)$.

Indépendance

Définition 5. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $(\mathcal{G}_1, \dots, \mathcal{G}_n)$ une famille de sous-tribus de \mathcal{F} . Ces sous-tribus sont dites indépendantes si

$$\mathbb{P}\left(\bigcap_{1 \leq i \leq n} A_i\right) = \prod_{1 \leq i \leq n} \mathbb{P}(A_i), \quad \forall A_i \in \mathcal{G}_i. \quad (5)$$

Des variables aléatoires X_1, \dots, X_n sont dites indépendantes si leurs tribus engendrées sont indépendantes. Une famille quelconque de tribus est dite indépendante lorsque toute sous-famille finie l'est.

Le théorème de classe monotone permet de restreindre le test d'indépendance d'événements à des π -systèmes lorsque les tribus sont engendrées par ceux-ci. Par exemple, la tribu borélienne de \mathbb{R} est engendrée par le π -système $(] - \infty, x], x \in \mathbb{R})$ ⁴, donc pour vérifier qu'une famille de variables aléatoires réelles est indépendante il suffit de vérifier la propriété de factorisation sur des éléments de ce π -système. Plus précisément : si X_1, \dots, X_n sont des variables aléatoires réelles, alors elles sont indépendantes si et seulement si

$$\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1) \cdots \mathbb{P}(X_n \leq x_n), \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n.$$

L'indépendance vérifie également le principe de regroupement par paquets : si $(\mathcal{G}_i)_{i \in I}$ une famille indépendante de sous-tribus de \mathcal{F} . soit I_1, \dots, I_p des sous-ensembles de I disjoints deux à deux. Alors les tribus $\sigma(\mathcal{G}_{i_1}, i_1 \in I_1), \dots, \sigma(\mathcal{G}_{i_p}, i_p \in I_p)$ sont indépendantes.

Le théorème suivant caractérise l'indépendance de variables aléatoires.

Théorème 2 (Caractérisation de l'indépendance). Soit des v.a. X_1, \dots, X_n respectivement à valeurs dans $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$. Les assertions suivantes sont équivalentes :

1. La famille $(X_i)_{1 \leq i \leq n}$ est indépendante ;
2. $\mathcal{P}_{(X_1, \dots, X_n)} = \mathcal{P}_{X_1} \otimes \cdots \otimes \mathcal{P}_{X_n}$ ⁵ ;
3. $\mathbb{E}[f_1(X_1) \cdots f_n(X_n)] = \mathbb{E}[f_1(X_1)] \cdots \mathbb{E}[f_n(X_n)]$ pour toutes fonctions $f_i : E_i \rightarrow \mathbb{R}$ mesurables positives (ou bornées).

En pratique, l'indépendance est un des deux seuls outils qui nous permettent de déterminer la loi d'un n -uplet aléatoire (X_1, \dots, X_n) – l'autre étant la donnée directe de sa loi, par exemple de sa densité. En effet, bien souvent, lorsqu'on va étudier un tel n -uplet, deux cas de figure se présenteront :

4. On peut même se restreindre à une famille dénombrable en prenant $x \in \mathbb{Q}$ par densité.

5. Cf. cours d'intégration pour la définition de la mesure produit. Il est intéressant de noter que la notation \otimes correspond bel et bien à un produit tensoriel au sens de l'algèbre linéaire : le théorème de Riesz–Markov nous dit qu'on peut représenter toute intégrale par rapport à une mesure de Borel sur \mathbb{R} comme une forme linéaire positive sur $\mathcal{C}_c(\mathbb{R})$, et l'intégrale du produit de deux telles mesures est le produit tensoriel des formes linéaires associées.

- Soit (X_1, \dots, X_n) est une famille indépendante et alors sa loi est la mesure produit des lois marginales de chacun des X_i : par exemple si X_i a pour densité f_{X_i} alors le n -uplet a pour densité

$$f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n), \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n.$$

- Soit (X_1, \dots, X_n) n'est pas une famille indépendante mais on peut la réexprimer à l'aide d'une famille indépendante (Y_1, \dots, Y_n) par une transformation Φ suffisamment régulière (en pratique, un \mathcal{C}^1 -difféomorphisme). Dans ce cas on applique le point précédent à (Y_1, \dots, Y_n) et on calcule $\mathbb{E}[f(X_1, \dots, X_n)] = \mathbb{E}[f \circ \Phi(Y_1, \dots, Y_n)]$ à l'aide d'un changement de variables.

On rappelle, à titre indicatif, la formule du changement de variables dans \mathbb{R}^n .

Théorème 3 (Changement de variables dans \mathbb{R}^n). *Soit f une fonction borélienne sur un ouvert V de \mathbb{R}^n et ϕ un \mathcal{C}^1 -difféomorphisme d'un ouvert U de \mathbb{R}^n vers V . Si f est intégrable sur V alors on a*

$$\int_V f(y) dy = \int_U (f \circ \phi)(x) |J_\phi(x)| dx, \quad (6)$$

où $J_\phi(x) = \det \left(\frac{\partial \phi_i(x)}{\partial x_j} \right)_{1 \leq i, j \leq n}$ est le déterminant de la jacobienne de ϕ , c'est-à-dire de la matrice de ses dérivées partielles.

(Co)variance

Si X et Y sont des v.a. réelles admettant des moments d'ordre 2, on définit leur *covariance* par

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \quad (7)$$

On définit de la même façon la *variance* de X par $\text{Var}(X) = \text{Cov}(X, X)$. La covariance est une forme bilinéaire symétrique, c'est-à-dire

$$\text{Cov}(X, Y) = \text{Cov}(Y, X)$$

et

$$\text{Cov}(aX_1 + bX_2, cY_1 + dY_2) = ac\text{Cov}(X_1, Y_1) + ad\text{Cov}(X_1, Y_2) + bc\text{Cov}(X_2, Y_1) + bd\text{Cov}(X_2, Y_2).$$

La variance de la somme de variables aléatoires s'exprime de la façon suivante à l'aide de leurs covariances respectives :

$$\text{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j). \quad (8)$$

Si (X_1, \dots, X_n) est un vecteur aléatoire dans \mathbb{R}^n , alors on définit sa *matrice de covariance* $D = (d_{ij})_{1 \leq i, j \leq n}$ par $d_{ij} = \text{Cov}(X_i, X_j)$. Par la propriété de la covariance, la matrice de covariance est une matrice symétrique, c'est-à-dire qu'elle est égale à sa transposée.

Proposition 1. *Soit X, Y deux variables aléatoires réelles.*

1. Si X et Y sont indépendantes, alors $\text{Cov}(X, Y) = 0$ et $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.
2. Si (X, Y) est un vecteur gaussien⁶ et $\text{Cov}(X, Y) = 0$ alors X et Y sont indépendantes.

6. C'est-à-dire que toute combinaison linéaire de X et Y est gaussienne; cette hypothèse est primordiale, sinon c'est faux!!!

Fonction caractéristique

Si $X : \Omega \rightarrow \mathbb{R}^n$ un vecteur aléatoire dans \mathbb{R}^n de loi \mathcal{P}_X , sa *fonction caractéristique* est la fonction

$$\phi_X : \begin{cases} \mathbb{R}^n & \rightarrow \mathbb{C} \\ t & \mapsto \mathbb{E}[e^{i\langle t, X \rangle}] \end{cases} .$$

Elle vérifie les propriétés suivantes :

1. Pour tout X , $\phi_X(0) = 1$ et $|\phi_X(t)| \leq 1$ pour tout $t \in \mathbb{R}^n$.
2. La fonction caractéristique est continue sur \mathbb{R}^n .
3. Si X est à densité, alors $\lim_{|t| \rightarrow \infty} \phi_X(t) = 0$.
4. Les variables aléatoires réelles X_1, \dots, X_n sont indépendantes si et seulement si

$$\phi_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = \phi_{X_1}(t_1) \cdots \phi_{X_n}(t_n), \quad \forall (t_1, \dots, t_n) \in \mathbb{R}^n.$$

5. Si X_1 et X_2 sont deux vecteurs aléatoires de même dimension n , alors⁷ $\phi_{X_1+X_2}(t) = \phi_{X_1}(t)\phi_{X_2}(t)$, $\forall t \in \mathbb{R}^n$.
6. Soit X une variable aléatoire réelle qui admet un moment d'ordre n . Alors ϕ_X est de classe \mathcal{C}^n et on a⁸

$$\mathbb{E}[X^n] = \frac{\phi_X^{(n)}(0)}{i^n}.$$

Outre la propriété de factorisation qui caractérise l'indépendance, le principal intérêt de la fonction caractéristique est – comme son nom l'indique – qu'elle *caractérise* une loi. En réalité, cela découle des propriétés fondamentales de la transformée de Fourier (cf. cours d'analyse fonctionnelle pour des détails sur cette dernière). Le théorème suivant résume cette caractérisation.

Théorème 4. *Soit X et Y deux vecteurs aléatoires réels sur \mathbb{R}^n .*

1. X et Y ont même loi, i.e. $\mathcal{P}_X = \mathcal{P}_Y$, si et seulement si $\phi_X = \phi_Y$.
2. Si ϕ_X est intégrable par rapport à la mesure de Lebesgue sur \mathbb{R}^n , alors X est à densité et sa densité f_X vérifie

$$f_X(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{-i\langle x, t \rangle} \phi_X(t) dt, \quad \forall x \in \mathbb{R}^n.$$

Suites et séries de variables aléatoires

Définition 6. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles⁹ et X une variable aléatoire réelle.*

1. X_n converge vers X presque-sûrement, et on note $X_n \rightarrow X$ p.s., s'il existe un événement $A \in \mathcal{F}$ de probabilité 1 tel que $X_n(\omega) \rightarrow X(\omega)$ pour tout $\omega \in A$;
2. X_n converge vers X en probabilité, et on note $X_n \xrightarrow{\mathbb{P}} X$, si pour tout $\varepsilon > 0$ on a $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$;
3. X_n converge vers X dans L^p , et on note $X_n \xrightarrow{L^p} X$, si pour tout n $X_n \in L^p$, $X \in L^p$ et si $\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^p] = 0$;

7. Attention, ici c'est une condition nécessaire mais **pas** suffisante !

8. Encore une fois, c'est une implication et pas une équivalence ! Le fait que ϕ_X soit de classe \mathcal{C}^n n'implique pas forcément que X admet un moment d'ordre n .

9. Note : l'essentiel des définitions et des résultats restent valables si l'on prend X_n à valeurs dans \mathbb{R}^n et pas seulement dans \mathbb{R} .

4. X_n converge vers X en loi, et on note $X_n \xrightarrow{(loi)} X$ (ou parfois $X_n \xrightarrow{(d)} X$, où le ‘d’ signifie “distribution”), si pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$.

Les différentes relations entre ces modes de convergence sont les suivantes :

- Les convergences p.s. et dans L^p impliquent chacune la convergence en probabilité.
- La convergence en probabilité implique la convergence p.s. d’une sous-suite.
- Une suite de v.a. converge en probabilité ssi pour toute sous-suite on peut extraire une sous-sous-suite qui converge p.s.
- La convergence en probabilité implique la convergence en loi, et la réciproque est vraie lorsque la limite est une constante.

Voyons à présent les résultats associés à chacun de ces modes de convergence.

Convergence en probabilité

Il n’y a pas grand-chose à savoir concernant la convergence en probabilité, c’est bien souvent la convergence la plus simple à montrer, et cela se fait en revenant à la définition. Néanmoins, la proposition suivante peut parfois se révéler utile.

Proposition 2. Si $X_n \xrightarrow{\mathbb{P}} X$ et si $f : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction continue, alors $f(X_n) \xrightarrow{\mathbb{P}} f(X)$.

Lorsqu’on a établi la convergence dans L^p on peut se servir d’inégalités du type de celle de Markov¹⁰ pour montrer la convergence en probabilité.

Proposition 3 (Inégalité de Markov). Soit $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ une variable aléatoire. Alors pour tout $\alpha > 0$, on a

$$\mathbb{P}(|X|^p \geq \alpha) \leq \frac{\mathbb{E}[|X|^p]}{\alpha^p}.$$

Le cas de $p = 1$ correspond à l’inégalité de Markov “traditionnelle”, et $p = 2$ correspond à l’inégalité de Bienaymé–Tchebychev. On voit assez facilement que si l’on remplace X par $X_n - X$ dans l’inégalité de Markov, alors la convergence dans L^p implique la convergence du membre de droite vers 0, donc a fortiori celle du membre de gauche, et cela implique bien la convergence en probabilité. Remarque : cela redémontre à peu de frais le fait que la convergence dans L^p implique la convergence en probabilité !

Une réciproque partielle s’obtient par le théorème de convergence dominée : Si $X_n \rightarrow X$ en probabilité, et s’il existe $Y \in L^1$ telle que $|X_n| \leq Y$ pour tout n , alors $\mathbb{E}[X_n] \rightarrow \mathbb{E}[X]$ (et en particulier $X_n \xrightarrow{L^1} X$).

Convergence presque-sûre

La convergence presque-sûre est souvent la plus difficile à montrer, car elle demande d’analyser un événement qui n’est pas évident, à savoir $A = \{\omega \mid X_n(\omega) \rightarrow X(\omega)\}$. Bien souvent on va en réalité considérer plutôt $A' = \{\omega \mid \limsup |X_n(\omega) - X(\omega)| = 0\} \subset A$ et montrer que $\mathbb{P}(A') = 1$. Pour ce faire on utilise en général le lemme de Borel–Cantelli.

Lemme 1 (Borel–Cantelli). Soit $(A_n)_n$ une suite d’événements.

1. Si $\sum_n \mathbb{P}(A_n) < \infty$ alors $\mathbb{P}(\limsup A_n) = 0$.
2. Si (A_n) est une famille indépendante et si $\sum_n \mathbb{P}(A_n) = \infty$ alors $\mathbb{P}(\limsup A_n) = 1$.

10. On appelle cela des inégalités de *concentration*, et il en existe de nombreuses, plus complexes que l’inégalité de Markov, dont l’inégalité de Hoeffding du devoir maison fait par exemple partie.

Notons que, par définition, $\limsup A_n = \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k$, et que l'on peut traduire cela par

$$\omega \in \limsup A_n \Leftrightarrow \forall n \in \mathbb{N}, \exists k \geq n : \omega \in A_k.$$

Autrement dit, ω est dans $\limsup A_n$ ssi ω appartient à une infinité de A_k . Inversement, on définit la limite inférieure de la façon suivante : $\liminf A_n = \bigcup_{n \in \mathbb{N}} \bigcap_{k \geq n} A_k$, et $\omega \in \liminf A_n$ ssi ω appartient à tous les A_k à partir d'un certain rang. On peut se convaincre du fait que la limite inférieure et la limite supérieure au sens ensembliste correspondent exactement aux définitions qu'on en donne sur \mathbb{R} , en remplaçant la relation d'ordre \leq sur \mathbb{R} par la relation d'ordre \subset sur les ensembles. Une autre façon de le voir réside dans l'application indicatrice $A \mapsto \mathbf{1}_A$. En effet,

$$\limsup \mathbf{1}_{A_n} = \mathbf{1}_{\limsup A_n} \text{ et } \liminf \mathbf{1}_{A_n} = \mathbf{1}_{\liminf A_n}.$$

L'utilisation du lemme de Borel–Cantelli pour montrer une convergence presque-sûre n'apparaît peut-être pas évidente, ne serait-ce que parce que l'événement A' vu plus haut ne s'écrit pas naturellement comme la limite supérieure ou inférieure d'une famille d'événements. Voici en réalité une proposition qui a de quoi nous rassurer.

Proposition 4. *Soit (X_n) une suite de v.a. réelles positives. Pour tout $\varepsilon > 0$, on a*

$$\{\limsup X_n > \varepsilon\} \subset \limsup \{X_n > \varepsilon\}.$$

Démonstration. Il suffit de montrer que tout $\omega \in \{\limsup X_n > \varepsilon\}$ appartient à $\limsup \{X_n > \varepsilon\}$. C'est immédiat : en effet, soit un tel ω , alors $\limsup X_n(\omega) > \varepsilon$ et donc pour tout n il existe $k \geq n$ tel que $X_k(\omega) > \varepsilon$, d'où $\omega \in \limsup \{X_n > \varepsilon\}$. \square

En particulier, en notant $B(\varepsilon) = \{\limsup |X_n - X| > \varepsilon\}$, $B_n(\varepsilon) = \{|X_n - X| > \varepsilon\}$, alors il vient que $A^c = \bigcap_{k \geq 1} B(1/k) \subset \bigcap_k \limsup B_n(1/k)$. Le lemme de Borel–Cantelli et la continuité à droite de la mesure \mathbb{P} impliquent donc que si $(\mathbb{P}(B_n))_n$ est une famille sommable, alors

$$\mathbb{P}(A^c) = \inf_{k \in \mathbb{N}} \mathbb{P}(B(1/k)) \leq \inf_{k \in \mathbb{N}} \mathbb{P}(\limsup B_n(1/k)) = 0.$$

On en déduit bien que $\mathbb{P}(A') = 1$.

Mentionnons également un résultat qui s'applique non plus aux suites mais aux séries de variables aléatoires.

Théorème 5. *Soit $(X_n)_n$ une suite de variables aléatoires indépendantes. On note $S_n = X_1 + \dots + X_n$ pour tout n . Si S_n converge en probabilité, alors S_n converge presque-sûrement.*

Convergence en loi

Contrairement aux convergences précédentes, on n'a pas besoin que les v.a. (X_n) soient définies sur le même espace de probabilité. En effet, par la formule de transfert, pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée, $\mathbb{E}[f(X_n)] = \int_{\mathbb{R}} f(x) d\mathcal{P}_{X_n}(x)$ et cette quantité ne dépend pas de $(\Omega, \mathcal{F}, \mathbb{P})$.

On a une caractérisation équivalente de la convergence en loi.

Proposition 5. *$(X_n)_n$ converge en loi vers X si et seulement si pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ continue à support compact,*

$$\mathbb{E}[f(X_n)] \longrightarrow \mathbb{E}[f(X)].$$

La convergence en loi provient en réalité de notions de convergence issues de la théorie de la mesure : en effet, une suite de mesures $(\mu_n)_n$ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ converge *étroitement* (resp. *vaguement*) vers une mesure μ si pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée (resp. à

support compact) on a $\int f d\mu_n \rightarrow \int f d\mu$. En général ces notions de convergences ne sont pas équivalentes¹¹ : par exemple la suite de mesures de Dirac en les entiers $(\delta_n)_n$ converge vaguement vers la mesure nulle (puisque pour toute fonction à support compact il existe un rang à partir duquel n sort du compact et donc l'intégrale vaut 0 à partir d'un certain rang) mais ne converge pas étroitement – exercice : le montrer par l'absurde!

Toutefois les deux notions de convergence de mesure deviennent équivalentes dès lors qu'on ajoute la *conservation de la masse* : si μ_n converge vaguement vers μ et si $\mu_n(\mathbb{R}) \rightarrow \mu(\mathbb{R})$, alors μ_n converge étroitement vers μ .

On a une caractérisation particulière de la convergence en loi dans le cas de variables aléatoires à valeurs entières.

Théorème 6. *Soit $(X_n)_n$ une suite de variables aléatoires à valeurs dans \mathbb{N} et X une variable aléatoire à valeurs dans \mathbb{N} . On a l'équivalence entre les points suivants :*

1. X_n converge vers X en loi ;
2. Pour tout $k \in \mathbb{N}$, on a $\lim_n \mathbb{P}(X_n = k) = \mathbb{P}(X = k)$;
3. La suite des fonctions génératrices de X_n converge simplement sur $[0, 1]$ vers la fonction génératrice de X :

$$\lim_{n \rightarrow \infty} G_{X_n}(s) = G_X(s), \quad \forall s \in [0, 1].$$

On a également une caractérisation de la convergence en loi dans le cas de variables aléatoires à valeurs réelles.

Théorème 7. *Soit $(X_n)_n$ une suite de variables aléatoires réelles et X une variable aléatoire réelle. On a l'équivalence entre les points suivants :*

1. X_n converge vers X en loi ;
2. Pour tout $x \in \mathbb{R}$ tel que $F_X(x) = F_X(x^-)$ (on parle de point de continuité de la fonction de répartition de X), $F_{X_n}(x) \rightarrow F_X(x)$;
3. La suite des fonctions caractéristiques de X_n converge simplement sur \mathbb{R} vers la fonction caractéristique de X :

$$\lim_{n \rightarrow \infty} \phi_{X_n}(t) = \phi_X(t), \quad \forall t \in \mathbb{R}.$$

On a par ailleurs le résultat suivant : si (ϕ_n) est une suite de fonctions caractéristiques de variables aléatoires X_n et si ϕ_n converge simplement vers une fonction ϕ qui est continue en 0, alors ϕ est la fonction caractéristique d'une variable aléatoire.

Terminons par un théorème qui permet d'effectuer quelques calculs avec des limites en loi.

Théorème 8 (Théorème de Slutsky). *Soit (X_n) et (Y_n) deux suites de v.a. réelles définies sur un même espace de probabilité. Si X_n converge en loi vers X et Y_n converge en loi vers $c \in \mathbb{R}$, alors :*

1. $X_n + Y_n \xrightarrow{(loi)} X + c$,
2. $X_n Y_n \xrightarrow{(loi)} cX$,
3. $\frac{X_n}{Y_n} \xrightarrow{(loi)} \frac{X}{c}$ si $c \neq 0$.

11. En revanche, toute fonction continue à support compact est bornée donc la convergence étroite implique la convergence vague.

Théorèmes limites

Ces théorèmes sont **fondamentaux** en statistique¹², et font partie des résultats les plus importants des probabilités. Leur énoncé et leur démonstration nécessitent une connaissance et une compréhension abouties de tout ce qui précède dans le cours, et ce n'est pas pour rien qu'un chapitre entier est consacré à chacun d'entre eux. De nombreux exemples figurent dans le poly et méritent d'être traités afin de comprendre les tenants et aboutissants de ces théorèmes, et il est important d'en maîtriser les limites. Par exemple, une suite de v.a. indépendantes de loi de Cauchy ne vérifie les hypothèses ni de la loi des grands nombres, ni du théorème central limite, et par conséquent la moyenne empirique de telles variables aléatoires admet une limite très particulière – on peut montrer (exercice!) qu'elle diverge presque-sûrement mais qu'elle converge en loi vers une v.a. de Cauchy.

Loi des grands nombres

Théorème 9 (Loi des grands nombres). *Soit (X_n) une suite de variables aléatoires réelles indépendantes et identiquement distribuées. On pose pour tout $n \in \mathbb{N}^*$ la suite des sommes partielles $(S_n)_n$:*

$$S_n = X_1 + \cdots + X_n.$$

Si $\mathbb{E}[|X_1|] < \infty$, alors $\frac{S_n}{n} \rightarrow \mathbb{E}[X_1]$ presque-sûrement (loi forte des grands nombres) et en probabilité (loi faible des grands nombres). Si $\mathbb{E}[|X_1|] = \infty$ alors la suite $(\frac{S_n}{n})_n$ diverge presque-sûrement.

Théorème central limite

Théorème 10 (Théorème central limite). *Soit $(X_n)_n$ une suite de variables aléatoires réelles indépendantes et identiquement distribuées, avec $\mathbb{E}[X_1^2] < \infty$. On pose pour tout $n \in \mathbb{N}^*$ la suite des sommes partielles $(S_n)_n$:*

$$S_n = X_1 + \cdots + X_n,$$

et on note également $\sigma^2 = \text{Var}(X_1)$. Alors

$$\frac{S_n - n\mathbb{E}[X_1]}{\sqrt{n}} \underset{n \rightarrow \infty}{\xrightarrow{\text{(loi)}}} X,$$

où $X \sim \mathcal{N}(0, \sigma^2)$ est une variable aléatoire gaussienne centrée de variance σ^2 .

12. À tel point que le théorème de Glivenko–Cantelli, qui est un corollaire de la loi forte des grands nombres, est souvent appelé “théorème fondamental de la statistique”.