

Probabilités et statistiques continues avancées

Michel Pain

KMAXPP03

Université Paul Sabatier

Licence 3

24 octobre 2024

Table des matières

1	Fondements de la théorie des probabilités	6
1.1	Espace de probabilité	6
1.1.1	Définition	6
1.1.2	Rappels de théorie de la mesure	7
1.2	Variables aléatoires	8
1.2.1	Cadre général	8
1.2.2	Variable aléatoire réelle	9
1.3	Espérance	10
1.3.1	Définition	10
1.3.2	Rappels de théorie de la mesure	11
1.3.3	Théorème de transfert	12
1.3.4	Moments	14
1.3.5	Espérance et queue de distribution	14
1.4	Fonctions caractérisant une loi sur \mathbb{R}	15
1.4.1	Fonction de répartition	15
1.4.2	Fonction caractéristique	18
1.4.3	Transformée de Laplace	21
1.5	Lois classiques	21
1.5.1	Lois discrètes	22
1.5.2	Lois continues	23
2	Vecteurs aléatoires et indépendance	25
2.1	Rappels de théorie de la mesure	25
2.1.1	Tribus produit	25
2.1.2	Mesures produit	25
2.1.3	Formule de changement de variable	26
2.2	Vecteurs aléatoires, loi jointe et marginale	27
2.2.1	Loi jointe, lois marginales	27
2.2.2	Lois continues sur \mathbb{R}^d	28
2.3	Indépendance	30
2.3.1	Indépendance d'événements	30
2.3.2	Indépendance de variables aléatoires	31
2.3.3	Indépendance de variables aléatoires continues	33
2.3.4	Indépendance de variables aléatoires discrètes	34
2.3.5	Somme de variables aléatoires indépendantes	35
2.3.6	Minimum ou maximum de variables aléatoires indépendantes	36
3	Inégalités de concentration et intervalles de confiance	38
3.1	Inégalités de concentration	38
3.1.1	Inégalité de Markov	38
3.1.2	Inégalité de Bienaymé–Tchebychev	39

3.1.3	Inégalité de Hoeffding	40
3.2	Intervalles de confiance	42
4	Convergence de variables aléatoires et loi des grands nombres	44
4.1	Suite de variables aléatoires indépendantes	44
4.2	Rappels sur les limites inférieure et supérieure	45
4.2.1	Limites inférieure et supérieure de réels	45
4.2.2	Limites inférieure et supérieure d'ensembles	45
4.3	Premières notions de convergence	46
4.3.1	Convergence presque sûre	46
4.3.2	Convergence L^p	47
4.3.3	Convergence en probabilité	49
4.3.4	Liens entre les notions	50
4.4	Lemmes de Borel–Cantelli et applications	51
4.4.1	Lemmes de Borel–Cantelli	51
4.4.2	Application à la convergence p.s.	54
4.4.3	Une caractérisation de la convergence en probabilité (bonus)	57
4.5	La loi forte des grands nombres	58
4.5.1	Démonstration	58
4.5.2	Applications	62
5	Convergence en loi et théorème central limite	64
5.1	Convergence en loi et convergence étroite	64
5.1.1	Définitions	64
5.1.2	Lien avec les autres convergences	66
5.1.3	Caractérisations de la convergence en loi	67
5.2	Théorème central limite et applications	71
5.2.1	Le théorème central limite	71
5.2.2	Précision dans la méthode de Monte–Carlo	74
5.2.3	Intervalles de confiance asymptotiques	75

Préambule

Ce cours introduit à la théorie des probabilités en se reposant sur la théorie de la mesure. L'usage de la théorie de la mesure permet d'unifier les probabilités discrètes et continues. Ainsi, le cours ne se limite pas au cas des variables aléatoires continues, mais présente des résultats généraux, qui peuvent ensuite s'appliquer dans les cas discrets et continus. Un objectif de cours est d'arriver à la présentation de deux des plus célèbres résultats en probabilité : la loi des grands nombres et le théorème central limite. Concernant les statistiques, elles ne seront que discrètement présentes dans ce cours, qui couvre uniquement la question des intervalles de confiance. Voici quelques références utilisées pour créer ces notes :

- *Probabilité* de Philippe Barbe et Michel Ledoux, livre avec exercices.
- *De l'intégration aux probabilités* d'Olivier Garet et Aline Kurtzmann, livre avec exercices corrigés (pour certains).
- *Intégration, Probabilités et Processus Aléatoires* de Jean-François Le Gall, notes de cours disponible [ici](#).

Notations

$\mathcal{P}(E)$	ensemble des parties de E (pour E un ensemble)
$\mathcal{B}(E)$	tribu borélienne de E (pour E un espace topologique)
$ E $	cardinal de E (pour E un ensemble)
$[[m, n]]$	ensemble des entiers de m à n inclus
δ_x	masse de Dirac en x
λ_d	mesure de Lebesgue sur \mathbb{R}^d

Chapitre 1

Fondements de la théorie des probabilités

1.1 Espace de probabilité

1.1.1 Définition

Définition 1.1.1. Un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ est un ensemble Ω muni d'une tribu \mathcal{A} sur laquelle est définie une mesure de probabilité \mathbb{P} (c'est-à-dire une mesure telle que $\mathbb{P}(\Omega) = 1$).

Il faut voir Ω comme l'ensemble de toutes les éventualités possibles, toutes les déterminations du hasard dans l'expérience considérée. Les éléments de \mathcal{A} sont appelés *événements*, ce sont les parties de Ω auxquelles on peut attribuer une probabilité. Contrairement au cas des probabilités discrètes, on ne peut pas a priori attribuer une probabilité à n'importe quelle partie de Ω .

Exemple 1.1.2. On veut modéliser un lancer de dé à 6 faces équilibré. Une première manière de faire est de considérer $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ l'ensemble des parties de Ω et \mathbb{P} la mesure uniforme sur Ω , que l'on peut écrire

$$\mathbb{P} = \frac{1}{6}(\delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5 + \delta_6).$$

Cependant, il y a plein d'autres manières de modéliser cette expérience aléatoire : on peut toujours "grossir" Ω en rajoutant des réalisations de l'expérience qui n'ont aucune chance d'arriver. On peut par exemple prendre $\Omega = \{1, 2, 3, 4, 5, 6, 7\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ et la même mesure \mathbb{P} que précédemment : l'issue 7 est alors présente dans l'espace des possibles, mais l'événement $\{7\}$ a une probabilité nulle d'avoir lieu. De la même manière, on peut prendre $\Omega = \mathbb{R}$, \mathcal{A} n'importe quelle tribu contenant toutes les parties de $\{1, 2, 3, 4, 5, 6\}$ (par exemple $\mathcal{P}(\mathbb{R})$ ou les boréliens de \mathbb{R} ou seulement les parties de $\{1, 2, 3, 4, 5, 6\}$) et \mathbb{P} comme précédemment.

Exemple 1.1.3. On tire aléatoirement un nombre dans $[0, 1]$ de manière uniforme. Cette expérience peut être modélisée par $([0, 1], \mathcal{B}([0, 1]), \lambda)$, où $\mathcal{B}([0, 1])$ est la tribu borélienne sur $[0, 1]$ et λ est la mesure de Lebesgue sur $[0, 1]$. Notons que, pour cette probabilité, chaque issue de l'expérience a une probabilité nulle : $\lambda(\{a\}) = 0$ pour tout $a \in [0, 1]$. Par contre, la probabilité d'obtenir un résultat inférieur ou égal à $1/3$ est $\lambda([0, 1/3]) = 1/3$.

Exemple 1.1.4 (Paradoxe de Bertrand). Joseph Bertrand pose en 1889 la question suivante : quelle est la probabilité qu'une corde tirée au hasard sur un cercle de rayon 1 soit plus longue que $\sqrt{3}$, c'est-à-dire que la longueur d'un côté d'un triangle équilatéral inscrit ? Bertrand propose trois manières de tirer cette corde :

- (i) On choisit les deux extrémités de la corde uniformément au hasard sur le cercle ;
- (ii) On choisit le centre de la corde uniformément au hasard sur le disque unité ;

- (iii) On choisit uniformément au hasard la direction du rayon orthogonal à la corde, puis le centre de la corde uniformément sur ce rayon.

Il montre alors que la probabilité recherchée est $1/3$ dans le cas (i), $1/4$ dans le cas (ii) et $1/2$ dans le cas (iii). Il présente cela comme un paradoxe, bien qu'il soit conscient qu'il n'y en a pas : cela met en évidence le fait qu'il y a plusieurs manières de "tirer une corde au hasard sur le cercle", qui se précisent en définissant l'espace de probabilité sur lequel on travaille et il est normal que selon l'espace de probabilité, le résultat diffère. Cet exemple est traité en détail dans le TD2.

Définition 1.1.5. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et $A \in \mathcal{A}$. On dit que l'événement A a lieu presque sûrement (abrégé en p.s.) si $\mathbb{P}(A) = 1$ (c'est-à-dire s'il a lieu \mathbb{P} -presque partout).

Exemple 1.1.6. On tire un nombre aléatoire uniformément dans $[0, 1]$ comme dans l'Exemple 1.1.3. Alors il est presque sûrement irrationnel : en effet, $\lambda([0, 1] \setminus \mathbb{Q}) = 1$. C'est a priori possible dans notre modèle de tirer un nombre rationnel puisque toutes les issues dans $[0, 1]$ sont incluses, mais cela a lieu avec probabilité nulle.

Définition 1.1.7. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et $A, B \in \mathcal{A}$. Si $\mathbb{P}(B) > 0$, on définit la probabilité conditionnelle de A sachant B par

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

1.1.2 Rappels de théorie de la mesure

La proposition suivante recense diverses propriétés de la mesure \mathbb{P} qui seront utilisées à répétition. Les points (i) et (vi) font partie de la définition de mesure de probabilité. Les autres sont des propriétés vues en cours de théorie de la mesure. En particulier, la continuité décroissante en (viii) est toujours vraie pour une mesure de probabilité car c'est une mesure finie.

Proposition 1.1.8. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Soit $A, B \in \mathcal{A}$ et $(A_n)_{n \geq 0} \in \mathcal{A}^{\mathbb{N}}$.

- (i) $\mathbb{P}(\emptyset) = 0$ et $\mathbb{P}(\Omega) = 1$.
- (ii) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
- (iii) Si $A \subset B$ alors $\mathbb{P}(A) \leq \mathbb{P}(B)$.
- (iv) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
- (v) $\mathbb{P}(\bigcup_{n \geq 0} A_n) \leq \sum_{n \geq 0} \mathbb{P}(A_n)$.
- (vi) Si les A_n sont deux à deux disjoints, alors $\mathbb{P}(\bigsqcup_{n \geq 0} A_n) = \sum_{n \geq 0} \mathbb{P}(A_n)$.
- (vii) Si $(A_n)_{n \geq 0}$ est une suite croissante, alors $\mathbb{P}(\bigcup_{n \geq 0} \uparrow A_n) = \lim_{n \rightarrow \infty} \uparrow \mathbb{P}(A_n)$.
- (viii) Si $(A_n)_{n \geq 0}$ est une suite décroissante, alors $\mathbb{P}(\bigcap_{n \geq 0} \downarrow A_n) = \lim_{n \rightarrow \infty} \downarrow \mathbb{P}(A_n)$.

Théorème 1.1.9 (Théorème d'unicité de mesure). Soit (E, \mathcal{E}) un espace mesurable. Soit P et Q deux mesures de probabilité sur E . Supposons qu'il existe $\mathcal{C} \subset \mathcal{E}$ tel que \mathcal{C} soit stable par intersections finies et

$$\forall C \in \mathcal{C}, \quad P(C) = Q(C).$$

Alors P et Q coïncident sur la tribu engendrée par \mathcal{C} .

Théorème 1.1.10 (Théorème d'unicité de mesure 2). Soit $d \geq 1$. Soit P et Q deux mesures de probabilité sur \mathbb{R}^d . Les propriétés suivantes sont équivalentes

- (i) $P = Q$;
- (ii) pour toute $f: \mathbb{R}^d \rightarrow \mathbb{R}_+$ mesurable, $\int_{\mathbb{R}^d} f(x) dP(x) = \int_{\mathbb{R}^d} f(x) dQ(x)$;
- (iii) pour toute $f: \mathbb{R}^d \rightarrow \mathbb{R}$ continue bornée, $\int_{\mathbb{R}^d} f(x) dP(x) = \int_{\mathbb{R}^d} f(x) dQ(x)$;
- (iv) pour toute $f: \mathbb{R}^d \rightarrow \mathbb{R}$ de classe \mathcal{C}^∞ à support compact¹, $\int_{\mathbb{R}^d} f(x) dP(x) = \int_{\mathbb{R}^d} f(x) dQ(x)$.

1. Une fonction est dite à support compact s'il existe un compact en dehors duquel elle est nulle.

1.2 Variables aléatoires

1.2.1 Cadre général

Définition 1.2.1. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et (E, \mathcal{E}) un espace mesurable. Une variable aléatoire (abrégé en v.a.) à valeurs dans E est une application mesurable $X: (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$, c'est-à-dire vérifiant

$$\forall B \in \mathcal{E}, X^{-1}(B) \in \mathcal{A}.$$

Exemple 1.2.2. On lance deux dés à 6 faces équilibrés, ce que l'on modélise par $\Omega = \llbracket 1, 6 \rrbracket^2$, $\mathcal{A} = \mathcal{P}(\Omega)$ et \mathbb{P} la mesure uniforme sur Ω , qui vérifie $\mathbb{P}(A) = |A|/36$ pour tout $A \subset \Omega$. Alors $X((i, j)) = i + j$, pour $(i, j) \in \Omega$, définit une variable aléatoire à valeurs dans $\llbracket 1, 12 \rrbracket$. La v.a. X représente la somme des deux dés. On peut aussi dire que X est une v.a. à valeurs dans \mathbb{R} .

Notation. Si (E, \mathcal{E}) et (F, \mathcal{F}) sont des espaces mesurables, X est une v.a. à valeurs dans E et $f: E \rightarrow F$ une fonction mesurable, alors on écrit $f(X)$ pour la variable aléatoire $f \circ X: \Omega \rightarrow F$, qui est bien mesurable comme composée de fonctions mesurables.

Définition 1.2.3. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité, (E, \mathcal{E}) un espace mesurable et X une v.a. à valeurs dans E . La loi de X , notée P_X , est la mesure image de \mathbb{P} par X , c'est-à-dire la mesure de probabilité définie par

$$\forall B \in \mathcal{E}, P_X(B) = \mathbb{P}(X^{-1}(B)).$$

De manière générale, une loi sur (E, \mathcal{E}) désigne une mesure de probabilité sur (E, \mathcal{E}) .

Le fait que P_X est bien une mesure de probabilité a été vu au TD1. Connaître la loi de X permet de calculer la probabilité des événements dépendants de la variable aléatoire X . À chaque détermination du hasard $\omega \in \Omega$ correspond une réalisation $X(\omega)$, et $P_X(B)$ nous donne la probabilité que cette réalisation tombe dans B .

Notation. En pratique, l'événement $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ est noté $\{X \in B\}$. De plus, quand on en prend la probabilité, on omet les accolades et on écrit $\mathbb{P}(X \in B)$. Ainsi,

$$P_X(B) = \mathbb{P}(X \in B).$$

De la même manière, on écrit par exemple $\{X = x\} = \{\omega \in \Omega : X(\omega) = x\}$ ou, si X et Y sont des v.a. réelles (c'est-à-dire à valeurs dans \mathbb{R}), $\{X < Y\} = \{\omega \in \Omega : X(\omega) < Y(\omega)\}$. On écrit également $\mathbb{P}(X = x)$ et $\mathbb{P}(X < Y)$ sans accolades pour les probabilités associées.

Exemple 1.2.4. Dans le cadre de l'Exemple 1.2.2, la loi de X est donnée par

$$P_X = \frac{1}{36}\delta_2 + \frac{2}{36}\delta_3 + \frac{3}{36}\delta_4 + \frac{4}{36}\delta_5 + \frac{5}{36}\delta_6 + \frac{6}{36}\delta_7 + \frac{5}{36}\delta_8 + \frac{4}{36}\delta_9 + \frac{3}{36}\delta_{10} + \frac{2}{36}\delta_{11} + \frac{1}{36}\delta_{12},$$

que l'on obtient en calculant la probabilité que $X = k$ pour chaque $k \in \llbracket 1, 12 \rrbracket$: par exemple

$$P_X(\{5\}) = \mathbb{P}(X = 5) = \mathbb{P}(\{(1, 4), (2, 3), (3, 2), (4, 1)\}) = \frac{4}{36}.$$

Connaître les $P_X(\{k\})$ pour tout $k \in E = \llbracket 1, 12 \rrbracket$ caractérise P_X car l'espace E est fini (vrai aussi si E est dénombrable).

Il y a alors deux manières de calculer la probabilité d'un événement dépendant de X . Soit on travaille sur l'espace $\Omega = \llbracket 1, 6 \rrbracket^2$ en utilisant la mesure \mathbb{P} :

$$\mathbb{P}(X \leq 4) = \mathbb{P}(\{(1, 1), (1, 2), (2, 1), (2, 2), (1, 3), (3, 1)\}) = \frac{6}{36} = \frac{1}{6}.$$

Soit on travaille sur l'espace $E = \llbracket 1, 12 \rrbracket$ en utilisant la mesure P_X :

$$\mathbb{P}(X \leq 4) = P_X(\{2, 3, 4\}) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} = \frac{1}{6}.$$

Notons que la loi d'une v.a. ne la caractérise pas : il peut y avoir plein de v.a. avec la même loi, voir l'exemple ci-dessous.

Exemple 1.2.5. Dans le cadre de l'exemple ci-dessus, considérons la v.a. $Y = 14 - X$. Alors elle a la même loi que X mais elle n'est pas égale à X (on a $X = Y$ ssi la somme des dés est 7).

Remarque 1.2.6. Dans ce cours, on travaillera principalement avec des variables aléatoires. Alors, l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ est principalement un prétexte pour pouvoir définir ces variables aléatoires. Ainsi, très souvent dans la suite, on ne spécifiera pas le choix de $(\Omega, \mathcal{A}, \mathbb{P})$, ce qui nous intéresse est uniquement l'espace d'arrivée (E, \mathcal{E}) et la loi P_X .

Par exemple, un énoncé peut commencer par "Soit X une v.a. de Poisson de paramètre 3" et cela est suffisant pour calculer la probabilité que X soit pair :

$$\mathbb{P}(X \in 2\mathbb{N}) = P_X(2\mathbb{N}) = \sum_{n=0}^{\infty} P_X(\{2n\}) = \sum_{n=0}^{\infty} \frac{e^{-3} 3^{2n}}{(2n)!} = e^{-3} \cosh(3) = \frac{1 + e^{-6}}{2},$$

sans que l'on ait à connaître le choix de $(\Omega, \mathcal{A}, \mathbb{P})$.

Notons que si μ est une mesure de probabilité sur (E, \mathcal{E}) , il existe toujours un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et une variable aléatoire à valeurs dans E de loi μ . En effet, il suffit de prendre $(\Omega, \mathcal{A}, \mathbb{P}) = (E, \mathcal{E}, \mu)$ et X la fonction identité. On sait donc qu'un espace $(\Omega, \mathcal{A}, \mathbb{P})$ convenable existe, mais le spécifier n'apporte généralement rien à la résolution des problèmes, donc on ne le fait pas (il y a quelques cas exceptionnels mais pas dans ce cours!).

En particulier, les énoncés qui suivront ne contiendront plus la mention "Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité", même s'il est toujours sous-entendu qu'on en a fixé un.

1.2.2 Variable aléatoire réelle

Définition 1.2.7. Une variable aléatoire réelle est une v.a. à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Il y a deux cas particuliers importants de v.a. réelles : les v.a. discrètes et les v.a. continues que nous allons à présent définir. L'intérêt de l'approche moderne des probabilités à l'aide de la théorie de la mesure est d'unifier ces deux cas et de les traiter de manière indifférenciée.

Définition 1.2.8. Une loi P sur \mathbb{R} est dite *discrète* si elle s'écrit de la forme

$$P = \sum_{i \in \mathcal{I}} p_i \delta_{x_i},$$

avec \mathcal{I} un ensemble non vide fini ou dénombrable, $(p_i)_{i \in \mathcal{I}} \in (\mathbb{R}_+^*)^{\mathcal{I}}$ et $x_i \in \mathbb{R}^{\mathcal{I}}$. Les x_i sont appelés les *atomes* de P . Une v.a. réelle X est dite *discrète* si sa loi est discrète.

Notons que si X a une loi discrète écrite sous la forme ci-dessus, alors $p_i = \mathbb{P}(X = x_i)$. On a aussi nécessairement

$$\sum_{i \in \mathcal{I}} p_i = 1.$$

Rappelons que pour une telle loi P , on a

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad P(B) = \sum_{i \in \mathcal{I}} p_i \mathbb{1}_B(x_i)$$

$$\forall f: \mathbb{R} \rightarrow \mathbb{R}_+ \text{ mesurable}, \quad \int_{\mathbb{R}} f(x) dP(x) = \sum_{i \in \mathcal{I}} p_i f(x_i)$$

De nombreux exemples de v.a. discrètes ont été vus en cours de probabilités discrètes : variables de Bernoulli, binomiale, géométrique, de Poisson. Voir la Section 1.5 pour un rappel.

De manière générale, on dira qu'une v.a. réelle X (ou que sa loi) a un *atome* en $x \in \mathbb{R}$ si $\mathbb{P}(X = x) > 0$. Une loi discrète est entièrement caractérisée par la connaissance de ses atomes et de leur mesure. À l'opposé, une loi est dite *diffuse* si elle n'a pas d'atome. Les lois continues définies ci-dessous sont un cas particulier de loi diffuse.

Définition 1.2.9. Une loi sur \mathbb{R} est dite *continue* si elle est à densité par rapport à la mesure de Lebesgue sur \mathbb{R} . Une v.a. réelle X est dite *continue* si sa loi est continue et on notera alors généralement $p_X: \mathbb{R} \rightarrow \mathbb{R}_+$ sa densité.

Rappelons qu'on dit que P a densité $p: \mathbb{R} \rightarrow \mathbb{R}_+$ par rapport à la mesure de Lebesgue sur \mathbb{R} si p est mesurable et

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad P(B) = \int_B p(x) dx,$$

ce qui est équivalent à

$$\forall f: \mathbb{R} \rightarrow \mathbb{R}_+ \text{ mesurable}, \quad \int_{\mathbb{R}} f(x) dP(x) = \int_{\mathbb{R}} f(x)p(x) dx$$

Le fait que P soit une mesure de probabilité implique que l'intégrale de p sur \mathbb{R} est égale à 1. On note $dP(x) = p(x) dx$ et on peut dire " X a loi $p(x) dx$ " si X a pour loi P (c'est un léger abus de notation où on amalgame P et $dP(x)$).

Exemple 1.2.10. Soit X une v.a. réelle de loi $\frac{x}{2} \mathbb{1}_{[0,2]}(x) dx$. Alors on peut par exemple calculer la probabilité suivante

$$\mathbb{P}(X \leq 1) = P_X(]-\infty, 1]) = \int_{-\infty}^1 \frac{x}{2} \mathbb{1}_{[0,2]}(x) dx = \int_0^1 \frac{x}{2} dx = \left[\frac{x^2}{4} \right]_0^1 = \frac{1}{4}.$$

En outre, comme la densité de X est nulle en dehors de $[0, 2]$, on peut dire que $X \in [0, 2]$ p.s.

Remarque 1.2.11. Il y a des v.a. réelles qui ne sont ni discrètes, ni continues. Il suffit de considérer un mélange de loi discrète et de loi continue, comme par exemple

$$\frac{1}{2} \delta_0 + \frac{1}{2} \lambda,$$

où λ est la mesure de Lebesgue sur $[0, 1]$. Cette loi correspond à l'expérience suivante : avec probabilité 1/2 on tire 0 et avec probabilité 1/2 on tire un nombre dans $[0, 1]$ uniformément. Il existe également des lois sur \mathbb{R} qui sont diffuses mais pas continues : on peut par exemple considérer la mesure uniforme sur l'ensemble triadique de Cantor (qu'on ne construira pas proprement ici) qui n'a aucun atome mais n'a pas de densité par rapport à la mesure de Lebesgue (car elle est supportée par l'ensemble de Cantor qui a mesure de Lebesgue nulle).

1.3 Espérance

1.3.1 Définition

Définition 1.3.1. Soit X une v.a. réelle. L'espérance de X est définie par

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega),$$

dès que cette intégrale a un sens, c'est-à-dire dans l'un des cas suivants :

- si $X \geq 0$ (alors $\mathbb{E}[X]$ peut potentiellement être infinie) ;
- si X est intégrable pour \mathbb{P} , c'est-à-dire $\mathbb{E}[|X|] < \infty$.

Exemple 1.3.2. Soit $n \in \mathbb{N}^*$. On considère $\Omega = \llbracket 1, n \rrbracket$ avec $\mathcal{A} = \mathcal{P}(\Omega)$ et \mathbb{P} la mesure uniforme sur Ω . Soit X une v.a. réelle. Alors

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \sum_{i=1}^n X(i) \mathbb{P}(\{i\}) = \frac{X(1) + X(2) + \dots + X(n)}{n}.$$

Dans ce cas, l'espérance est donc la moyenne des valeurs de X au sens classique du terme.

Remarque 1.3.3. Soit $A \in \mathcal{A}$ un événement. Rappelons que la fonction indicatrice $\mathbb{1}_A$ est définie par

$$\forall \omega \in \Omega, \quad \mathbb{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{sinon.} \end{cases}$$

Comme $\mathbb{1}_A$ est mesurable (car A l'est), c'est une v.a. Son espérance est $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$.

1.3.2 Rappels de théorie de la mesure

Par définition, l'espérance est l'intégrale par rapport à \mathbb{P} . Les résultats sur les intégrales par rapport à une mesure quelconque s'appliquent donc directement. Rappelons-en quelques-uns.

Proposition 1.3.4 (Linéarité de l'espérance). Soit X, Y des v.a. réelles et $a, b \in \mathbb{R}$.

- Si X et Y sont intégrables pour \mathbb{P} , alors $aX + bY$ est intégrable pour \mathbb{P} .
- Si $X, Y, a, b \geq 0$, alors $aX + bY \geq 0$.

Dans chacun des cas précédents, on a $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.

Pour montrer la convergence de l'espérance d'une suite de v.a., on dispose des trois mêmes outils que pour les intégrales. Rappelons la définition suivante, pour $(a_n)_{n \geq 0} \in \overline{\mathbb{R}}^{\mathbb{N}}$,

$$\limsup_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \downarrow \sup_{k \geq n} a_k \quad \text{et} \quad \liminf_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \uparrow \inf_{k \geq n} a_k.$$

Théorème 1.3.5. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. réelles.

- (Convergence monotone). Si $X_n \geq 0$ pour tout $n \geq 0$ et $X_n \uparrow X$, alors $\mathbb{E}[X_n] \uparrow \mathbb{E}[X]$.
- (Lemme de Fatou). Si $X_n \geq 0$ pour tout $n \geq 0$, alors $\mathbb{E}[\liminf X_n] \leq \liminf \mathbb{E}[X_n]$.
- (Convergence dominée). Si X_n converge p.s. vers une v.a. réelle X et qu'il existe une v.a. réelle Z telle que $\mathbb{E}[|Z|] < \infty$ et $|X_n| \leq Z$ p.s. pour tout $n \geq 0$, alors $\mathbb{E}[X_n] \rightarrow \mathbb{E}[X]$.

Rappelons les théorèmes de continuité et de dérivabilité des intégrales à paramètre.

Théorème 1.3.6 (Continuité d'espérances à paramètre). Soit X une v.a. à valeurs dans (E, \mathcal{E}) . Soit I un intervalle de \mathbb{R} et $f: I \times E \rightarrow \mathbb{R}$. Supposons

- pour tout $t \in I$, l'application $x \in E \mapsto f(t, x)$ est mesurable ;
- presque sûrement, l'application $t \in I \mapsto f(t, X)$ est continue sur I ;
- il existe une v.a. réelle Z telle que $\mathbb{E}[|Z|] < \infty$ et, pour tout $t \in I$, $|f(t, X)| \leq Z$ p.s.

Alors la fonction $t \in I \mapsto \mathbb{E}[f(t, X)]$ est continue sur I .

Théorème 1.3.7 (Dérivabilité d'espérances à paramètre). Soit X une v.a. à valeurs dans (E, \mathcal{E}) . Soit I un intervalle de \mathbb{R} et $f: I \times E \rightarrow \mathbb{R}$. Supposons

- pour tout $t \in I$, l'application $x \in E \mapsto f(t, x)$ est mesurable ;
- presque sûrement, l'application $t \in I \mapsto f(t, X)$ est dérivable sur I ;
- il existe une v.a. réelle Z telle que $\mathbb{E}[|Z|] < \infty$ et, pour tout $t \in I$, $|\frac{\partial f}{\partial t}(t, X)| \leq Z$ p.s.

Alors la fonction $t \in I \mapsto \mathbb{E}[f(t, X)]$ est dérivable sur I , de dérivée $t \mapsto \mathbb{E}[\frac{\partial f}{\partial t}(t, X)]$.

On peut également considérer les espaces $L^p(\Omega, \mathcal{A}, \mathbb{P})$ pour $p \in [1, \infty]$. Pour X une v.a. réelle, rappelons que $\|X\|_p = \mathbb{E}[|X|^p]^{1/p}$ quand $p < \infty$ et $\|X\|_\infty = \inf\{C > 0 : |X| \leq C \text{ p.s.}\}$. L'espace $L^p(\Omega, \mathcal{A}, \mathbb{P})$ est défini comme l'ensemble des v.a. réelles X telles que $\|X\|_p < \infty$, que l'on quotiente par la relation d'équivalence "être égal p.s." (i.e. deux variables aléatoires qui sont égales p.s. sont considérées comme identiques). Alors $(L^p(\Omega, \mathcal{A}, \mathbb{P}), \|\cdot\|_p)$ est un espace de Banach.

Proposition 1.3.8 (Inégalité de Hölder). Soit X et Y des v.a. réelles. Soit $p, q \in [1, \infty]$ tels que $\frac{1}{p} + \frac{1}{q} = 1$. Alors

$$\|XY\|_1 \leq \|X\|_p \|Y\|_q.$$

En particulier, si $X \in L^p(\Omega, \mathcal{A}, \mathbb{P})$ et $Y \in L^q(\Omega, \mathcal{A}, \mathbb{P})$, alors $XY \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. Dans le cas où $p, q \in]1, \infty[$, on peut réécrire l'inégalité de Hölder comme

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{1/p} \cdot \mathbb{E}[|Y|^q]^{1/q}.$$

Le cas $p = q = 2$ donne l'inégalité de Cauchy-Schwarz :

$$\mathbb{E}[|XY|] \leq \mathbb{E}[X^2]^{1/2} \cdot \mathbb{E}[Y^2]^{1/2}.$$

1.3.3 Théorème de transfert

Le théorème de transfert (ou de transport) est une formule exprimant une espérance d'une fonction d'une v.a. X comme une intégrale contre la loi de X . C'est une conséquence simple de la définition de la loi de X comme mesure image, mais c'est un résultat très important : c'est l'outil principal pour calculer des espérances.

Théorème 1.3.9 (Théorème de transfert). Soit X une v.a. à valeurs dans un espace mesurable (E, \mathcal{E}) . Soit $f: E \rightarrow \overline{\mathbb{R}}$ une fonction mesurable (rappelons que $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$). On a l'équivalence

$$f \in L^1(E, \mathcal{E}, P_X) \Leftrightarrow f(X) \in L^1(\Omega, \mathcal{A}, \mathbb{P}).$$

En outre, si $f \in L^1(E, \mathcal{E}, P_X)$ ou si f est positive, alors

$$\mathbb{E}[f(X)] = \int_E f(x) dP_X(x). \tag{1.1}$$

Démonstration. C'est une propriété générale de la mesure image. La démonstration consiste à traiter des fonctions de plus en plus générales, de manière similaire à la construction de l'intégrale en théorie de la mesure.

Étape 1. On commence par remarquer que la formule (1.1) est vraie par définition de P_X quand $f = \mathbb{1}_B$ avec $B \in \mathcal{B}(\mathbb{R})$:

$$\begin{aligned} \mathbb{E}[f(X)] &= \mathbb{E}[\mathbb{1}_{X^{-1}(B)}] && \text{(car } f = \mathbb{1}_B) \\ &= \mathbb{P}(X^{-1}(B)) && \text{(cf Remarque 1.3.3)} \\ &= P_X(B) && \text{(par définition de } P_X) \\ &= \int_E f(x) dP_X(x) && \text{(car } f = \mathbb{1}_B). \end{aligned}$$

Étape 2. Il en suit par linéarité que la formule (1.1) est vraie pour toute fonction étagée positive, i.e. pour $f = \sum_{i=1}^n c_i \mathbb{1}_{B_i}$, pour $n \in \mathbb{N}$, $c_1, \dots, c_n \geq 0$ et $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$.

Étape 3. Considérons à présent $f: E \rightarrow \overline{\mathbb{R}}$ mesurable et positive. Il a été vu en cours de théorie de la mesure qu'une telle fonction f est limite croissante d'une suite de fonctions étagées positives $(f_n)_{n \geq 0}$. Ainsi, on a

$$\begin{aligned} \mathbb{E}[f(X)] &= \lim_{n \rightarrow \infty} \mathbb{E}[f_n(X)] && \text{(convergence monotone)} \\ &= \lim_{n \rightarrow \infty} \int_E f_n(x) dP_X(x) && \text{(par l'étape 2)} \\ &= \int_E f(x) dP_X(x) && \text{(convergence monotone)}. \end{aligned}$$

Cela montre la formule (1.1) pour les fonctions f positives.

Étape 4. Soit $f: E \rightarrow \overline{\mathbb{R}}$ mesurable. Alors, par l'étape 3, la formule (1.1) est vraie pour $|f|$ donc les deux côtés de l'égalité sont soit finis tous les deux soit infinis tous les deux. Cela montre que $f \in L^1(E, \mathcal{E}, P_X)$ si et seulement si $f(X) \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. En outre, si $f \in L^1(E, \mathcal{E}, P_X)$, on décompose $f = f_+ - f_-$ avec $f_+ = \max(f, 0)$ et $f_- = \max(-f, 0)$. Comme f_+ est positive, on a, par l'étape 3,

$$\mathbb{E}[f_+(X)] = \int_E f_+(x) dP_X(x)$$

et les deux côtés de cette égalité sont finis car $f_+ \leq |f|$. La même chose étant vraie pour f_- , on en conclut par linéarité que la formule (1.1) est vraie pour f . \square

Exemple 1.3.10. Considérons une v.a. X de loi uniforme sur $[0, \pi]$, on veut calculer $\mathbb{E}[\sin(X)]$. Notons tout d'abord que X a pour densité $p_X(x) = \frac{1}{\pi} \mathbb{1}_{[0, \pi]}(x)$. Pour appliquer le théorème de transfert, on vérifie que $\sin \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$: en effet,

$$\int_{\mathbb{R}} |\sin| dP_X \leq \int_{\mathbb{R}} 1 dP_X = P_X(\mathbb{R}) = 1 < \infty.$$

On peut donc appliquer le théorème de transfert et on obtient

$$\mathbb{E}[\sin(X)] = \int_{\mathbb{R}} \sin(x) dP_X(x) = \int_{\mathbb{R}} \sin(x) p_X(x) dx = \frac{1}{\pi} \int_0^\pi \sin(x) dx = \frac{1}{\pi}.$$

Méthode (Déterminer une loi). Une méthode pour déterminer la loi d'une v.a. réelle X consiste à exprimer $\mathbb{E}[f(X)]$, pour $f: \mathbb{R} \rightarrow \mathbb{R}_+$ mesurable quelconque, sous la forme $\int_{\mathbb{R}} f(x) d\mu(x)$. Alors on peut en conclure que $P_X = \mu$: en effet, par la formule de transfert, cela implique

$$\int_{\mathbb{R}} f(x) dP_X(x) = \int_{\mathbb{R}} f(x) d\mu(x),$$

pour tout $f: \mathbb{R} \rightarrow \mathbb{R}_+$ mesurable, et donc $P_X = \mu$ (en prenant $f = \mathbb{1}_B$ pour $B \in \mathcal{B}(\mathbb{R})$).

Cette méthode est particulièrement utile dans le cas où X est de la forme $\psi(Y)$ où ψ est une fonction explicite et Y est une v.a. réelle continue dont on connaît la densité p_Y . Alors, on peut écrire

$$\mathbb{E}[f(X)] = \mathbb{E}[f(\psi(Y))] = \int_{\mathbb{R}} f(\psi(y)) p_Y(y) dy,$$

d'après le théorème de transfert. Il faut ensuite faire le changement de variable $x = \psi(y)$ pour se ramener à une intégrale de la forme $\int_{\mathbb{R}} f(x) p(x) dx$, ce qui permet de conclure que X a densité p . Ce changement de variable peut nécessiter de découper le domaine de ψ en parties sur lesquelles ψ est injective et dérivable, ce qui n'est pas toujours possible. En particulier, une telle v.a. X n'est pas forcément continue : par exemple, si ψ ne prend qu'un nombre fini de valeurs, alors X est discrète. Dans ce cas, la méthode présentée dans le 1er paragraphe peut quand même fonctionner, mais il faut s'attendre à trouver une mesure μ qui contient des masses de Dirac.

Exemple 1.3.11. Soit Y une v.a. réelle de loi $e^{-y} \mathbb{1}_{[0, \infty[}(y) dy$. On veut déterminer la loi de $X = Y^2$. Pour cela on écrit, pour $f: \mathbb{R} \rightarrow \mathbb{R}_+$ mesurable quelconque,

$$\mathbb{E}[f(X)] = \mathbb{E}[f(Y^2)] = \int_0^\infty f(y^2) e^{-y} dy.$$

On effectue alors le changement de variable $x = y^2$ qui est bijectif de \mathbb{R}_+^* dans \mathbb{R}_+^* , ce qui donne

$$\mathbb{E}[f(X)] = \int_0^\infty f(x) e^{-\sqrt{x}} \frac{1}{2\sqrt{x}} dx.$$

Cela montre que X est continue, de densité $x \mapsto \frac{e^{-\sqrt{x}}}{2\sqrt{x}} \mathbb{1}_{[0, \infty[}(x)$.

1.3.4 Moments

Définition 1.3.12. Soit X une v.a. réelle et $n \in \mathbb{N}$. Le **moment d'ordre n** de X (ou n -ième moment) est $\mathbb{E}[X^n]$, dès que cette quantité est bien définie (i.e. si $\mathbb{E}[|X|^n] < \infty$ ou $X \geq 0$).

Une conséquence importante de l'inégalité de Hölder est la propriété d'inclusion suivante pour les espaces L^p . Attention, elle est particulière aux mesures finies et n'est pas vraie pour des espaces L^p associés à des mesures infinies. En particulier, ce résultat implique que si une v.a. réelle a un moment d'ordre n fini, alors tous ses moments d'ordres plus petits sont finis aussi.

Proposition 1.3.13. Soit $1 \leq p < q \leq \infty$. Alors $L^q(\Omega, \mathcal{A}, \mathbb{P}) \subset L^p(\Omega, \mathcal{A}, \mathbb{P})$.

Démonstration. Soit X une v.a. réelle. En appliquant l'inégalité de Hölder aux v.a. $|X|^p$ et 1 avec $p' = q/p > 1$ et q' tel que $\frac{1}{p'} + \frac{1}{q'} = 1$, on obtient

$$\mathbb{E}[|X|^p] = \mathbb{E}[|X|^p \cdot 1] \leq \mathbb{E}\left[(|X|^p)^{p'} \right]^{1/p'} \mathbb{E}\left[(1)^{q'} \right]^{1/q'} = \mathbb{E}[|X|^q]^{p/q}.$$

Ainsi, si $X \in L^q(\Omega, \mathcal{A}, \mathbb{P})$, alors $\mathbb{E}[|X|^q] < \infty$ et donc $\mathbb{E}[|X|^p] < \infty$ par l'inégalité ci-dessus, ce qui montre que $X \in L^p(\Omega, \mathcal{A}, \mathbb{P})$. \square

Définition 1.3.14. Soit $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. La **variance** de X est

$$\text{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$

et l'écart type de X est $\sigma_X = \sqrt{\text{Var}(X)}$.

Proposition 1.3.15. Soit $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. Alors

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Démonstration. Notons $m = \mathbb{E}[X]$. On a alors en développant le carré et en utilisant la linéarité de l'espérance :

$$\text{Var}(X) = \mathbb{E}\left[(X - m)^2\right] = \mathbb{E}\left[X^2\right] - 2m\mathbb{E}[X] + m^2 = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2,$$

en remplaçant m par $\mathbb{E}[X]$ dans la dernière égalité. \square

1.3.5 Espérance et queue de distribution

La queue de distribution d'une v.a. réelle X fait référence au comportement asymptotique de $\mathbb{P}(X > x)$ et de $\mathbb{P}(X < -x)$ quand $x \rightarrow \infty$. Pour une v.a. positive, la formule suivante relie queue de distribution et espérance et peut se révéler très utile.

Proposition 1.3.16. Soit X une v.a. positive. Alors

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > x) dx = \int_0^\infty \mathbb{P}(X \geq x) dx.$$

Démonstration. Notons que l'espérance est bien définie car $X \geq 0$. Par le théorème de transfert (et le fait que P_X est supportée sur $[0, \infty)$ car $X \geq 0$), on a

$$\mathbb{E}[X] = \int_0^\infty x dP_X(x) = \int_0^\infty \left(\int_0^\infty \mathbb{1}_{t < x} dt \right) dP_X(x) = \int_0^\infty \left(\int_0^\infty \mathbb{1}_{t < x} dP_X(x) \right) dt,$$

par le théorème de Fubini–Tonelli. Notons alors que $\int_0^\infty \mathbb{1}_{t < x} dP_X(x) = P_X(]t, \infty[) = \mathbb{P}(X > t)$, ce qui montre la première égalité de l'énoncé. La deuxième égalité se montre de manière similaire, en remarquant que l'on peut remplacer ' $t < x$ ' par ' $t \leq x$ ' au début du calcul. \square

Corollaire 1.3.17. Soit X une v.a. positive et $p > 0$. Alors

$$\mathbb{E}[X^p] = \int_0^\infty \mathbb{P}(X > x) p x^{p-1} dx = \int_0^\infty \mathbb{P}(X \geq x) p x^{p-1} dx.$$

Démonstration. Voir TD 3. □

Ces formules permettent d'avoir des critères simples en terme de la vitesse de décroissance de la queue de distribution pour savoir si le moment d'ordre n de X est fini.

1.4 Fonctions caractérisant une loi sur \mathbb{R}

1.4.1 Fonction de répartition

Définition 1.4.1. Soit P une loi sur \mathbb{R} . La fonction de répartition de P est la fonction $F: \mathbb{R} \rightarrow [0, 1]$ définie par

$$\forall x \in \mathbb{R}, \quad F(x) = P(] - \infty, x]).$$

Soit X une v.a. réelle. La fonction de répartition de X est la fonction de répartition de sa loi. On la notera souvent $F_X: \mathbb{R} \rightarrow [0, 1]$ et elle vérifie

$$\forall x \in \mathbb{R}, \quad F_X(x) = \mathbb{P}(X \leq x).$$

Exemple 1.4.2. Soit X une variable réelle continue. Sa fonction de répartition est donnée, pour $x \in \mathbb{R}$, par

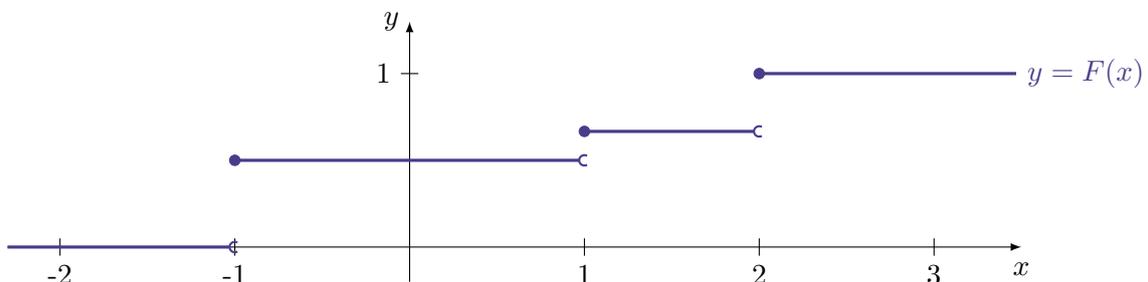
$$F_X(x) = \int_{-\infty}^x p_X(y) dy.$$

En particulier, si p_X est continue sur \mathbb{R} alors F_X est de classe \mathcal{C}^1 sur \mathbb{R} et $F'_X = p_X$.

Exemple 1.4.3. On considère la mesure discrète $P = \frac{1}{2}\delta_{-1} + \frac{1}{6}\delta_1 + \frac{1}{3}\delta_2$. Alors sa fonction de répartition est donnée, pour $x \in \mathbb{R}$, par

$$F(x) = \frac{1}{2} \mathbb{1}_{[-1, \infty[}(x) + \frac{1}{6} \mathbb{1}_{[1, \infty[}(x) + \frac{1}{3} \mathbb{1}_{[2, \infty[}(x) = \begin{cases} 0 & \text{if } x < -1, \\ 1/2 & \text{if } x \in [-1, 1[, \\ 2/3 & \text{if } x \in [1, 2[, \\ 1 & \text{if } x \geq 2, \end{cases}$$

et est représentée sur la figure suivante. On peut noter que la fonction de répartition est constante par morceaux et effectue un saut à chaque atome de P de la taille du poids de cet atome.



Proposition 1.4.4. Soit X une v.a. réelle. Sa fonction de répartition F_X est croissante, continue à droite et satisfait

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad \text{et} \quad \lim_{x \rightarrow \infty} F_X(x) = 1.$$

Démonstration. Voir TD1. □

Théorème 1.4.5. Une loi sur \mathbb{R} est caractérisée par sa fonction de répartition. Autrement dit, si deux v.a. ont la même fonction de répartition, alors elles ont la même loi.

Démonstration. Soit X et Y des v.a. telles que $F_X = F_Y$. Alors, pour tout $a \in \mathbb{R}$,

$$P_X(] - \infty, a]) = \mathbb{P}(X \leq a) = F_X(a) = F_Y(a) = \mathbb{P}(Y \leq a) = P_Y(] - \infty, a]).$$

Donc les mesures de probabilité P_X et P_Y coïncident sur $\mathcal{C} = \{] - \infty, a] : a \in \mathbb{R}\}$. Comme \mathcal{C} est stable par intersections finies, par le Théorème 1.1.9, P_X et P_Y coïncident sur la tribu engendrée par \mathcal{C} , qui est $\mathcal{B}(\mathbb{R})$. \square

Le théorème précédent est un résultat d'unicité : une fonction de répartition est associée à au plus une loi. Notons qu'une fonction de répartition ne caractérise pas la v.a. : des v.a. qui ont la même loi ont la même fonction de répartition. Le théorème suivant est un résultat d'existence : pour toute fonction ressemblant à une fonction de répartition (i.e. satisfaisant les propriétés de la Proposition 1.4.4), il existe une loi dont c'est la fonction de répartition, et donc en particulier il existe une v.a. dont c'est la fonction de répartition.

Théorème 1.4.6. Soit $F: \mathbb{R} \rightarrow [0, 1]$. Supposons que F soit croissante, continue à droite et satisfasse

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{et} \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

Alors il existe une loi P sur \mathbb{R} telle que F soit la fonction de répartition de P .

Démonstration. Admise. La construction de P suit une méthode proche de celle utilisée pour construire la mesure de Lebesgue en cours de théorie de la mesure : on introduit une mesure extérieure appropriée, on montre que tous les boréliens sont mesurables pour cette mesure extérieure et on vérifie que sa restriction aux boréliens a pour fonction de répartition F . Voir les notes de Le Gall, Section 3.5 pour les détails. \square

Au delà de ce résultat théorique, il y a de nombreux cas pratiques où l'on peut déterminer la loi à partir de la fonction de répartition F . Deux cas particuliers sont importants, déjà évoqués aux Exemples 1.4.2 et 1.4.3 : si F est constante par morceaux, alors la loi est discrète avec des atomes aux points de discontinuité dont la masse est donnée par la taille du saut de F en ce point ; si F est de classe \mathcal{C}^1 sur \mathbb{R} , alors la loi est continue de densité F' . Le résultat suivant couvre un mélange de ces deux cas. Cela exclut certaines lois, comme des lois discrètes avec accumulation d'atomes en certains points, des lois continues avec une densité trop irrégulière ou des lois diffuses qui ne sont pas continues (cf. Remarque 1.2.11).

Proposition 1.4.7. Soit P une loi sur \mathbb{R} et F sa fonction de répartition. Supposons que F est de classe \mathcal{C}^1 sur $\mathbb{R} \setminus \{a_i, i \in \mathcal{I}\}$ avec $(a_i)_{i \in \mathcal{I}}$ une suite strictement croissante dans l'un des cas suivants

- $\mathcal{I} = \llbracket 1, n \rrbracket$ avec $n \in \mathbb{N}$;
- $\mathcal{I} = \mathbb{N}$ et $a_n \rightarrow \infty$ quand $n \rightarrow \infty$;
- $\mathcal{I} = \mathbb{Z}_-$ et $a_n \rightarrow -\infty$ quand $n \rightarrow -\infty$;
- $\mathcal{I} = \mathbb{Z}$ et $a_n \rightarrow \pm\infty$ quand $n \rightarrow \pm\infty$.

On définit f par $f = F'$ sur $\mathbb{R} \setminus \{a_i, i \in \mathcal{I}\}$ et $f = 0$ en dehors. Alors

$$dP(x) = f(x) dx + \sum_{i \in \mathcal{I}} (F(a_i) - F(a_i-)) d\delta_{a_i}(x),$$

où $F(a-)$ désigne la limite à gauche de F en a .

Démonstration. Notons Q la mesure définie par la formule ci-dessus, de sorte que l'on cherche à montrer $P = Q$. Considérons des réels $a < b$, on va montrer que $F(b) - F(a) = Q(]a, b])$.

Cas 1. Supposons $]a, b] \cap \{a_i, i \in \mathcal{I}\} = \emptyset$. Alors F est continue sur $]a, b]$ (on utilise ici que F est continue à droite en a) et dérivable sur $]a, b[$, donc, par le théorème fondamental de l'analyse,

$$F(b) - F(a) = \int_a^b f(x) dx = Q(]a, b]),$$

car aucun des atomes de Q n'est dans $]a, b]$.

Cas 2. Supposons $]a, b] \cap \{a_i, i \in \mathcal{I}\} = \{b\}$. Alors, cette fois, on a $F(b-) - F(a) = \int_a^b f(x) dx$ et donc

$$F(b) - F(a) = F(b-) - F(b) + \int_a^b f(x) dx = Q(\{b\}) + Q(]a, b]) = Q(]a, b]),$$

car l'unique atome de Q sur $]a, b]$ est en b et sa masse est $F(b-) - F(b)$.

Cas 3. Supposons que $]a, b] \cap \{a_i, i \in \mathcal{I}\} \neq \emptyset$. Notons que, par nos hypothèses sur la suite $(a_i)_{i \in \mathcal{I}}$, $]a, b]$ ne peut contenir qu'un nombre fini de ses éléments, que l'on note $a_k < \dots < a_\ell$ avec $k \leq \ell$. On écrit alors, grâce à une somme télescopique

$$\begin{aligned} F(b) - F(a) &= F(b) - F(a_\ell) + \sum_{i=k}^{\ell-1} (F(a_{i+1}) - F(a_i)) + F(a_k) - F(a) \\ &= Q(]a_\ell, b]) + \sum_{i=k}^{\ell-1} Q(]a_i, a_{i+1}]) + Q(]a, a_k]), \end{aligned}$$

en utilisant le Cas 1 sur l'intervalle $]a_\ell, b]$ et le Cas 2 sur les autres intervalles. En prenant l'union de ces intervalles disjoints, on en déduit que $F(b) - F(a) = Q(]a, b])$.

Conclusion. Fixons b et prenons $a \rightarrow -\infty$. Alors $F(a) \rightarrow 0$ par la Proposition 1.4.4 et $Q(]a, b]) \rightarrow Q(]-\infty, b])$ par continuité croissante de la mesure. On a donc montré que $F(b) = Q(]-\infty, b])$, donc F est la fonction de répartition de Q . Comme la fonction de répartition caractérise la loi, on a donc $P = Q$. \square

Exemple 1.4.8. Pour $x \in \mathbb{R}$, on définit

$$F(x) = \mathbb{1}_{x \geq 0} \left(1 - \frac{1}{x+2} \right).$$

Cette fonction est croissante, continue à droite, $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow \infty} F(x) = 1$. c'est donc la fonction de répartition d'une loi P sur \mathbb{R} par le Théorème 1.4.6. Comme F est \mathcal{C}^1 sur $\mathbb{R} \setminus \{0\}$, par la Proposition 1.4.4, on a

$$dP(x) = \frac{\mathbb{1}_{x > 0}}{(x+2)^2} dx + \frac{1}{2} \delta_0.$$

Remarque 1.4.9. Dans la Proposition 1.4.4, les points a_i ne sont pas forcément des points de discontinuité de F , seulement des points où F n'est pas \mathcal{C}^1 . Si F est continue en a_i , alors $F(a_i) - F(a_i-) = 0$ donc il n'y a pas d'atome en a_i . En particulier, si F vérifie les hypothèses de la Proposition 1.4.4 et est continue sur \mathbb{R} , alors la loi P associée est continue de densité $f = F'$ qui est définie Lebesgue-presque partout (voir par exemple la loi exponentielle en Section 1.5).

Méthode (Déterminer une loi). Une autre méthode pour déterminer la loi d'une v.a. réelle X consiste à calculer sa fonction de répartition et utiliser la Proposition 1.4.7. Cette méthode est particulièrement appropriée si X est définie comme un minimum ou un maximum d'autres v.a.

2. Pour s'en convaincre on peut considérer la fonction \tilde{F} définie par $\tilde{F} = F$ sur $]a, b[$ et $\tilde{F}(b) = F(b-)$. Alors \tilde{F} est continue sur $]a, b]$ et dérivable sur $]a, b[$, donc on peut lui appliquer le théorème fondamental de l'analyse et cela donne la formule voulue.

Fonction quantile (bonus). La fin de cette section concernant la fonction quantile n'a pas été traitée en cours.

Définition 1.4.10. Soit P une loi sur \mathbb{R} et F sa fonction de répartition. La fonction quantile de P est la fonction $q: [0, 1] \rightarrow \overline{\mathbb{R}}$ définie par

$$\forall \alpha \in [0, 1], \quad q(\alpha) = \inf\{x \in \overline{\mathbb{R}} : \alpha \leq F(x)\},$$

où l'on prolonge par continuité $F(-\infty) = 0$ et $F(\infty) = 1$. On appelle $q(\alpha)$ le α -quantile de P et $q(1/2)$ la médiane.

Notons que si, pour un intervalle $I \subset \overline{\mathbb{R}}$, la fonction $F: I \rightarrow [0, 1]$ est bijective, alors on peut écrire plus simplement $q = F^{-1}$, où F^{-1} est l'inverse de la fonction F restreinte à I . De manière générale, on a toujours $F(q(\alpha)) \geq \alpha$ (par continuité à droite de F) mais pas forcément d'égalité.

La fonction q permet de générer une v.a. avec la loi P à partir d'une v.a. de loi uniforme sur $[0, 1]$ (c'est-à-dire dont la loi est la mesure de Lebesgue sur $[0, 1]$), comme montré dans le résultat suivant.

Proposition 1.4.11. Soit P une loi sur \mathbb{R} et Q sa fonction quantile. Soit U une v.a. uniforme sur $[0, 1]$. Alors $q(U)$ suit la loi P .

Démonstration. On va montrer le résultat en calculant la fonction de répartition de la v.a. $X = q(U)$. Pour $x \in \mathbb{R}$, on a $F_X(x) = \mathbb{P}(q(U) \leq x)$. De plus, pour $\alpha \in [0, 1]$, on a $\alpha \leq F(x)$ ssi $q(\alpha) \leq x$:

- Si $\alpha \leq F(x)$, alors $q(\alpha) \leq x$ par définition de q .
- Si $q(\alpha) \leq x$, alors $F(q(\alpha)) \leq F(x)$ car F est croissante. Mais, par continuité à droite de F , on a $F(q(\alpha)) \geq \alpha$. On a donc $\alpha \leq F(x)$

On en déduit que

$$F_X(x) = \mathbb{P}(q(U) \leq x) = \mathbb{P}(U \leq F(x)) = P_U([0, F(x)]) = F(x),$$

car P_U est la mesure de Lebesgue sur $[0, 1]$. Comme la fonction de répartition caractérise la loi, on en déduit que $P_X = P$. \square

1.4.2 Fonction caractéristique

Définition 1.4.12. Soit P une loi sur \mathbb{R} . La fonction caractéristique de P est la fonction $\phi: \mathbb{R} \rightarrow \mathbb{C}$ définie par

$$\forall \theta \in \mathbb{R}, \quad \phi(\theta) = \int_{\mathbb{R}} e^{i\theta x} dP(x).$$

Soit X une v.a. réelle. La fonction caractéristique de X est la fonction caractéristique de sa loi. On la notera souvent ϕ_X et elle vérifie

$$\forall \theta \in \mathbb{R}, \quad \phi_X(\theta) = \mathbb{E}\left[e^{i\theta X}\right].$$

Comme $|e^{i\theta x}| \leq 1$ (et que 1 est intégrable pour une mesure de probabilité!), la fonction caractéristique est bien définie pour tout $\theta \in \mathbb{R}$ et $|\phi(\theta)| \leq 1$. La reformulation en terme d'espérance est une conséquence du théorème de transfert.

En analyse, la transformée de Fourier d'une mesure finie μ sur \mathbb{R} est la fonction

$$\theta \in \mathbb{R} \longmapsto \int_{\mathbb{R}} e^{-i\theta x} d\mu(x).$$

Ainsi, au signe de θ près, elle correspond à la fonction caractéristique de μ quand μ est une mesure de probabilité.

Le résultat suivant montre qu'une fonction caractéristique est toujours continue et que, plus la v.a. a de moments finis, plus la fonction caractéristique est régulière.

Proposition 1.4.13 (Régularité). Soit $n \in \mathbb{N}$. Soit X une v.a. réelle de moment d'ordre n fini. Alors ϕ_X est de classe C^n sur \mathbb{R} et

$$\forall k \in \llbracket 0, n \rrbracket, \quad \forall \theta \in \mathbb{R}, \quad \phi_X^{(k)}(\theta) = i^k \mathbb{E}[X^k e^{i\theta X}].$$

Démonstration. Voir TD3. □

En particulier, il est intéressant de noter que les moments de X s'expriment alors en termes des dérivées en 0 de sa fonction caractéristique :

$$\forall k \in \llbracket 0, n \rrbracket, \quad \mathbb{E}[X^k] = (-i)^k \phi_X^{(k)}(0).$$

La réciproque suivante est vraie : si, pour $m \in \mathbb{N}$, ϕ_X est $2m$ fois dérivable en 0, alors le moment d'ordre $2m$ de X est fini.

Finalement, le résultat le plus important de cette section est le suivant, qui justifie le nom de "fonction caractéristique".

Théorème 1.4.14. Une loi sur \mathbb{R} est caractérisée par sa fonction caractéristique.

Démonstration (non vue en cours, donc non requise). Soit P une loi sur \mathbb{R} et ϕ sa fonction caractéristique. On pose, pour $x \in \mathbb{R}$ et $\sigma > 0$,

$$g_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right),$$

de sorte que g_σ est la densité de la loi gaussienne $\mathcal{N}(0, \sigma^2)$, voir Section 1.5. On définit également, pour tout $x \in \mathbb{R}$,

$$(g_\sigma * P)(x) = \int_{\mathbb{R}} g_\sigma(x - y) dP(y),$$

qui est bien définie par g_σ est bornée, donc P -intégrable. Les étapes sont les suivantes :

1. Écrire g_σ comme une fonction caractéristique d'une autre gaussienne.
2. Utiliser l'étape 1, pour écrire $g_\sigma * P$ uniquement en terme de ϕ , ce qui montre en particulier que $g_\sigma * P$ est caractérisée par ϕ .
3. Montrer que, pour toute $f: \mathbb{R} \rightarrow \mathbb{R}$ continue bornée, $\int_{\mathbb{R}} f(x)(g_\sigma * P)(x) dx \rightarrow \int_{\mathbb{R}} f(x) dP(x)$ quand $\sigma \rightarrow 0$.

On conclut alors ainsi. Soit Q une loi sur \mathbb{R} de fonction caractéristique ϕ . Alors l'étape 2 montre que $g_\sigma * P = g_\sigma * Q$. Donc, par l'étape 3, on en déduit que $\int_{\mathbb{R}} f(x) dP(x) = \int_{\mathbb{R}} f(x) dQ(x)$ pour tout f continue bornée. Cela implique que $P = Q$ d'après le Théorème 1.1.10.

Étape 1. Par le calcul de la fonction caractéristique de la gaussienne (énoncé en Section 1.5 et fait au TD 3), on a, pour tout $\sigma > 0$ et $\theta \in \mathbb{R}$,

$$\int_{\mathbb{R}} e^{i\theta x} g_\sigma(x) dx = \exp\left(-\frac{\sigma^2 \theta^2}{2}\right) = \frac{\sqrt{2\pi}}{\sigma} g_{1/\sigma}(\theta).$$

En remplaçant σ par $1/\sigma$, on obtient

$$g_\sigma(\theta) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\theta x} g_{1/\sigma}(x) dx,$$

ce qui écrit g_σ (à un préfacteur près) comme une fonction caractéristique.

Étape 2. Pour tout $x \in \mathbb{R}$, on a

$$\begin{aligned}
(g_\sigma * P)(x) &= \int_{\mathbb{R}} g_\sigma(x-y) dP(y) && \text{(définition)} \\
&= \int_{\mathbb{R}} \frac{1}{\sigma\sqrt{2\pi}} \left(\int_{\mathbb{R}} e^{i(x-y)t} g_{1/\sigma}(t) dt \right) dP(y) && \text{(par l'étape 1)} \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{ixt} g_{1/\sigma}(t) \left(\int_{\mathbb{R}} e^{-iyt} dP(y) \right) dt && \text{(par Fubini)} \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{ixt} g_{1/\sigma}(t) \phi(-t) dt, && \text{(par définition de } \phi)
\end{aligned}$$

où le théorème de Fubini est justifié en bornant $|e^{i(x-y)t} g_{1/\sigma}(t)| \leq g_{1/\sigma}(t)$ qui est intégrable par rapport à dt et donc par rapport à $dt \otimes P(dy)$ (car P est une mesure de probabilité). On a donc exprimé $g_\sigma * P$ en terme de ϕ uniquement.

Étape 3. Soit $f: \mathbb{R} \rightarrow \mathbb{R}$ continue bornée. Alors, on a

$$\int_{\mathbb{R}} f(x)(g_\sigma * P)(x) dx = \int_{\mathbb{R}} f(x) \left(\int_{\mathbb{R}} g_\sigma(x-y) dP(y) \right) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x) g_\sigma(x-y) dx \right) dP(y),$$

en utilisant le théorème de Fubini justifié en bornant $|f(x)g_\sigma(x-y)| \leq \|f\|_\infty g_\sigma(x-y)$ qui est intégrable par rapport à dx , d'intégrale indépendante de y (par changement de variable $x-y \rightarrow x$) et qui est donc elle-même intégrable par rapport à $dP(y)$. En utilisant que g_σ est paire puis la définition de la convolution de deux fonctions, on obtient

$$\int_{\mathbb{R}} f(x)(g_\sigma * P)(x) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x) g_\sigma(y-x) dx \right) dP(y) = \int_{\mathbb{R}} (f * g_\sigma)(y) dP(y).$$

Mais la famille $(g_\sigma)_{\sigma>0}$, quand $\sigma \rightarrow 0$, est une approximation de Dirac au sens où elle vérifie

$$\forall \sigma > 0, \quad \int_{\mathbb{R}} g_\sigma(x) dx = 1 \quad \text{et} \quad \forall \varepsilon > 0, \quad \int_{\mathbb{R} \setminus [-\varepsilon, \varepsilon]} g_\sigma(x) dx \xrightarrow{\sigma \rightarrow 0} 0.$$

Une conséquence vue en cours de théorie de la mesure est que, pour f continue bornée, $f * g_\sigma$ converge simplement vers f quand $\sigma \rightarrow 0$. Ainsi, par le théorème de convergence dominée (en bornant $|(f * g_\sigma)(y)| \leq \|f\|_\infty$ qui est P -intégrable), on obtient

$$\int_{\mathbb{R}} f(x)(g_\sigma * P)(x) dx = \int_{\mathbb{R}} (f * g_\sigma)(y) dP(y) \xrightarrow{\sigma \rightarrow 0} \int_{\mathbb{R}} f(y) dP(y),$$

qui est ce que l'on voulait démontrer à cette étape. \square

En général, il est difficile de retrouver une loi P à partir de sa fonction caractéristique. Mais dans le cas où la fonction caractéristique est intégrable sur \mathbb{R} on peut utiliser cette version du théorème d'inversion de Fourier.

Théorème 1.4.15. *Soit X une v.a. réelle. Si sa fonction caractéristique ϕ_X est intégrable par rapport à la mesure de Lebesgue sur \mathbb{R} , alors X est continue de densité donnée par*

$$\forall x \in \mathbb{R}, \quad p_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\theta x} \phi_X(\theta) d\theta.$$

Démonstration (non vue en cours, donc non requise). On reprend les notations de la démonstration du Théorème 1.4.14, avec $P = P_X$. Posons $p(x) := \frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\theta x} \phi_X(\theta) d\theta$ (on ne sait pas encore que c'est p_X). On veut montrer que X a pour loi $p(x) dx$. Pour cela, il suffit de montrer que, pour tout $f: \mathbb{R} \rightarrow \mathbb{R}$ continue à support compact

$$\int_{\mathbb{R}} f(x) dP_X(x) = \int_{\mathbb{R}} f(x) p(x) dx, \tag{1.2}$$

car cela caractérise P_X par le Théorème 1.1.10.

Dans l'étape 3 de la démonstration du Théorème 1.4.14, on a montré que

$$\int_{\mathbb{R}} f(x)(g_{\sigma} * P)(x) dx \xrightarrow{\sigma \rightarrow 0} \int_{\mathbb{R}} f(x) dP_X(x). \quad (1.3)$$

En calculant la limite autrement on va obtenir (1.2). Dans l'étape 2, on a aussi montré, pour tout $\sigma > 0$ et tout $x \in \mathbb{R}$,

$$(g_{\sigma} * P_X)(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{ixt} g_{1/\sigma}(t) \phi_X(-t) dt = \frac{1}{2\pi} \int_{\mathbb{R}} e^{ixt} e^{-\sigma^2 t^2/2} \phi_X(-t) dt,$$

où l'on a utilisé la définition de $g_{1/\sigma}$ dans la 2ème égalité. On a $e^{-\sigma^2 t^2/2} \rightarrow 0$ quand $\sigma \rightarrow 0$, ainsi que la domination $|e^{ixt} e^{-\sigma^2 t^2/2} \phi_X(-t)| \leq |\phi_X(-t)|$ qui est intégrable sur \mathbb{R} par hypothèse du théorème. Donc, par le théorème de convergence dominée,

$$(g_{\sigma} * P_X)(x) \xrightarrow{\sigma \rightarrow 0} \frac{1}{2\pi} \int_{\mathbb{R}} e^{ixt} \phi_X(-t) dt = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ixt} \phi_X(t) dt = p(x). \quad (1.4)$$

Donc, $f(x)(g_{\sigma} * P_X)(x) \rightarrow f(x)p(x)$ quand $\sigma \rightarrow 0$ et on a la domination suivante, pour tout $\sigma > 0$ et tout $x \in \mathbb{R}$,

$$|f(x)(g_{\sigma} * P_X)(x)| \leq \|f\|_{\infty} \mathbb{1}_{[-M, M]}(x) \cdot \frac{1}{2\pi} \int_{\mathbb{R}} |\phi_X(t)| dt,$$

où $M > 0$ est tel que f est nulle en dehors de $[-M, M]$. La fonction qui domine est bien intégrable en x sur \mathbb{R} , donc, par convergence dominée, on obtient

$$\int_{\mathbb{R}} f(x)(g_{\sigma} * P)(x) dx \xrightarrow{\sigma \rightarrow 0} \int_{\mathbb{R}} f(x)p(x) dx.$$

En combinant cela à (1.3), on obtient (1.2) et cela conclut la démonstration. \square

1.4.3 Transformée de Laplace

Définition 1.4.16. Soit P une loi sur \mathbb{R} . La transformée de Laplace de P est la fonction $L: \mathbb{R} \rightarrow [0, \infty]$ définie par

$$\forall t \in \mathbb{R}, \quad L(t) = \int_{\mathbb{R}} e^{-tx} dP(x).$$

Soit X une v.a. réelle. La transformée de Laplace de X est la transformée de Laplace de sa loi. On la notera souvent L_X et elle vérifie

$$\forall t \in \mathbb{R}, \quad L_X(t) = \mathbb{E}[e^{-tX}].$$

La transformée de Laplace est bien définie sur tout \mathbb{R} car $e^{-tx} \geq 0$, mais elle peut être infinie. Pour certaines lois, elle est même infinie partout sauf en 0 (voir la loi de Cauchy en Section 1.5). En particulier, elle ne caractérise pas la loi en général. On peut montrer que si elle est finie autre part qu'en 0, alors elle caractérise la loi. Cependant nous n'énoncerons pas et ne montrerons pas ce résultat ici. Plus globalement, la transformée de Laplace sera peu utilisée dans ce cours, même si c'est une notion importante.

1.5 Lois classiques

Dans cette section sont regroupées certaines lois classiques, ainsi que certaines de leurs propriétés. Les notations pour ces lois ne sont pas universellement utilisées (sauf pour la gaussienne). Pour certaines de ces lois, d'autres choix de paramètres ou même de définition sont parfois faits. Il est important de connaître la définition de ces lois, mais pas toutes leurs caractéristiques, à l'exception de la gaussienne qui joue un rôle particulièrement important en probabilités.

Notation. Si P est une loi, on note $X \sim P$ pour dire que X est une v.a. de loi P .

1.5.1 Lois discrètes

Loi de Bernoulli.

La *loi de Bernoulli* de paramètre $p \in [0, 1]$ est la loi $\mathcal{B}(p) := (1-p)\delta_0 + p\delta_1$, modélisant une expérience à deux issues (échec et succès). Toute indicatrice d'événement suit une loi de Bernoulli puisqu'elle ne prend que les valeurs 0 et 1.

Si $X \sim \mathcal{B}(p)$, alors $\mathbb{E}[X] = p$ et $\text{Var}(X) = p(1-p)$. De plus, $\phi_X(\theta) = (1-p) + pe^{i\theta}$ pour tout $\theta \in \mathbb{R}$ et $L_X(t) = (1-p) + pe^{-t}$ pour tout $t \in \mathbb{R}$.

Loi binomiale.

La *loi binomiale* de paramètres $n \in \mathbb{N}$ et $p \in [0, 1]$, notée $\mathcal{B}(n, p)$, est la loi supportée par $\llbracket 0, n \rrbracket$ telle que, si $X \sim \mathcal{B}(n, p)$,

$$\forall k \in \llbracket 0, n \rrbracket, \quad \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Elle modélise le nombre de succès obtenus en répétant n fois une expérience qui résulte en un succès avec probabilité p . Notons que $\mathcal{B}(1, p) = \mathcal{B}(p)$.

On a $\mathbb{E}[X] = np$ et $\text{Var}(X) = np(1-p)$. En outre, $\phi_X(\theta) = ((1-p) + pe^{i\theta})^n$ pour tout $\theta \in \mathbb{R}$ et $L_X(t) = ((1-p) + pe^{-t})^n$ pour tout $t \in \mathbb{R}$.

Loi géométrique.

La *loi géométrique* de paramètre $p \in]0, 1]$, notée $\mathcal{G}(p)$, est la loi supportée par \mathbb{N}^* telle que, si $X \sim \mathcal{G}(p)$,

$$\forall k \in \mathbb{N}^*, \quad \mathbb{P}(X = k) = p(1-p)^{k-1}.$$

Elle modélise le nombre de fois qu'il faut répéter une expérience résultant en un succès avec probabilité p afin d'obtenir un premier succès.

On a $\mathbb{E}[X] = 1/p$ et $\text{Var}(X) = (1-p)/p^2$. Sa fonction caractéristique est donnée par

$$\forall \theta \in \mathbb{R}, \quad \phi_X(\theta) = \frac{pe^{i\theta}}{1 - (1-p)e^{i\theta}}$$

et sa transformée de Laplace

$$\forall t \in \mathbb{R}, \quad L_X(t) = \begin{cases} \frac{pe^{-t}}{1 - (1-p)e^{-t}} & \text{si } t > \log(1-p), \\ \infty & \text{sinon.} \end{cases}$$

Loi de Poisson.

La *loi de Poisson* de paramètre (ou de taux) $\lambda > 0$, notée $\mathcal{P}(\lambda)$, est la loi supportée par \mathbb{N} telle que, si $X \sim \mathcal{P}(\lambda)$,

$$\forall k \in \mathbb{N}, \quad \mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Elle est généralement introduite comme la limite (en un sens que l'on formalisera plus tard dans ce cours) de la loi binomiale $\mathcal{B}(n, p)$ dans le régime où $n \rightarrow \infty$ et $p \rightarrow 0$ simultanément de sorte que $pn \rightarrow \lambda$.

On a $\mathbb{E}[X] = \lambda$ et $\text{Var}(X) = \lambda$. De plus, $\phi_X(\theta) = e^{\lambda(e^{i\theta}-1)}$ pour tout $\theta \in \mathbb{R}$ et $L_X(t) = e^{\lambda(e^{-t}-1)}$ pour tout $t \in \mathbb{R}$.

1.5.2 Loïs continues

Loi uniforme.

Soit $a < b$ des réels. La *loi uniforme* sur $[a, b]$, notée $\mathcal{U}([a, b])$, est la mesure de densité $\frac{1}{b-a} \mathbb{1}_{[a, b]}$ par rapport à la mesure de Lebesgue. Elle représente un point tiré uniformément au hasard sur $[a, b]$.

Si $X \sim \mathcal{U}([a, b])$, alors

$$\mathbb{E}[X] = \frac{a+b}{2} \quad \text{et} \quad \text{Var}(X) = \frac{(b-a)^2}{12}.$$

Notons que $\mathbb{E}[X]$ est le milieu du segment $[a, b]$. Sa fonction de répartition est donnée par

$$\forall x \in \mathbb{R}, \quad F_X(x) = \begin{cases} 0 & \text{si } x \leq a, \\ \frac{x-a}{b-a} & \text{si } a < x < b, \\ 1 & \text{si } x \geq b. \end{cases}$$

Sa fonction caractéristique et sa transformée de Laplace sont

$$\forall \theta \in \mathbb{R}, \quad \phi_X(\theta) = \frac{e^{i\theta b} - e^{i\theta a}}{i\theta(b-a)} \quad \text{et} \quad \forall t \in \mathbb{R}, \quad L_X(t) = \frac{e^{-ta} - e^{-tb}}{t(b-a)}$$

Loi gaussienne.

Soit $m \in \mathbb{R}$ et $\sigma > 0$. La *loi gaussienne* (ou *loi normale*) de moyenne m et variance σ^2 , notée $\mathcal{N}(m, \sigma^2)$, est la mesure

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx.$$

Dans le cas particulier $m = 0$ et $\sigma = 1$, la loi $\mathcal{N}(0, 1)$ est appelée *loi gaussienne standard* (ou *loi normale standard*, ou *loi normale centrée réduite*). Une propriété importante est que l'on peut réécrire toutes les lois gaussiennes à partir de la loi gaussienne standard : si $X \sim \mathcal{N}(0, 1)$, alors $m + \sigma X \sim \mathcal{N}(m, \sigma^2)$, voir TD2.

Si $Z \sim \mathcal{N}(m, \sigma^2)$, on a $\mathbb{E}[Z] = m$ et $\text{Var}(Z) = \sigma^2$. De plus, $\phi_Z(\theta) = e^{im\theta - \sigma^2\theta^2/2}$ pour tout $\theta \in \mathbb{R}$, comme on le verra au TD3 en vérifiant que ϕ_Z est solution d'une équation différentielle. Par calcul direct, on peut montrer que $L_Z(t) = e^{-mt + \sigma^2 t^2/2}$ pour tout $t \in \mathbb{R}$. La fonction de répartition de Z n'est pas explicite en terme des fonctions usuelles. On note généralement Φ la fonction de répartition de $\mathcal{N}(0, 1)$. Alors, on a $F_Z(x) = \Phi\left(\frac{x-m}{\sigma}\right)$ pour tout $x \in \mathbb{R}$.

Loi exponentielle.

La *loi exponentielle* de paramètre $\lambda > 0$, notée $\mathcal{E}(\lambda)$, est la mesure $\lambda e^{-\lambda x} \mathbb{1}_{[0, \infty[}(x) dx$. Elle représente un temps d'attente avec la propriété de perte de mémoire suivante : si $X \sim \mathcal{E}(\lambda)$, alors

$$\forall s, t \geq 0, \quad \mathbb{P}(X > t+s | X > t) = \mathbb{P}(X > s).$$

Autrement dit, à l'instant t , si l'on sait que l'attente n'est pas terminée (i.e. $X > t$), alors le temps d'attente restant est toujours de loi $\mathcal{E}(\lambda)$. Les lois exponentielles sont les seules lois sur $[0, \infty[$ à satisfaire cette propriété de perte de mémoire (à l'exception de la loi δ_0 qui est un cas dégénéré), voir TD3.

Si $X \sim \mathcal{E}(\lambda)$, on a $\mathbb{E}[X] = 1/\lambda$ et $\text{Var}(X) = 1/\lambda^2$. Sa fonction de répartition est donnée par

$$\forall x \in \mathbb{R}, \quad F_X(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ 1 - e^{-\lambda x} & \text{si } x \geq 0. \end{cases}$$

Sa fonction caractéristique et sa transformée de Laplace sont

$$\forall \theta \in \mathbb{R}, \quad \phi_X(\theta) = \frac{\lambda}{\lambda - i\theta} \quad \text{et} \quad \forall t \in \mathbb{R}, \quad L_X(t) = \begin{cases} \frac{\lambda}{\lambda + t} & \text{si } t > -\lambda, \\ \infty & \text{si } t \leq -\lambda. \end{cases}$$

Loi de Cauchy.

La *loi de Cauchy* de paramètre de position m et de paramètre d'échelle $a > 0$, notée $\mathcal{C}(m, a)$, est la mesure

$$\frac{a}{\pi(a^2 + (x - m)^2)} dx.$$

Le paramètre m est le centre de symétrie de la densité et le point où elle atteint son maximum. La densité forme une bosse dont la largeur est proportionnelle à a et la hauteur à $1/a$. Dans le cas particulier $m = 0$ et $a = 1$, la loi $\mathcal{C}(0, 1)$ est appelée *loi de Cauchy standard*. Si $Y \sim \mathcal{C}(0, 1)$, alors $aY + m \sim \mathcal{C}(m, a)$.

Si $X \sim \mathcal{C}(m, a)$, alors $\mathbb{E}[|X|] = \infty$, donc ni $\mathbb{E}[X]$ ni $\text{Var}(X)$ ne sont bien définis. Sa fonction de répartition est donnée par $F_X(x) = \frac{1}{\pi} \arctan\left(\frac{x-m}{a}\right) + \frac{1}{2}$ pour tout $x \in \mathbb{R}$. Sa fonction caractéristique et sa transformée de Laplace sont

$$\forall \theta \in \mathbb{R}, \quad \phi_X(\theta) = e^{im\theta - a|\theta|} \quad \text{et} \quad \forall t \in \mathbb{R}, \quad L_X(t) = \begin{cases} 1 & \text{si } t = 0, \\ \infty & \text{si } t \neq 0. \end{cases}$$

Cette fonction caractéristique peut s'obtenir en utilisant la formule d'inversion de Fourier, voir le TD3 pour les détails.

Chapitre 2

Vecteurs aléatoires et indépendance

2.1 Rappels de théorie de la mesure

2.1.1 Tribus produit

Définition 2.1.1. Soit $(E_1, \mathcal{E}_1), \dots, (E_d, \mathcal{E}_d)$ des espaces mesurables. On peut munir $E_1 \times \dots \times E_d$ de la tribu produit, notée $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_d$, qui est la tribu engendrée par l'ensemble des pavés

$$\{B_1 \times \dots \times B_d : B_1 \in \mathcal{E}_1, \dots, B_d \in \mathcal{E}_d\}.$$

On peut montrer que la tribu produit $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_d$ est la plus petite tribu sur $E_1 \times \dots \times E_d$ rendant les projections

$$\begin{aligned} \pi_i : E_1 \times \dots \times E_d &\longrightarrow E_i \\ (x_1, \dots, x_d) &\longmapsto x_i \end{aligned}$$

mesurables pour chaque $i \in \{1, \dots, d\}$.

Nous serons particulièrement intéressé par l'espace produit \mathbb{R}^d , pour lequel le produit des tribus boréliennes de \mathbb{R} coïncide avec la tribu borélienne de \mathbb{R}^d , comme énoncé dans le résultat suivant. Par défaut, \mathbb{R}^d sera toujours supposé être muni de sa tribu borélienne $\mathcal{B}(\mathbb{R}^d)$.

Proposition 2.1.2. On a $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R})^{\otimes d}$. De plus, cette tribu est engendrée par

$$\{] - \infty, a_1] \times \dots \times] - \infty, a_d] : a_1, \dots, a_d \in \mathbb{R}\}.$$

Il en découle le résultat d'unicité suivant pour les mesures de probabilités sur \mathbb{R}^d .

Corollaire 2.1.3. Soit P et Q deux mesures de probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Supposons que

$$\forall a_1, \dots, a_d \in \mathbb{R}, \quad P(] - \infty, a_1] \times \dots \times] - \infty, a_d]) = Q(] - \infty, a_1] \times \dots \times] - \infty, a_d]).$$

Alors $P = Q$.

Démonstration. C'est une conséquence directe du Théorème 1.1.9 et de la Proposition 2.1.2, car l'ensemble des $] - \infty, a_1] \times \dots \times] - \infty, a_d]$ est stable par intersection finie. \square

2.1.2 Mesures produit

Les mesures produit sont un cas particulier de mesures construites sur un espace produit. Leur construction a été vue en théorie de la mesure.

Théorème 2.1.4. Soit $(E_1, \mathcal{E}_1, \mu_1), \dots, (E_d, \mathcal{E}_d, \mu_d)$ des espaces mesurés, avec μ_1, \dots, μ_d des mesures σ -finies. Il existe une unique mesure $\mu_1 \otimes \dots \otimes \mu_d$ sur $(E_1 \times \dots \times E_d, \mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_d)$, appelée mesure produit, telle que

$$\forall B_1 \in \mathcal{E}_1, \dots, B_d \in \mathcal{E}_d, \quad (\mu_1 \otimes \dots \otimes \mu_d)(B_1 \times \dots \times B_d) = \mu_1(B_1) \cdots \mu_d(B_d).$$

Sur \mathbb{R}^d , on notera λ_d la mesure de Lebesgue. Elle coïncide avec le produit de d mesures de Lebesgue sur \mathbb{R} , i.e. $\lambda_d = \lambda_1^{\otimes d}$.

Pour travailler avec des intégrales par rapport à des mesures produits, les résultats les plus importants sont les deux théorèmes de Fubini.

Théorème 2.1.5 (Fubini–Tonelli). *Soit (E, \mathcal{E}, μ) et (F, \mathcal{F}, ν) des espaces mesurés, avec μ et ν des mesures σ -finies. Soit $f: E \times F \rightarrow [0, \infty]$ une fonction mesurable. Alors, les fonctions $x \in E \mapsto \int_F f(x, y) d\nu(y)$ et $y \in F \mapsto \int_E f(x, y) d\mu(x)$ sont mesurables et on a*

$$\int_{E \times F} f d(\mu \otimes \nu) = \int_E \left(\int_F f(x, y) d\nu(y) \right) d\mu(x) = \int_F \left(\int_E f(x, y) d\mu(x) \right) d\nu(y).$$

Théorème 2.1.6 (Fubini–Lebesgue). *Soit (E, \mathcal{E}, μ) et (F, \mathcal{F}, ν) des espaces mesurés, avec μ et ν des mesures σ -finies. Soit $f: E \times F \rightarrow \mathbb{R}$ une fonction mesurable. Supposons que f est intégrable par rapport à $\mu \otimes \nu$. Alors, la fonction $x \in E \mapsto \int_F f(x, y) d\nu(y)$ est bien définie μ -p.p., mesurable et intégrable par rapport à μ , la fonction $y \in F \mapsto \int_E f(x, y) d\mu(x)$ est bien définie ν -p.p., mesurable et intégrable par rapport à ν , et on a*

$$\int_{E \times F} f d(\mu \otimes \nu) = \int_E \left(\int_F f(x, y) d\nu(y) \right) d\mu(x) = \int_F \left(\int_E f(x, y) d\mu(x) \right) d\nu(y).$$

2.1.3 Formule de changement de variable

La formule de changement de variable est un outil spécifique à la mesure de Lebesgue sur \mathbb{R}^d . Elle nous sera particulièrement utile pour travailler avec des vecteurs aléatoires de loi continue.

Soit U et V deux ouverts de \mathbb{R}^d . Soit $\varphi: U \rightarrow V$, on note $\varphi = (\varphi_1, \dots, \varphi_d)$. Rappelons qu'on dit que φ est un \mathcal{C}^1 -difféomorphisme si φ est bijective, de classe \mathcal{C}^1 sur U et φ^{-1} est de classe \mathcal{C}^1 sur V . Cela implique en particulier que la *matrice jacobienne* de φ en x , donnée par

$$\left(\frac{\partial \varphi_i}{\partial x_j}(x) \right)_{1 \leq i, j \leq d} = \begin{pmatrix} \frac{\partial \varphi_1}{\partial x_1}(x) & \cdots & \frac{\partial \varphi_1}{\partial x_d}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial \varphi_d}{\partial x_1}(x) & \cdots & \frac{\partial \varphi_d}{\partial x_d}(x) \end{pmatrix}$$

est inversible en tout point de $x \in U$. On note $\text{Jac}_\varphi(x)$ le déterminant de la matrice jacobienne et on l'appelle le *Jacobien* de φ en x .

Théorème 2.1.7 (Changement de variable). *Soit U et V deux ouverts de \mathbb{R}^d . Soit $\varphi: U \rightarrow V$ un \mathcal{C}^1 -difféomorphisme. Alors, pour toute fonction mesurable $f: V \rightarrow \mathbb{R}_+$,*

$$\int_V f(y) dy = \int_U f(\varphi(x)) |\text{Jac}_\varphi(x)| dx.$$

Rappelons que, grâce au théorème d'inversion globale, une manière de montrer que $\varphi: U \rightarrow V$ est un \mathcal{C}^1 -difféomorphisme est de vérifier que

- $\varphi: U \rightarrow V$ est bijective ;
- Les dérivées partielles de φ sont continues sur U ;
- Pour tout $x \in U$, $\text{Jac}_\varphi(x) \neq 0$.

Il est aussi suffisant de vérifier ces propriétés pour φ^{-1} , ce qui est intéressant si ses dérivées partielles sont plus faciles à calculer. On peut alors utiliser $\text{Jac}_\varphi(x) = [\text{Jac}_{\varphi^{-1}}(\varphi(x))]^{-1}$.

2.2 Vecteurs aléatoires, loi jointe et marginale

2.2.1 Loi jointe, lois marginales

Si X est une v.a. à valeurs dans \mathbb{R}^d (que l'on munira toujours de $\mathcal{B}(\mathbb{R}^d)$) et qu'on note (X_1, \dots, X_d) ses coordonnées, alors les X_i sont des v.a. réelles. Réciproquement, si X_1, \dots, X_d sont des v.a. réelles, alors $X = (X_1, \dots, X_d)$ est une v.a. à valeurs dans \mathbb{R}^d . Ainsi, il est équivalent de considérer un vecteur de d variables aléatoires réelles ou de considérer une seule variable aléatoire à valeur dans \mathbb{R}^d .

Définition 2.2.1. Soit $X = (X_1, \dots, X_d)$ une v.a. à valeur dans \mathbb{R}^d .

- Les lois P_{X_i} pour $i \in \llbracket 1, d \rrbracket$ sont appelées les lois marginales de X .
- La loi P_X est appelée loi jointe de (X_1, \dots, X_d) .

Notons que la notion de loi jointe n'apporte rien de nouveau : la loi jointe de (X_1, \dots, X_d) est simplement la loi de (X_1, \dots, X_d) . Le mot "jointe" est là pour souligner le fait que l'on considère les v.a. réelles X_1, \dots, X_d comme un vecteur aléatoire, en opposition aux lois marginales où on considère les X_i individuellement. Comme nous allons le voir, connaître la loi jointe permet de connaître les lois marginales, mais la réciproque est fautive !

Proposition 2.2.2. Soit $X = (X_1, \dots, X_d)$ une v.a. à valeur dans \mathbb{R}^d . La loi de X caractérise ses lois marginales. Plus précisément, pour $i \in \llbracket 1, d \rrbracket$, en notant $\pi_i: \mathbb{R}^d \rightarrow \mathbb{R}$ la projection sur la i -ème coordonnée, P_{X_i} est la mesure image de P_X par la fonction π_i .

Démonstration. Pour tout $B \in \mathcal{B}(\mathbb{R})$, on a

$$P_{X_i}(B) = \mathbb{P}(X_i \in B) = \mathbb{P}(\pi_i(X) \in B) = \mathbb{P}(X \in \pi_i^{-1}(B)) = P_X(\pi_i^{-1}(B)).$$

Cela montre que P_{X_i} est la mesure image de P_X par la fonction π_i et, en particulier, que P_{X_i} est caractérisée par P_X . \square

Exemple 2.2.3. Considérons une v.a. $X = (X_1, X_2)$ à valeurs dans \mathbb{R}^2 de loi

$$P_X = \frac{1}{4}\delta_{(0,0)} + \frac{1}{4}\delta_{(0,1)} + \frac{1}{4}\delta_{(1,0)} + \frac{1}{4}\delta_{(1,1)}.$$

On peut noter que X_1 est à valeurs dans $\{0, 1\}$, donc sa loi est déterminée par la connaissance de

$$\mathbb{P}(X_1 = 0) = \mathbb{P}(X \in \{(0,0), (0,1)\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2},$$

ce qui implique que $\mathbb{P}(X_1 = 1) = \frac{1}{2}$ et donc X_1 est de loi $\mathcal{B}(1/2)$ (i.e. la loi de Bernoulli de paramètre $1/2$). De la même manière, X_2 est aussi de loi $\mathcal{B}(1/2)$.

Considérons à présent $Y = (Y_1, Y_2)$ à valeurs dans \mathbb{R}^2 de loi

$$P_Y = \frac{1}{2}\delta_{(0,0)} + \frac{1}{2}\delta_{(1,1)}.$$

Alors, on peut montrer de manière similaire que Y_1 et Y_2 sont de loi $\mathcal{B}(1/2)$.

Ainsi X et Y ont les mêmes lois marginales, mais pas la même loi. Autrement dit, dire "considérons deux v.a. de loi $\mathcal{B}(1/2)$ " ne permet pas de décrire leur loi jointe. Ici, X_1 et X_2 correspondent au cas où les deux v.a. sont indépendantes (voir la prochaine section) alors que Y_1 et Y_2 correspondent au cas où les deux v.a. sont identiques (on a $Y_1 = Y_2$ p.s., car $Y \in \{(0,0), (1,1)\}$ p.s.).

On conclut cette section en définissant la covariance de deux v.a. réelles. Cette quantité quantifie à quel point X et Y prennent leurs valeurs au-dessus de la moyenne en même temps.

Définition 2.2.4. Soit X et Y des v.a. réelles dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$. La **covariance** de X et Y est

$$\text{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right].$$

Remarque 2.2.5. Par l'inégalité de Cauchy–Schwarz, on a

$$|\text{Cov}(X, Y)| \leq \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]^{1/2} \cdot \mathbb{E}\left[(Y - \mathbb{E}[Y])^2\right]^{1/2} = \sigma_X \cdot \sigma_Y,$$

où on rappelle que σ_X est l'écart-type de X . Quelques cas particuliers important sont :

- Si $\text{Cov}(X, Y) = \sigma_X \cdot \sigma_Y$, alors $Y = cX + a$ p.s. pour $c \geq 0$ et $a \in \mathbb{R}$ ou $X = a$ p.s. pour $a \in \mathbb{R}$. Voir le TD4 pour la démonstration.
- Si $\text{Cov}(X, Y) = -\sigma_X \cdot \sigma_Y$, alors $Y = -cX + a$ p.s. pour $c \geq 0$ et $a \in \mathbb{R}$ ou $X = a$ p.s. pour $a \in \mathbb{R}$.
- Quand X et Y sont indépendantes, on verra que leur covariance est nulle. Cependant, la réciproque est fausse.

Proposition 2.2.6. Soit X et Y des v.a. réelles dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$. On a

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

Démonstration. Cela s'obtient en développant le produit $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])$, puis en prenant l'espérance. \square

2.2.2 Lois continues sur \mathbb{R}^d

Définition 2.2.7. Une loi sur \mathbb{R}^d est dite **continue** si elle est à densité par rapport à la mesure de Lebesgue sur \mathbb{R}^d . Une v.a. X à valeurs dans \mathbb{R}^d est dite **continue** si sa loi est continue et on notera alors généralement $p_X : \mathbb{R}^d \rightarrow \mathbb{R}_+$ sa densité.

On notera alors $dP_X(x) = p_X(x) dx$, où $x = (x_1, \dots, x_d)$ est ici un élément de \mathbb{R}^d . Il est équivalent de dire “ X est une v.a. continue de densité p_X ” et “ X est de loi $p_X(x) dx$ ”.

Le résultat suivant décrit les marginales d'une v.a. continue à valeurs dans \mathbb{R}^d .

Proposition 2.2.8. Soit $X = (X_1, \dots, X_d)$ une v.a. continue à valeur dans \mathbb{R}^d . Alors ses lois marginales sont continues. Plus précisément, pour $i \in \llbracket 1, d \rrbracket$, X_i a pour densité

$$p_{X_i} : x_i \in \mathbb{R} \mapsto \int_{\mathbb{R}^{d-1}} p_X(x_1, \dots, x_d) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_d.$$

Par exemple, si $d = 2$, on a

$$p_{X_1}(x_1) = \int_{\mathbb{R}} p_X(x_1, x_2) dx_2 \quad \text{et} \quad p_{X_2}(x_2) = \int_{\mathbb{R}} p_X(x_1, x_2) dx_1.$$

Démonstration. Soit $f : \mathbb{R} \rightarrow \mathbb{R}_+$ une fonction mesurable. Notons π_i la projection sur la i -ième coordonnée, de sorte que $X_i = \pi_i(X)$. Alors, par le théorème de transfert,

$$\mathbb{E}[f(X_i)] = \mathbb{E}[f(\pi_i(X))] = \int_{\mathbb{R}^d} f(\pi_i(x)) p_X(x) dx = \int_{\mathbb{R}^d} f(x_i) p_X(x_1, \dots, x_d) dx_1 \cdots dx_d.$$

Par Fubini-Tonelli, on peut intégrer dans l'ordre que l'on souhaite. On intègre d'abord par rapport à toutes les variables sauf x_i , ce qui donne

$$\mathbb{E}[f(X_i)] = \int_{\mathbb{R}} f(x_i) \left(\int_{\mathbb{R}^{d-1}} p_X(x_1, \dots, x_d) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_d \right) dx_i.$$

Cela implique que la loi de X_i est continue, de densité donnée par la quantité entre parenthèses dans l'intégrale ci-dessus, vue comme une fonction de x_i . \square

Exemple 2.2.9. On veut tirer au hasard un point du plan, de manière uniforme dans le triangle de sommets $(0, 0)$, $(0, 1)$, $(1, 0)$. Notons

$$T = \{(x, y) \in \mathbb{R}^2 : x \geq 0, y \geq 0, x + y \leq 1\}$$

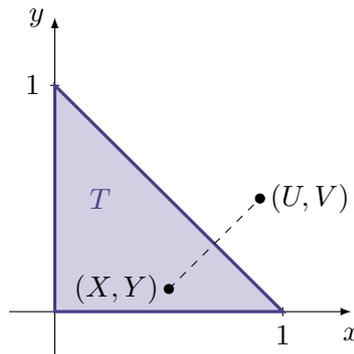
ce triangle. Sa mesure de Lebesgue est $\lambda_2(T) = 1/2$ (c'est son aire!). La loi uniforme sur T est donc $2\mathbb{1}_T(x, y) dx dy$. On considère donc une v.a. (X, Y) continue de densité $2\mathbb{1}_T$, qui représente un point du plan tiré uniformément au hasard dans T . Ses marginales sont donc également continues et on a, pour tout $x \in \mathbb{R}$,

$$p_X(x) = \int_{\mathbb{R}} p_{(X,Y)}(x, y) dy = \int_{\mathbb{R}} 2\mathbb{1}_T(x, y) dy = \begin{cases} 0 & \text{si } x \notin [0, 1], \\ \int_0^{1-x} 2 dy & \text{si } x \in [0, 1]. \end{cases}$$

Donc X a pour densité $p_X : x \mapsto 2(1-x)\mathbb{1}_{[0,1]}(x)$. Par symétrie du problème, Y a la même loi.

Méthode (Déterminer une loi sur \mathbb{R}^d). La méthode présentée en Section 1.3.3 pour déterminer la loi d'une v.a. réelle fonctionne aussi en dimension supérieure. Ainsi, pour déterminer la loi d'une v.a. X à valeurs dans \mathbb{R}^d , on peut essayer d'exprimer $\mathbb{E}[f(X)]$, pour $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$ mesurable quelconque, sous la forme $\int_{\mathbb{R}^d} f(x) d\mu(x)$ avec μ une mesure sur \mathbb{R}^d . Alors on peut en conclure que $P_X = \mu$. Dans le cas où X est définie à partir d'une autre v.a. continue, cela nécessite généralement de faire un changement de variable en dimension d , voir le Théorème 2.1.7 pour un rappel.

Exemple 2.2.10. On reprend l'exemple précédent. On s'intéresse maintenant au point (U, V) qui est le symétrique de (X, Y) par rapport à l'hypoténuse de T , voir le dessin. On veut déterminer la loi de (U, V) .



Pour cela, notons que $(U, V) = (1 - Y, 1 - X)$. Donc, pour $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ mesurable, on a

$$\begin{aligned} \mathbb{E}[f(U, V)] &= \mathbb{E}[f(1 - Y, 1 - X)] \\ &= \int_{\mathbb{R}^2} f(1 - y, 1 - x) p_{(X,Y)}(x, y) dx dy \\ &= \int_{\mathbb{R}^2} f(1 - y, 1 - x) \cdot 2\mathbb{1}_{x \geq 0, y \geq 0, x+y \leq 1} dx dy. \end{aligned}$$

On veut alors faire le changement de variables $(u, v) = (1 - y, 1 - x) \Leftrightarrow (x, y) = (1 - v, 1 - u)$. Pour cela on considère la fonction $\varphi : (u, v) \in \mathbb{R}^2 \mapsto (1 - v, 1 - u) \in \mathbb{R}^2$ qui est bijective (elle est sa propre réciproque), a des dérivées partielles continues sur \mathbb{R}^2 (car constantes) et son Jacobien est

$$\text{Jac}_\varphi(u, v) = \begin{vmatrix} 0 & -1 \\ -1 & 0 \end{vmatrix} = -1,$$

il est donc non nul partout. Donc $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ est un \mathcal{C}^1 -difféomorphisme et on peut appliquer la formule de changement de variables pour obtenir

$$\begin{aligned} \mathbb{E}[f(U, V)] &= \int_{\mathbb{R}^2} f(1 - (1 - u), 1 - (1 - v)) \cdot 2\mathbb{1}_{1-v \geq 0, 1-u \geq 0, 2-u-v \leq 1} \cdot |\text{Jac}_\varphi(u, v)| du dv \\ &= \int_{\mathbb{R}^2} f(u, v) \cdot 2\mathbb{1}_{v \leq 1, u \leq 1, u+v \geq 1} du dv. \end{aligned}$$

Donc (U, V) a pour loi $2\mathbb{1}_{v \leq 1, u \leq 1, u+v \geq 1}$ $du dv$, qui est la loi uniforme sur le symétrique de T par rapport à son hypoténuse.

2.3 Indépendance

2.3.1 Indépendance d'événements

Définition 2.3.1. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Soit $A, B \in \mathcal{A}$ des événements. On dit que A et B sont **indépendants** si $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

Cette définition peut s'interpréter, si $\mathbb{P}(B) > 0$, en notant qu'on a alors $\mathbb{P}(A|B) = \mathbb{P}(A)$, ce qui signifie que le fait de savoir que l'événement B est réalisé ne modifie pas la probabilité de l'événement A .

Remarque 2.3.2. Si A et B sont indépendants, alors A et B^c sont indépendants. En effet

$$\mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A) \cdot \mathbb{P}(B) = \mathbb{P}(A) \cdot (1 - \mathbb{P}(B)) = \mathbb{P}(A) \cdot \mathbb{P}(B^c).$$

Définition 2.3.3. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Soit $A_1, \dots, A_n \in \mathcal{A}$ des événements. On dit que A_1, \dots, A_n sont **indépendants** si, pour tout $p \in \llbracket 1, n \rrbracket$ et tous $1 \leq k_1 < \dots < k_p \leq n$, $\mathbb{P}(A_{k_1} \cap \dots \cap A_{k_p}) = \mathbb{P}(A_{k_1}) \cdot \dots \cdot \mathbb{P}(A_{k_p})$.

Il est important de noter qu'il n'est pas suffisant de montrer que les événements A_1, \dots, A_n sont deux à deux indépendants pour obtenir l'indépendance des n événements ensemble, comme montré dans l'exemple suivant.

Exemple 2.3.4. Considérons l'expérience consistant à lancer deux dés à 6 faces équilibrés, modélisée par $\Omega = \llbracket 1, 6 \rrbracket^2$ muni de $\mathcal{A} = \mathcal{P}(\Omega)$ et de la mesure uniforme \mathbb{P} , i.e. $\mathbb{P}(A) = |A|/36$ pour $A \in \mathcal{A}$. On considère les événements suivants :

$$\begin{aligned} A &= \text{'les deux dés donnent le même résultat'} \\ B &= \text{'le 1er dé donne un résultat pair'} \\ C &= \text{'le 2ème dé donne un 1'} \\ D &= \text{'le 2ème dé donne un 1 ou un 2'}. \end{aligned}$$

On calcule alors $\mathbb{P}(A) = 1/6$, $\mathbb{P}(B) = 1/2$, $\mathbb{P}(C) = 1/6$ et $\mathbb{P}(D) = 1/3$. On a ensuite

$$\mathbb{P}(A \cap B) = \mathbb{P}(\{(2, 2), (4, 4), (6, 6)\}) = \frac{3}{36} = \frac{1}{12} = \mathbb{P}(A) \cdot \mathbb{P}(B),$$

donc A et B sont indépendants. On montre similairement que A et C sont indépendants, ainsi que A et D , B et C , B et D . Mais, on a

$$\mathbb{P}(C \cap D) = \mathbb{P}(C) = \frac{1}{6} \neq \frac{1}{18} = \mathbb{P}(C) \cdot \mathbb{P}(D),$$

donc C et D ne sont pas indépendants. D'autre part, même si A , B et C sont deux à deux indépendants, ils ne sont pas indépendants tous les trois ensemble car

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\emptyset) = 0 \neq \frac{1}{72} = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C).$$

Mais, on peut dire que A , B et D sont indépendants car

$$\mathbb{P}(A \cap B \cap D) = \mathbb{P}(\{(2, 2)\}) = \frac{1}{36} = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(D)$$

et on a déjà vérifié les autres conditions de la définition pour $p = 2$ (et pour $p = 1$, elles sont toujours vraies).

2.3.2 Indépendance de variables aléatoires

Définition 2.3.5. Pour $1 \leq i \leq n$, soit X_i une v.a. à valeurs dans (E_i, \mathcal{E}_i) . On dit que X_1, \dots, X_n sont **indépendantes** si pour tous $B_1 \in \mathcal{E}_1, \dots, B_n \in \mathcal{E}_n$,

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \cdots \mathbb{P}(X_n \in B_n).$$

L'indépendance entre X_1 et X_2 signifie que le fait de savoir quelque chose sur X_1 n'apporte pas d'information sur X_2 (et vice-versa) : on a $\mathbb{P}(X_2 \in B_2 | X_1 \in B_1) = \mathbb{P}(X_2 \in B_2)$ dès que $\mathbb{P}(X_1 \in B_1) > 0$.

Comme pour les événements, ils ne suffit pas que les v.a. X_1, \dots, X_n soient indépendantes deux à deux pour qu'elles soient indépendantes dans leur ensemble.

Remarque 2.3.6. Cette définition peut sembler différente de celle pour les événements, car on n'a pas besoin de considérer des sous-familles. Cependant les deux définitions sont cohérentes car, pour $A_1, \dots, A_n \in \mathcal{A}$, on a l'équivalence suivante : les événements A_1, \dots, A_n sont indépendants si et seulement si les v.a. $\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n}$ sont indépendantes. Voir le TD4 pour la démonstration.

Remarque 2.3.7. On a les deux conséquences suivantes de l'indépendance. Supposons que X_1, \dots, X_n sont indépendantes.

- Pour n'importe quel $\mathcal{I} \subset \llbracket 1, n \rrbracket$, les v.a. X_i pour $i \in \mathcal{I}$ sont indépendantes. En effet, il suffit d'appliquer la définition de l'indépendance pour X_1, \dots, X_n en prenant $B_j = \Omega$ pour $j \notin \mathcal{I}$.
- Pour tous $B_1 \in \mathcal{E}_1, \dots, B_n \in \mathcal{E}_n$, les événements $\{X_1 \in B_1\}, \dots, \{X_n \in B_n\}$ sont indépendants. Pour le vérifier, on considère une sous-famille $1 \leq i_1 < \dots < i_p \leq n$ et on utilise le fait que X_{i_1}, \dots, X_{i_p} sont indépendantes par le point précédent pour obtenir $\mathbb{P}(X_{i_1} \in B_{i_1}, \dots, X_{i_p} \in B_{i_p}) = \mathbb{P}(X_{i_1} \in B_{i_1}) \cdots \mathbb{P}(X_{i_p} \in B_{i_p})$.

Proposition 2.3.8 (Fonctions de v.a. indépendantes). Pour $1 \leq i \leq n$, soit X_i une v.a. à valeurs dans (E_i, \mathcal{E}_i) et $f_i: (E_i, \mathcal{E}_i) \rightarrow (F_i, \mathcal{F}_i)$ mesurable. Supposons que X_1, \dots, X_n sont indépendantes. Alors $f_1(X_1), \dots, f_n(X_n)$ sont indépendantes.

Démonstration. Pour tous $B_1 \in \mathcal{F}_1, \dots, B_n \in \mathcal{F}_n$,

$$\begin{aligned} \mathbb{P}(f_1(X_1) \in B_1, \dots, f_n(X_n) \in B_n) &= \mathbb{P}(X_1 \in f_1^{-1}(B_1), \dots, X_n \in f_n^{-1}(B_n)) \\ &= \mathbb{P}(X_1 \in f_1^{-1}(B_1)) \cdots \mathbb{P}(X_n \in f_n^{-1}(B_n)) \\ &= \mathbb{P}(f_1(X_1) \in B_1) \cdots \mathbb{P}(f_n(X_n) \in B_n), \end{aligned}$$

où l'on a utilisé l'indépendance de X_1, \dots, X_n dans la 2ème égalité, en notant que $f_i^{-1}(B_i) \in \mathcal{E}_i$ car f_i est mesurable. \square

Exemple 2.3.9. Soit X et Y des v.a. réelles indépendantes. Alors X^2 et $2Y+1$ sont indépendantes.

La prochaine proposition présente quelques caractérisations de l'indépendance. La caractérisation (ii) est très importante : X_1, \dots, X_n sont indépendantes si et seulement si leur loi jointe est le *produit* des lois marginales. Cela implique en particulier que, dans le cas de v.a. indépendantes, connaître les lois marginales suffit à connaître la loi jointe.

Proposition 2.3.10 (Caractérisations de l'indépendance). Pour $1 \leq i \leq n$, soit X_i une v.a. à valeurs dans (E_i, \mathcal{E}_i) . Les propositions suivantes sont équivalentes :

- X_1, \dots, X_n sont indépendantes ;
- $P_{(X_1, \dots, X_n)} = P_{X_1} \otimes \cdots \otimes P_{X_n}$;
- pour toutes $f_1: E_1 \rightarrow \mathbb{R}_+, \dots, f_n: E_n \rightarrow \mathbb{R}_+$ mesurables, $\mathbb{E}[\prod_{i=1}^n f_i(X_i)] = \prod_{i=1}^n \mathbb{E}[f_i(X_i)]$;

- (iv) pour toutes $f_1: E_1 \rightarrow \mathbb{R}, \dots, f_n: E_n \rightarrow \mathbb{R}$ mesurables telles que $\mathbb{E}[|f_i(X_i)|] < \infty$ pour chaque $i \in \llbracket 1, n \rrbracket$, on a $\mathbb{E}[\prod_{i=1}^n |f_i(X_i)|] < \infty$ et $\mathbb{E}[\prod_{i=1}^n f_i(X_i)] = \prod_{i=1}^n \mathbb{E}[f_i(X_i)]$;

Démonstration. On va d'abord montrer (i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (i), ce qui montrera que les trois premières propositions sont équivalentes :

- (i) \Rightarrow (ii). Soit $B_1 \in \mathcal{E}_1, \dots, B_n \in \mathcal{E}_n$. On a

$$\begin{aligned} P_{(X_1, \dots, X_n)}(B_1 \times \dots \times B_n) &= \mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) && \text{(définition de la loi)} \\ &= \mathbb{P}(X_1 \in B_1) \cdots \mathbb{P}(X_n \in B_n) && \text{(par (i))} \\ &= P_{X_1}(B_1) \cdots P_{X_n}(B_n) && \text{(définition de la loi)} \\ &= (P_{X_1} \otimes \dots \otimes P_{X_n})(B_1 \times \dots \times B_n) && \text{(déf de la mesure produit)} \end{aligned}$$

ce qui montre (ii) par le Théorème 1.1.9 car l'ensemble des pavés est stable par intersections finies et engendre la tribu produit.

- (ii) \Rightarrow (iii). Pour $f_1, \dots, f_n: \mathbb{R} \rightarrow \mathbb{R}_+$ mesurables, on a

$$\begin{aligned} \mathbb{E} \left[\prod_{i=1}^n f_i(X_i) \right] &= \int_{E_1 \times \dots \times E_n} \left(\prod_{i=1}^n f_i(x_i) \right) dP_{(X_1, \dots, X_n)}(x_1, \dots, x_n) && \text{(théorème de transfert)} \\ &= \int_{E_1 \times \dots \times E_n} \left(\prod_{i=1}^n f_i(x_i) \right) d(P_{X_1} \otimes \dots \otimes P_{X_n})(x_1, \dots, x_n) && \text{(par (ii))} \\ &= \prod_{i=1}^n \int_{E_i} f_i(x_i) dP_{X_i}(x_i) && \text{(par Fubini–Tonelli)} \\ &= \prod_{i=1}^n \mathbb{E}[f_i(X_i)]. && \text{(théorème de transfert)} \end{aligned}$$

- (iii) \Rightarrow (i). La définition de l'indépendance correspond à un cas particulier de (iii) où on prend $f_i = \mathbb{1}_{B_i}$: en effet $f_i(X_i) = \mathbb{1}_{\{X_i \in B_i\}}$ et $\prod_{i=1}^n f_i(X_i) = \mathbb{1}_{\{X_1 \in B_1, \dots, X_n \in B_n\}}$.

Il reste à montrer que (iv) est équivalente aux autres propositions. Montrons (ii) \Rightarrow (iv). Le fait que (ii) \Rightarrow (iii) appliqué aux fonctions $|f_i|$ justifie que $\mathbb{E}[\prod_{i=1}^n |f_i(X_i)|] = \prod_{i=1}^n \mathbb{E}[|f_i(X_i)|] < \infty$. Puis on peut répéter le calcul fait dans la partie (ii) \Rightarrow (iii), en utilisant Fubini–Lebesgue à la place de Fubini–Tonelli. Cela montre que (ii) \Rightarrow (iv). Puis la démonstration de (iv) \Rightarrow (i) est identique à celle de (iii) \Rightarrow (i) car $\mathbb{E}[\mathbb{1}_{B_i}(X_i)] < \infty$. Donc (iv) est équivalent aux propriétés précédentes. \square

Exemple 2.3.11. Soit X une v.a. uniforme sur $[0, 1]$ et Y une v.a. de loi de Cauchy standard. Supposons que X et Y sont indépendantes. Alors leur loi jointe est

$$dP_{(X,Y)}(x, y) = d(P_X \otimes P_Y)(x, y) = dP_X(x) dP_Y(y) = \mathbb{1}_{[0,1]}(x) \frac{1}{\pi(1+y^2)} dx dy.$$

Remarque 2.3.12. Soit $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. Si X et Y sont indépendantes, alors $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ par la partie (iv) de la proposition précédente, car $\mathbb{E}[|X|] < \infty$ et $\mathbb{E}[|Y|] < \infty$, parce que $L^2(\Omega, \mathcal{A}, \mathbb{P}) \subset L^1(\Omega, \mathcal{A}, \mathbb{P})$. On a donc $\text{Cov}(X, Y) = 0$. La réciproque est fautive : une covariance nulle n'implique pas l'indépendance, voir TD5.

Corollaire 2.3.13 (Regroupement par paquets). *Pour $1 \leq i \leq n$, soit X_i une v.a. à valeurs dans (E_i, \mathcal{E}_i) . Supposons que X_1, \dots, X_n sont indépendantes. Soit $k \geq 1$ et $0 = i_0 < i_1 < \dots < i_k = n$. Alors $(X_{i_0+1}, \dots, X_{i_1}), (X_{i_1+1}, \dots, X_{i_2}), \dots, (X_{i_{k-1}+1}, \dots, X_{i_k})$ sont indépendantes.*

Démonstration. Cela découle de la caractérisation de l'indépendance en terme de mesure produit ainsi que de l'associativité du produit de mesures. En effet, on a

$$\begin{aligned}
P_{((X_{i_0+1}, \dots, X_{i_1}), \dots, (X_{i_{k-1}+1}, \dots, X_{i_k}))} &= P_{(X_1, \dots, X_n)} \\
&= P_{X_1} \otimes \dots \otimes P_{X_n} \\
&= (P_{X_{i_0+1}} \otimes \dots \otimes P_{X_{i_1}}) \otimes \dots \otimes (P_{X_{i_{k-1}+1}} \otimes \dots \otimes P_{X_{i_k}}) \\
&= P_{(X_{i_0+1}, \dots, X_{i_1})} \otimes \dots \otimes P_{(X_{i_{k-1}+1}, \dots, X_{i_k})},
\end{aligned}$$

où on a utilisé l'indépendance de X_1, \dots, X_n dans la 2ème égalité (avec le critère (ii) de la Proposition 2.3.10), l'associativité du produit de mesures dans la 3ème et le fait que $X_{i_{\ell-1}+1}, \dots, X_{i_\ell}$ sont indépendantes pour chaque $\ell \in \llbracket 1, k \rrbracket$ (voir Remarque 2.3.7). Par le critère (ii) de la Proposition 2.3.10, la relation ci-dessus montre que $(X_{i_0+1}, \dots, X_{i_1}), \dots, (X_{i_{k-1}+1}, \dots, X_{i_k})$ sont indépendantes. \square

Exemple 2.3.14. Soit X_1, \dots, X_4 des v.a. indépendantes. Alors $X_1^2 X_2$ et $X_3 - X_4$ sont indépendantes. En effet, par le Corollaire 2.3.13, (X_1, X_2) et (X_3, X_4) sont indépendantes, puis par la Proposition 2.3.8, appliquée aux fonctions $f: (x, y) \in \mathbb{R}^2 \rightarrow x^2 y$ et $g: (x, y) \in \mathbb{R}^2 \rightarrow x - y$, les v.a. $f(X_1, X_2)$ et $g(X_3, X_4)$ sont indépendantes.

2.3.3 Indépendance de variables aléatoires continues

On s'intéresse à présent au cas de variables dont la loi jointe est continue. La proposition suivante permet de déterminer, au vu de la densité de la loi jointe, si les v.a. sont indépendantes et de trouver facilement leurs marginales.

Proposition 2.3.15. Soit X_1, \dots, X_n des v.a. réelles telles que (X_1, \dots, X_n) soit continue. Supposons que sa densité s'écrive sous la forme

$$p_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n q_i(x_i),$$

avec $q_1, \dots, q_n: \mathbb{R} \rightarrow \mathbb{R}_+$ mesurables. Alors X_1, \dots, X_n sont indépendantes et, pour chaque i , X_i a pour densité $p_{X_i} = \frac{1}{c_i} q_i$ où $c_i = \int_{\mathbb{R}} q_i(x) dx > 0$.

Démonstration. On remarque d'abord que, par Fubini–Tonelli,

$$\prod_{i=1}^n \left(\int_{\mathbb{R}} q_i(x_i) dx_i \right) = \int_{\mathbb{R}^n} \left(\prod_{i=1}^n q_i(x_i) \right) dx_1 \cdots dx_n = \int_{\mathbb{R}^n} p_{(X_1, \dots, X_n)}(x_1, \dots, x_n) dx_1 \cdots dx_n = 1.$$

En particulier, on a $c_i = \int_{\mathbb{R}} q_i(x) dx > 0$ pour tout $i \in \llbracket 1, n \rrbracket$. Par la Proposition 2.2.8, on sait que, pour chaque $i \in \llbracket 1, n \rrbracket$, X_i est continue de densité donnée, pour tout $x_i \in \mathbb{R}$, par

$$\begin{aligned}
p_{X_i}(x_i) &= \int_{\mathbb{R}^{n-1}} p_{(X_1, \dots, X_n)} dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n \\
&= q_i(x_i) \int_{\mathbb{R}^{n-1}} \left(\prod_{j \neq i} q_j(x_j) \right) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n \\
&= q_i(x_i) \prod_{j \neq i} \left(\int_{\mathbb{R}} q_j(x_j) dx_j \right) && \text{(par Fubini–Tonelli)} \\
&= \frac{1}{c_i} q_i(x_i) \prod_{j=1}^n \left(\int_{\mathbb{R}} q_j(x_j) dx_j \right) \\
&= \frac{1}{c_i} q_i(x_i),
\end{aligned}$$

par l'égalité démontrée précédemment. On a donc montré la formule pour la densité de X_i . Il reste à montrer l'indépendance. Pour cela, on remarque que

$$p_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n q_i(x_i) = \prod_{i=1}^n c_i p_{X_i}(x_i) = \left(\prod_{i=1}^n c_i \right) \left(\prod_{i=1}^n p_{X_i}(x_i) \right) = \prod_{i=1}^n p_{X_i}(x_i),$$

en utilisant de nouveau que $\prod_{i=1}^n c_i = 1$. Cela implique que $P_{(X_1, \dots, X_n)} = P_{X_1} \otimes \dots \otimes P_{X_n}$, et donc que X_1, \dots, X_n sont indépendantes par la Proposition 2.3.10. \square

Exemple 2.3.16. Soit (X, Y) une v.a. à valeurs dans \mathbb{R}^2 de loi

$$e^{-2x} \mathbb{1}_{\mathbb{R}_+ \times [0, 2]}(x, y) dx dy.$$

La densité de (X, Y) s'écrit sous la forme d'un produit d'une fonction de x et d'une fonction de y :

$$p_{(X, Y)}(x, y) = \left(e^{-2x} \mathbb{1}_{\mathbb{R}_+}(x) \right) \cdot \left(\mathbb{1}_{[0, 2]}(y) \right).$$

Donc X et Y sont indépendantes. Pour trouver leurs densités, on calcule $\int_{\mathbb{R}} \mathbb{1}_{[0, 2]}(y) dy = 2$ (donc nécessairement $\int_{\mathbb{R}} e^{-2x} \mathbb{1}_{\mathbb{R}_+}(x) dx = 1/2$) et on en déduit que

$$p_X(x) = 2e^{-2x} \mathbb{1}_{\mathbb{R}_+}(x) \quad \text{et} \quad p_Y(y) = \frac{1}{2} \mathbb{1}_{[0, 2]}(y).$$

On peut donc conclure en disant que X et Y sont indépendantes avec X de loi exponentielle de paramètre 2 et Y de loi uniforme sur $[0, 2]$.

Exemple 2.3.17. Reprenons le cadre de l'Exemple 2.2.9 : soit (X, Y) un point tiré de manière uniforme dans le triangle $T = \{(x, y) \in \mathbb{R}^2 : x \geq 0, y \geq 0, x + y \leq 1\}$. On a vu que la loi de (X, Y) est continue de densité

$$p_{(X, Y)}(x, y) = 2 \cdot \mathbb{1}_T(x, y) = 2 \cdot \mathbb{1}_{x \geq 0} \mathbb{1}_{y \geq 0} \mathbb{1}_{x+y \leq 1}.$$

Il ne semble pas possible de factoriser $\mathbb{1}_{x+y \leq 1}$ sous la forme d'un produit d'une fonction de x et d'une fonction de y , ce qui suggère que X et Y ne sont pas indépendantes. Mais cela ne fournit pas une démonstration. Pour le montrer, on va trouver un contre-exemple à la définition de l'indépendance. Par exemple on peut montrer que

$$\mathbb{P}(X > 1/2, Y > 1/2) \neq \mathbb{P}(X > 1/2) \mathbb{P}(Y > 1/2).$$

En effet, on a $\mathbb{P}(X > 1/2, Y > 1/2) = 0$ car $X + Y \leq 1$ p.s. mais, d'autre part, $\mathbb{P}(X > 1/2) > 0$ (car la région $\{(x, y) \in \mathbb{R}^2 : x > 1/2, y \geq 0, x + y \leq 1\}$ est un triangle d'aire non nulle) et de même $\mathbb{P}(Y > 1/2) > 0$. Notons qu'on a $\mathbb{P}(X > 1/2) = \mathbb{P}(Y > 1/2) = 1/4$ mais qu'on n'a pas besoin de les calculer précisément pour conclure.

2.3.4 Indépendance de variables aléatoires discrètes

On rappelle ici le critère pour vérifier que des v.a. discrètes sont indépendantes.

Proposition 2.3.18. Soit X_1, \dots, X_n des v.a. à valeurs dans des ensembles finis ou dénombrables E_1, \dots, E_n . Si on a

$$\forall x_1 \in X_1, \dots, x_n \in X_n, \quad \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n),$$

alors X_1, \dots, X_n sont indépendantes.

Notons que la réciproque est clairement vraie en prenant $B_i = \{x_i\}$ dans la définition de l'indépendance.

Démonstration. La v.a. $X = (X_1, \dots, X_n)$ prend ses valeurs dans $E_1 \times \dots \times E_n$ qui est dénombrable. Sa loi est donc caractérisée par sa valeur sur les singletons. Soit $x = (x_1, \dots, x_n) \in E_1 \times \dots \times E_n$. On a

$$\begin{aligned} P_X(\{x\}) &= \mathbb{P}(X = x) && \text{(par définition de la loi)} \\ &= \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \\ &= \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n) && \text{(par l'hypothèse)} \\ &= P_{X_1}(\{x_1\}) \cdots P_{X_n}(\{x_n\}) && \text{(par définition de la loi)} \\ &= (P_{X_1} \otimes \cdots \otimes P_{X_n})(\{x\}) && \text{(car } \{x\} = \{x_1\} \times \cdots \times \{x_n\}\text{)}. \end{aligned}$$

Donc $P_{(X_1, \dots, X_n)} = P_{X_1} \otimes \cdots \otimes P_{X_n}$ et X_1, \dots, X_n sont indépendantes par la Proposition 2.3.10. \square

Exemple 2.3.19. Considérons de nouveau l'expérience consistant à lancer deux dés à 6 faces équilibrés. On s'intéresse à la v.a. X qui est la différence absolue entre les dés, i.e. $X(i, j) = |i - j|$ pour $(i, j) \in \Omega$, et la v.a. $Y = \mathbb{1}_B$, où B est l'événement 'le 1er dé donne un résultat pair'. Montrons que X et Y sont indépendantes. Notons que X est à valeurs dans $\{0, 1, 2, 3, 4, 5\}$ et Y à valeurs dans $\{0, 1\}$. On vérifie les égalités suivantes :

$$\begin{aligned} \mathbb{P}(X = 0, Y = 1) &= \mathbb{P}(\{(2, 2), (4, 4), (6, 6)\}) = \frac{3}{36} = \frac{6}{36} \cdot \frac{1}{2} = \mathbb{P}(X = 0) \cdot \mathbb{P}(Y = 1), \\ \mathbb{P}(X = 1, Y = 1) &= \mathbb{P}(\{(2, 1), (2, 3), (4, 3), (4, 5), (6, 5)\}) = \frac{5}{36} = \frac{10}{36} \cdot \frac{1}{2} = \mathbb{P}(X = 1) \cdot \mathbb{P}(Y = 1), \\ \mathbb{P}(X = 2, Y = 1) &= \mathbb{P}(\{(2, 4), (4, 2), (4, 6), (6, 4)\}) = \frac{4}{36} = \frac{8}{36} \cdot \frac{1}{2} = \mathbb{P}(X = 2) \cdot \mathbb{P}(Y = 1), \end{aligned}$$

et de même $\mathbb{P}(X = i, Y = 1) = \mathbb{P}(X = i) \cdot \mathbb{P}(Y = 1)$ pour $i = 3, 4, 5$. Notons qu'on n'a pas besoin de faire le calcul pour $\mathbb{P}(X = i, Y = 0)$: en effet, pour $i \in \{0, \dots, 5\}$, on vient de montrer que les événements $\{X = i\}$ et $\{Y = 1\}$ sont indépendants, mais donc $\{X = i\}$ et $\{Y = 1\}^c = \{Y = 0\}$ sont aussi indépendants (voir Remarque 2.3.2), donc on a bien

$$\mathbb{P}(X = i, Y = 0) = \mathbb{P}(X = i) \cdot \mathbb{P}(Y = 0).$$

On a vérifié les hypothèses de la Proposition 2.3.18 donc X et Y sont indépendantes.

2.3.5 Somme de variables aléatoires indépendantes

Cette section regroupe quelques résultats concernant les sommes de v.a. indépendantes. Le premier montre que la densité de la somme de deux v.a. continues est la convolution de leur densité.

Proposition 2.3.20. *Soit X et Y des v.a. réelles continues indépendantes. Alors $X + Y$ est continue et sa densité est $p_{X+Y} = p_X * p_Y$, i.e.*

$$\forall s \in \mathbb{R}, \quad p_{X+Y}(s) = \int_{\mathbb{R}} p_X(x) p_Y(s - x) dx.$$

Démonstration. Soit $f: \mathbb{R} \rightarrow \mathbb{R}_+$ mesurable. Comme X et Y sont indépendantes, la loi de (X, Y) est $p_X(x) p_Y(y) dx dy$. Ainsi,

$$\begin{aligned} \mathbb{E}[f(X + Y)] &= \int_{\mathbb{R}^2} f(x + y) p_X(x) p_Y(y) dx dy \\ &= \int_{\mathbb{R}} p_X(x) \left(\int_{\mathbb{R}} f(x + y) p_Y(y) dy \right) dx && \text{(Fubini-Tonelli)} \\ &= \int_{\mathbb{R}} p_X(x) \left(\int_{\mathbb{R}} f(s) p_Y(s - x) ds \right) dx && \text{(changement } s = x + y\text{)} \\ &= \int_{\mathbb{R}} f(s) \left(\int_{\mathbb{R}} p_X(x) p_Y(s - x) dx \right) ds, && \text{(Fubini-Tonelli)} \end{aligned}$$

ce qui montre le résultat. \square

Un autre outil très utile lorsqu'on travaille avec des sommes de v.a. indépendantes est la fonction caractéristique.

Méthode (Déterminer la loi d'une somme de v.a. indépendantes). Si X_1, \dots, X_n sont des v.a. indépendantes, alors leur somme a la fonction caractéristique suivante, pour tout $\theta \in \mathbb{R}$,

$$\phi_{X_1+\dots+X_n}(\theta) = \mathbb{E}\left[e^{i\theta(X_1+\dots+X_n)}\right] = \mathbb{E}\left[\prod_{k=1}^n e^{i\theta X_k}\right] = \prod_{k=1}^n \mathbb{E}\left[e^{i\theta X_k}\right] = \prod_{k=1}^n \phi_{X_k}(\theta),$$

où, pour la 3ème égalité, on a utilisé l'indépendance et le critère (iv) de la Proposition 2.3.10 (qui se généralise facilement au cas de fonctions à valeurs complexes). Ainsi la fonction caractéristique de $X_1 + \dots + X_n$ se calcule facilement. Si on arrive à l'identifier avec la fonction caractéristique d'une loi connue, on peut alors conclure que $X_1 + \dots + X_n$ suit cette loi.

Exemple 2.3.21 (Somme de v.a. de Poisson). Soit X_1, \dots, X_n des v.a. indépendantes telles que X_k suive la loi de Poisson $\mathcal{P}(\lambda_k)$. Rappelons que $\phi_{X_k}(\theta) = \exp(\lambda_k(e^{i\theta} - 1))$, on a donc

$$\phi_{X_1+\dots+X_n}(\theta) = \prod_{k=1}^n \phi_{X_k}(\theta) = \prod_{k=1}^n \exp(\lambda_k(e^{i\theta} - 1)) = \exp\left(\left(\sum_{k=1}^n \lambda_k\right)(e^{i\theta} - 1)\right).$$

Donc $X_1 + \dots + X_n$ suit la loi de Poisson de paramètre $\sum_{k=1}^n \lambda_k$.

Exemple 2.3.22 (Somme de v.a. gaussiennes). Soit X_1, \dots, X_n des v.a. indépendantes telles que X_k suive la loi gaussienne $\mathcal{N}(m_k, \sigma_k^2)$. Rappelons que $\phi_{X_k}(\theta) = \exp(im_k\theta - \frac{1}{2}\sigma_k^2\theta^2)$, on a donc

$$\phi_{X_1+\dots+X_n}(\theta) = \prod_{k=1}^n \phi_{X_k}(\theta) = \prod_{k=1}^n \exp\left(im_k\theta - \frac{1}{2}\sigma_k^2\theta^2\right) = \exp\left(i\left(\sum_{k=1}^n m_k\right)\theta - \frac{1}{2}\left(\sum_{k=1}^n \sigma_k^2\right)\theta^2\right).$$

Donc $X_1 + \dots + X_n$ suit la loi gaussienne $\mathcal{N}(\sum_{k=1}^n m_k, \sum_{k=1}^n \sigma_k^2)$.

2.3.6 Minimum ou maximum de variables aléatoires indépendantes

Une autre quantité que l'on croise régulièrement est le maximum ou le minimum de v.a. indépendantes. Dans ce cas, l'outil le plus adapté est la fonction de répartition.

Méthode (Déterminer la loi du minimum ou du maximum de v.a. indépendantes). Si X_1, \dots, X_n sont des variables aléatoires réelles indépendantes, et qu'on s'intéresse à la loi de $\max(X_1, \dots, X_n)$ ou de $\min(X_1, \dots, X_n)$, alors la méthode la plus efficace est de calculer leur fonction de répartition. En effet, pour $a \in \mathbb{R}$, on a

$$\mathbb{P}(\max(X_1, \dots, X_n) \leq a) = \mathbb{P}(X_1 \leq a, \dots, X_n \leq a) = \mathbb{P}(X_1 \leq a) \cdots \mathbb{P}(X_n \leq a)$$

et, en notant que $\mathbb{P}(\min(X_1, \dots, X_n) \leq a) = 1 - \mathbb{P}(\min(X_1, \dots, X_n) > a)$,

$$\mathbb{P}(\min(X_1, \dots, X_n) \leq a) = 1 - \mathbb{P}(X_1 > a, \dots, X_n > a) = 1 - \mathbb{P}(X_1 > a) \cdots \mathbb{P}(X_n > a),$$

où l'on peut ensuite écrire $\mathbb{P}(X_k > a) = 1 - \mathbb{P}(X_k \leq a)$ si l'on veut se ramener à la fonction de répartition de X_k .

Exemple 2.3.23 (Minimum de v.a. exponentielles). Soit X_1, \dots, X_n des v.a. indépendantes de lois exponentielles de paramètres $\lambda_1, \dots, \lambda_n$ et $M_n = \min(X_1, \dots, X_n)$. Rappelons que la fonction de répartition de la loi exponentielle de paramètre $\lambda > 0$ est donnée, pour $a \geq 0$, par

$$F(a) = \int_0^a \lambda e^{-\lambda x} dx = 1 - e^{-\lambda a}$$

et, pour $a < 0$, par $F(a) = 0$. En procédant comme dans la méthode ci-dessus, on obtient, pour $a \geq 0$,

$$\begin{aligned} F_{M_n}(a) &= \mathbb{P}(\min(X_1, \dots, X_n) \leq a) \\ &= 1 - \mathbb{P}(X_1 > a) \cdots \mathbb{P}(X_n > a) \\ &= 1 - (1 - F_{X_1}(a)) \cdots (1 - F_{X_n}(a)) \\ &= 1 - e^{-\lambda_1 a} \cdots e^{-\lambda_n a} \\ &= 1 - e^{-(\lambda_1 + \cdots + \lambda_n)a} \end{aligned}$$

et, pour $a < 0$, $F_{M_n}(a) = 0$ car $M_n > 0$ p.s. On reconnaît que F_{M_n} est la fonction de répartition de la loi exponentielle de paramètre $\lambda_1 + \cdots + \lambda_n$. Donc M_n suit la loi exponentielle de paramètre $\lambda_1 + \cdots + \lambda_n$.

Cette propriété de la loi exponentielle est très importante. Par exemple, si l'on sait qu'un atome radioactif se désintègre au bout d'un temps aléatoire de loi exponentielle de paramètre λ (on peut raisonnablement penser que c'est le cas, car on s'attend à avoir la propriété de perte de mémoire, voir TD3), alors le temps avant d'observer une désintégration parmi N atomes suit la loi exponentielle de paramètre $N\lambda$.

Chapitre 3

Inégalités de concentration et intervalles de confiance

3.1 Inégalités de concentration

3.1.1 Inégalité de Markov

La première inégalité de ce chapitre est très basique mais aussi très utile. C'est en particulier la base pour démontrer les autres inégalités présentées dans cette section.

Proposition 3.1.1 (Inégalité de Markov). *Soit X une v.a. positive. Alors, pour tout $a > 0$,*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Démonstration. En distinguant les cas $X \geq a$ et $0 \leq X < a$, on remarque que $\mathbb{1}_{X \geq a} \leq X/a$. En prenant l'espérance de chaque côté, on obtient l'inégalité désirée. \square

Notons que cette inégalité est inutile si $\mathbb{E}[X] = \infty$ ou si $a \leq \mathbb{E}[X]$. Elle est aussi optimale au sens où, pour chaque $a > 0$ il existe une v.a. X telle qu'il y ait égalité. En effet, en considérant X qui vaut a avec probabilité $p \in [0, 1]$ et 0 sinon, on a $\mathbb{E}[X] = pa$ et donc

$$\mathbb{P}(X \geq a) = p = \frac{\mathbb{E}[X]}{a}.$$

Dans le cas d'une v.a. réelle X pas forcément positive, on a, pour tout $a > 0$,

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[|X|]}{a},$$

ce qui permet de borner à la fois $\mathbb{P}(X \geq a)$ et $\mathbb{P}(X \leq -a)$.

Remarque 3.1.2. L'inégalité de Markov ne s'utilise pas forcément telle quelle. Il peut être judicieux de l'appliquer à une fonction de X pour obtenir une meilleure décroissance de la queue de distribution à l'infini. Par exemple, si X est une v.a. positive telle que $\mathbb{E}[X^p] < \infty$ pour un certain $p > 1$, on peut écrire

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

Pour a suffisamment grand, cette inégalité sera meilleure que celle obtenue en appliquant directement l'inégalité de Markov.

3.1.2 Inégalité de Bienaymé–Tchebychev

La seconde inégalité de cette section est une conséquence directe de l'inégalité de Markov. Elle est particulièrement utile pour étudier des sommes de v.a. indépendantes, comme on va le voir.

Proposition 3.1.3 (Inégalité de Bienaymé–Tchebychev). *Soit $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. Alors, pour tout $a > 0$,*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\text{Var}(X)}{a^2}.$$

Démonstration. On passe au carré l'inégalité à l'intérieur de la probabilité, puis on applique l'inégalité de Markov :

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) = \mathbb{P}\left((X - \mathbb{E}[X])^2 \geq a^2\right) \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{a^2} = \frac{\text{Var}(X)}{a^2},$$

ce qui montre l'inégalité désirée. \square

Afin d'appliquer l'inégalité de Bienaymé–Tchebychev à des sommes de v.a. indépendantes, on va tout d'abord montrer ce résultat concernant leur variance.

Proposition 3.1.4 (Variance d'une somme). *Soit $X_1, \dots, X_n \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. On a*

$$\text{Var}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \sum_{\ell=1}^n \text{Cov}(X_k, X_\ell).$$

Si en outre X_1, \dots, X_n sont indépendantes, alors

$$\text{Var}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \text{Var}(X_k).$$

Démonstration. Montrons la première identité. On a

$$\text{Var}\left(\sum_{k=1}^n X_k\right) = \mathbb{E}\left[\left(\sum_{k=1}^n X_k - \mathbb{E}\left[\sum_{k=1}^n X_k\right]\right)^2\right] = \mathbb{E}\left[\left(\sum_{k=1}^n (X_k - \mathbb{E}[X_k])\right)^2\right].$$

Puis en développant le carré, on obtient

$$\text{Var}\left(\sum_{k=1}^n X_k\right) = \mathbb{E}\left[\sum_{k=1}^n \sum_{\ell=1}^n (X_k - \mathbb{E}[X_k])(X_\ell - \mathbb{E}[X_\ell])\right] = \sum_{k=1}^n \sum_{\ell=1}^n \text{Cov}(X_k, X_\ell),$$

en sortant les sommes de l'espérance. Si X_1, \dots, X_n sont indépendantes, alors $\text{Cov}(X_k, X_\ell) = 0$ dès que $k \neq \ell$ et donc on obtient la seconde identité (car $\text{Cov}(X_k, X_k) = \text{Var}(X_k)$). \square

On en déduit l'inégalité de concentration suivante pour des sommes de variables aléatoires *indépendantes et identiquement distribuées* (abrégé en *i.i.d.*), c'est-à-dire indépendantes et ayant toutes la même loi.

Corollaire 3.1.5 (Bienaymé–Tchebychev pour une somme de v.a. i.i.d.). *Soit X_1, \dots, X_n des v.a. réelles i.i.d. telles que $\mathbb{E}[X_1^2] < \infty$. Alors, pour tout $\varepsilon > 0$,*

$$\mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n} - \mathbb{E}[X_1]\right| \geq \varepsilon\right) \leq \frac{\text{Var}(X_1)}{n\varepsilon^2}.$$

Cette inégalité nous montre que, si l'on répète plusieurs fois la même expérience aléatoire, la moyenne empirique des résultats $(X_1 + \dots + X_n)/n$ se *concentre* autour de la vraie moyenne $\mathbb{E}[X_1]$. En effet, pour tout $\varepsilon > 0$, la probabilité que la moyenne empirique soit dans l'intervalle $[\mathbb{E}[X_1] - \varepsilon, \mathbb{E}[X_1] + \varepsilon]$ est aussi proche de 1 que l'on veut en choisissant n suffisamment grand.

Démonstration. Comme X_1, \dots, X_n sont identiquement distribuées, elles ont les mêmes moments : en particulier, elles sont toutes dans L^2 et, pour tout $i \in \llbracket 1, n \rrbracket$, $\mathbb{E}[X_i] = \mathbb{E}[X_1]$ et $\text{Var}(X_i) = \text{Var}(X_1)$. On pose $\bar{X} = (X_1 + \dots + X_n)/n$. On a alors

$$\mathbb{E}[\bar{X}] = \frac{1}{n}(\mathbb{E}[X_1] + \dots + \mathbb{E}[X_n]) = \mathbb{E}[X_1],$$

et, par indépendance de X_1, \dots, X_n et la Proposition 3.1.4,

$$\text{Var}(\bar{X}) = \frac{1}{n^2} \text{Var}(X_1 + \dots + X_n) = \frac{1}{n^2}(\text{Var}(X_1) + \dots + \text{Var}(X_n)) = \frac{1}{n} \text{Var}(X_1).$$

En appliquant l'inégalité de Bienaymé–Tchebychev à \bar{X} , on obtient

$$\mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n} - \mathbb{E}[X_1]\right| \geq \varepsilon\right) = \mathbb{P}\left(|\bar{X} - \mathbb{E}[\bar{X}]| \geq \varepsilon\right) \leq \frac{\text{Var}(\bar{X})}{\varepsilon^2} = \frac{\text{Var}(X_1)}{n\varepsilon^2},$$

qui est le résultat désiré. \square

3.1.3 Inégalité de Hoeffding

L'application de l'inégalité de Bienaymé–Tchebychev aux sommes de v.a. i.i.d. vue ci-dessus a le mérite d'être vraie dès que les v.a. ont un moment d'ordre 2 fini, mais la borne qu'elle donne sur les probabilités ne décroît pas très vite quand n tend vers l'infini. En général, plus l'on suppose que les v.a. considérées ont une queue de distribution qui décroît vite, meilleures seront les inégalités que l'on peut obtenir. On traite ici le meilleur cas possible concernant la queue de distribution, c'est-à-dire celui de v.a. bornées.

Proposition 3.1.6 (Inégalité de Hoeffding). *Soit $a < b$. Soit X_1, \dots, X_n des v.a. réelles i.i.d. à valeurs dans $[a, b]$. Alors, pour tout $\varepsilon > 0$,*

$$\mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n} - \mathbb{E}[X_1]\right| > \varepsilon\right) \leq 2 \exp\left(-\frac{2\varepsilon^2 n}{(b-a)^2}\right).$$

Pour un ε fixé, cette inégalité montre que la probabilité que la moyenne empirique dévie de plus de ε de la vraie moyenne décroît au moins exponentiellement vite en n . C'est une décroissance bien plus rapide que celle en $1/n$ obtenue par l'inégalité de Bienaymé–Tchebychev.

La démonstration de l'inégalité de Hoeffding repose sur une méthode classique en probabilité : au lieu d'appliquer l'inégalité de Markov à $|S|$ avec $S = X_1 + \dots + X_n - n\mathbb{E}[X_1]$ (ou à S^2 comme on l'a fait pour Bienaymé–Tchebychev), on va l'appliquer à e^{tS} pour $t \geq 0$ quelconque, puis on optimisera en t pour obtenir la meilleure borne possible. Afin de majorer $\mathbb{E}[e^{tS}]$, nous aurons besoin du lemme préliminaire suivant.

Lemme 3.1.7. *Soit $a < b$ et Y une variable aléatoire à valeur dans $[a, b]$ telle que $\mathbb{E}[Y] = 0$. Alors, pour tout $t \geq 0$,*

$$\mathbb{E}[e^{tY}] \leq \exp\left(\frac{t^2(b-a)^2}{8}\right).$$

Démonstration (non vue en cours, donc non requise). Par convexité de la fonction $y \in [a, b] \mapsto e^{ty}$, en écrivant $Y = \frac{b-Y}{b-a}a + \frac{Y-a}{b-a}b$, on a

$$e^{tY} \leq \frac{b-Y}{b-a}e^{ta} + \frac{Y-a}{b-a}e^{tb}$$

et donc, en passant à l'espérance et en utilisant que $\mathbb{E}[Y] = 0$,

$$\mathbb{E}[e^{tY}] \leq \frac{b}{b-a}e^{ta} + \frac{-a}{b-a}e^{tb}.$$

En notant $g(t) = \log\left(\frac{b}{b-a}e^{ta} + \frac{-a}{b-a}e^{tb}\right)$, on veut maintenant montrer que $g(t) \leq \frac{1}{8}t^2(b-a)^2$. On calcule les dérivées :

$$g'(t) = \frac{ab(e^{ta} - e^{tb})}{be^{ta} - ae^{tb}} \quad \text{et} \quad g''(t) = ab \frac{(b-a)^2 e^{t(a+b)}}{(be^{ta} - ae^{tb})^2}.$$

Or par formule de Taylor avec reste intégral, on a

$$g(t) = g(0) + g'(0)t + \int_0^t g''(u)u \, du.$$

Remarquons alors que $g(0) = g'(0) = 0$ et

$$g''(u) = (b-a)^2 \frac{abe^{u(a+b)}}{(be^{ua} - ae^{ub})^2} = -(b-a)^2 \frac{xy}{(x+y)^2}$$

avec $x = be^{ua} \geq 0$ et $y = -ae^{ub} \geq 0$ (en effet on a $a \leq 0 \leq b$ car $\mathbb{E}[Y] = 0$). Finalement, on en déduit que $|g''(u)| \leq (b-a)^2/4$ et donc

$$|g(t)| = \left| \int_0^t g''(u)u \, du \right| \leq \frac{t^2(b-a)^2}{8},$$

ce qui donne le résultat souhaité. □

Démonstration de la Proposition 3.1.6. On pose $S = X_1 + \dots + X_n - n\mathbb{E}[X_1]$. Alors, on a

$$\mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n} - \mathbb{E}[X_1]\right| > \varepsilon\right) = \mathbb{P}(|S| > \varepsilon n) = \mathbb{P}(S > \varepsilon n) + \mathbb{P}(S < -\varepsilon n).$$

Pour majorer $\mathbb{P}(S > \varepsilon n)$, on considère un paramètre $t \geq 0$ et on écrit

$$\mathbb{P}(S \geq \varepsilon n) = \mathbb{P}(e^{tS} \geq e^{t\varepsilon n}) \leq \frac{\mathbb{E}[e^{tS}]}{e^{t\varepsilon n}},$$

par l'inégalité de Markov appliquée à e^{tS} . On veut à présent majorer $\mathbb{E}[e^{tS}]$. Pour cela, on pose $Y_i = X_i - \mathbb{E}[X_i]$. En utilisant que $S = Y_1 + \dots + Y_n$, puis l'indépendance de Y_1, \dots, Y_n , on obtient

$$\mathbb{E}[e^{tS}] = \prod_{i=1}^n \mathbb{E}[e^{tY_i}] \leq \prod_{i=1}^n \exp\left(\frac{t^2(b-a)^2}{8}\right),$$

en utilisant le Lemme 3.1.7 car $\mathbb{E}[Y_i] = 0$ et Y_i est à valeurs dans $[a - \mathbb{E}[X_i], b - \mathbb{E}[X_i]]$. On a donc montré que

$$\mathbb{P}(S \geq \varepsilon n) \leq \exp\left(-t\varepsilon n + \frac{t^2(b-a)^2}{8}n\right).$$

Finalement, on cherche le t qui minimise cette expression et c'est $t = 4\varepsilon/(b-a)^2$. En choisissant ce t , on obtient

$$\mathbb{P}(S \geq \varepsilon n) \leq \exp\left(-\frac{2\varepsilon^2 n}{(b-a)^2}\right).$$

En appliquant ce résultat aux v.a. $-X_1, \dots, -X_n$, on obtient la même borne pour $\mathbb{P}(-S \geq \varepsilon n)$ et ça conclut la démonstration. □

3.2 Intervalles de confiance

Dans cette section, on va présenter l'application des inégalités de concentration à l'obtention d'intervalles de confiance en statistique. Soulignons tout d'abord la différence de point de vue entre la statistique et les probabilités. En probabilités, on suppose connue la loi des variables aléatoires et on en déduit des résultats sur leurs comportements. En statistique, on observe des réalisations de variables aléatoires de lois inconnues et on tente d'en déduire des informations sur ces lois.

On se place ici dans le cadre où l'on observe la réalisation de v.a. réelles X_1, \dots, X_n i.i.d. dont la loi appartient à une certaine famille $(P_\theta)_{\theta \in \Theta}$ (on dit alors que X_1, \dots, X_n est un *échantillon de taille n*). L'objectif est d'estimer la valeur du paramètre $\theta \in \Theta$ tel que la loi des X_i soit P_θ . On suppose pour simplifier que $\Theta \subset \mathbb{R}$ et on se concentre sur le fait d'établir un intervalle de confiance pour θ : on veut construire un intervalle $I(\omega)$ à partir des réalisations $X_1(\omega), \dots, X_n(\omega)$ tel que l'on pourra affirmer que l'intervalle aléatoire I contient θ avec probabilité au moins 95% ou 99%.

Définition 3.2.1. Soit $\alpha \in]0, 1[$. Un intervalle de confiance pour θ de niveau $1 - \alpha$ est un intervalle aléatoire I de la forme $[F(X_1, \dots, X_n), G(X_1, \dots, X_n)]$ avec $F, G: \mathbb{R}^n \rightarrow \mathbb{R}$ mesurables (ne dépendant pas de θ), tel que pour tout $\theta \in \Theta$, si X_1, \dots, X_n ont loi P_θ , alors

$$\mathbb{P}(\theta \in I) \geq 1 - \alpha.$$

Il est très important que I soit défini à partir de X_1, \dots, X_n sans faire intervenir θ , car on ne connaît pas θ (sinon on prendrait $I = [\theta - \varepsilon, \theta + \varepsilon]$ avec ε aussi petit que l'on veut et on aurait toujours $\mathbb{P}(\theta \in I) = 1$). Il faut aussi bien noter que dans l'écriture " $\theta \in I$ ", c'est I qui est aléatoire et pas θ . En particulier, il ne faut pas confondre l'intervalle de confiance, qui est aléatoire, et sa réalisation : ça n'a pas de sens d'écrire $\mathbb{P}(\theta \in [0.74, 0.78]) = 0.95$ car cette probabilité vaut 0 si $\theta \notin [0.74, 0.78]$ et 1 si $\theta \in [0.74, 0.78]$.

Pour un certain niveau α donné, il n'y a pas de choix unique de l'intervalle de confiance mais on essaie d'en trouver un le plus court possible. En particulier, selon l'inégalité de concentration utilisée, on obtiendra des résultats différents. Il est également important de noter qu'il y a toujours un compromis entre le niveau et la longueur : plus on veut un niveau proche de 1, plus l'intervalle sera long.

Exemple 3.2.2. Considérons un sondage où l'on interroge n individus et chacun répond 1 avec probabilité θ et 0 avec probabilité $1 - \theta$. On suppose de plus que leurs réponses sont indépendantes. Cela correspond au cas où $\Theta = [0, 1]$, $P_\theta = \mathcal{B}(\theta)$ (la loi de Bernoulli de paramètre θ) et X_i est la réponse de i -ème individu. Si X_1, \dots, X_n ont loi P_θ , alors $\mathbb{E}[X_i] = \theta$ donc on sait que la moyenne empirique $\bar{X}_n = (X_1 + \dots + X_n)/n$ va nous fournir un bon estimateur de la valeur de θ . Il reste à quantifier cela pour obtenir un intervalle de confiance pour θ de niveau $1 - \alpha$. On va le faire de deux manières différentes : avec l'inégalité de Bienaymé–Tchebychev puis avec l'inégalité de Hoeffding.

- Par l'inégalité de Bienaymé–Tchebychev, on a, pour tout $\varepsilon > 0$,

$$\mathbb{P}(|\bar{X}_n - \theta| \geq \varepsilon) \leq \frac{\text{Var}(X_1)}{n\varepsilon^2} = \frac{\theta(1 - \theta)}{n\varepsilon^2},$$

où l'on a utilisé que la variance de la loi $\mathcal{B}(\theta)$ vaut $\theta(1 - \theta)$. On en déduit que

$$\mathbb{P}(\theta \in [\bar{X}_n - \varepsilon, \bar{X}_n + \varepsilon]) \geq 1 - \frac{\theta(1 - \theta)}{n\varepsilon^2}.$$

On veut à présent choisir ε le plus petit possible tel que, pour tout $\theta \in \Theta$,

$$1 - \frac{\theta(1 - \theta)}{n\varepsilon^2} \geq 1 - \alpha \quad \Leftrightarrow \quad \varepsilon \geq \sqrt{\frac{\theta(1 - \theta)}{n\alpha}}.$$

On rappelle que l'intervalle de confiance ne doit pas dépendre de θ donc on ne peut pas choisir $\varepsilon = \sqrt{\theta(1-\theta)/(n\alpha)}$. On remarque plutôt que $\theta(1-\theta) \leq 1/4$ pour tout $\theta \in \Theta$, donc on peut prendre

$$\varepsilon = \sqrt{\frac{1}{4n\alpha}},$$

qui satisfait l'inégalité précédente pour tout $\theta \in \Theta$. Ainsi, l'intervalle

$$I = \left[\bar{X}_n - \sqrt{\frac{1}{4n\alpha}}, \bar{X}_n + \sqrt{\frac{1}{4n\alpha}} \right]$$

est un intervalle de confiance pour θ de niveau $1 - \alpha$.

- Si l'on utilise l'inégalité de Hoeffding (car les X_i sont à valeurs dans $[0, 1]$), on obtient, pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(|\bar{X}_n - \theta| \geq \varepsilon\right) \leq 2 \exp(-2\varepsilon^2 n)$$

et donc

$$\mathbb{P}\left(\theta \in [\bar{X}_n - \varepsilon, \bar{X}_n + \varepsilon]\right) \geq 1 - 2 \exp(-2\varepsilon^2 n),$$

où ici la borne obtenue ne dépend pas de θ . On choisit donc ε tel que $2 \exp(-2\varepsilon^2 n) = \alpha$. Cela donne l'intervalle

$$I = \left[\bar{X}_n - \sqrt{\frac{\log(2/\alpha)}{2n}}, \bar{X}_n + \sqrt{\frac{\log(2/\alpha)}{2n}} \right],$$

qui est donc aussi un intervalle de confiance pour θ de niveau $1 - \alpha$.

On peut à présent comparer la longueur de ces intervalles de confiance pour voir lequel est le meilleur. Il est intéressant de voir que pour les deux la dépendance en n est la même : leur longueur décroît en $1/\sqrt{n}$. On ne peut pas espérer mieux, car le théorème central limite nous dit que \bar{X}_n fluctue autour de $\mathbb{E}[X_1]$ à l'ordre $1/\sqrt{n}$.

Concernant la dépendance en α , on voit que le premier intervalle est plus court ssi $\alpha > 0.2322\dots$. Quand α est proche de 0, le second intervalle de confiance est vraiment meilleur : par exemple pour $\alpha = 0.01$ et $n = 1000$, on obtient environ les intervalles

$$\left[\bar{X}_n - 0.16, \bar{X}_n + 0.16 \right] \quad \text{et} \quad \left[\bar{X}_n - 0.0026, \bar{X}_n + 0.0026 \right].$$

Chapitre 4

Convergence de variables aléatoires et loi des grands nombres

Dans ce chapitre seront présentées trois premières notions de convergence pour des suites de variables aléatoires : la convergence presque sûre, la convergence dans L^p et la convergence en probabilité. Les deux premières ont déjà été considérées en théorie de la mesure. En effet, une suite $(X_n)_{n \geq 1}$ de v.a. réelles n'est rien d'autre qu'une suite de fonctions mesurables de (Ω, \mathcal{A}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et donc on peut transposer les notions de convergences vues pour des suites de fonctions. Une quatrième notion de convergence sera vue au chapitre suivant : la convergence en loi.

4.1 Suite de variables aléatoires indépendantes

Pour fournir de la matière aux exemples qui seront couverts dans ce chapitre, on a besoin de généraliser la notion d'indépendance à des suites.

Définition 4.1.1 (Indépendance de suites).

- Soit $A_1, A_2, \dots \in \mathcal{A}$. On dit que $(A_k)_{k \geq 1}$ est une suite d'événements indépendants si, pour tout $n \geq 1$, A_1, \dots, A_n sont indépendants.
- Soit X_1, X_2, \dots des v.a. On dit que $(X_k)_{k \geq 1}$ est une suite de variables aléatoires indépendantes si, pour tout $n \geq 1$, X_1, \dots, X_n sont indépendantes.

Le choix de faire commencer ces suites à $k = 1$ est bien évidemment arbitraire et la définition se généralise aisément aux autres cas. Certaines propriétés vues pour un nombre fini d'événements ou de v.a. indépendants se transfèrent directement aux suites comme conséquence de cette définition. Par exemple, si $(A_k)_{k \geq 1}$ est une suite d'événements indépendants, alors on peut remplacer n'importe quelle partie de ces événements par leur complémentaires et la suite reste indépendante. Si $(X_k)_{k \geq 1}$ est une suite de v.a. indépendantes, alors :

- Toute sous-suite $(X_{\varphi(k)})_{k \geq 1}$, avec $\varphi: \mathbb{N}^* \rightarrow \mathbb{N}^*$ strictement croissante, est une suite de v.a. indépendantes.
- Pour toutes fonctions f_k mesurables, $(f_k(X_k))_{k \geq 1}$ est une suite de v.a. indépendantes.
- Toute suite de paquets disjoints de v.a. parmi les X_k est aussi une suite de v.a. indépendantes. Par exemple, $((X_{2k-1}, X_{2k}))_{k \geq 1}$ est une suite de v.a. indépendantes.

Le résultat suivant garantit l'existence d'une suite de variables aléatoires indépendante avec des lois données. Il nous assure que l'on ne travaille pas avec un objet inexistant dans les exemples.

Théorème 4.1.2 (Existence de suite de v.a. indépendantes). *Pour chaque $k \geq 1$, soit P_k une loi sur \mathbb{R} . Alors il existe un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et des v.a. réelles X_1, X_2, \dots définies sur cet espace tels que $(X_k)_{k \geq 1}$ soit une suite de variables aléatoires indépendantes et, pour tout $k \geq 1$, X_k ait pour loi P_k .*

Démonstration (admise). Pour les curieux, vous pouvez consulter le Théorème IV.3.1 dans le livre de Barbe et Ledoux. \square

Remarque 4.1.3. Dans le cas d'un nombre fini de variables aléatoires, ce résultat découle du théorème d'existence des mesures produits (Théorème 2.1.4). En effet, pour construire des v.a. réelles X_1, \dots, X_n indépendantes et de lois P_1, \dots, P_n , on peut considérer $\Omega = \mathbb{R}^n$, $\mathcal{A} = \mathcal{B}(\mathbb{R}^n)$, $\mathbb{P} = P_1 \otimes \dots \otimes P_n$ et

$$X_k: \omega = (\omega_1, \dots, \omega_n) \in \mathbb{R}^n \mapsto \omega_k \in \mathbb{R}.$$

Alors la fonction $X = (X_1, \dots, X_n): \mathbb{R}^n \rightarrow \mathbb{R}^n$ est l'identité et donc la loi de X (qui est la mesure image de \mathbb{P} par X) est égale à $\mathbb{P} = P_1 \otimes \dots \otimes P_n$. Il en découle que X_1, \dots, X_n sont indépendantes et de lois P_1, \dots, P_n .

Le théorème ci-dessus revient donc à montrer qu'il existe une mesure sur $\mathbb{R}^{\mathbb{N}^*}$ qui est la mesure produit de toutes les lois P_k pour $k \in \mathbb{N}^*$. Cette construction d'un produit infini de mesures est spécifique aux mesures de probabilité, alors que l'on peut définir le produit fini de mesures dès qu'elles sont σ -finies. Cela nécessite également de préciser la définition de la tribu produit dont on munit $\mathbb{R}^{\mathbb{N}^*}$. Nous n'aborderons pas ces détails dans ce cours.

Remarque 4.1.4. Un cas particulier de ce théorème peut être déduit de l'existence de la mesure de Lebesgue : c'est l'existence d'une suite $(X_n)_{n \geq 1}$ de v.a. indépendantes de loi de Bernoulli de paramètre $1/2$. Pour cela, on considère $\Omega = [0, 1[$ muni de $\mathcal{A} = \mathcal{B}([0, 1[)$ et de \mathbb{P} la mesure de Lebesgue sur $[0, 1[$, puis, pour $\omega \in \Omega$, on définit $X_k(\omega)$ comme le k -ième coefficient dans le développement dyadique de ω . Voir le TD6 pour les détails (en particulier, pour certains $\omega \in [0, 1[$, le développement dyadique n'est pas unique et il faut donc préciser comment on le choisit pour être rigoureux).

4.2 Rappels sur les limites inférieure et supérieure

4.2.1 Limites inférieure et supérieure de réels

Rappelons les définitions et propriétés des limites inférieure et supérieure de réels. Pour $(x_n)_{n \geq 1}$ une suite de réels, on définit

$$\begin{aligned} \liminf_{n \rightarrow \infty} x_n &= \lim_{n \rightarrow \infty} \uparrow \inf_{k \geq n} x_k \in \overline{\mathbb{R}}, \\ \limsup_{n \rightarrow \infty} x_n &= \lim_{n \rightarrow \infty} \downarrow \sup_{k \geq n} x_k \in \overline{\mathbb{R}}, \end{aligned}$$

où $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. Ces deux limites sont toujours bien définies par monotonie, ce qui permet de les étudier sans savoir encore que la vraie limite existe (ou même si la vraie limite n'existe pas).

De plus, $\liminf_{n \rightarrow \infty} x_n$ est la plus petite valeur d'adhérence de $(x_n)_{n \geq 1}$ et $\limsup_{n \rightarrow \infty} x_n$ est la plus grande valeur d'adhérence de $(x_n)_{n \geq 1}$. Finalement, la convergence de la suite peut être caractérisée en termes des limites inférieure et supérieure :

$$\lim_{n \rightarrow \infty} x_n \text{ existe dans } \overline{\mathbb{R}} \iff \liminf_{n \rightarrow \infty} x_n = \limsup_{n \rightarrow \infty} x_n,$$

et dans ce cas les trois limites coïncident.

4.2.2 Limites inférieure et supérieure d'ensembles

Soit $(A_n)_{n \geq 1}$ une suite de parties de Ω . Ses limites inférieure et supérieure sont définies ainsi :

$$\begin{aligned} \liminf_{n \rightarrow \infty} A_n &= \bigcup_{m \geq 1} \bigcap_{n \geq m} A_n = \{\omega \in \Omega : \omega \text{ est dans tous les } A_n \text{ à partir d'un certain rang}\}, \\ \limsup_{n \rightarrow \infty} A_n &= \bigcap_{m \geq 1} \bigcup_{n \geq m} A_n = \{\omega \in \Omega : \omega \text{ appartient à une infinité de } A_n\}. \end{aligned}$$

Si les A_n sont des événements (c'est-à-dire appartiennent à la tribu \mathcal{A}) alors leurs limites inférieure et supérieure aussi.

Rappelons que l'on a vu au TD1 les inégalités

$$\limsup_{n \rightarrow \infty} \mathbb{P}(A_n) \leq \mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) \quad \text{et} \quad \liminf_{n \rightarrow \infty} \mathbb{P}(A_n) \geq \mathbb{P}\left(\liminf_{n \rightarrow \infty} A_n\right),$$

l'une pouvant se déduire de l'autre grâce au fait que $(\limsup_{n \rightarrow \infty} A_n)^c = \liminf_{n \rightarrow \infty} A_n^c$.

4.3 Premières notions de convergence

4.3.1 Convergence presque sûre

La convergence presque sûre est l'équivalent de la convergence presque partout pour les probabilistes.

Définition 4.3.1. Soit X, X_1, X_2, \dots des v.a. réelles. On dit que $(X_n)_{n \geq 1}$ converge presque sûrement vers X , que l'on note

$$X_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} X,$$

si $(X_n)_{n \geq 1}$ converge \mathbb{P} -presque partout vers X , i.e. si l'événement

$$\left\{ \lim_{n \rightarrow \infty} X_n = X \right\} = \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right\}$$

a probabilité 1.

Cette définition repose sur le fait que l'ensemble $\{\lim_{n \rightarrow \infty} X_n = X\}$ est mesurable. Rappelons que cela se montre ainsi :

$$\left\{ \lim_{n \rightarrow \infty} X_n = X \right\} = \left\{ \forall k \geq 1, \exists m \geq 1, \forall n \geq m, |X_n - X| \leq \frac{1}{k} \right\} = \bigcap_{k \geq 1} \bigcup_{m \geq 1} \bigcap_{n \geq m} \left\{ |X_n - X| \leq \frac{1}{k} \right\},$$

où chacun des ensembles $\{|X_n - X| \leq 1/k\}$ est dans \mathcal{A} et donc, par stabilité de \mathcal{A} par unions et intersections dénombrables, on a $\{\lim_{n \rightarrow \infty} X_n = X\} \in \mathcal{A}$.

Cela a également un sens de dire " $(X_n)_{n \geq 1}$ converge p.s." sans préciser la limite. En effet, l'ensemble $\{(X_n)_{n \geq 1} \text{ converge}\}$ est mesurable : on peut le montrer ainsi, en utilisant la complétude de \mathbb{R} ,

$$\{(X_n)_{n \geq 1} \text{ converge}\} = \{(X_n)_{n \geq 1} \text{ est de Cauchy}\} = \bigcap_{k \geq 1} \bigcup_{\ell \geq 1} \bigcap_{m, n \geq \ell} \left\{ |X_n - X_m| \leq \frac{1}{k} \right\},$$

dont on déduit que $\{(X_n)_{n \geq 1} \text{ converge}\} \in \mathcal{A}$.

Exemple 4.3.2. Soit U une v.a. uniforme dans $[0, 1]$. On a

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

En effet, on a $U \in [0, 1[$ p.s. et, pour tout $\omega \in \{U \in [0, 1[\}$, on a $U(\omega)^n \rightarrow 0$.

La convergence p.s. satisfait les mêmes propriétés de stabilité par opérations que la convergence de suites de réels. Par exemple, si $(X_n)_{n \geq 1}$ et $(Y_n)_{n \geq 1}$ sont des suites de v.a. réelles convergeant p.s. vers X et Y respectivement, alors

$$X_n + Y_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} X + Y, \quad X_n - Y_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} X - Y \quad \text{et} \quad X_n Y_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} XY.$$

En effet, il suffit de considérer $\omega \in \{\lim_{n \rightarrow \infty} X_n = X\} \cap \{\lim_{n \rightarrow \infty} Y_n = Y\}$, qui est un événement de probabilité 1, puis d'appliquer les résultats connus pour les suites de réels $X_n(\omega)$ et $Y_n(\omega)$. Si on a en outre $\mathbb{P}(Y = 0) = 0$, alors

$$X_n/Y_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} X/Y,$$

en considérant $\omega \in \{\lim_{n \rightarrow \infty} X_n = X\} \cap \{\lim_{n \rightarrow \infty} Y_n = Y\} \cap \{Y \neq 0\}$, qui est aussi un événement de probabilité 1 car $Y \neq 0$ p.s. Mentionnons finalement que, si $f: \mathbb{R} \rightarrow \mathbb{R}$ est continue, alors on a

$$f(X_n) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} f(X).$$

4.3.2 Convergence L^p

Soit $p \in [1, \infty[$. Rappelons que, pour X une v.a. réelle, on définit $\|X\|_p = \mathbb{E}[|X|^p]^{1/p}$. L'espace $L^p(\Omega, \mathcal{A}, \mathbb{P})$, que l'on abrégera en L^p quand il n'y a pas d'ambiguïté possible, est l'ensemble des v.a. réelles X telles que $\|X\|_p < \infty$, que l'on quotiente par la relation d'équivalence "être égal p.s.". En outre, $\|\cdot\|_p$ définit bien une norme sur L^p et donc naturellement d'une notion de convergence.

Définition 4.3.3. Soit $p \in [1, \infty[$. Soit $X, X_1, X_2, \dots \in L^p$. On dit que $(X_n)_{n \geq 1}$ converge dans L^p vers X , que l'on note

$$X_n \xrightarrow[n \rightarrow \infty]{L^p} X,$$

si $\|X_n - X\|_p \rightarrow 0$ quand $n \rightarrow \infty$, i.e. si $\mathbb{E}[|X_n - X|^p] \rightarrow 0$ quand $n \rightarrow \infty$.

Remarque 4.3.4. Soit $p \in [1, \infty[$. On rappelle que si $(X_n)_{n \geq 1}$ converge dans L^p vers X , alors

$$\|X_n\|_p \xrightarrow[n \rightarrow \infty]{} \|X\|_p.$$

De plus, si p est un entier (de sorte que x^p ait un sens pour tout $x \in \mathbb{R}$), alors

$$\mathbb{E}[X_n^p] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[X^p].$$

Méthode (Montrer une convergence dans L^p). Voici quelques approches possibles pour montrer qu'une suite $(X_n)_{n \geq 1}$ converge dans L^p vers X :

- À partir d'autres convergences dans L^p , en utilisant les inégalités de Minkowski et d'Hölder (voir par exemple l'exercice 1 du TD8).
- En calculant directement $\mathbb{E}[|X_n - X|^p]$ pour montrer qu'elle tend vers zéro. On peut le faire si l'on connaît la loi explicitement. Un autre cas important est celui où $p = 2$ et X_n est une somme de v.a. indépendantes : on utilise alors le fait que la variance d'une somme de v.a. indépendantes est égale à la somme des variances.
- Si l'on sait que $(X_n)_{n \geq 1}$ converge p.s. vers X , on peut appliquer le théorème de convergence dominée pour en déduire une convergence dans L^p (si l'on arrive à justifier la domination!).

Exemple 4.3.5. Soit U une v.a. uniforme dans $[0, 1]$. On a vu à l'Exemple 4.3.2 que $(U^n)_{n \geq 0}$ converge p.s. vers 0. En outre, on a $|(U^n)^p| \leq 1$ p.s. et $\mathbb{E}[1] < \infty$. Donc, par le théorème de convergence dominée,

$$\mathbb{E}[|U^n - 0|^p] = \mathbb{E}[(U^n)^p] \xrightarrow[n \rightarrow \infty]{} 0.$$

Donc $(U^n)_{n \geq 0}$ converge dans L^p vers 0, pour tout $p \geq 1$.

On peut également montrer cette convergence par calcul direct : on a

$$\mathbb{E}[|U^n - 0|^p] = \mathbb{E}[(U^n)^p] = \int_0^1 x^{np} dx = \frac{1}{np+1} \xrightarrow[n \rightarrow \infty]{} 0,$$

ce qui montre que $(U^n)_{n \geq 0}$ converge dans L^p vers 0.

On omet dans ce cours le cas de la convergence L^∞ , qui est la convergence pour la norme $\|\cdot\|_\infty$, dont la définition est rappelée en Section 1.3.2. En effet, cette convergence est rarement utilisée pour des variables aléatoires.

On a déjà vu que si $1 \leq p < q < \infty$, alors $L^q \subset L^p$ (voir Proposition 1.3.13). De la même manière il est plus fort de converger dans L^q que dans L^p , comme montré dans le résultat suivant.

Proposition 4.3.6. *Soit $X, X_1, X_2, \dots \in L^q$. Soit $1 \leq p < q < \infty$. Si $(X_n)_{n \geq 1}$ converge dans L^q vers X , alors $(X_n)_{n \geq 1}$ converge dans L^p vers X .*

Démonstration. On applique l'inégalité de Hölder aux v.a. $|X_n - X|^p$ et 1 avec $p' = q/p > 1$ et q' tel que $\frac{1}{p'} + \frac{1}{q'} = 1$:

$$\mathbb{E}[|X_n - X|^p] = \mathbb{E}[|X_n - X|^p \cdot 1] \leq \mathbb{E}\left[(|X_n - X|^p)^{p'} \right]^{1/p'} \mathbb{E}\left[(1)^{q'} \right] = \mathbb{E}[|X_n - X|^{q}]^{p/q}.$$

Comme $\mathbb{E}[|X_n - X|^q] \rightarrow 0$, on en déduit que $\mathbb{E}[|X_n - X|^p] \rightarrow 0$ et donc que $(X_n)_{n \geq 1}$ converge dans L^p vers X . \square

Concernant la stabilité par opérations de la convergence L^p , tout ne marche pas aussi bien que pour la convergence p.s. On a la stabilité par combinaison linéaire qui découle directement du fait que l'on considère une convergence par rapport à une norme : si $(X_n)_{n \geq 1}$ et $(Y_n)_{n \geq 1}$ sont des suites de v.a. réelles convergeant dans L^p vers X et Y respectivement et si $a, b \in \mathbb{R}$, alors

$$aX_n + bY_n \xrightarrow[n \rightarrow \infty]{L^p} aX + bY,$$

Mais pour le produit c'est plus compliqué : les v.a. $X_n Y_n$ n'appartiennent pas forcément à L^p (et même si c'est le cas elles ne convergent pas forcément dans L^p). De même, il n'y a pas non plus forcément stabilité par composition par une fonction continue f générale : en effet, $f(X_n)$ n'est pas forcément dans L^p (par exemple avec $f(x) = x^2$). Cependant, certains résultats de stabilité pour le produit et la composition sont vrais avec des hypothèses appropriées (voir l'exercice 1 du TD8).

Remarque 4.3.7. Rappelons que L^p muni de $\|\cdot\|_p$ est *complet* (ce qui en fait un espace de Banach). Cela signifie que si une suite $(X_n)_{n \geq 1}$ de v.a. dans L^p est *de Cauchy*, i.e.

$$\forall \varepsilon > 0, \quad \exists n_0 \geq 1, \quad \forall n > m \geq n_0, \quad \|X_n - X_m\|_p \leq \varepsilon,$$

alors il existe $X \in L^p$ telle que $(X_n)_{n \geq 1}$ converge dans L^p vers X . Utiliser la complétude est utile pour montrer qu'une suite converge sans en connaître la limite. C'est par exemple souvent le cas pour des séries, voir ci-dessous.

Exemple 4.3.8. Soit $(Y_k)_{k \geq 1}$ une suite de v.a. indépendantes dans L^2 telles que

$$\mathbb{E}[Y_k] = 0 \quad \text{et} \quad \sum_{k \geq 1} \mathbb{E}[Y_k^2] < \infty.$$

On va montrer que la série $\sum_{k \geq 1} Y_k$ converge dans L^2 , c'est-à-dire que la suite des sommes partielles

$$X_n := \sum_{k=1}^n Y_k$$

converge dans L^2 . Pour cela, on va vérifier que $(X_n)_{n \geq 1}$ est de Cauchy dans L^2 . Soit $n > m \geq n_0 \geq 1$. On a

$$\|X_n - X_m\|_2^2 = \mathbb{E}\left[\left(\sum_{k=m+1}^n Y_k \right)^2 \right] = \text{Var}\left(\sum_{k=m+1}^n Y_k \right) = \sum_{k=m+1}^n \text{Var}(Y_k) = \sum_{k=m+1}^n \mathbb{E}[Y_k^2],$$

où l'on utilise que $\mathbb{E}[Y_k] = 0$ (pour passer du moment d'ordre 2 à la variance et vice versa), et l'indépendance des Y_k à la 3ème égalité (pour appliquer la Proposition 3.1.4). Soit $\varepsilon > 0$. Comme la série $\sum_{k \geq 1} \mathbb{E}[Y_k^2]$ est convergente, son reste tend vers 0 donc il existe $n_0 \geq 1$ tel que

$$\sum_{k \geq n_0} \mathbb{E}[Y_k^2] \leq \varepsilon.$$

Alors, par le calcul précédent, on voit que pour tout $n > m \geq n_0$, on a $\|X_n - X_m\|_2^2 \leq \varepsilon$. Cela montre que $(X_n)_{n \geq 1}$ est de Cauchy dans L^2 et donc convergente par complétude.

4.3.3 Convergence en probabilité

La nouvelle notion de convergence vue dans ce chapitre est la suivante.

Définition 4.3.9. Soit X, X_1, X_2, \dots des v.a. réelles. On dit que $(X_n)_{n \geq 1}$ converge en probabilité vers X , que l'on note

$$X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X,$$

si, pour tout $\varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0$ quand $n \rightarrow \infty$.

Exemple 4.3.10 (Loi faible des grands nombres). Soit $(X_k)_{k \geq 1}$ une suite de v.a. réelles i.i.d. telles que $\mathbb{E}[X_1^2] < \infty$. Alors

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mathbb{E}[X_1].$$

En effet, pour tout $\varepsilon > 0$, par l'inégalité de Bienaymé–Chebychev, on a

$$\mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n} - \mathbb{E}[X_1]\right| \geq \varepsilon\right) \leq \frac{\text{Var}(X_1)}{n\varepsilon^2} \xrightarrow[n \rightarrow \infty]{} 0,$$

ce qui montre la convergence en probabilité énoncée ci-dessus. C'est ce qu'on appelle la loi faible des grands nombres, en opposition à la loi forte que l'on verra plus loin dans ce chapitre. La loi forte a une hypothèse plus faible et une conclusion plus forte, c'est donc un résultat strictement meilleur, mais sa preuve est aussi (beaucoup) plus compliquée.

Proposition 4.3.11. Une limite en probabilité est unique à égalité p.s. près. Autrement dit, si une suite de v.a. réelles $(X_n)_{n \geq 1}$ converge en probabilité à la fois vers X et vers Y , alors $X = Y$ p.s.

Démonstration. Soit $\varepsilon > 0$. Si $|X - Y| > \varepsilon$, alors on a nécessairement $|X_n - X| > \varepsilon/2$ ou $|X_n - Y| > \varepsilon/2$ (par inégalité triangulaire). Donc, on a

$$\mathbb{P}(|X - Y| > \varepsilon) \leq \mathbb{P}(|X_n - X| > \varepsilon/2) + \mathbb{P}(|X_n - Y| > \varepsilon/2) \xrightarrow[n \rightarrow \infty]{} 0,$$

car $(X_n)_{n \geq 1}$ converge en probabilité vers X et vers Y . On en déduit que $\mathbb{P}(|X - Y| > \varepsilon) = 0$. On a alors

$$\mathbb{P}(X \neq Y) = \mathbb{P}\left(\bigcup_{n \geq 1} \uparrow \left\{ |X - Y| > \frac{1}{n} \right\}\right) = \lim_{n \rightarrow \infty} \uparrow \mathbb{P}\left(|X - Y| > \frac{1}{n}\right) = 0,$$

ce qui montre que $X = Y$ p.s. □

La convergence en probabilité satisfait les mêmes propriétés de stabilité par transformation que la convergence p.s., comme énoncé dans le résultat suivant.

Proposition 4.3.12. Soit $(X_n)_{n \geq 1}$ et $(Y_n)_{n \geq 1}$ des suites de v.a. réelles convergeant en probabilité vers X et Y respectivement. Alors, on a les convergences suivantes :

$$\bullet \quad X_n + Y_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X + Y, \quad X_n - Y_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X - Y \quad \text{et} \quad X_n Y_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} XY;$$

- Si $\mathbb{P}(Y = 0) = 0$, alors $X_n/Y_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X/Y$;
- Si $f: \mathbb{R} \rightarrow \mathbb{R}$ est continue, alors $f(X_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} f(X)$.

Démonstration. Voir le TD9 pour une démonstration à partir de la définition de la convergence en probabilité. Une autre démonstration (plus courte) sera donnée plus loin en s'appuyant sur une caractérisation de la convergence en probabilité. \square

4.3.4 Liens entre les notions

La convergence en probabilité est la plus faible des trois notions de convergence introduites plus haut : comme montré dans les deux prochaines propositions, la convergence en probabilité est impliqué par la convergence p.s. ou par la convergence dans L^p .

Proposition 4.3.13. Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles convergeant p.s. vers une v.a. réelle X . Alors $(X_n)_{n \geq 1}$ converge en probabilité vers X .

Démonstration. Soit $\varepsilon > 0$. On a

$$\left\{ \lim_{n \rightarrow \infty} X_n = X \right\} \subset \bigcup_{m \geq 1} \bigcap_{n \geq m} \{|X_n - X| \leq \varepsilon\} = \liminf_{n \rightarrow \infty} \{|X_n - X| \leq \varepsilon\}.$$

Comme $(X_n)_{n \geq 1}$ converge p.s. vers X , l'événement de gauche a probabilité 1 et donc celui de droite aussi. Donc son complémentaire a probabilité nulle :

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} \{|X_n - X| > \varepsilon\}\right) = \mathbb{P}\left(\bigcap_{m \geq 1} \bigcup_{n \geq m} \{|X_n - X| > \varepsilon\}\right) = 0.$$

On a vu en Section 4.2.2 que $\limsup_{n \rightarrow \infty} \mathbb{P}(A_n) \leq \mathbb{P}(\limsup_{n \rightarrow \infty} A_n)$ pour toute suite d'événements $(A_n)_{n \geq 1}$. Donc

$$\limsup_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

Comme $\mathbb{P}(|X_n - X| > \varepsilon) \geq 0$, cela montre que $\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0$ quand $n \rightarrow \infty$. Donc $(X_n)_{n \geq 1}$ converge en probabilité vers X . \square

Proposition 4.3.14. Soit $p \geq 1$. Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles convergeant dans L^p vers une v.a. réelle X . Alors $(X_n)_{n \geq 1}$ converge en probabilité vers X .

Démonstration. Par la Proposition 4.3.6, on sait que $(X_n)_{n \geq 1}$ converge dans L^1 vers X . Donc $\mathbb{E}[|X_n - X|] \rightarrow 0$ quand $n \rightarrow \infty$. Soit $\varepsilon > 0$. Par l'inégalité de Markov, on a

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

Donc $(X_n)_{n \geq 1}$ converge en probabilité vers X . \square

On peut en déduire qu'une suite de v.a. réelles ne peut pas avoir deux limites différentes pour deux notions de convergence (à égalité p.s. près).

Corollaire 4.3.15. Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles convergeant vers une v.a. réelle X pour la convergence p.s. ou L^p ou en probabilité. Supposons que $(X_n)_{n \geq 1}$ converge aussi vers une v.a. réelle Y pour la convergence p.s. ou L^p ou en probabilité (pas forcément pour la même notion). Alors $X = Y$ p.s.

Démonstration. Par les deux propositions précédentes, on sait que $(X_n)_{n \geq 1}$ converge en probabilité vers X et vers Y . Donc par la Proposition 4.3.11, on a $X = Y$ p.s. \square

On va montrer dans les exemples suivants qu'il n'y a pas d'autre implication vraie en général entre ces notions de convergence.

Exemple 4.3.16. Soit U une v.a. uniforme dans $[0, 1]$. Comme $U \in [0, 1[$ p.s., on a

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

Mais, d'autre part,

$$\mathbb{E}[|nU^n - 0|] = \mathbb{E}[nU^n] = \int_0^1 nx^n dx = \frac{n}{n+1} \xrightarrow[n \rightarrow \infty]{} 1.$$

Donc $(nU^n)_{n \geq 1}$ ne converge dans L^1 pas vers 0 et donc pas non plus dans L^p pour $p > 1$. Cela montre que la convergence p.s. n'implique pas la convergence L^1 (ni L^p). En particulier, la convergence en probabilité n'implique pas la convergence L^1 (car on a vu que la convergence p.s. est plus forte que la convergence en probabilité).

Exemple 4.3.17. On considère $\Omega = [0, 1]$ muni de $\mathcal{A} = \mathcal{B}([0, 1])$ et de \mathbb{P} la mesure de Lebesgue. Soit $n \geq 1$. Soit k l'unique entier tel que $2^k \leq n < 2^{k+1}$. On définit X_n comme l'indicatrice de l'intervalle

$$I_n = \left[\frac{n - 2^k}{2^k}, \frac{n - 2^k + 1}{2^k} \right].$$

On a $X_1 = \mathbb{1}_{[0,1]}$, $X_2 = \mathbb{1}_{[0,1/2]}$, $X_3 = \mathbb{1}_{[1/2,1]}$, $X_4 = \mathbb{1}_{[0,1/4]}$, $X_5 = \mathbb{1}_{[1/4,1/2]}$, et ainsi de suite. Notons que la longueur de I_n est $2^{-k} < 2/n$, avec l'entier k introduit ci-dessus. Donc, pour $p \geq 1$, on a

$$\mathbb{E}[|X_n - 0|^p] = \mathbb{E}[\mathbb{1}_{I_n}] = \mathbb{P}(I_n) < \frac{2}{n} \xrightarrow[n \rightarrow \infty]{} 0.$$

Cela montre que $(X_n)_{n \geq 1}$ converge dans L^p vers 0, pour tout $p \geq 1$. D'autre part, pour $\omega \in [0, 1]$ fixé, on a à la fois $\omega \in I_n$ pour une infinité de n et aussi $\omega \notin I_n$ pour une infinité de n . Donc

$$\liminf_{n \rightarrow \infty} X_n(\omega) = 0 \quad \text{et} \quad \limsup_{n \rightarrow \infty} X_n(\omega) = 1.$$

Cela montre que $(X_n(\omega))_{n \geq 1}$ ne converge pour aucun $\omega \in \Omega$. En particulier, $(X_n)_{n \geq 1}$ ne converge pas pour la convergence p.s. On a donc montré que, pour tout $p \in [1, \infty[$, la convergence L^p n'implique pas la convergence p.s.

4.4 Lemmes de Borel–Cantelli et applications

4.4.1 Lemmes de Borel–Cantelli

Les lemmes de Borel–Cantelli fournissent des critères pour montrer que la limite supérieure d'une suite d'événements a probabilité 0 (premier lemme) ou 1 (deuxième lemme).

Lemme 4.4.1 (Lemme de Borel–Cantelli). Soit $(A_n)_{n \geq 1}$ une suite d'événements. Si

$$\sum_{n \geq 1} \mathbb{P}(A_n) < \infty,$$

alors

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = 0.$$

Rappelons qu'on a vu en Section 4.2.2 que $\limsup_{n \rightarrow \infty} \mathbb{P}(A_n) \leq \mathbb{P}(\limsup_{n \rightarrow \infty} A_n)$, donc en particulier $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0$ implique que $\mathbb{P}(A_n) \rightarrow 0$ quand $n \rightarrow \infty$. Le lemme de Borel–Cantelli fournit un contrôle dans l'autre sens : si $\mathbb{P}(A_n)$ tend suffisamment vite vers 0 pour être sommable, alors $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0$.

Démonstration. Par définition de la limsup ensembliste, on a

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = \mathbb{P}\left(\bigcap_{m \geq 1} \bigcup_{n \geq m} A_n\right) = \lim_{m \rightarrow \infty} \mathbb{P}\left(\bigcup_{n \geq m} A_n\right) \leq \lim_{m \rightarrow \infty} \sum_{n \geq m} \mathbb{P}(A_n) = 0,$$

car comme la série $\sum_{n \geq 1} \mathbb{P}(A_n)$ est convergente, son reste tend vers 0. \square

Exemple 4.4.2 (Retours en 0 de la marche aléatoire simple asymétrique). Soit $(X_n)_{n \geq 1}$ une suite de v.a. indépendantes avec la loi suivante :

$$\mathbb{P}(X_n = 1) = p \quad \text{et} \quad \mathbb{P}(X_n = -1) = 1 - p,$$

pour $p \in [0, 1]$. On pose $S_0 := 0$ et, pour $n \geq 1$,

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

La suite de v.a. $(S_n)_{n \geq 0}$ est appelée marche aléatoire simple. On suppose que $p \neq 1/2$ (c'est le cas asymétrique). On va montrer que, presque sûrement, la marche $(S_n)_{n \geq 0}$ ne passe qu'un nombre fini de fois en 0.

Pour $n \geq 0$, on pose $A_n := \{S_n = 0\}$, de sorte que $\limsup_{n \rightarrow \infty} A_n$ soit l'événement " $(S_n)_{n \geq 0}$ passe une infinité de fois en 0". D'après le lemme de Borel-Cantelli, il suffit de montrer que la série $\sum_{n \geq 0} \mathbb{P}(A_n)$ converge pour obtenir le résultat voulu. Calculons donc $\mathbb{P}(A_n)$:

- *Cas n impair.* On montre facilement par récurrence que S_n a la même parité que n . Donc, si n est impair, alors S_n est impair et donc ne peut pas être nul. On a donc $\mathbb{P}(A_n) = 0$.
- *Cas n pair.* Si n est pair, on l'écrit $n = 2m$ avec $m \in \mathbb{N}$. Soit N le nombre de X_k égaux à 1 pour $k \in [1, 2m]$. Alors on a $S_{2m} = 0$ si et seulement si $N = m$, et donc

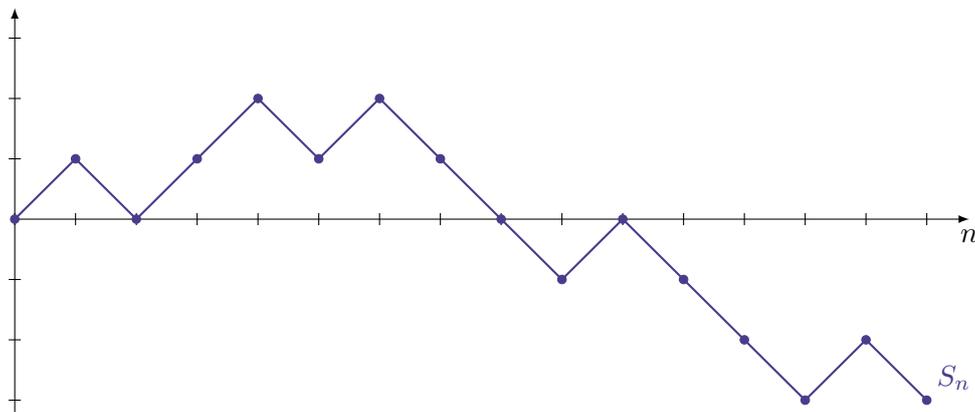
$$\mathbb{P}(A_n) = \mathbb{P}(S_{2m} = 0) = \mathbb{P}(N = m) = \binom{2m}{m} p^m (1-p)^m,$$

en remarquant que N suit la loi binomiale de paramètre $(2m, p)$.

Par ce qui précède, on obtient donc

$$\sum_{n \geq 0} \mathbb{P}(A_n) = \sum_{m \geq 0} \mathbb{P}(A_{2m}) < \infty,$$

car $\mathbb{P}(A_{2m+2})/\mathbb{P}(A_{2m}) = \frac{(2m+1)(2m+2)}{(m+1)^2} p(1-p) \rightarrow 4p(1-p)$ quand $m \rightarrow \infty$ et $4p(1-p) < 1$ puisque $p \neq 1/2$.



On peut se demander si la condition $\sum_{n \geq 1} \mathbb{P}(A_n) < \infty$ est une condition nécessaire pour avoir $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0$. En général, la réponse est non, mais le lemme suivant montre que c'est le cas pour une suite d'événements indépendants.

Lemme 4.4.3 (Second lemme de Borel–Cantelli). *Soit $(A_n)_{n \geq 1}$ une suite d'événements. Si*

$$\sum_{n \geq 1} \mathbb{P}(A_n) = \infty$$

et si les événements de la suite $(A_n)_{n \geq 1}$ sont indépendants, alors

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = 1.$$

Démonstration. Par définition de la limsup ensembliste, on a

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = \mathbb{P}\left(\bigcap_{m \geq 1} \downarrow \bigcup_{n \geq m} A_n\right) = \lim_{m \rightarrow \infty} \downarrow \mathbb{P}\left(\bigcup_{n \geq m} A_n\right) = 1 - \lim_{m \rightarrow \infty} \uparrow \mathbb{P}\left(\bigcap_{n \geq m} A_n^c\right).$$

Mais, pour tout $m \geq 1$, on a

$$\mathbb{P}\left(\bigcap_{n \geq m} A_n^c\right) = \mathbb{P}\left(\bigcap_{N \geq m} \downarrow \bigcap_{n=m}^N A_n^c\right) = \lim_{N \rightarrow \infty} \downarrow \mathbb{P}\left(\bigcap_{n=m}^N A_n^c\right)$$

Pour $N \geq m$, en utilisant que A_m^c, \dots, A_N^c sont indépendants, on obtient

$$\mathbb{P}\left(\bigcap_{n=m}^N A_n^c\right) = \prod_{n=m}^N \mathbb{P}(A_n^c) = \prod_{n=m}^N (1 - \mathbb{P}(A_n)) \leq \exp\left(-\sum_{n=m}^N \mathbb{P}(A_n)\right)$$

où l'on a utilisé que $1 - x \leq e^{-x}$ pour tout $x \in \mathbb{R}$. Mais comme $\sum_{n \geq 1} \mathbb{P}(A_n) = \infty$, on a $\sum_{n=m}^N \mathbb{P}(A_n) \rightarrow \infty$ quand $N \rightarrow \infty$ et donc

$$\mathbb{P}\left(\bigcap_{n \geq m} A_n^c\right) \leq \lim_{N \rightarrow \infty} \exp\left(-\sum_{n=m}^N \mathbb{P}(A_n)\right) = 0.$$

En revenant à la première équation de la démonstration, cela montre le résultat désiré. \square

Exemple 4.4.4. Soit $(X_n)_{n \geq 1}$ une suite de v.a. indépendantes de loi de Bernoulli de paramètre $p \in]0, 1]$. On veut montrer que, presque sûrement, elle contient une infinité de fois quatre 1 d'affilée. Pour cela, on considère les événements

$$A_k = \{X_{4k+1} = X_{4k+2} = X_{4k+3} = X_{4k+4} = 1\}$$

pour $k \in \mathbb{N}$. Par regroupement par paquets, la suite des v.a. $(X_{4k+1}, X_{4k+2}, X_{4k+3}, X_{4k+4})$ pour $k \in \mathbb{N}$ est indépendante et donc les événements A_k pour $k \in \mathbb{N}$ le sont aussi. De plus, $\mathbb{P}(A_k) = p^4 > 0$ pour tout $k \in \mathbb{N}$, donc

$$\sum_{k \geq 0} \mathbb{P}(A_k) = \infty.$$

Par le 2nd lemme de Borel–Cantelli, on en déduit que

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = 1,$$

ce qui signifie que, p.s., pour une infinité de k , on a $X_{4k+1} = X_{4k+2} = X_{4k+3} = X_{4k+4} = 1$.

Notons que les événements A_k ne comptabilisent pas toutes les séquences de quatre 1 d'affilée mais seulement celles commençant à un indice de la forme $4k + 1$. On aurait pu être tenté de considérer les événements

$$B_k = \{X_{k+1} = X_{k+2} = X_{k+3} = X_{k+4} = 1\},$$

qui comptabilisent toutes les séquences de quatre 1 d'affilée. Mais ces événements ne sont pas indépendants : par exemple B_0 et B_1 dépendent tous les deux du résultat de X_2 .

4.4.2 Application à la convergence p.s.

On utilise souvent le premier lemme de Borel–Cantelli pour montrer des convergences p.s., par exemple à travers le critère suivant.

Proposition 4.4.5. *Soit X, X_1, X_2, \dots des v.a. réelles. Supposons que, pour tout $\varepsilon > 0$, on ait*

$$\sum_{n \geq 1} \mathbb{P}(|X_n - X| > \varepsilon) < \infty.$$

Alors $(X_n)_{n \geq 1}$ converge p.s. vers X .

Il est intéressant de comparer ce critère à la définition de la convergence en probabilité. Pour montrer que $(X_n)_{n \geq 1}$ converge en probabilité vers X , il suffit de vérifier que $\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0$ pour tout $\varepsilon > 0$. Pour montrer une convergence p.s., ce critère nous dit qu’il suffit de vérifier que $\mathbb{P}(|X_n - X| > \varepsilon)$ tend suffisamment vite vers 0 pour être sommable.

Démonstration. Soit $\varepsilon > 0$. Comme $\sum_{n \geq 1} \mathbb{P}(|X_n - X| > \varepsilon) < \infty$, par le lemme de Borel–Cantelli, on a

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} \{|X_n - X| > \varepsilon\}\right) = 0$$

et donc p.s. il n’y a qu’un nombre fini de n tel que $|X_n - X| > \varepsilon$. Autrement dit, en prenant $\varepsilon = 1/k$ avec $k \in \mathbb{N}^*$, on a montré que

$$\forall k \in \mathbb{N}^*, \text{ p.s., pour tout } n \text{ à partir d'un certain rang, } |X_n - X| \leq 1/k.$$

Mais on peut échanger le “pour tout $k \in \mathbb{N}^*$ ” avec le “p.s.” (car \mathbb{N}^* est dénombrable, voir la remarque ci-dessous), donc on a

$$\text{p.s., } \forall k \in \mathbb{N}^*, \text{ pour tout } n \text{ à partir d'un certain rang, } |X_n - X| \leq 1/k.$$

Mais cela signifie exactement que, p.s., X_n converge vers X . □

Remarque 4.4.6. On a utilisé dans la démonstration le fait important suivant : dans une proposition, on peut échanger un “p.s.” avec un “pour tout”, si ce “pour tout” porte sur un ensemble au plus dénombrable : si $(A_k)_{k \geq 1}$ est une suite d’événements alors

$$\forall k \geq 1, \text{ p.s., } A_k \text{ a lieu} \iff \text{p.s., } \forall k \geq 1, A_k \text{ a lieu.}$$

En d’autres termes, si $\mathbb{P}(A_k) = 1$ pour tout $k \geq 1$, alors on a $\mathbb{P}(\bigcap_{k \geq 1} A_k) = 1$. En effet, on peut vérifier que

$$\mathbb{P}\left(\left(\bigcap_{k \geq 1} A_k\right)^c\right) = \mathbb{P}\left(\bigcup_{k \geq 1} A_k^c\right) \leq \sum_{k \geq 1} \mathbb{P}(A_k^c) = 0.$$

Exemple 4.4.7 (Loi forte des grands nombres pour des v.a. bornées). Soit X_1, \dots, X_n des v.a. réelles i.i.d. à valeurs dans un intervalle borné $[a, b]$. Montrons que

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbb{E}[X_1].$$

Pour cela, on vérifie le critère de la proposition précédente : pour $\varepsilon > 0$, on a, par l’inégalité de Hoeffding,

$$\mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n} - \mathbb{E}[X_1]\right| > \varepsilon\right) \leq 2 \exp\left(-\frac{2\varepsilon^2 n}{(b-a)^2}\right),$$

qui est bien sommable en n . Notons que, pour des v.a. dans L^2 , l’inégalité de Bienaymé–Chebychev donne une borne qui décroît en $1/n$ et qui n’est donc pas sommable. On ne peut donc pas montrer simplement qu’il y a convergence p.s. dans ce cas, mais c’est vrai comme nous le verrons dans en Section 4.5.

Le second lemme de Borel–Cantelli peut (parfois) permettre de montrer qu’une suite $(X_n)_{n \geq 1}$ de v.a. réelles n’est pas p.s. convergente, voire qu’elle est p.s. pas convergente¹. Une stratégie pour cela consiste à montrer que les limites inférieures et supérieures de la suite ne sont pas p.s. égales, voire sont p.s. différentes.

Exemple 4.4.8. Soit $(Y_n)_{n \geq 1}$ une suite de v.a. indépendantes de loi exponentielle de paramètre 1. Pour $n \geq 2$, on pose $X_n = Y_n / \log n$. On va montrer que, pour tout $p \in [1, \infty[$,

$$X_n \xrightarrow[n \rightarrow \infty]{L^p} 0,$$

mais qu’il n’y a pas convergence p.s. Cela fournit un exemple plus probabiliste que celui vu à l’Exemple 4.3.17, du fait que la convergence L^p n’implique pas la convergence p.s.

Soit $p \in [1, \infty[$. Comme Y_n et Y_1 ont la même loi, on a

$$\mathbb{E}[|X_n - 0|^p] = \frac{\mathbb{E}[|Y_n|^p]}{(\log n)^p} = \frac{\mathbb{E}[|Y_1|^p]}{(\log n)^p} \xrightarrow[n \rightarrow \infty]{} 0,$$

car $\mathbb{E}[|Y_1|^p] = \int_0^\infty x^p e^{-x} dx < \infty$. Cela montre la convergence dans L^p (et donc aussi en probabilité).

À présent on va montrer qu’il n’y a pas convergence p.s. en vérifiant que

$$\limsup_{n \rightarrow \infty} X_n \geq 1 \text{ p.s.} \quad \text{et} \quad \liminf_{n \rightarrow \infty} X_n \leq 0 \text{ p.s.}$$

Commençons par étudier la limite supérieure. Comme Y_n suit une loi exponentielle de paramètre 1, on a, pour tout $a \geq 0$, $\mathbb{P}(Y_n \geq a) = \int_a^\infty e^{-x} dx = e^{-a}$ donc

$$\mathbb{P}(X_n \geq 1) = \mathbb{P}(Y_n \geq \log n) = e^{-\log n} = \frac{1}{n}.$$

Donc $\sum_{n \geq 1} \mathbb{P}(X_n \geq 1) = \infty$ et, comme les événements $\{X_n \geq 1\}$ pour $n \geq 1$ sont indépendants, le second lemme de Borel–Cantelli nous dit que

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} \{X_n \geq 1\}\right) = 1.$$

Autrement dit : p.s., il existe une infinité de n tels que $X_n \geq 1$. Mais si pour un certain $\omega \in \Omega$, la suite $X_n(\omega)$ prend une infinité de valeurs supérieure à 1, alors

$$\limsup_{n \rightarrow \infty} X_n(\omega) = \lim_{m \rightarrow \infty} \sup_{n \geq m} X_n(\omega) \geq 1.$$

Ainsi, on a montré que

$$\limsup_{n \rightarrow \infty} X_n \geq 1 \text{ p.s.}$$

Étudions à présent la limite inférieure. On ne peut pas montrer que p.s. il existe une infinité de n tels que $X_n \leq 0$, car $\mathbb{P}(X_n \leq 0) = \mathbb{P}(Y_n \leq 0) = 0$. On considère donc d’abord $\varepsilon > 0$ et on va montrer que la limite inférieure de X_n est plus petite que ε p.s. Pour cela, on remarque que

$$\mathbb{P}(X_n \leq \varepsilon) = 1 - \mathbb{P}(Y_n > \varepsilon \log n) = 1 - e^{-\varepsilon \log n} = 1 - n^{-\varepsilon} \xrightarrow[n \rightarrow \infty]{} 1.$$

Donc $\sum_{n \geq 1} \mathbb{P}(X_n \leq \varepsilon) = \infty$ et, comme les événements $\{X_n \leq \varepsilon\}$ sont indépendants, le second lemme de Borel–Cantelli nous donne

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} \{X_n \leq \varepsilon\}\right) = 1.$$

1. La négation d’un presque sûr peut toujours être ambigu. Ici on dit que $(X_n)_{n \geq 1}$ n’est pas p.s. convergente si $\mathbb{P}(\lim_{n \rightarrow \infty} X_n \text{ existe}) < 1$. On dit qu’elle est p.s. pas convergente si $\mathbb{P}(\lim_{n \rightarrow \infty} X_n \text{ existe}) = 0$.

Donc p.s., il existe une infinité de n tels que $X_n \leq \varepsilon$, ce qui implique que la limite inférieure est plus petite que ε . Ainsi on a montré que

$$\forall \varepsilon > 0, \text{ p.s.}, \liminf_{n \rightarrow \infty} X_n \leq \varepsilon.$$

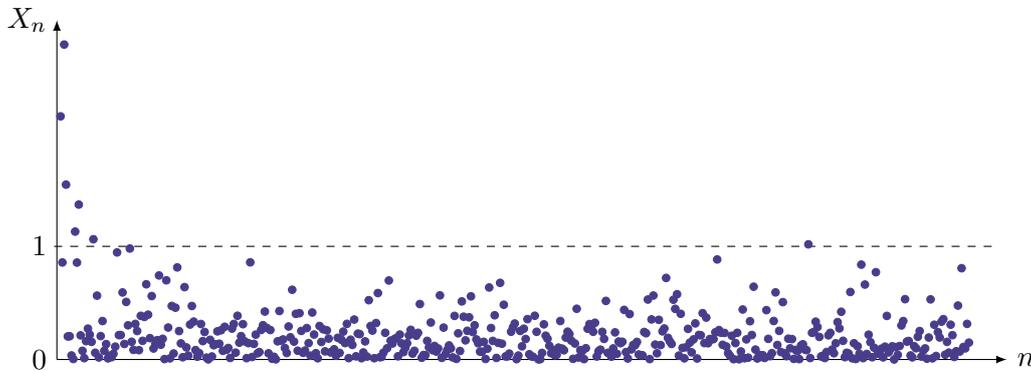
On a envie d'échanger le " $\forall \varepsilon > 0$ " avec le "p.s." pour pouvoir conclure. Pour cela on le transforme en une "pour tout" dénombrable : on a $\forall k \in \mathbb{N}^*$, p.s., $\liminf_{n \rightarrow \infty} X_n \leq 1/k$. On peut alors échanger (voir Remarque 4.4.6) : on a p.s., $\forall k \in \mathbb{N}^*$, $\liminf_{n \rightarrow \infty} X_n \leq 1/k$. Mais, pour $\omega \in \Omega$ fixé, on a

$$\forall k \in \mathbb{N}^*, \liminf_{n \rightarrow \infty} X_n(\omega) \leq \frac{1}{k} \Rightarrow \liminf_{n \rightarrow \infty} X_n(\omega) \leq 0.$$

On a donc montré que

$$\liminf_{n \rightarrow \infty} X_n \leq 0 \text{ p.s.}$$

Ainsi, p.s. les limites inférieure et supérieure de $(X_n)_{n \geq 2}$ sont différentes, on en conclut donc que p.s. $(X_n)_{n \geq 2}$ ne converge pas.



Notons qu'on peut transformer les inégalités sur les limites supérieures et inférieures en égalités et montrer que

$$\limsup_{n \rightarrow \infty} X_n = 1 \text{ p.s.} \quad \text{et} \quad \liminf_{n \rightarrow \infty} X_n = 0 \text{ p.s.}$$

Pour la limite inférieure, c'est facile, on sait que pour tout $n \geq 2$, p.s. $X_n \geq 0$ donc p.s., pour tout $n \geq 2$, $X_n \geq 0$ et donc p.s. $\liminf_{n \rightarrow \infty} X_n \geq 0$. Pour la limite supérieure, cela nécessite un peu plus de travail. On fixe $\varepsilon > 0$ et on remarque que $\mathbb{P}(X_n \geq 1 + \varepsilon) = n^{-1-\varepsilon}$ est sommable en n . Par le 1er lemme de Borel–Cantelli, on en déduit que p.s., il n'y a qu'un nombre fini de n tels que $X_n \geq 1 + \varepsilon$. Cela implique que

$$\limsup_{n \rightarrow \infty} X_n \leq 1 + \varepsilon \text{ p.s.}$$

En remplaçant ε par $1/k$ pour $k \in \mathbb{N}^*$, puis en échangeant le "pour tout" et le "p.s." on en déduit : p.s., $\forall k \geq 1$, $\limsup_{n \rightarrow \infty} X_n \leq 1 + 1/k$. On en conclut que $\limsup_{n \rightarrow \infty} X_n \leq 1$ p.s.

Commentons intuitivement ce qu'il se passe, pour comprendre la différence entre convergence en probabilité et convergence p.s. La convergence en probabilité vers 0 signifie que pour n grand il est de plus en plus probable que X_n soit proche de 0 : une grande majorité des réalisations ω sont telles que $|X_n(\omega)| \leq \varepsilon$. Mais ce ne sont pas forcément les mêmes ω pour X_n et pour X_{n+1} puisqu'elles sont indépendantes. Au contraire, la convergence p.s. vers 0 signifierait que, si l'on fixe ω dans un certain événement de probabilité 1, on a convergence de $X_n(\omega)$ vers 0, donc $|X_n(\omega)| \leq \varepsilon$ pour tous les n à partir d'un certain rang. Ici ce n'est pas le cas : pour un ω fixé, même si la grande majorité des $X_n(\omega)$ sont proches de 0, il y en a quand même une infinité qui sont plus grand que 1 (pour des n de plus en plus espacés, mais pour une infinité de n quand même). Voir la figure pour une illustration.

Remarque 4.4.9. Il est important de ne pas mélanger limites supérieures de suites et ensemblistes. L'événement $\limsup_{n \rightarrow \infty} \{X_n \geq a\}$ n'est pas égal à $\{\limsup_{n \rightarrow \infty} X_n \geq a\}$ (mais pas loin) et l'événement $\limsup_{n \rightarrow \infty} \{X_n \leq a\}$ n'est pas égal à $\{\limsup_{n \rightarrow \infty} X_n \leq a\}$ (alors là pas du tout). De même avec les limites inférieures. Voir l'exercice 4 du TD8 pour les détails.

4.4.3 Une caractérisation de la convergence en probabilité (bonus)

Cette section ne sera pas traitée en cours et n'est donc pas à connaître pour les examens. On y présente la caractérisation suivante de la convergence en probabilité en terme de convergence p.s. ainsi qu'une application.

Rappelons que pour une suite $(x_n)_{n \geq 1}$ dans un espace quelconque, une sous-suite est une suite de la forme $(x_{\varphi(n)})_{n \geq 1}$, où $\varphi: \mathbb{N}^* \rightarrow \mathbb{N}^*$ est une extractrice, i.e. une fonction strictement croissante. Si $(x_n)_{n \geq 1}$ converge vers x pour quelque notion que ce soit, alors toutes ses sous-suites convergent vers x .

Proposition 4.4.10. *Soit X, X_1, X_2, \dots des v.a. réelles. La suite $(X_n)_{n \geq 1}$ converge en probabilité vers X si et seulement si, de toute sous-suite de $(X_n)_{n \geq 1}$, on peut extraire une sous-sous-suite qui converge p.s. vers X .*

On va clarifier la preuve en présentant tout d'abord deux lemmes.

Lemme 4.4.11. *Soit X, X_1, X_2, \dots des v.a. réelles. Si la suite $(X_n)_{n \geq 1}$ converge en probabilité vers X , alors on peut en extraire une sous-suite qui converge p.s. vers X .*

Démonstration. Voir TD9. □

Lemme 4.4.12. *Soit $x, x_1, x_2, \dots \in \mathbb{R}$. La suite $(x_n)_{n \geq 1}$ converge vers x si et seulement si, de toute sous-suite de $(x_n)_{n \geq 1}$, on peut extraire une sous-sous-suite qui converge vers x .*

Démonstration. (\Rightarrow) Cela découle du fait que si $(x_n)_{n \geq 1}$ converge vers x alors toute sous-suite de $(x_n)_{n \geq 1}$ converge vers x .

(\Leftarrow) On va montrer la contraposée : supposons que $(x_n)_{n \geq 1}$ ne converge pas vers x . Alors on a forcément $\limsup_{n \rightarrow \infty} x_n \neq x$ ou $\liminf_{n \rightarrow \infty} x_n \neq x$. Supposons qu'on soit dans le 1er cas (l'autre étant identique). La limite supérieure étant une valeur d'adhérence de $(x_n)_{n \geq 1}$, il existe une extractrice $\varphi: \mathbb{N}^* \rightarrow \mathbb{N}^*$ telle que $(x_{\varphi(n)})_{n \geq 1}$ converge vers $\limsup_{n \rightarrow \infty} x_n$. Alors, toute sous-suite de $(x_{\varphi(n)})_{n \geq 1}$ converge vers $\limsup_{n \rightarrow \infty} x_n \neq x$ donc ne converge pas vers x . On a donc construit une sous-suite de $(x_n)_{n \geq 1}$ dont on ne peut pas extraire de sous-sous-suite qui converge vers x . La contraposée est donc montrée. □

Démonstration de la Proposition 4.4.10. (\Rightarrow) Supposons que $(X_n)_{n \geq 1}$ converge en probabilité vers X . Alors, toute sous-suite de $(X_n)_{n \geq 1}$ converge aussi en probabilité vers X , donc, par le Lemme 4.4.11, on peut en extraire une sous-sous-suite qui converge p.s. vers X .

(\Leftarrow) Supposons que de toute sous-suite de $(X_n)_{n \geq 1}$, on peut extraire une sous-sous-suite qui converge p.s. vers X . Soit $\varepsilon > 0$. Pour montrer que $\mathbb{P}(|X_n - X| \geq \varepsilon)$ tend vers 0, on va utiliser le Lemme 4.4.12 : il suffit de montrer que de toute sous-suite de $\mathbb{P}(|X_n - X| \geq \varepsilon)$, il existe une sous-sous-suite qui tend vers 0. Considérons donc une sous-suite $\mathbb{P}(|X_{\varphi(n)} - X| \geq \varepsilon)$ avec $\varphi: \mathbb{N}^* \rightarrow \mathbb{N}^*$ une extractrice. Alors $(X_{\varphi(n)})_{n \geq 1}$ est une sous-suite de $(X_n)_{n \geq 1}$ donc il existe une sous-sous-suite $(X_{\varphi(\psi(n))})_{n \geq 1}$ convergeant p.s. vers X . Alors, $(X_{\varphi(\psi(n))})_{n \geq 1}$ converge aussi en probabilité vers X , donc

$$\mathbb{P}(|X_{\varphi(\psi(n))} - X| \geq \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0.$$

On a donc bien montré qu'il existe une sous-suite de $\mathbb{P}(|X_{\varphi(n)} - X| \geq \varepsilon)$ tendant vers 0. Donc $\mathbb{P}(|X_n - X| \geq \varepsilon)$ tend vers 0 et cela montre que $(X_n)_{n \geq 1}$ converge en probabilité vers X . □

Comme application de cette caractérisation, on va montrer la stabilité de la convergence en probabilité par les différentes opérations.

Démonstration de la Proposition 4.3.12. L'idée est de transférer les propriétés de stabilité de la convergence p.s. grâce à la Proposition 4.3.12. On va montrer la stabilité par somme, les autres étant identiques. Pour montrer que

$$X_n + Y_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X + Y,$$

on va montrer que de toute sous-suite de $(X_n + Y_n)_{n \geq 1}$ on peut extraire une sous-sous-suite qui converge p.s. Considérons donc une sous-suite $(X_{\varphi(n)} + Y_{\varphi(n)})_{n \geq 1}$ avec $\varphi: \mathbb{N}^* \rightarrow \mathbb{N}^*$ une extractrice. Comme $(X_n)_{n \geq 1}$ converge en probabilité vers X , de la sous-suite $(X_{\varphi(n)})_{n \geq 1}$ on peut extraire une sous-sous-suite $(X_{\varphi(\psi(n))})_{n \geq 1}$ convergeant p.s. vers X . Puis, comme $(Y_n)_{n \geq 1}$ converge en probabilité vers Y , de la sous-suite $(Y_{\varphi(\psi(n))})_{n \geq 1}$ on peut extraire une sous-sous-suite $(Y_{\varphi(\psi(\xi(n)))})_{n \geq 1}$ convergeant p.s. vers Y . Mais $(X_{\varphi(\psi(\xi(n)))})_{n \geq 1}$ convergeant p.s. vers X (en tant que sous-suite de $(X_{\varphi(\psi(n))})_{n \geq 1}$ qui converge p.s. vers X). Donc, par stabilité par somme de la convergence p.s.,

$$X_{\varphi(\psi(\xi(n)))} + Y_{\varphi(\psi(\xi(n)))} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} X + Y.$$

On a donc montré qu'il existe une sous-suite de $(X_{\varphi(n)} + Y_{\varphi(n)})_{n \geq 1}$ convergeant p.s. vers 0 et cela conclut la preuve. \square

4.5 La loi forte des grands nombres

La loi des grands nombres stipule que la moyenne empirique d'un grand nombre de v.a. i.i.d. se rapproche de leur vraie moyenne. On parle généralement de loi forte quand on obtient une convergence p.s. et de loi faible quand on obtient une convergence en probabilité. Nous avons vu deux versions de loi des grands nombres dans des exemples de ce chapitre :

- À l'Exemple 4.3.10, on a montré une loi faible pour des v.a. dans L^2 .
- À l'Exemple 4.4.7, on a montré une loi forte pour des v.a. bornées (hypothèse très forte).

Voici l'énoncé unique pour les unifier tous. Il a à la fois la conclusion la plus forte (la convergence p.s.) et une hypothèse plus faible que les versions mentionnées plus haut (v.a. dans L^1). C'est donc l'unique résultat à retenir.

Théorème 4.5.1 (Loi forte des grands nombres). *Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles i.i.d. Si $\mathbb{E}[|X_1|] < \infty$, alors*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbb{E}[X_1].$$

On peut également montrer que l'hypothèse $\mathbb{E}[|X_1|] < \infty$ est nécessaire pour avoir une convergence p.s. vers un réel. Plus précisément, si $\mathbb{E}[|X_1|] = \infty$, différents cas sont possibles : soit les moyennes empiriques tendent vers $+\infty$ p.s., soit elles tendent vers $-\infty$ p.s., soit elles oscillent p.s. entre les deux, c'est-à-dire

$$\liminf_{n \rightarrow \infty} \frac{X_1 + \cdots + X_n}{n} = -\infty \quad \text{et} \quad \limsup_{n \rightarrow \infty} \frac{X_1 + \cdots + X_n}{n} = +\infty \quad \text{p.s.}$$

Cependant, il y a des cas où $\mathbb{E}[|X_1|] = \infty$, mais où la moyenne empirique converge en probabilité vers un réel.

4.5.1 Démonstration

La démonstration que l'on présente ici est due à Etemadi en 1981. La première démonstration de la loi forte des grands nombres est due à Kolmogorov en 1930. Deux idées importantes sont présentes dans cette preuve : le fait de tronquer des variables aléatoires pour pouvoir ensuite leur appliquer l'inégalité de Bienaymé–Chebychev et le fait de d'abord montrer la convergence d'une sous-suite pour ensuite comparer les termes restants de la suite à ceux de cette sous-suite.

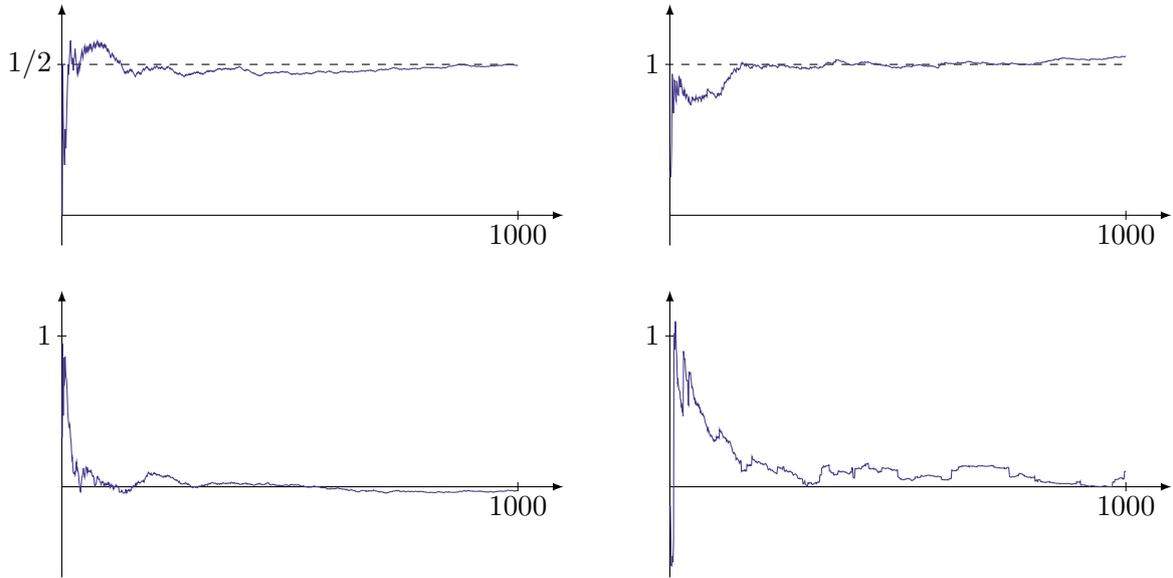


FIGURE 4.1 – Représentation de $(X_1 + \dots + X_n)/n$ pour $n = 1, \dots, 1000$ pour $(X_n)_{n \geq 1}$ une suite de v.a. réelles i.i.d. de différentes lois. En haut à gauche, X_1 a loi $\mathcal{B}(1/2)$. En haut à droite, X_1 a loi $\mathcal{E}(1)$. En bas à gauche, X_1 a loi $\mathcal{N}(0, 1)$. En bas à droite, X_1 a la loi $\frac{3}{4}(1 + |x|)^{-5/2} dx$. Cette dernière loi a une queue lourde (par exemple $X \notin L^2$) et on peut remarquer que la convergence est plus difficile, car même tardivement il y a des grands sauts.

Étape 1 : Se ramener au cas de v.a. positives.

Supposons le résultat montré dans le cas où X_1 est positive. Montrons que l'on peut en déduire le cas où X_1 est de signe quelconque : supposons que $\mathbb{E}[|X_1|] < \infty$ et montrons la loi forte des grands nombres.

Pour cela, on décompose $X_n = X_n^+ - X_n^-$ pour tout $n \geq 1$, où

$$X_n^+ = \max(X_n, 0) \quad \text{et} \quad X_n^- = -\min(X_n, 0).$$

Alors, $(X_n^+)_{n \geq 1}$ forme une suite de v.a. i.i.d. avec $\mathbb{E}[X_1^+] \leq \mathbb{E}[|X_1|] < \infty$ et X_1^+ positive donc, comme on a supposé la loi forte des grands nombres montré pour les v.a. positives, on a

$$\frac{X_1^+ + \dots + X_n^+}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbb{E}[X_1^+].$$

De manière identique, on a la même convergence pour $(X_n^-)_{n \geq 1}$. En combinant les deux, on obtient

$$\frac{X_1 + \dots + X_n}{n} = \frac{X_1^+ + \dots + X_n^+}{n} - \frac{X_1^- + \dots + X_n^-}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbb{E}[X_1^+] - \mathbb{E}[X_1^-] = \mathbb{E}[X_1].$$

ce qui montre la loi forte des grands nombres pour $(X_n)_{n \geq 1}$.

Conclusion. On suppose donc dans la suite de la démonstration que X_1 est une v.a. positive et on cherche à montrer le théorème dans ce cas.

Étape 2 : Argument de troncature pour se ramener à des v.a. dans L^2 .

L'idée ici est de remplacer X_n par $Y_n := X_n \mathbb{1}_{X_n \leq n}$, qui a l'avantage d'avoir un second moment fini, ce qui nous permettra d'utiliser l'inégalité de Bienaymé–Chebychev dans la suite. Notons que les Y_n pour $n \geq 1$ sont indépendantes mais pas identiquement distribuées.

On va montrer ici que

$$\frac{X_1 + \dots + X_n}{n} - \frac{Y_1 + \dots + Y_n}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0. \quad (4.1)$$

Pour cela, on commence par remarquer que

$$\begin{aligned} \sum_{n \geq 1} \mathbb{P}(X_n \neq Y_n) &= \sum_{n \geq 1} \mathbb{P}(X_n > n) = \sum_{n \geq 1} \mathbb{P}(X_1 > n) = \sum_{n \geq 1} \int_{n-1}^n \mathbb{P}(X_1 > n) dx \\ &\leq \sum_{n \geq 1} \int_{n-1}^n \mathbb{P}(X_1 > x) dx = \int_0^\infty \mathbb{P}(X_1 > x) dx = \mathbb{E}[X_1] < \infty, \end{aligned}$$

où la dernière égalité vient de la Proposition 1.3.16. Ainsi, par le 1er lemme de Borel–Cantelli,

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} \{X_n \neq Y_n\}\right) = 0,$$

ce qui signifie que, p.s., il n’y a qu’un nombre fini de n tels que $X_n \neq Y_n$. Fixons un tel ω . Il existe m (qui dépend de ω) tel que, pour tout $n \geq m$, $X_n(\omega) = Y_n(\omega)$. Alors, pour tout $n \geq m$,

$$\frac{X_1(\omega) + \cdots + X_n(\omega)}{n} - \frac{Y_1(\omega) + \cdots + Y_n(\omega)}{n} = \frac{X_1(\omega) + \cdots + X_m(\omega)}{n} - \frac{Y_1(\omega) + \cdots + Y_m(\omega)}{n},$$

qui tend vers 0 quand $n \rightarrow \infty$ (car m est fixé). Cela montre (4.1).

Conclusion. Il suffit donc à présent de montrer que

$$\frac{Y_1 + \cdots + Y_n}{n} \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[X_1], \quad \text{presque sûrement.}$$

Étape 3 : Majorer les variances des Y_n .

En vue d’appliquer l’inégalité de Bienaymé–Chebychev, on va montrer ici la borne suivante sur les variances des Y_n :

$$\sum_{n=1}^{\infty} \frac{\text{Var}(Y_n)}{n^2} \leq 4\mathbb{E}[X_1]. \quad (4.2)$$

Par le Corollaire 1.3.17, on a

$$\mathbb{E}[Y_n^2] = \int_0^\infty 2y\mathbb{P}(Y_n > y) dy \leq \int_0^n 2y\mathbb{P}(X_1 > y) dy.$$

car $Y_n = X_n \mathbb{1}_{X_n \leq n}$ donc $\mathbb{P}(Y_n > y) = 0$ si $y \geq n$, et $\mathbb{P}(Y_n > y) \leq \mathbb{P}(X_n > y) = \mathbb{P}(X_1 > y)$ si $y \in [0, n]$. Comme $\text{Var}(Y_n) \leq \mathbb{E}[Y_n^2]$, on en déduit que

$$\sum_{n=1}^{\infty} \frac{\text{Var}(Y_n)}{n^2} \leq \sum_{n=1}^{\infty} \frac{1}{n^2} \int_0^n 2y\mathbb{P}(X_1 > y) dy = \int_0^\infty 2y \left(\sum_{n=1}^{\infty} \frac{1}{n^2} \mathbb{1}_{n > y} \right) \mathbb{P}(X_1 > y) dy,$$

où l’on a utilisé Fubini–Tonelli pour échanger la somme et l’intégrale. On veut à présent majorer la dernière somme sur n . D’une part, pour $y \geq 1$, par décroissance de la fonction $x \mapsto 1/x^2$ sur \mathbb{R}_+^* , on a

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \mathbb{1}_{n > y} = \sum_{n=\lfloor y \rfloor + 1}^{\infty} \frac{1}{n^2} \leq \sum_{n=\lfloor y \rfloor + 1}^{\infty} \int_{n-1}^n \frac{1}{x^2} dx \leq \int_{\lfloor y \rfloor}^{\infty} \frac{1}{x^2} dx = \frac{1}{\lfloor y \rfloor} \leq \frac{2}{y},$$

où l’on utilise que $y \geq 1$ pour la dernière inégalité. D’autre part, pour $y < 1$, on a

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \mathbb{1}_{n > y} \leq \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \leq 2 \leq \frac{2}{y}.$$

On peut donc majorer

$$\sum_{n=1}^{\infty} \frac{\text{Var}(Y_n)}{n^2} \leq \int_0^\infty 2y \cdot \frac{2}{y} \cdot \mathbb{P}(X_1 > y) dy = 4 \int_0^\infty \mathbb{P}(X_1 > y) dy = 4\mathbb{E}[X_1],$$

par la Proposition 1.3.16. On a donc montré (4.2).

Étape 4 : Convergence d'une sous-suite.

Pour $n \geq 1$, on note $S_n := Y_1 + \dots + Y_n$. Rappelons que l'on veut montrer que S_n/n converge p.s. vers $\mathbb{E}[X_1]$ quand $n \rightarrow \infty$. On va utiliser une technique classique pour montrer des convergences p.s. : on va d'abord montrer la convergence d'une sous-suite explicite, avant de comparer au reste de la suite dans l'étape suivante. Travailler avec une sous-suite permet d'avoir moins de termes dans la série lorsque l'on essaie d'appliquer le critère de convergence vu à la Proposition 4.4.5.

Soit $\alpha > 1$. On définit $n_k = \lfloor \alpha^k \rfloor$, où $\lfloor \cdot \rfloor$ est la partie entière². On va d'abord montrer la convergence

$$\frac{S_{n_k} - \mathbb{E}[S_{n_k}]}{n_k} \xrightarrow[k \rightarrow \infty]{\text{p.s.}} 0. \quad (4.3)$$

Pour cela on veut utiliser le critère vu à la Proposition 4.4.5. Soit $\varepsilon > 0$. Par l'inégalité de Bienaymé-Chebychev,

$$\mathbb{P}\left(\left|\frac{S_{n_k} - \mathbb{E}[S_{n_k}]}{n_k}\right| \geq \varepsilon\right) \leq \frac{\text{Var}(S_{n_k})}{(\varepsilon n_k)^2} = \frac{1}{(\varepsilon n_k)^2} \sum_{n=1}^{n_k} \text{Var}(Y_n),$$

par indépendance des Y_n . On veut vérifier que cette probabilité est sommable : on a

$$\sum_{k \geq 1} \mathbb{P}\left(\left|\frac{S_{n_k} - \mathbb{E}[S_{n_k}]}{n_k}\right| \geq \varepsilon\right) \leq \sum_{k \geq 1} \sum_{n=1}^{n_k} \frac{1}{(\varepsilon n_k)^2} \text{Var}(Y_n) = \frac{1}{\varepsilon^2} \sum_{n \geq 1} \text{Var}(Y_n) \sum_{k \text{ tel que } n_k \geq n} \frac{1}{n_k^2}.$$

On majore la somme sur k , en notant que $n_k = \lfloor \alpha^k \rfloor \geq \frac{\alpha^k}{2}$ car $\alpha^k \geq 1$ et que

$$n_k \geq n \Leftrightarrow \lfloor \alpha^k \rfloor \geq n \Leftrightarrow \alpha^k \geq n \Leftrightarrow k \geq \frac{\log n}{\log \alpha} \Leftrightarrow k \geq \left\lceil \frac{\log n}{\log \alpha} \right\rceil,$$

ce qui nous donne

$$\sum_{k \text{ tel que } n_k \geq n} \frac{1}{n_k^2} \leq \sum_{k \geq \lceil \log n / \log \alpha \rceil} 4 \cdot \alpha^{-2k} \leq 4 \cdot \frac{\alpha^{-2 \log n / \log \alpha}}{1 - \alpha^{-2}} = \frac{4}{1 - \alpha^{-2}} \cdot \frac{1}{n^2}.$$

On revient à ce qui précède et on obtient

$$\sum_{k \geq 1} \mathbb{P}\left(\left|\frac{S_{n_k} - \mathbb{E}[S_{n_k}]}{n_k}\right| \geq \varepsilon\right) \leq \frac{4}{\varepsilon^2(1 - \alpha^{-2})} \sum_{n \geq 1} \frac{\text{Var}(Y_n)}{n^2} \leq \frac{16\mathbb{E}[X_1]}{\varepsilon^2(1 - \alpha^{-2})} < \infty,$$

où l'on a utilisé l'étape 3 dans la 2ème inégalité. Par la Proposition 4.4.5, cela montre (4.3).

Finalement, on veut vérifier que $\mathbb{E}[S_{n_k}]/n_k$ tend vers $\mathbb{E}[X_1]$. Pour cela, notons que

$$\mathbb{E}[Y_n] = \mathbb{E}[X_1 \mathbb{1}_{X_1 \leq n}] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[X_1],$$

par convergence dominée car $X_1 \mathbb{1}_{X_1 \leq n} \rightarrow X_1$ p.s. quand $n \rightarrow \infty$ et $|X_1 \mathbb{1}_{X_1 \leq n}| \leq X_1 \in L^1$. Donc, par le théorème de Cesàro³,

$$\frac{\mathbb{E}[S_{n_k}]}{n_k} = \frac{\mathbb{E}[Y_1] + \dots + \mathbb{E}[Y_{n_k}]}{n_k} \xrightarrow[k \rightarrow \infty]{} \mathbb{E}[X_1].$$

En combinant cela avec (4.3), on obtient

$$\frac{S_{n_k}}{n_k} \xrightarrow[k \rightarrow \infty]{\text{p.s.}} \mathbb{E}[X_1].$$

On a donc montré la convergence que l'on souhaitait, mais le long de la sous-suite indexée par n_n seulement.

2. On a fait un abus de langage en disant que $(S_{n_k})_{k \geq 1}$ est une sous-suite de $(S_n)_{n \geq 1}$ car la suite $(n_k)_{k \geq 1}$ n'est pas strictement croissante (seulement croissante). Mais peu importe : la convergence (4.3) a bien un sens et on peut aussi vérifier que la suite $(n_k)_{k \geq 1}$ est strictement croissante à partir d'un certain rang.

3. Rappelons le théorème de Cesàro : si une suite $(a_n)_{n \geq 1}$ de réels converge vers $a \in \mathbb{R}$, alors $(a_1 + \dots + a_n)/n \rightarrow a$ quand $n \rightarrow \infty$.

Étape 5 : En déduire la convergence de toute la suite.

On veut à présent montrer la convergence de toute la suite $(S_n/n)_{n \geq 1}$. C'est ici principalement que l'on utilise la positivité de X_1 car ainsi $(S_n)_{n \geq 1}$ est une suite croissante et on peut donc comparer tous les termes de la suite à ceux de la sous-suite.

Soit $n \geq n_1$. Comme $n_k \rightarrow \infty$ quand $k \rightarrow \infty$, il existe k tel que $n_k \leq n \leq n_{k+1}$. Comme $(S_n)_{n \geq 1}$ est croissante, on peut encadrer S_n/n ainsi :

$$\frac{S_{n_k}}{n_k} \frac{n_k}{n_{k+1}} = \frac{S_{n_k}}{n_{k+1}} \leq \frac{S_n}{n} \leq \frac{S_{n_{k+1}}}{n_k} = \frac{S_{n_{k+1}}}{n_{k+1}} \frac{n_{k+1}}{n_k}.$$

Soit $\omega \in \Omega$ tel que $S_{n_k}(\omega)/n_k$ tend vers $\mathbb{E}[X_1]$. On fait tendre $n \rightarrow \infty$ (et donc le k associé tend aussi vers l'infini) dans l'encadrement ci-dessus : comme $n_{k+1}/n_k \rightarrow \alpha$, on obtient

$$\frac{\mathbb{E}[X_1]}{\alpha} \leq \liminf_{n \rightarrow \infty} \frac{S_n(\omega)}{n} \leq \limsup_{n \rightarrow \infty} \frac{S_n(\omega)}{n} \leq \alpha \mathbb{E}[X_1].$$

Comme $S_{n_k}/n_k \rightarrow \mathbb{E}[X_1]$ p.s. par l'étape 4, on a montré que

$$\forall \alpha > 1, \text{ p.s.}, \quad \frac{\mathbb{E}[X_1]}{\alpha} \leq \liminf_{n \rightarrow \infty} \frac{S_n}{n} \leq \limsup_{n \rightarrow \infty} \frac{S_n}{n} \leq \alpha \mathbb{E}[X_1].$$

En se restreignant à $\alpha \in \mathbb{Q} \cap]1, \infty[$, on peut échanger le "pour tout α " (qui est devenu dénombrable) avec le "p.s.", et on en déduit que

$$\text{p.s.}, \quad \mathbb{E}[X_1] \leq \liminf_{n \rightarrow \infty} \frac{S_n}{n} \leq \limsup_{n \rightarrow \infty} \frac{S_n}{n} \leq \mathbb{E}[X_1].$$

Cela montre que

$$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbb{E}[X_1],$$

ce qui conclut la démonstration par l'étape 2.

4.5.2 Applications

Méthode de Monte-Carlo.

La loi des grands nombres fournit une manière d'approcher numériquement la valeur d'une espérance, en simulant de nombreuses fois la variable aléatoire sous-jacente et en calculant la moyenne empirique. Plus précisément, soit X une v.a. réelle et $f: \mathbb{R} \rightarrow \mathbb{R}$ mesurable telle que $\mathbb{E}[|f(X)|] < \infty$. Supposons que l'on veuille estimer $\mathbb{E}[f(X)]$ numériquement. Pour cela, on génère une suite $(X_n)_{n \geq 1}$ de v.a. réelles i.i.d. de même loi que X et on utilise que

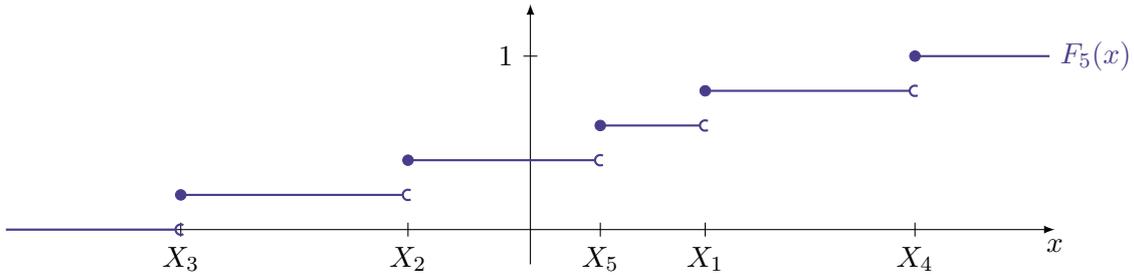
$$\frac{f(X_1) + \cdots + f(X_n)}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbb{E}[f(X_1)] = \mathbb{E}[f(X)],$$

par la loi forte des grands nombres, car les v.a. de la suite $(f(X_n))_{n \geq 1}$ sont i.i.d. et $\mathbb{E}[|f(X_1)|] = \mathbb{E}[|f(X)|] < \infty$. Le principal défaut de cette méthode est de savoir quel n on doit choisir pour quelle précision, ce que ne dit pas la loi des grands nombres. Pour cela, il faut utiliser des inégalités de concentration, ou le théorème central limite que l'on verra au prochain chapitre.

Un cas particulier important est celui du calcul de probabilités. Si $B \in \mathcal{B}(\mathbb{R})$, on peut estimer $\mathbb{P}(X \in B)$ par la méthode de Monte-Carlo avec $f = \mathbb{1}_B$. On a alors

$$\frac{|\{k \in \llbracket 1, n \rrbracket : X_k \in B\}|}{n} = \frac{\mathbb{1}_B(X_1) + \cdots + \mathbb{1}_B(X_n)}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbb{E}[\mathbb{1}_B(X)] = \mathbb{P}(X \in B).$$

C'est la définition fréquentiste de la probabilité : $\mathbb{P}(X \in B)$ est la proportion asymptotique de X_k qui tombent dans B .



Fonction de répartition empirique.

Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles i.i.d. Pour $n \geq 1$, la *fonction de répartition empirique* de l'échantillon X_1, \dots, X_n est la fonction définie par

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

Il est important de noter que c'est une fonction aléatoire : pour chaque $x \in \mathbb{R}$, $F_n(x)$ est une v.a. réelle. C'est en fait la fonction de répartition de la mesure empirique de l'échantillon, qui est la mesure de probabilité sur \mathbb{R} suivante (qui est aléatoire aussi!) :

$$\frac{1}{n} \sum_{k=1}^n \delta_{X_k}.$$

Pour chaque $x \in \mathbb{R}$ fixé, on a vu dans le paragraphe précédent que

$$F_n(x) = \frac{|\{k \in \llbracket 1, n \rrbracket : X_k \leq x\}|}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbb{P}(X_1 \leq x) = F(x),$$

où F est la fonction de répartition de X_1 . Ainsi, la fonction de répartition empirique approche la fonction de répartition théorique quand la taille de l'échantillon tend vers l'infini.

On vient de voir que :

$$\forall x \in \mathbb{R}, \text{ p.s.}, \quad F_n(x) \xrightarrow[n \rightarrow \infty]{} F(x).$$

On peut alors se demander si on peut échanger le "pour tout" et le "p.s." Pour cela on se restreint d'abord aux x rationnels, de sorte à rendre le "pour tout" dénombrable. On peut alors l'échanger avec le p.s. et on obtient : p.s., $\forall x \in \mathbb{Q}, F_n(x) \rightarrow F(x)$. Mais on a le résultat suivant pour des fonctions déterministes (laissé en exercice) : si f, f_1, f_2, \dots sont des fonctions croissantes continues à droite telles que $\forall x \in \mathbb{Q}, f_n(x) \rightarrow f(x)$, alors $\forall x \in \mathbb{R}, f_n(x) \rightarrow f(x)$. On en conclut

$$\text{p.s.}, \forall x \in \mathbb{R}, \quad F_n(x) \xrightarrow[n \rightarrow \infty]{} F(x).$$

Autrement dit, presque sûrement, la suite de fonctions $(F_n)_{n \geq 1}$ converge simplement vers F .

Chapitre 5

Convergence en loi et théorème central limite

5.1 Convergence en loi et convergence étroite

5.1.1 Définitions

Définition 5.1.1. Soit X, X_1, X_2, \dots des v.a. réelles. On dit que $(X_n)_{n \geq 1}$ converge en loi vers X , que l'on note

$$X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X,$$

si, pour toute fonction $f: \mathbb{R} \rightarrow \mathbb{R}$ continue bornée,

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

Remarque 5.1.2. Cette notion de convergence ne dépend que de la loi des v.a. considérées : le choix précis des v.a. (leurs valeurs pour chaque ω) n'importe pas, seule leur loi compte. En effet, si X et Y ont la même loi, alors pour toute fonction $f: \mathbb{R} \rightarrow \mathbb{R}$ continue bornée, on a $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$. En particulier, il est clair que l'on peut remplacer X par Y de même loi à la limite et la convergence a toujours lieu.

Réciproquement, par le Théorème 1.1.10, si, pour toute fonction $f: \mathbb{R} \rightarrow \mathbb{R}$ continue bornée, on a $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$, alors X et Y ont la même loi. Il y a donc unicité de la limite en loi. C'est à comparer aux convergences vues au chapitre précédent pour lesquelles il y a unicité de la limite à égalité p.s. près.

Comme cette convergence ne concerne que les lois des variables aléatoires, on parle parfois directement de la convergence des lois : on introduit pour cela la notion de convergence étroite.

Définition 5.1.3. Soit P, P_1, P_2, \dots des lois sur \mathbb{R} . On dit que $(P_n)_{n \geq 1}$ converge étroitement vers P , que l'on note

$$P_n \xrightarrow[n \rightarrow \infty]{\text{étroit.}} P,$$

si, pour toute fonction $f: \mathbb{R} \rightarrow \mathbb{R}$ continue bornée,

$$\int_{\mathbb{R}} f(x) dP_n(x) \xrightarrow[n \rightarrow \infty]{} \int_{\mathbb{R}} f(x) dP(x).$$

Si X, X_1, X_2, \dots sont des v.a. réelles, il est clair que

$$X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X \Leftrightarrow P_{X_n} \xrightarrow[n \rightarrow \infty]{\text{étroit.}} P_X.$$

Quand on montre la convergence en loi d'une suite de v.a., comme la définition précise de la variable aléatoire limite n'importe pas mais seule sa loi compte, on écrit parfois

$$X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} P,$$

où P est une loi. C'est un léger abus de notation car les quantités ne sont pas du même type des deux côtés de la flèche : cela signifie que P_{X_n} converge étroitement vers P ou, de manière équivalente, que X_n converge en loi vers X , où X est n'importe quelle v.a. de loi P .

Exemple 5.1.4. Si X, X_1, X_2, \dots sont des v.a. réelles de même loi, alors il est clair que $(X_n)_{n \geq 1}$ converge en loi vers X . Si ces v.a. sont toutes égales, alors il y a aussi convergence p.s. et en probabilité (et dans L^p si elles sont dans L^p). Mais si on les prend toutes indépendantes alors aucune de ces autres convergences ne sera vraie (sauf si les v.a. sont constantes p.s.).

Exemple 5.1.5. Soit $(\lambda_n)_{n \geq 1}$ une suite décroissante de réels convergeant vers $\lambda > 0$. Pour tout $n \geq 1$, soit X_n une v.a. de loi $\mathcal{E}(\lambda_n)$. Soit X une v.a. de loi $\mathcal{E}(\lambda)$. Montrons que

$$X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X.$$

Pour cela, on considère $f: \mathbb{R} \rightarrow \mathbb{R}$ continue bornée quelconque. Par le théorème de transfert, on a

$$\mathbb{E}[f(X_n)] = \int_0^\infty f(x) \lambda_n e^{-\lambda_n x} dx$$

Comme $\lambda_n \rightarrow \lambda$, on a $f(x) \lambda_n e^{-\lambda_n x} \rightarrow f(x) \lambda e^{-\lambda x}$ pour tout $x \geq 0$. En outre, par décroissance de $(\lambda_n)_{n \geq 1}$, on peut majorer

$$\forall x \geq 0, \quad \left| f(x) \lambda_n e^{-\lambda_n x} \right| \leq \|f\|_\infty \lambda_1 e^{-\lambda x},$$

la fonction de droite étant indépendante de n et intégrable par rapport à la mesure de Lebesgue sur $[0, \infty[$. Par le théorème de convergence dominée, on obtient

$$\mathbb{E}[f(X_n)] = \int_0^\infty f(x) \lambda_n e^{-\lambda_n x} dx \xrightarrow[n \rightarrow \infty]{} \int_0^\infty f(x) \lambda e^{-\lambda x} dx = \mathbb{E}[f(X)].$$

Cela montre que $(X_n)_{n \geq 1}$ converge en loi vers X .

Exemple 5.1.6 (Convergence en loi de v.a. aléatoires constantes p.s.). Soit $(x_n)_{n \geq 1}$ une suite de réels et, pour chaque $n \geq 1$, une v.a. X_n telle que $X_n = x_n$ p.s. Soit $x \in \mathbb{R}$ et X une v.a. telle que $X = x$ p.s. Alors, on montrera au TD11 que

$$x_n \xrightarrow[n \rightarrow \infty]{} x \quad \Leftrightarrow \quad X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X.$$

En terme de convergence de lois, cela se réécrit

$$x_n \xrightarrow[n \rightarrow \infty]{} x \quad \Leftrightarrow \quad \delta_{x_n} \xrightarrow[n \rightarrow \infty]{\text{étroit.}} \delta_x.$$

Remarque 5.1.7. Il est important de noter que la convergence en loi ne requiert la convergence $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ que pour les fonctions $f: \mathbb{R} \rightarrow \mathbb{R}$ continues bornées, et pas toutes les fonctions mesurables bornées. Et, si $(X_n)_{n \geq 1}$ converge en loi vers X , il peut y avoir des fonctions $f: \mathbb{R} \rightarrow \mathbb{R}$ mesurables bornées telles que $\mathbb{E}[f(X_n)] \not\rightarrow \mathbb{E}[f(X)]$. Il peut même y avoir des boréliens $B \in \mathcal{B}(\mathbb{R})$ tels que $\mathbb{P}(X_n \in B) \not\rightarrow \mathbb{P}(X \in B)$.

Montrons le sur un exemple. Prenons, $X_n = 1/n$ p.s. et $X = 0$ p.s. Alors $(X_n)_{n \geq 1}$ converge en loi vers X par l'exemple précédent. Mais, on a

$$\begin{aligned} \mathbb{P}(X_n \leq 0) &= 0 \xrightarrow[n \rightarrow \infty]{} 1 = \mathbb{P}(X \leq 0) \\ \mathbb{P}(X_n > 0) &= 1 \xrightarrow[n \rightarrow \infty]{} 0 = \mathbb{P}(X > 0), \end{aligned}$$

ce qui montre que la convergence $\mathbb{P}(X_n \in B) \rightarrow \mathbb{P}(X \in B)$ n'a pas lieu en général, même pour B fermé ou ouvert ou un intervalle.

5.1.2 Lien avec les autres convergences

Le résultat suivant montre que la convergence en loi est plus faible que la convergence en probabilité, et donc est plus faible que toutes les convergences vues au Chapitre 4.

Proposition 5.1.8. *Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles convergeant en probabilité vers une v.a. réelle X . Alors $(X_n)_{n \geq 1}$ converge en loi vers X .*

Démonstration. Soit f continue bornée. Montrons que $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$. Pour cela on considère $\varepsilon > 0$ et on écrit

$$\begin{aligned} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| &\leq \mathbb{E}[|f(X_n) - f(X)|] \\ &= \mathbb{E}\left[|f(X_n) - f(X)| \mathbb{1}_{|f(X_n) - f(X)| \leq \varepsilon}\right] + \mathbb{E}\left[|f(X_n) - f(X)| \mathbb{1}_{|f(X_n) - f(X)| > \varepsilon}\right] \\ &\leq \varepsilon + 2\|f\|_\infty \mathbb{P}(|f(X_n) - f(X)| > \varepsilon). \end{aligned}$$

Comme f est continue et $(X_n)_{n \geq 1}$ converge en probabilité vers X , par la Proposition 4.3.12, on sait que $(f(X_n))_{n \geq 1}$ converge en probabilité vers $f(X)$. Ainsi $\mathbb{P}(|f(X_n) - f(X)| > \varepsilon) \rightarrow 0$ et donc, pour tout n à partir d'un certain rang,

$$|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq 2\varepsilon.$$

Cela montre que $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ et donc que $(X_n)_{n \geq 1}$ converge en loi vers X . \square

En général, la réciproque est fautive : converger en loi n'implique pas converger en probabilité. En effet, pour la convergence en loi, on peut remplacer la limite X d'une suite $(X_n)_{n \geq 1}$ par n'importe quelle autre v.a. X' de même loi. Alors $(X_n)_{n \geq 1}$ converge en loi aussi vers X' . Mais, si on peut choisir X et X' qui ne sont pas égales p.s. alors $(X_n)_{n \geq 1}$ ne peut pas converger en probabilité à la fois vers X et vers X' (par la Proposition 4.3.11).

Il n'y a qu'un seul cas où on ne peut pas choisir X et X' de même loi mais pas égales p.s. : c'est dans le cas où leur loi est une masse de Dirac, c'est-à-dire dans le cas de v.a. constantes p.s. Il se trouve que dans ce cas particulier, la convergence en loi implique la convergence en probabilité comme montré ci-dessous.

Proposition 5.1.9. *Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles convergeant en loi vers une v.a. réelle X . Si X est constante p.s., alors $(X_n)_{n \geq 1}$ converge en probabilité vers X .*

Démonstration. Comme X est constante p.s., il existe $x \in \mathbb{R}$ tel que $X = x$ p.s. On considère la fonction

$$f(t) = \min(1, |t - x|), \quad t \in \mathbb{R}.$$

C'est une fonction continue et bornée, donc par convergence en loi de $(X_n)_{n \geq 1}$ vers X ,

$$\mathbb{E}[f(X_n)] \xrightarrow{n \rightarrow \infty} \mathbb{E}[f(X)] = f(x) = 0.$$

Montrons que $X_n \rightarrow x$ en probabilité. Soit $\varepsilon \in]0, 1]$. Alors

$$\mathbb{P}(|X_n - x| \geq \varepsilon) = \mathbb{P}(\min(1, |X_n - x|) \geq \varepsilon) = \mathbb{P}(f(X_n) \geq \varepsilon) \leq \frac{\mathbb{E}[f(X_n)]}{\varepsilon} \xrightarrow{n \rightarrow \infty} 0,$$

où l'on a utilisé l'inégalité de Markov. Si $\varepsilon \geq 1$, alors $\mathbb{P}(|X_n - x| \geq \varepsilon) \leq \mathbb{P}(|X_n - x| \geq 1) \rightarrow 0$. Cela montre que $X_n \rightarrow x$ en probabilité et donc que $X_n \rightarrow X$ en probabilité. \square

5.1.3 Caractérisations de la convergence en loi

Fonctions test plus régulières.

On montre ici que pour montrer une convergence en loi, on peut se restreindre à des fonctions test f qui sont de classe \mathcal{C}^∞ et à support compact. On rappelle qu'une fonction est dite à support compact s'il existe un compact en dehors duquel elle est nulle.

Théorème 5.1.10. *Soit X, X_1, X_2, \dots des v.a. réelles. Alors $(X_n)_{n \geq 1}$ converge en loi vers X si et seulement si, pour toute fonction $f: \mathbb{R} \rightarrow \mathbb{R}$ de classe \mathcal{C}^∞ et à support compact,*

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

Pour la démonstration, on utilisera le lemme de densité suivant (probablement vu en théorie de la mesure).

Lemme 5.1.11. *Soit $f: \mathbb{R} \rightarrow \mathbb{R}$ continue à support compact. Soit $\varepsilon > 0$. Il existe une fonction $g: \mathbb{R} \rightarrow \mathbb{R}$ de classe \mathcal{C}^∞ et à support compact telle que $\|f - g\|_\infty \leq \varepsilon$, c'est-à-dire*

$$\forall x \in \mathbb{R}, \quad |f(x) - g(x)| \leq \varepsilon.$$

Démonstration du Théorème 5.1.10. L'implication directe est évidente car toute fonction \mathcal{C}^∞ à support compact est aussi continue bornée. On se concentre donc sur le fait de montrer l'implication réciproque : on suppose que $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ pour toute fonction f de classe \mathcal{C}^∞ à support compact. On va procéder en deux étapes : on montre d'abord que la convergence reste vraie pour f continue à support compact, puis on étend à toute fonction f continue bornée.

Première étape. Soit $f: \mathbb{R} \rightarrow \mathbb{R}$ continue à support compact. Soit $\varepsilon > 0$. Par le lemme 5.1.11, il existe $g: \mathbb{R} \rightarrow \mathbb{R}$ de classe \mathcal{C}^∞ et à support compact telle que $\|f - g\|_\infty \leq \varepsilon$. On borne alors, par inégalité triangulaire,

$$\begin{aligned} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| &\leq |\mathbb{E}[f(X_n)] - \mathbb{E}[g(X_n)]| + |\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]| + |\mathbb{E}[g(X)] - \mathbb{E}[f(X)]| \\ &\leq \varepsilon + |\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]| + \varepsilon. \end{aligned}$$

Comme g est \mathcal{C}^∞ à support compact, on sait que $\mathbb{E}[g(X_n)] \rightarrow \mathbb{E}[g(X)]$ par hypothèse. Donc, pour tout n à partir d'un certain rang, on a $|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq 3\varepsilon$. Cela montre que $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ quand $n \rightarrow \infty$.

Deuxième étape. Soit $f: \mathbb{R} \rightarrow \mathbb{R}$ continue bornée. Soit $\varepsilon > 0$. Par continuité décroissante de la mesure,

$$\lim_{N \rightarrow \infty} \mathbb{P}(|X| \geq N) = \mathbb{P}\left(\bigcap_{N \geq 1} \downarrow \{|X| \geq N\}\right) = \mathbb{P}(|X| = \infty) = 0,$$

donc il existe $N \geq 1$ tel que $\mathbb{P}(|X| \geq N) \leq \varepsilon$. Soit $h: \mathbb{R} \rightarrow \mathbb{R}$ continue à support compact telle que $\mathbb{1}_{[-N, N]} \leq h \leq 1$. On a, en écrivant $f = fh + f(1 - h)$ et en utilisant l'inégalité triangulaire,

$$\begin{aligned} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| &\leq |\mathbb{E}[fh(X_n)] - \mathbb{E}[fh(X)]| + |\mathbb{E}[f(1 - h)(X_n)] - \mathbb{E}[f(1 - h)(X)]| \\ &\leq |\mathbb{E}[fh(X_n)] - \mathbb{E}[fh(X)]| + \|f\|_\infty (\mathbb{E}[(1 - h)(X_n)] + \mathbb{E}[(1 - h)(X)]), \end{aligned}$$

car $1 - h \geq 0$. On a $\mathbb{E}[(1 - h)(X_n)] = 1 - \mathbb{E}[h(X_n)]$ et, en appliquant le résultat de la première étape à fh et h qui sont continues à support compact, on obtient

$$\limsup_{n \rightarrow \infty} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq 0 + \|f\|_\infty (1 - \mathbb{E}[h(X)] + \mathbb{E}[(1 - h)(X)]) \leq 2\varepsilon \|f\|_\infty,$$

car $\mathbb{E}[(1 - h)(X)] \leq \mathbb{P}(|X| > N) \geq \varepsilon$. Cela montre que $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ quand $n \rightarrow \infty$. \square

Fonction de répartition.

Une des manières de montrer la convergence en loi d'une suite de v.a. réelles est de montrer la convergence de leurs fonctions de répartition. Notons qu'elle n'a pas forcément lieu en tout point, comme on l'avait vu dans la Remarque 5.1.7 avec l'exemple $X_n = 1/n$ et $X = 0$.

Théorème 5.1.12. *Soit X, X_1, X_2, \dots des v.a. réelles. Alors $(X_n)_{n \geq 1}$ converge en loi vers X si et seulement si*

$$\forall a \in \mathbb{R} \text{ point de continuité de } F_X, \quad F_{X_n}(a) \xrightarrow[n \rightarrow \infty]{} F_X(a).$$

Remarque 5.1.13. L'ensemble des points de discontinuité de F_X est au plus dénombrable (c'est le cas de toute fonction croissante). En effet, comme F_X est croissante et $0 \leq F_X \leq 1$, il ne peut y avoir qu'au plus k discontinuités de tailles $1/k$ pour $k \in \mathbb{N}^*$, donc l'ensemble des points de discontinuité de F_X peut s'écrire

$$\bigcup_{k \geq 1} \left\{ a \in \mathbb{R} : F_X(a) - F_X(a-) \geq \frac{1}{k} \right\},$$

qui est une union dénombrable d'ensembles finis et est donc au plus dénombrable.

Démonstration. Sens direct. Supposons que $(X_n)_{n \geq 1}$ converge en loi vers X . Soit $a \in \mathbb{R}$ un point de continuité de F_X . Alors

$$F_{X_n}(a) = \mathbb{P}(X_n \leq a) = \mathbb{E}[\mathbb{1}_{]-\infty, a]}(X_n)].$$

Cependant on ne peut pas passer à la limite car la fonction $\mathbb{1}_{]-\infty, a]}$ n'est pas continue. Soit $\varepsilon > 0$. Soit $\delta > 0$ qu'on choisira plus tard en fonction de $\varepsilon > 0$. Il existe des fonctions continues f_1 et f_2 (on peut les prendre affines par morceaux) telles que

$$\mathbb{1}_{]-\infty, a-\delta]} \leq f_1 \leq \mathbb{1}_{]-\infty, a]} \leq f_2 \leq \mathbb{1}_{]-\infty, a+\delta]}.$$

Alors on a l'encadrement

$$\mathbb{E}[f_1(X_n)] \leq F_{X_n}(a) \leq \mathbb{E}[f_2(X_n)]$$

Mais pour $i \in \{1, 2\}$, on a $\mathbb{E}[f_i(X_n)] \rightarrow \mathbb{E}[f_i(X)]$ quand $n \rightarrow \infty$, car f_i est continue bornée. Donc, pour tout n à partir d'un certain rang,

$$\mathbb{E}[f_1(X)] - \varepsilon \leq F_{X_n}(a) \leq \mathbb{E}[f_2(X)] + \varepsilon$$

Mais $\mathbb{E}[f_2(X)] \leq \mathbb{E}[\mathbb{1}_{]-\infty, a+\delta]}(X)] = F_X(a + \delta)$ et de même $\mathbb{E}[f_1(X)] \geq F_X(a - \delta)$. On a donc montré que, pour tout n à partir d'un certain rang,

$$F_X(a - \delta) - \varepsilon \leq F_{X_n}(a) \leq F_X(a + \delta) + \varepsilon.$$

En choisissant δ suffisamment petit pour que $|F_X(a \pm \delta) - F_X(a)| \leq \varepsilon$ (possible par continuité de F_X en a), cela donne $|F_{X_n}(a) - F_X(a)| \leq 2\varepsilon$. On a donc montré que $F_{X_n}(a) \rightarrow F_X(a)$.

Sens réciproque. Supposons que, pour tout point de continuité a de F_X , on ait $F_{X_n}(a) \rightarrow F_X(a)$. On va montrer la convergence en loi grâce au Théorème 5.1.10. On considère donc une fonction test $f: \mathbb{R} \rightarrow \mathbb{R}$ de classe C^∞ à support compact. Pour une telle fonction, on a $f(x) = \int_{-\infty}^x f'(y) dy$ et donc

$$\begin{aligned} \mathbb{E}[f(X_n)] &= \int_{\mathbb{R}} f(x) P_{X_n}(dx) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbb{1}_{y < x} f'(y) dy \right) P_{X_n}(dx) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbb{1}_{y < x} P_{X_n}(dx) \right) f'(y) dy = \int_{\mathbb{R}} \mathbb{P}(X_n > y) f'(y) dy \\ &= \int_{\mathbb{R}} (1 - F_{X_n}(y)) f'(y) dy, \end{aligned}$$

où la 3ème égalité est justifiée par le théorème de Fubini–Lebesgue car, comme P_{X_n} a masse totale 1, on a $\int_{\mathbb{R}} (\int_{\mathbb{R}} \mathbb{1}_{y < x} f'(y) |P_{X_n}(dx)) dy \leq \int_{\mathbb{R}} |f'(y)| dy < \infty$ (car f' est nulle en dehors d'un compact et bornée sur ce compact car continue). Par hypothèse, $(1 - F_{X_n}(y))f'(y) \rightarrow (1 - F_X(y))f'(y)$ Lebesgue-presque partout (car les points de discontinuité de F_X sont au plus dénombrables par la Remarque 5.1.13). En outre, on peut dominer $|(1 - F_{X_n}(y))f'(y)| \leq |f'(y)|$ et $\int_{\mathbb{R}} |f'(y)| dy < \infty$. Donc, par le théorème de convergence dominée,

$$\mathbb{E}[f(X_n)] \xrightarrow{n \rightarrow \infty} \int_{\mathbb{R}} (1 - F_X(y))f'(y) dy = \mathbb{E}[f(X)],$$

par le calcul précédent appliqué à X au lieu de X_n . Cela implique la convergence en loi par le Théorème 5.1.10. \square

Montrer une convergence en loi en utilisant les fonction de répartition est particulièrement utile quand on étudie un minimum ou un maximum de v.a. indépendantes, comme dans l'exemple suivant.

Exemple 5.1.14. Soit $(U_n)_{n \geq 1}$ une suite de v.a. i.i.d. de loi uniforme sur $[0, 1]$. On considère $M_n = \min(U_1, \dots, U_n)$. Pour $a \in \mathbb{R}$, on a

$$F_{M_n}(a) = 1 - \mathbb{P}(M_n > a) = 1 - \mathbb{P}(U_1 > a, \dots, U_n > a) = 1 - \mathbb{P}(U_1 > a) \cdots \mathbb{P}(U_n > a),$$

par indépendance des U_k et donc

$$F_{M_n}(a) = \begin{cases} 0 & \text{si } a < 0, \\ 1 - (1 - a)^n & \text{si } a \in [0, 1], \\ 1 & \text{si } a > 1. \end{cases}$$

À partir de ce calcul, on peut vérifier aisément que $(M_n)_{n \geq 1}$ converge en probabilité vers 0, et avec un peu plus d'effort on peut montrer que la convergence a lieu p.s. On se demande alors à quelle vitesse $(M_n)_{n \geq 1}$ tend vers 0. Une manière d'avoir une idée est de calculer $\mathbb{E}[M_n]$: comme M_n est positive, on a

$$\mathbb{E}[M_n] = \int_0^\infty \mathbb{P}(M_n > x) dx = \int_0^1 (1 - x)^n dx = \frac{1}{n+1} \sim \frac{1}{n},$$

quand $n \rightarrow \infty$. Ainsi, une quantité que l'on peut espérer être d'ordre 1 est nM_n et, en effet, on va montrer la convergence en loi de $(nM_n)_{n \geq 1}$. Pour cela, on calcule la fonction de répartition de nM_n : pour tout $a \in \mathbb{R}$,

$$F_{nM_n}(a) = F_{M_n}\left(\frac{a}{n}\right) = \begin{cases} 0 & \text{si } a < 0, \\ 1 - \left(1 - \frac{a}{n}\right)^n & \text{si } a \in [0, n], \\ 1 & \text{si } a > n. \end{cases}$$

On rappelle que $\left(1 - \frac{a}{n}\right)^n \rightarrow e^{-a}$ donc, pour tout $a \in \mathbb{R}$,

$$F_{nM_n}(a) \xrightarrow{n \rightarrow \infty} (1 - e^{-a}) \mathbb{1}_{a \geq 0}.$$

C'est la fonction de répartition d'une v.a. X de loi exponentielle de paramètre 1. Donc

$$nM_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X.$$

Remarque 5.1.15. Pour étudier la convergence en loi ou non d'une suite $(X_n)_{n \geq 1}$ à l'aide des fonctions de répartition, on procède ainsi. Tout d'abord, on calcule F_{X_n} pour tout $n \geq 1$ et on

étudie la convergence de $F_{X_n}(a)$ pour $a \in \mathbb{R}$. Si $F_{X_n}(a)$ ne converge pas pour un ensemble non-dénombrable de a alors il n'y a pas convergence en loi (d'après le Théorème 5.1.12 et la Remarque 5.1.13). Sinon, on a une limite pour tout $a \in \mathbb{R}$ sauf un sous-ensemble au plus dénombrable :

$$F_{X_n}(a) \xrightarrow[n \rightarrow \infty]{} F(a),$$

où la fonction F est nécessairement croissante et est choisie continue à droite aux points de discontinuité (quitte à ce que $F(a)$ ne soit pas égal à la limite de $F_{X_n}(a)$). Il est essentiel pour cette méthode de bien choisir F continue à droite. Il y a alors deux cas :

- Si $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow \infty} F(x) = 1$, alors F est bien la fonction de répartition d'une v.a. X par le Théorème 1.4.6 et on a convergence en loi de $(X_n)_{n \geq 1}$ vers X . De plus, on peut souvent déterminer la loi de X avec la Proposition 1.4.7.
- Sinon, F n'est pas une fonction de répartition et donc $(X_n)_{n \geq 1}$ ne converge pas en loi. Pour le montrer, supposons que $(X_n)_{n \geq 1}$ converge en loi vers X . Alors $F_{X_n}(a) \rightarrow F_X(a)$ en tout point de continuité de F_X . Alors on a $F(a) = F_X(a)$ pour tout $a \in A$ où A est un ensemble de complémentaire au plus dénombrable. En particulier, A est dense dans \mathbb{R}^1 et donc pour tout $a \in \mathbb{R}$, il existe une suite de $a_n \in A$ telle que $a_n \rightarrow a$ avec $a_n \geq a$. Comme $F(a_n) = F_X(a_n)$ pour tout n et F et F_X sont continues à droite, cela implique que $F(a) = F_X(a)$. Donc $F = F_X$ sur \mathbb{R} . Cela contredit le fait que F ne soit pas une fonction de répartition.

Quand la fonction de répartition limite est continue, ce qui arrive par exemple si la limite est une v.a. continue, on peut en déduire la convergence des probabilités d'être dans n'importe quel intervalle.

Corollaire 5.1.16. *Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles convergeant en loi vers X . On suppose que F_X est continue sur \mathbb{R} (ou, de manière équivalente, que X n'a pas d'atome). Alors, pour tout intervalle $I \subset \mathbb{R}$,*

$$\mathbb{P}(X_n \in I) \xrightarrow[n \rightarrow \infty]{} \mathbb{P}(X \in I).$$

Démonstration. Voir le TD12. □

Fonction caractéristique.

Une autre manière de montrer une convergence en loi est de montrer la convergence des fonctions caractéristiques. Cette approche est particulièrement utile lorsqu'on travaille avec une somme de v.a. indépendantes, comme dans la démonstration du théorème central limite dans la section suivante.

Théorème 5.1.17 (Théorème de Lévy faible). *Soit X, X_1, X_2, \dots des v.a. réelles. Alors $(X_n)_{n \geq 1}$ converge en loi vers X si et seulement si*

$$\forall \theta \in \mathbb{R}, \quad \phi_{X_n}(\theta) \xrightarrow[n \rightarrow \infty]{} \phi_X(\theta).$$

Démonstration (admise). □

Le cas des v.a. à valeurs entières.

Pour des v.a. à valeurs entières, il suffit de montrer la convergence des probabilités de prendre chaque valeur possible pour montrer la convergence en loi. On se restreint ici au cas de v.a. à valeurs entières, mais cela se généralise de manière similaire au cas d'une suite de v.a. à valeurs dans le même ensemble $E \subset \mathbb{R}$ dénombrable sans point d'accumulation (pour tout K compact inclus dans \mathbb{R} , $E \cap K$ est fini).

1. Notons λ la mesure de Lebesgue sur \mathbb{R} . Comme A^c est au plus dénombrable, on a $\lambda(A^c) = 0$. Donc pour tout $a < b$, on a $\lambda(]a, b[\cap A) = \lambda(]a, b[) = b - a > 0$. En particulier, cela implique que $]a, b[\cap A \neq \emptyset$. Donc A est dense dans \mathbb{R} .

Proposition 5.1.18. Soit X, X_1, X_2, \dots des v.a. à valeurs dans \mathbb{Z} . Alors $(X_n)_{n \geq 1}$ converge en loi vers X si et seulement si

$$\forall k \in \mathbb{Z}, \quad \mathbb{P}(X_n = k) \xrightarrow{n \rightarrow \infty} \mathbb{P}(X = k).$$

Démonstration. Sens direct. Supposons que $(X_n)_{n \geq 1}$ converge en loi vers X . Soit $k \in \mathbb{Z}$. On considère une fonction $f: \mathbb{R} \rightarrow \mathbb{R}$ continue bornée telle que $f(k) = 1$ et f est nulle sur tous les autres entiers (par exemple, $f(x) = \max(1 - |x - k|, 0)$). Alors

$$\mathbb{E}[f(X_n)] = \sum_{i \in \mathbb{Z}} \mathbb{P}(X_n = i) f(i) = \mathbb{P}(X_n = k)$$

et de même $\mathbb{E}[f(X)] = \mathbb{P}(X = k)$. Or, par convergence en loi, on a $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ quand $n \rightarrow \infty$ et donc $\mathbb{P}(X_n = k) \rightarrow \mathbb{P}(X = k)$.

Sens réciproque. Supposons que, pour tout $k \in \mathbb{Z}$, on ait $\mathbb{P}(X_n = k) \rightarrow \mathbb{P}(X = k)$ quand $n \rightarrow \infty$. Soit $f: \mathbb{R} \rightarrow \mathbb{R}$ à support compact. Soit $M > 0$ tel que $f(x) = 0$ pour tout $x \notin [-M, M]$. Alors

$$\begin{aligned} \mathbb{E}[f(X_n)] &= \sum_{k \in \mathbb{Z}} \mathbb{P}(X_n = k) f(k) = \sum_{k \in \mathbb{Z} \cap [-M, M]} \mathbb{P}(X_n = k) f(k) \\ &\xrightarrow{n \rightarrow \infty} \sum_{k \in \mathbb{Z} \cap [-M, M]} \mathbb{P}(X = k) f(k) = \mathbb{E}[f(X)], \end{aligned}$$

où la convergence est justifiée car la somme est finie. Cela implique la convergence en loi par le Théorème 5.1.10. \square

Exemple 5.1.19 (Limite poissonnienne des lois binomiales). Soit $(X_n)_{n \geq 1}$ une suite de v.a. telle que X_n ait une loi binomiale de paramètre (n, p_n) , où $(p_n)_{n \geq 1}$ est une suite de réels dans $[0, 1]$ telle que

$$np_n \xrightarrow{n \rightarrow \infty} \lambda,$$

pour un certain $\lambda > 0$. Alors, pour tout $k \in \mathbb{N}$ fixé, on a

$$\mathbb{P}(X_n = k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{n!}{k!(n-k)!} \left(\frac{p_n}{1-p_n} \right)^k \exp(n \log(1-p_n)).$$

En utilisant la formule de Stirling, le fait que $\frac{p_n}{1-p_n} \sim \frac{\lambda}{n}$ et que $n \log(1-p_n) \sim -\lambda$, on obtient

$$\mathbb{P}(X_n = k) \sim \frac{1}{k!} \cdot \frac{\sqrt{2\pi n} \left(\frac{n}{e}\right)^n}{\sqrt{2\pi(n-k)} \left(\frac{n-k}{e}\right)^{n-k}} \cdot \left(\frac{\lambda}{n}\right)^k \exp(-\lambda) \sim \frac{1}{k!} \cdot e^{-k} \left(\frac{n}{n-k}\right)^n \cdot \lambda^k e^{-\lambda}.$$

Finalement, comme $\left(\frac{n-k}{n}\right)^n = \exp(n \log(1 - \frac{k}{n})) \rightarrow e^{-k}$, on obtient

$$\mathbb{P}(X_n = k) \xrightarrow{n \rightarrow \infty} \frac{\lambda^k}{k!} e^{-\lambda}.$$

Notons aussi que, pour tout $k < 0$, $\mathbb{P}(X_n = k) = 0 \rightarrow 0$. Par la proposition précédente, cela montre que $(X_n)_{n \geq 1}$ converge en loi vers une v.a. de loi de Poisson de paramètre λ .

5.2 Théorème central limite et applications

5.2.1 Le théorème central limite

Nous avons vu la loi des grands nombres qui garantit que, pour une suite $(X_n)_{n \geq 1}$ une suite de v.a. réelles i.i.d. dans L^1 , on a

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbb{E}[X_1].$$

Autrement dit, au premier ordre, la v.a. $X_1 + \dots + X_n$ se comporte de manière déterministe comme $n\mathbb{E}[X_1]$. On s'intéresse ici à l'ordre suivant dans ce comportement asymptotique ou, en termes probabilistes, aux fluctuations autour de cette loi des grands nombres. C'est le théorème central limite qui décrit ces fluctuations dans le cas où les v.a. X_k sont dans L^2 . Il nous dit que, typiquement, la différence entre $X_1 + \dots + X_n$ et $n\mathbb{E}[X_1]$ est d'ordre \sqrt{n} et se distribue approximativement comme une loi gaussienne à cette échelle.

Théorème 5.2.1. *Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles i.i.d. telle que $\mathbb{E}[X_1^2] < \infty$. On pose $m = \mathbb{E}[X_1]$ et $\sigma^2 = \text{Var}(X_1)$. Si $\sigma^2 > 0$, alors*

$$\frac{X_1 + \dots + X_n - nm}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\text{loi}} Z,$$

où Z est une v.a. de loi $\mathcal{N}(0, \sigma^2)$.

Démonstration. Quitte à considérer $X_k - m$ à la place de X_k (ce qui ne change pas sa variance), on peut supposer que $m = 0$. On veut alors montrer la convergence de

$$Z_n := \frac{X_1 + \dots + X_n}{\sqrt{n}}$$

et pour cela on va calculer sa fonction caractéristique. Soit $\theta \in \mathbb{R}$. On a

$$\phi_{Z_n}(\theta) = \mathbb{E} \left[\exp \left(\frac{i\theta}{\sqrt{n}} \sum_{k=1}^n X_k \right) \right] = \mathbb{E} \left[\prod_{k=1}^n \exp \left(\frac{i\theta}{\sqrt{n}} X_k \right) \right] = \prod_{k=1}^n \mathbb{E} \left[\exp \left(\frac{i\theta}{\sqrt{n}} X_k \right) \right] = \phi_{X_1} \left(\frac{\theta}{\sqrt{n}} \right)^n,$$

où l'on a utilisé l'indépendance des X_k dans la 3ème égalité et le fait qu'ils aient la même loi dans la 4ème. On remarque à présent que, comme $\mathbb{E}[X_1^2] < \infty$, par la Proposition 1.4.13, on sait que ϕ_{X_1} est de classe \mathcal{C}^2 sur \mathbb{R} et

$$\phi_{X_1}(0) = 1, \quad \phi'_{X_1}(0) = i\mathbb{E}[X_1] = 0 \quad \text{et} \quad \phi''_{X_1}(0) = -\mathbb{E}[X_1^2] = -\sigma^2.$$

On a donc le développement, quand $n \rightarrow \infty$ (pour θ fixé),

$$\phi_{X_1} \left(\frac{\theta}{\sqrt{n}} \right) = 1 - \frac{\sigma^2 \theta^2}{2n} + o \left(\frac{1}{n} \right) = e^{-\sigma^2 \theta^2 / (2n)} + o \left(\frac{1}{n} \right).$$

Ainsi, on a

$$\left| \phi_{Z_n}(\theta) - e^{-\sigma^2 \theta^2 / 2} \right| = \left| \phi_{X_1} \left(\frac{\theta}{\sqrt{n}} \right)^n - \left(e^{-\sigma^2 \theta^2 / (2n)} \right)^n \right| \leq n \left| \phi_{X_1} \left(\frac{\theta}{\sqrt{n}} \right) - e^{-\sigma^2 \theta^2 / (2n)} \right| \xrightarrow[n \rightarrow \infty]{} 0,$$

où l'on a utilisé l'inégalité $|z^n - w^n| \leq n|z - w|$ pour $z, w \in \mathbb{C}$ de modules inférieurs à 1², puis le développement mentionné précédemment pour obtenir la convergence vers 0. On a donc montré que

$$\phi_{Z_n}(\theta) \xrightarrow[n \rightarrow \infty]{} e^{-\sigma^2 \theta^2 / 2} = \phi_Z(\theta),$$

pour tout $\theta \in \mathbb{R}$ et cela montre la convergence en loi désirée. □

Remarque 5.2.2. Le théorème central limite est un exemple de convergence en loi qui ne peut pas être améliorée en une des autres convergences vues au Chapitre 4, pour lesquelles le choix précis de la v.a limite Z est important (i.e. sa définition pour \mathbb{P} -presque tout $\omega \in \Omega$, pas seulement sa loi). Une manière non rigoureuse de voir cela est la Figure 5.1.

2. Pour montrer cette inégalité, on écrit $z^n - w^n = (z - w) \sum_{k=0}^{n-1} z^k w^{n-1-k}$, puis on utilise l'inégalité triangulaire.

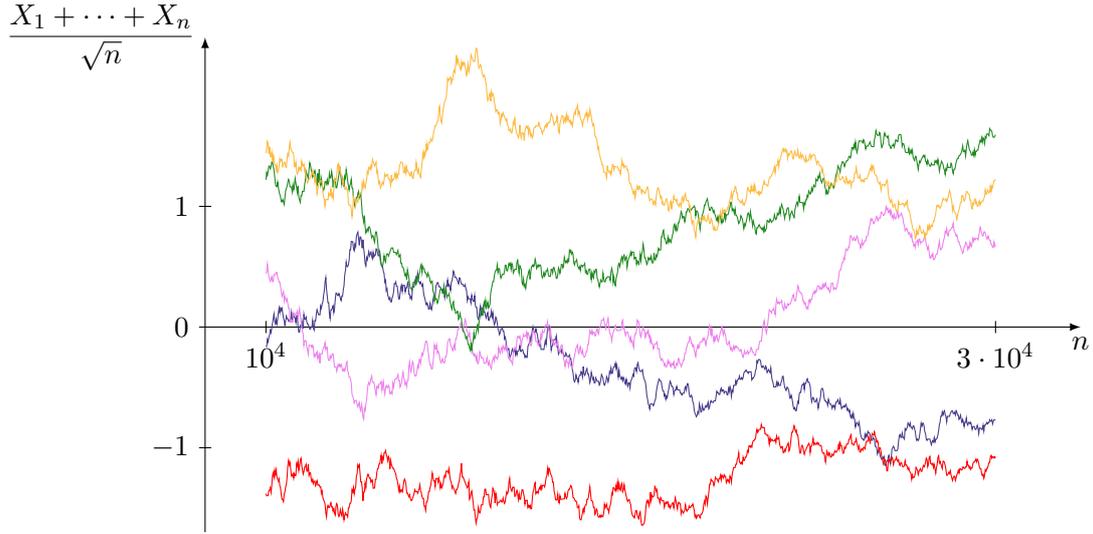


FIGURE 5.1 – Représentation de $(X_1 + \dots + X_n)/\sqrt{n}$ où $(X_k)_{k \geq 1}$ est une suite de v.a. i.i.d. de loi $\frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$, qui a moyenne nulle. Cinq réalisations sont représentées dans des couleurs différentes. On peut voir que cette quantité reste d'ordre 1, mais ne converge pas vers une limite pour un ω fixé (c'est-à-dire le long d'une des courbes représentées). La convergence n'a lieu qu'en loi : si l'on représentait un grand nombre de ses réalisations (plus que cinq!) alors à chaque n , elles seraient distribuées approximativement comme une gaussienne.

Plus rigoureusement, supposons que la convergence vers Z ait lieu en probabilité. On peut décomposer

$$\frac{X_1 + \dots + X_{2n} - 2nm}{\sqrt{2n}} = \frac{1}{\sqrt{2}} \left(\frac{X_1 + \dots + X_n - nm}{\sqrt{n}} + \frac{X_{n+1} + \dots + X_{2n} - nm}{\sqrt{n}} \right),$$

et donc écrire

$$\begin{aligned} \frac{X_{n+1} + \dots + X_{2n} - nm}{\sqrt{n}} &= \sqrt{2} \cdot \frac{X_1 + \dots + X_{2n} - 2nm}{\sqrt{2n}} - \frac{X_1 + \dots + X_n - nm}{\sqrt{n}} \\ &\xrightarrow[n \rightarrow \infty]{\mathbb{P}} (\sqrt{2} - 1)Z, \end{aligned}$$

par stabilité par somme de la convergence en probabilité. Cela implique en particulier la convergence en loi. Mais d'autre part,

$$\frac{X_{n+1} + \dots + X_{2n} - nm}{\sqrt{n}} \text{ a la même loi que } \frac{X_1 + \dots + X_n - nm}{\sqrt{n}}.$$

Donc, comme la convergence en loi ne dépend que de la loi, on a également

$$\frac{X_{n+1} + \dots + X_{2n} - nm}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\text{loi}} Z.$$

Comme $(\sqrt{2}-1)Z$ et Z n'ont pas la même loi, c'est une contradiction car la limite d'une convergence en loi est unique à égalité en loi près.

Remarque 5.2.3. Plaçons nous sous les hypothèses du théorème, on donne ici d'autres manières d'écrire la conclusion du théorème. Tout d'abord, comme Z/σ a la loi $\mathcal{N}(0, 1)$ et que la convergence en loi est stable par composition par une fonction continue (ici $x \mapsto x/\sigma$), on a

$$\frac{X_1 + \dots + X_n - nm}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\text{loi}} Y,$$

où Y est une v.a. de loi $\mathcal{N}(0, 1)$. On peut alors en déduire que, pour tout intervalle $I \subset \mathbb{R}$,

$$\mathbb{P}\left(\frac{X_1 + \dots + X_n - nm}{\sigma\sqrt{n}} \in I\right) \xrightarrow[n \rightarrow \infty]{} \int_I \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx. \quad (5.1)$$

En effet, cela découle du fait que la fonction de répartition de Y est continue et du Corollaire 5.1.16.

5.2.2 Précision dans la méthode de Monte–Carlo

Le théorème central limite permet d'estimer combien de variables aléatoires il faut considérer dans la méthode de Monte–Carlo pour obtenir une certaine précision. Rappelons que pour estimer $\mathbb{E}[f(X)]$ avec X une v.a. réelle et $f: \mathbb{R} \rightarrow \mathbb{R}$ mesurable telle que $\mathbb{E}[|f(X)|] < \infty$, on génère une suite $(X_n)_{n \geq 1}$ de v.a. réelles i.i.d. de même loi que X et on calcule

$$\frac{f(X_1) + \cdots + f(X_n)}{n}$$

qui approche $\mathbb{E}[f(X)]$ p.s. par la loi forte des grands nombres. Supposons que l'on veuille approcher $\mathbb{E}[f(X)]$ à ε près avec probabilité au moins $1 - \alpha$, pour $\varepsilon > 0$ et $\alpha \in]0, 1[$ donnés. Cela signifie que l'on veut trouver n tel que

$$\mathbb{P}\left(\left|\frac{f(X_1) + \cdots + f(X_n)}{n} - \mathbb{E}[f(X)]\right| \leq \varepsilon\right) \geq 1 - \alpha$$

On peut trouver une valeur de n telle que cette inégalité soit satisfaite (de manière exacte) en utilisant les inégalités de concentrations vues au Chapitre 3. Mais ici, on va voir comment choisir n de sorte que cette inégalité soit asymptotiquement satisfaite (quand $n \rightarrow \infty$, i.e. $\varepsilon \rightarrow 0$) et pour cela on utilise le théorème central limite. Heuristiquement, à α fixé, il nous dit qu'il faut que n soit de l'ordre de $1/\varepsilon^2$, car les fluctuations de $(f(X_1) + \cdots + f(X_n))/n$ autour de la vraie moyenne sont d'ordre $1/\sqrt{n}$.

Plus précisément, supposons que $\mathbb{E}[f(X)^2] < \infty$ (sans quoi on ne peut pas appliquer le théorème central limite). Alors, en notant $\sigma^2 = \text{Var}(f(X))$ et en utilisant (5.1), on a, pour tout $a > 0$ fixé,

$$\begin{aligned} \mathbb{P}\left(\left|\frac{f(X_1) + \cdots + f(X_n)}{n} - \mathbb{E}[f(X)]\right| \leq \frac{a\sigma}{\sqrt{n}}\right) &= \mathbb{P}\left(\frac{f(X_1) + \cdots + f(X_n) - n\mathbb{E}[f(X)]}{\sigma\sqrt{n}} \in [-a, a]\right) \\ &\xrightarrow{n \rightarrow \infty} \int_{-a}^a \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx. \end{aligned}$$

Rappelons que l'on note Φ la fonction de répartition de la loi $\mathcal{N}(0, 1)$. On choisit a en terme de $\alpha \in]0, 1[$ de sorte que

$$\begin{aligned} 1 - \alpha &= \int_{-a}^a \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \quad \Leftrightarrow \quad 1 - \alpha = 2\Phi(a) - 1 \\ &\Leftrightarrow \quad a = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \end{aligned} \tag{5.2}$$

où l'on a utilisé dans la première équivalence le calcul suivant

$$\begin{aligned} \int_{-a}^a \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx &= \int_{-\infty}^a \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx - \int_{-\infty}^{-a} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \Phi(a) - \int_a^{\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\ &= \Phi(a) - (1 - \Phi(a)) = 2\Phi(a) - 1. \end{aligned}$$

Puis on choisit n tel que

$$\varepsilon = \frac{a\sigma}{\sqrt{n}} \quad \Leftrightarrow \quad n = \frac{a^2\sigma^2}{\varepsilon^2}.$$

Alors, la convergence ci-dessus se réécrit en l'approximation

$$\mathbb{P}\left(\left|\frac{f(X_1) + \cdots + f(X_n)}{n} - \mathbb{E}[f(X)]\right| \leq \varepsilon\right) \approx 1 - \alpha.$$

On a donc bien trouvé comment choisir n en fonction de α et ε pour répondre approximativement à la question (où l'approximation se justifie dans la limite $\varepsilon \rightarrow 0$ avec α fixé).

Remarque 5.2.4. On peut voir que le choix de n est proportionnel à la variance σ^2 . En particulier, plus la variance est grande, moins on est précis dans une méthode de Monte–Carlo et donc plus on doit prendre n grand (sans grande surprise). Une difficulté ici est que généralement on ne connaît pas la valeur de σ^2 : en effet, on essaie d’estimer numériquement $\mathbb{E}[f(X)]$ car on ne sait pas le calculer de manière théorique, mais il y a alors peu de chance que l’on sache calculer $\text{Var}(f(X))$. Une solution peut être d’obtenir une borne supérieure σ_{\max}^2 pour $\text{Var}(f(X))$ et prendre $n = a^2 \sigma_{\max}^2 / \varepsilon^2$, qui correspond au “pire des cas”. Si on ne voit pas comment majorer $\text{Var}(f(X))$, il est bon au moins d’avoir en tête que n doit être proportionnel à $1/\varepsilon^2$: pour gagner une décimale de plus en précision, il faut prendre n cent fois plus grand.

5.2.3 Intervalles de confiance asymptotiques

On a vu en Section 3.2 comment construire des intervalles de confiance à n fixé à partir des inégalités de concentration. Ici on utilise le théorème central limite pour construire des intervalles de confiance *asymptotiques*, c’est-à-dire qu’ils satisfont le niveau de confiance désiré dans la limite $n \rightarrow \infty$. On se place dans le même cadre qu’en Section 3.2 : soit $(X_k)_{k \geq 1}$ une suite de v.a. réelles X_1, \dots, X_n i.i.d. dont la loi appartient à une certaine famille $(P_\theta)_{\theta \in \Theta}$, où $\Theta \subset \mathbb{R}$.

Définition 5.2.5. Soit $\alpha \in]0, 1[$. Un intervalle de confiance asymptotique pour θ de niveau $1 - \alpha$ est la donnée, pour chaque $n \geq 1$, d’un intervalle aléatoire $I_n = [F_n(X_1, \dots, X_n), G_n(X_1, \dots, X_n)]$ avec $F_n, G_n : \mathbb{R}^n \rightarrow \mathbb{R}$ mesurables (ne dépendant pas de θ), tel que pour tout $\theta \in \Theta$, si les v.a. X_k ont loi P_θ , alors

$$\liminf_{n \rightarrow \infty} \mathbb{P}(\theta \in I_n) \geq 1 - \alpha.$$

Contrairement au cas de l’intervalle de confiance non-asymptotique vu en Section 3.2 où la définition a un sens à n fixé, ici la définition n’a de sens que si l’on est capable de définir une suite d’intervalles pour pouvoir parler du comportement limite. En particulier, c’est pour cela qu’ici on souligne dans la notation I_n la dépendance en n de l’intervalle.

Ce type d’intervalle de confiance est moins satisfaisant, car on ne connaît pas la vitesse de convergence de $\mathbb{P}(\theta \in I_n)$, donc, en pratique, quand on veut l’appliquer à des données pour un certain n fixé, on n’est pas sûr de si ce n est suffisamment grand pour justifier l’approximation de $\mathbb{P}(\theta \in I_n)$ par sa limite.

Exemple 5.2.6. Reprenons le cas d’un sondage traité dans l’exemple 5.2.6. On a $\Theta = [0, 1]$, $P_\theta = \mathcal{B}(\theta)$ (la loi de Bernoulli de paramètre θ). Soit $(X_k)_{k \geq 1}$ est une suite de v.a. i.i.d. de loi P_θ . Comme $\mathbb{E}[X_k] = \theta$, il est naturel d’utiliser la moyenne empirique $\bar{X}_n = (X_1 + \dots + X_n)/n$ pour construire notre intervalle de confiance asymptotique. Les v.a. X_k sont dans L^2 et $\text{Var}(X_k) = \theta(1 - \theta)$, donc, par le théorème central limite (la version (5.1)), pour tout $a > 0$ fixé,

$$\mathbb{P}\left(\left|\bar{X}_n - \theta\right| \leq \frac{a\sqrt{\theta(1-\theta)}}{\sqrt{n}}\right) = \mathbb{P}\left(\left|\frac{X_1 + \dots + X_n - n\theta}{\sqrt{\theta(1-\theta)}\sqrt{n}}\right| \leq a\right) \xrightarrow{n \rightarrow \infty} \int_{-a}^a \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx.$$

On choisit alors a en terme de $\alpha \in]0, 1[$ de sorte que

$$1 - \alpha = \int_{-a}^a \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \quad \Leftrightarrow \quad a = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right),$$

en procédant comme en (5.2). Pour cette valeur de a , on a donc

$$\mathbb{P}\left(\theta \in \left[\bar{X}_n - \frac{a\sqrt{\theta(1-\theta)}}{\sqrt{n}}, \bar{X}_n + \frac{a\sqrt{\theta(1-\theta)}}{\sqrt{n}}\right]\right) \xrightarrow{n \rightarrow \infty} 1 - \alpha. \quad (5.3)$$

Mais l’intervalle ci-dessus dépend de θ ! Il faut donc choisir I_n de sorte qu’il contienne l’intervalle ci-dessus pour tout $\theta \in [0, 1]$. Pour cela, on remarque que $\max_{\theta \in [0, 1]} \theta(1 - \theta) = 1/4$, donc on prend

$$I_n := \left[\bar{X}_n - \frac{a}{2\sqrt{n}}, \bar{X}_n + \frac{a}{2\sqrt{n}}\right].$$

Alors, comme $\mathbb{P}(\theta \in I_n)$ est supérieure à la probabilité dans (5.3), on en déduit que

$$\liminf_{n \rightarrow \infty} \mathbb{P}(\theta \in I_n) \geq 1 - \alpha.$$

Donc, I_n est un intervalle de confiance asymptotique pour θ de niveau $1 - \alpha$.