



Notes de cours de
MAT7070 - Mesure et probabilités

Hélène Guérin
Université du Québec à Montréal
Département de Mathématiques
guerin.helene@uqam.ca

Automne 2024

Table des matières

1	Rappels et compléments en analyse	5
1.1	Rappels sur les suites et séries numériques	5
1.1.a)	Convergence	5
1.1.b)	Suites extraites	6
1.1.c)	Critère de Cauchy	6
1.1.d)	Limite supérieure et limite inférieure d'une suite	6
1.1.e)	Séries numériques	6
1.1.f)	Exercices	7
1.2	Suites de fonctions	8
1.2.a)	Notions de convergence	8
1.2.b)	Exercices	8
1.3	Un peu de topologie	9
1.3.a)	Exercices	9
1.4	Nombres complexes	9
1.4.a)	Définition	9
1.4.b)	forme trigonométrique	10
1.4.c)	Forme exponentielle	11
1.4.d)	Exercices	11
2	Théorie de la mesure	13
2.1	Notion de tribu	13
2.1.a)	Tribu engendrée	14
2.1.b)	Image réciproque d'une tribu	14
2.1.c)	Tribu produit	15
2.1.d)	La tribu borélienne	15
2.1.e)	Exercices sur les tribus	16
2.2	Applications mesurables	18
2.2.a)	Les fonctions boréliennes	19
2.2.b)	Propriétés de stabilité	19
2.2.c)	Exercices sur les fonctions mesurables	21
2.3	Approximation des fonctions mesurables	23
2.4	Mesures positives et mesure de Lebesgue	24
2.4.a)	Mesures discrètes	27
2.4.b)	Mesure de Lebesgue	28
2.4.c)	Exercices sur les mesures	30
2.5	Construction de l'intégrale	32
2.5.a)	Intégration des fonctions étagées positives	32

2.5.b)	Intégration des fonctions mesurables positives	33
2.5.c)	Intégration des fonctions mesurables	36
2.5.d)	Intégrale par rapport à une mesure discrète	37
2.5.e)	Intégrale de Lebesgue et comparaison avec l'intégrale de Riemann	37
2.5.f)	Exercices sur l'intégrale	40
2.6	Un aperçu de l'intégrale multiple	41
2.6.a)	Exercices sur l'intégrale multiple	47
3	Variables aléatoires, Espérance et Indépendance	51
3.1	Espaces de probabilité	51
3.1.a)	Définitions et propriétés	51
3.1.b)	Exercices sur les espaces de probabilités	53
3.2	Variables aléatoires	54
3.2.a)	Définition	54
3.2.b)	Loi d'une variable aléatoire	54
3.2.c)	Classe monotone	55
3.2.d)	Fonction de répartition	56
3.2.e)	Exercices sur les variables aléatoires	58
3.3	Exemples de variables aléatoires	59
3.3.a)	Variables discrètes	59
3.3.b)	Théorème de Radon-Nikodym	60
3.3.c)	Variables à densité	61
3.3.d)	Lois marginales d'un vecteur aléatoire	62
3.4	Espérance d'une variable aléatoire	64
3.4.a)	Exemples de calcul d'espérance	65
3.4.b)	Inégalités importantes	66
3.4.c)	Identification de la loi d'une variable aléatoire par l'espérance	67
3.4.d)	Exercices sur l'espérance	69
3.5	Moments d'une variable aléatoire	71
3.5.a)	Les espaces L^p	71
3.5.b)	Variance, Covariance	74
3.5.c)	Exercices sur les moments	75
3.6	Fonction caractéristique	76
3.6.a)	Théorème de Fourier	77
3.6.b)	Théorème d'inversion de Fourier	79
3.6.c)	Formulaire des lois classiques	80
3.6.d)	Exercices sur les fonctions caractéristiques	80
3.7	Notion d'indépendance	82
3.7.a)	Critères d'indépendance	84
3.7.b)	Événements asymptotiques	85
3.7.c)	Lemmes de Borel-Cantelli	87
3.7.d)	Exercices sur l'indépendance	88
4	Différentes notions de convergence et théorèmes limites	93
4.1	Rappels sur les ensembles négligeables	93
4.2	Convergences trajectorielles	93
4.2.a)	Convergence presque-sûre	93

4.2.b)	Convergence dans L^p	95
4.2.c)	Exercices	98
4.2.d)	Convergence en probabilité	99
4.2.e)	Uniforme intégrabilité	101
4.2.f)	Critères de convergence	104
4.3	Convergence en loi	105
4.4	Résumé des liens entre les différentes convergences	109
4.5	Exercices sur les différentes convergence	109
4.6	Théorèmes limites classiques	110
4.6.a)	Loi des grands nombres : LGN	110
4.6.b)	Théorème Central Limite (TCL)	113
4.6.c)	Preuve de la loi forte des grands nombres	116
4.6.d)	Extensions	119
5	Esperance Conditionnelle et Martingales	121
5.1	L'espérance conditionnelle	121
5.1.a)	Heuristique : signification intuitive et utilisation de l'espérance conditionnelle	121
5.1.b)	Définitions	122
5.1.c)	Propriétés fondamentales de l'espérance conditionnelle	124
5.2	La notion de loi conditionnelle	125
5.2.a)	Cas gaussien	126
5.3	Exercices sur l'espérance conditionnelle	128
5.4	Les martingales	129
5.4.a)	Définition et Premières propriétés	129
5.4.b)	Quelques exemples de martingales	130
5.5	Martingales et temps d'arrêt	132
5.6	Inégalités sur les martingales	134
5.7	Décomposition de Doob	136
5.8	Théorèmes limites et Martingales	136
5.8.a)	Convergence p.s.	136
5.8.b)	Convergence dans L^1	137
5.8.c)	Martingales dans L^2	138
5.9	Exercices sur les martingales	138
	Bibliographie	139

Chapitre 1

Rappels et compléments en analyse

1.1 Rappels sur les suites et séries numériques

Dans cette section, quelques résultats et définitions sur les suites et les séries sont rappelés. Il vous est fortement conseillé de relire vos cours précédents d'analyse.

Définition 1.1.1. Une suite réelle (ou complexe) est une famille de réels (complexes) indexée par $\mathbb{N}$. La suite $u \in \mathbb{R}^{\mathbb{N}}$ ($u \in \mathbb{C}^{\mathbb{N}}$) est notée $u = (u_n)_{n \in \mathbb{N}}$.

On dit que la suite $(u_n)_{n \in \mathbb{N}}$ est bornée s'il existe $K > 0$ tel que $\forall n \in \mathbb{N}, |u_n| \leq K$.

On dit que la suite réelle $(u_n)_{n \in \mathbb{N}}$ est majorée s'il existe $M \in \mathbb{R}$ tel que $\forall n \in \mathbb{N}, u_n \leq M$.

On dit que la suite réelle $(u_n)_{n \in \mathbb{N}}$ est minorée s'il existe $m \in \mathbb{R}$ tel que $\forall n \in \mathbb{N}, u_n \geq m$.

La suite réelle $(u_n)_{n \in \mathbb{N}}$ est dite croissante si $\forall n \in \mathbb{N}, u_n \leq u_{n+1}$.

La suite réelle $(u_n)_{n \in \mathbb{N}}$ est dite décroissante si $\forall n \in \mathbb{N}, u_n \geq u_{n+1}$.

Une suite sera notée indifféremment $(u_n)_{n \in \mathbb{N}}$ ou $(u_n)_{n \geq 0}$ ou (u_n) .

1.1.a) Convergence

Définition 1.1.2. Une suite $(u_n)_{n \in \mathbb{N}}$ est **convergente** s'il existe un réel (ou un complexe) ℓ , tel que

$$\forall \varepsilon > 0 \exists n_0 \in \mathbb{N} : \forall n \geq n_0, |u_n - \ell| \leq \varepsilon.$$

ℓ est appelé **limite** de la suite (u_n) . Si une suite n'est pas convergente, elle est dite divergente.

On dit que la suite admet la limite $+\infty$ (resp. $-\infty$) si

$$\forall M > 0 \exists n_0 \geq 0, \forall n \geq n_0, \text{ on a } u_n \geq M \text{ (resp. } u_n \leq -M).$$

Une suite qui a pour limite $\pm\infty$ est alors divergente.

Propriété 1.1.3. Si une suite converge, alors sa limite est unique.

Toute suite convergente est bornée.

Théorème 1.1.4. Toute suite réelle monotone (i.e., croissante ou décroissante) et bornée est convergente.

Théorème 1.1.5. Soit $(u_n)_{n \geq 0}$ une suite convergente vers ℓ et f une fonction continue au point ℓ . Alors la suite $(f(u_n))_{n \geq 0}$ converge vers $f(\ell)$.

1.1.b) Suites extraites

Définition 1.1.6. Soit $(u_n)_{n \in \mathbb{N}}$ une suite. La suite $(v_n)_{n \in \mathbb{N}}$ est appelée **suite extraite** ou **sous-suite** de $(u_n)_{n \in \mathbb{N}}$ s'il existe une application $\varphi : \mathbb{N} \rightarrow \mathbb{N}$ strictement croissante telle que $\forall n \in \mathbb{N}, v_n = u_{\varphi(n)}$.

Théorème 1.1.7. Toute suite extraite d'une suite convergente est convergente et de même limite.

Conséquence. Si on peut extraire deux suites convergentes ayant des limites distinctes d'une suite alors la suite diverge.

Si on peut extraire une sous-suite divergente d'une suite, alors la suite diverge.

Remarque. Soit $(u_n)_{n \in \mathbb{N}}$ une suite avec $(u_{2n})_{n \in \mathbb{N}}$ et $(u_{2n+1})_{n \in \mathbb{N}}$ qui convergent vers la même limite ℓ . Alors la suite $(u_n)_{n \in \mathbb{N}}$ converge vers ℓ .

Théorème 1.1.8 (Théorème de Bolzano-Weierstrass). De toute suite bornée, on peut extraire une suite convergente.

1.1.c) Critère de Cauchy

Définition 1.1.9. On dit qu'une suite $(u_n)_{n \in \mathbb{N}}$ est de Cauchy si

$$\forall \varepsilon > 0 \quad \exists n_0 \in \mathbb{N} : \forall p, q \geq n_0, |u_p - u_q| \leq \varepsilon.$$

Théorème 1.1.10. Une suite réelle (ou complexe) $(u_n)_{n \in \mathbb{N}}$ converge si et seulement si elle est de Cauchy.

L'espace des réels est par conséquent un espace complet.

1.1.d) Limite supérieure et limite inférieure d'une suite

Notation : On note $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$.

Définition 1.1.11. Soit $(u_n)_{n \geq 0}$ une suite à valeurs dans $\overline{\mathbb{R}}$.

On appelle **point d'accumulation** (aussi appelé **valeur d'adhérence**) une quantité $\ell \in \overline{\mathbb{R}}$ tel qu'il existe une sous-suite $(u_{\varphi(n)})_{n \geq 0}$ avec $\lim_{n \rightarrow +\infty} u_{\varphi(n)} = \ell$.

Le plus grand (resp. petit) point d'accumulation de la suite (u_n) est notée $\limsup u_n$ (resp. $\liminf u_n$).

Les quantités $\limsup u_n$ et $\liminf u_n$ sont des éléments de $\overline{\mathbb{R}}$, on a

$$\limsup_{n \rightarrow +\infty} u_n = \inf_{n \geq 0} \sup_{k \geq n} u_k \quad \text{et} \quad \liminf_{n \rightarrow +\infty} u_n = \sup_{n \geq 0} \inf_{k \geq n} u_k.$$

Si $(u_{\varphi(n)})$ est une sous-suite de limite ℓ de la suite (u_n) alors $\limsup u_n \leq \ell \leq \liminf u_n$.

Par ailleurs, une suite (u_n) admet une limite quand $n \rightarrow +\infty$ si et seulement si $\limsup u_n = \liminf u_n$.

1.1.e) Séries numériques

Définition 1.1.12. Soit $(u_n)_{n \in \mathbb{N}}$ une suite réelle ou complexe. On appelle **série de terme général** u_n , notée $\sum u_n$, la suite des sommes partielles $(S_n)_{n \in \mathbb{N}}$ définie par

$$S_n = \sum_{k=0}^n u_k.$$

On parle de série de nombres réels si tous les u_n sont réels.

On dit que la série converge si la suite des sommes partielles $(S_n)_{n \in \mathbb{N}}$ converge. Sinon, on dit que la série diverge.

En cas de convergence, la limite de $(S_n)_{n \in \mathbb{N}}$ est appelée somme de la série et est notée $\sum_{n=0}^{+\infty} u_n$.

Remarque. Soit $n_0 \in \mathbb{N}$. Si $|q| < 1$, on a

$$\sum_{n=n_0}^{+\infty} q^n = \frac{q^{n_0}}{1-q}.$$

On remarque que les premiers termes d'une série ont une influence sur la valeur de la somme, alors que pour une suite convergente, les premiers termes n'ont aucune influence sur la valeur de la limite.

Proposition 1.1.13 (Condition nécessaire de convergence).

Si $\sum u_n$ est une série convergente alors $\lim_{n \rightarrow +\infty} u_n = 0$.

Théorème 1.1.14. Soient $\sum u_n$ et $\sum v_n$ deux séries à termes positifs avec pour tout $n \in \mathbb{N}$, $u_n \leq v_n$. Si $\sum v_n$ converge, alors $\sum u_n$ converge et

$$0 \leq \sum_{n=0}^{+\infty} u_n \leq \sum_{n=0}^{+\infty} v_n.$$

Si $\sum u_n$ diverge, alors $\sum v_n$ diverge.

1.1.f) Exercices

Exercice 1.1.1. Montrer la Propriété 1.1.3.

Exercice 1.1.2. Montrer le Théorème 1.1.4

Exercice 1.1.3. Montrer le Théorème 1.1.5

Exercice 1.1.4. Montrer que si $|q| < 1$, on a

$$\sum_{n=n_0}^{+\infty} q^n = \frac{q^{n_0}}{1-q}.$$

Exercice 1.1.5. Trouver la limite supérieure et la limite inférieure des suites suivantes

1. $u_n = n^2$;
2. $v_n = \frac{n+2}{\cos(n\pi)n+3}$.
3. $w_n = a_n$ si n est impair et $w_n = b_n$ si n est pair, avec $a_n = \frac{2}{3} - \frac{1}{n}$ et $b_n = 1 + \frac{1}{n}$.

Exercice 1.1.6. Pour chacune des suites $(a_{ij})_{i \geq 0, j \geq 0}$ suivantes calculer $\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_{ij}$ et $\sum_{j=0}^{\infty} \sum_{i=0}^{\infty} a_{ij}$:

- $\forall j \geq 0, a_{0j} = 1$ et pour $i \geq 1, a_{ij} = -2^{-i}$;
- $\forall i, j \geq 0, a_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i > j \\ -2^{i-j} & \text{si } i < j \end{cases}$

Que remarquez-vous ?

1.2 Suites de fonctions

On considère une suite de fonctions $(f_n)_{n \in \mathbb{N}}$ définies sur un espace E à valeurs dans $\mathbb{R}$ ou $\mathbb{C}$, i.e. $\forall n \in \mathbb{N}$ $f_n : E \rightarrow \mathbb{R}$ (ou $\mathbb{C}$) et f une fonction définie sur E à valeurs dans $\mathbb{R}$ ou $\mathbb{C}$.

1.2.a) Notions de convergence

Définition 1.2.1. Soit f une fonction de E dans $\mathbb{R}$ ou $\mathbb{C}$.

1. On dit que la suite (f_n) converge ponctuellement (ou simplement) vers f sur E si $\forall x \in E$ la suite $(f_n(x))_{n \in \mathbb{N}}$ converge, i.e.

$$\forall x \in E, \forall \varepsilon > 0, \exists n_0 \in \mathbb{N} : \forall n \geq n_0, \quad |f_n(x) - f(x)| < \varepsilon.$$

2. On dit que la suite $(f_n)_{n \in \mathbb{N}}$ converge uniformément vers f sur E si

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} : \forall n \geq n_0, \forall x \in E, \quad |f_n(x) - f(x)| < \varepsilon.$$

Remarque. On voit facilement d'après la définition que la convergence uniforme implique la convergence simple, mais la réciproque est fausse.

La convergence uniforme est liée à la norme uniforme.

Définition 1.2.2. Soit f une fonction définie sur E . On appelle norme uniforme de f sur E , et on note $\|f\|_\infty$ la quantité

$$\|f\|_\infty = \sup_{x \in E} |f(x)| = \sup \{|f(x)| : x \in E\}.$$

Cette norme est aussi appelée norme infini.

Lorsqu'on veut spécifier l'espace sur lequel la fonction f est étudiée, on note $\|f\|_{E,\infty}$ la norme infinie sur E .

Proposition 1.2.3. La suite de fonctions $(f_n)_{n \in \mathbb{N}}$ converge uniformément vers f sur E si et seulement si la suite de réels $(\|f_n - f\|_\infty)_{n \geq 0}$ converge vers 0, i.e.

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} : \forall n \geq n_0, \quad \|f_n - f\|_\infty < \varepsilon.$$

Proposition 1.2.4. La limite d'une suite de fonctions continues qui converge simplement n'est pas forcément continue. Par contre, si la convergence est uniforme alors la limite d'une suite de fonctions continues est continue.

1.2.b) Exercices

Exercice 1.2.1. Montrer que la suite (f_n) définie par $f_n(x) = \frac{x}{n}$ converge uniformément sur $E = [0, 1]$, ponctuellement sur $E = \mathbb{R}^+$, mais ne converge pas uniformément sur $E = \mathbb{R}^+$.

Exercice 1.2.2. Étudier la convergence (simple et uniforme) de la suite de fonctions $(f_n)_{n \geq 0}$ définie sur $[0, 1]$ par

$$f_n(x) = x^n.$$

1.3 Un peu de topologie

Soit E un espace (la plupart du temps on considérera les espaces $\mathbb{R}$, $\mathbb{R}^d$).

Soient A, B deux sous ensembles de E : $A \subset E$, $B \subset E$ (dans ce cours, le symbole $\subset$ symbolise l'inclusion au sens large, i.e. $\subseteq$).

On note A^c (souvent noté $\bar{A}$ en probabilité) le complémentaire de A dans E : $A^c = \{x \in E : x \notin A\}$.

Si $B \subset A$, on définit la différence $A \setminus B = A \cap B^c$.

Définition 1.3.1. Soit $O \subset E$, avec $E = \mathbb{R}$ ou $\mathbb{R}^d$. On dit que O est un ensemble ouvert si $\forall x \in O$, il existe $r > 0$ tel que $B(x, r) \subset O$, où $B(x, r) = \{y \in E : |x - y| < r\}$ est la boule centrée en x de rayon r .

Un sous-ensemble F de E est un fermé si F^c est un ouvert.

On définit l'intérieur d'un ensemble $A \subset E$ de la façon suivante : $\mathring{A} = \{x \in A : \exists r > 0, B(x, r) \subset A\}$.

Un ensemble A est dit borné s'il existe $M > 0$ tel que $\forall x \in A$ $|x| \leq M$.

Un ensemble fermé borné est dit compact.

Toute partie majorée de $\mathbb{R}$ admet une borne supérieure finie, qui est définie comme le plus petit des majorants, et par convention la borne supérieure d'une partie non majorée est $+\infty$.

Toute partie minorée de $\mathbb{R}$ admet une borne inférieure finie, qui est définie comme le plus grand des minorants, et par convention la borne inférieure d'une partie non minorée est $-\infty$.

Un ensemble A est dit **dénombrable** s'il existe une bijection entre l'ensemble $\mathbb{N}$ des entiers naturels et A . Par abus, on parlera parfois d'ensemble dénombrable pour des ensembles finis ou dénombrables.

1.3.a) Exercices

Exercice 1.3.1. Montrer que l'ensemble des entiers relatifs $\mathbb{Z} = \{p : p \in \mathbb{N} \text{ ou } -p \in \mathbb{N}\}$ est dénombrable.

Exercice 1.3.2. (*) Montrer que l'ensemble des nombres rationnels $\mathbb{Q} = \left\{\frac{p}{q} : p \in \mathbb{Z}, q \in \mathbb{N}^*\right\}$ est dénombrable.

Exercice 1.3.3. (*) Soit $(A_n)_{n \geq 1}$ une suite d'ensembles dénombrables. Montrer que $\bigcup_{n=1}^{\infty} A_n$ est un ensemble dénombrable.

Exercice 1.3.4. Montrer qu'une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ est continue si et seulement si l'image réciproque par f de tout ouvert est un ouvert.

Rappel : une fonction f est continue sur $\mathbb{R}$ si $\forall a \in \mathbb{R}$, $\forall \varepsilon > 0$, il existe $\eta > 0$ tel que $\forall x \in \mathbb{R} : |x - a| < \eta$, on a $|f(x) - f(a)| < \varepsilon$.

1.4 Nombres complexes

1.4.a) Définition

On définit l'ensemble des nombres complexes, noté $\mathbb{C}$, l'ensemble des nombres de la forme $a + ib$, avec a et b deux réels quelconques :

$$\mathbb{C} = \{a + ib, (a, b) \in \mathbb{R}^2\}.$$

où i est un nombre (imaginaire) tel que $i^2 = -1$.

Deux nombres complexes $z = a + ib$ et $z' = c + id$ sont égaux si et seulement si

$$a = c \quad \text{et} \quad b = d.$$

Soit $z = a + ib$ un nombre complexe. Le réel a est appelé **partie réelle** et le réel b est appelé **partie imaginaire** de z . On note

$$z = \operatorname{Re}(z) + i\operatorname{Im}(z)$$

Si $z \in \mathbb{C}$ avec $\operatorname{Im}(z) = 0$, alors z est réel. Si $z \in \mathbb{C}$ avec $\operatorname{Re}(z) = 0$, alors on dit que z est un imaginaire pur.

L'ensemble des nombres complexes $\mathbb{C}$ contient l'ensemble des réels $\mathbb{R}$.

Définition 1.4.1. Soit $z = a + ib$ un nombre complexe. On définit

- le **module** de z , noté $|z|$, le réel positif $|z| = \sqrt{a^2 + b^2}$,
- le **conjugué** de z , noté $\bar{z}$, le nombre complexe $\bar{z} = a - ib$.

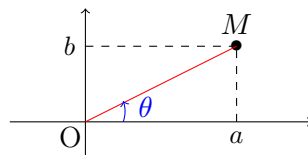
Propriété 1.4.2. Soit z, z' deux nombres complexes.

- $z \cdot \bar{z} = |z|^2$,
- si $z \neq 0$, l'inverse de z est $z^{-1} = \frac{1}{z} = \frac{\bar{z}}{|z|^2}$,
- $\operatorname{Re}(z) = \frac{z + \bar{z}}{2}$ et $\operatorname{Im}(z) = \frac{z - \bar{z}}{2i}$,
- $\overline{zz'} = \bar{z}\bar{z'}$, $\overline{\left(\frac{z}{z'}\right)} = \frac{\bar{z}}{\bar{z'}}$ et $\overline{\bar{z}} = z$,
- $|zz'| = |z||z'|$, $\left|\frac{z}{z'}\right| = \frac{|z|}{|z'|}$ et $|\bar{z}| = |z|$,
- $z \in \mathbb{R} \Leftrightarrow \operatorname{Im}(z) = 0 \Leftrightarrow \bar{z} = z$.

Cette écriture du complexe $z = a + ib$ s'appelle la forme **algébrique** du complexe z .

1.4.b) forme trigonométrique

Au nombre complexe $z = a + ib$, on peut associer le point $M(a, b)$ dans le plan munit du repère orthonormé $(O, \vec{i}, \vec{j})$.



Par le théorème de pythagore, on remarque que $OM = |z| = \sqrt{a^2 + b^2}$.

Supposons que $|z| = 1$. Par définition du sinus et du cosinus, il existe un unique angle $\theta \in [0, 2\pi[$ tel que

$$z = a + ib = \cos(\theta) + i \sin(\theta).$$

De plus, si un nombre complexe est de la forme $z = \cos(\theta) + i \sin(\theta)$ alors $|z| = 1$ (car $\cos^2(\theta) + \sin^2(\theta) = 1$).

Proposition 1.4.3. Soit $z \in \mathbb{C}^*$, il existe un unique $\theta \in [0, 2\pi[$ tel que

$$z = |z|(\cos(\theta) + i \sin(\theta)).$$

Le nombre θ s'appelle l'**argument** de z , et noté $\arg(z)$.

Cette écriture du nombre complexe z s'appelle **forme trigonométrique** de z .

1.4.c) Forme exponentielle

Définition 1.4.4. On définit l'exponentielle complexe du nombre complexe de module 1 par

$$e^{i\theta} := \cos(\theta) + i \sin(\theta).$$

Par conséquent, pour un nombre complexe z de module $r = |z|$ et d'argument θ , on a

$$z = re^{i\theta}.$$

Cette écriture s'appelle la **forme exponentielle** du nombre complexe z et le couple (r, θ) s'appelle **coordonnées polaires** de z .

Soit $z = re^{i\theta}$ et $z' = r'e^{i\theta'}$ deux nombres complexes sous forme exponentielle. Alors,

$$zz' = rr'e^{i(\theta+\theta')} \text{ et } z^n = r^n e^{in\theta} \text{ pour } n \in \mathbb{Z}.$$

1.4.d) Exercices

Exercice 1.4.1. Faire la preuve de la propriété 1.4.2

Exercice 1.4.2. Trouver la partie réelle et la partie imaginaire du nombre complexe

$$z = \frac{2 + 4i}{1 - 2i}.$$

Exercice 1.4.3. On considère le nombre complexe $z = 2 + 2i$. Donner sa forme trigonométrique. Chercher les nombres complexes w tels que $w^2 = z$.

Exercice 1.4.4. On considère le nombre complexe $z = (1 - \sqrt{3}) - i(1 + \sqrt{3})$.

1. Écrire z^2 sous forme algébrique.
2. En déduire le module et l'argument de z .
3. En déduire les valeurs exactes de

$$\cos\left(\frac{5\pi}{12}\right) \quad \text{et} \quad \sin\left(\frac{5\pi}{12}\right).$$

Chapitre 2

Théorie de la mesure

La théorie de la mesure est une théorie générale sur l'intégration. Dans ce chapitre nous verrons les notions de la théorie de la mesure, l'intégrale de Lebesgue et quelques théorèmes classiques de convergence.

On se place sur un espace général E , mais la plupart du temps E sera $\mathbb{R}$, $\mathbb{R}^d$ (avec $d \geq 1$), ou un ensemble dénombrable $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, ...

2.1 Notion de tribu

On introduit tout d'abord la notion de tribu. Une mesure sera définie sur une tribu, et les seuls espaces que l'on pourra mesurer seront les éléments de la tribu. Il s'agit donc d'une notion essentielle de la théorie de la mesure.

Soit E un ensemble. On note $\mathcal{P}(E)$ l'ensemble des sous-ensembles (des parties) de E .

Exemple. Soit $E = \{1, 2, 3\}$. Alors $\mathcal{P}(E) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, E\}$.

Si E est un ensemble fini de cardinal n , alors $\mathcal{P}(E)$ est un ensemble fini de cardinal 2^n .

Définition 2.1.1. Une **tribu** (aussi appelée σ -algèbre) $\mathcal{A}$ sur E est un sous-ensemble non vide de $\mathcal{P}(E)$ tel que

1. l'ensemble vide appartient à $\mathcal{A}$: $\emptyset \in \mathcal{A}$,
2. $\mathcal{A}$ est stable par passage au complémentaire : si $A \in \mathcal{A}$ alors $A^c \in \mathcal{A}$,
3. $\mathcal{A}$ est stable par union dénombrable : si $(A_n)_{n \geq 0}$ suite d'éléments de $\mathcal{A}$ alors $\bigcup_{n \geq 0} A_n \in \mathcal{A}$.

Conséquence. Si $\mathcal{A}$ est une tribu de E , alors

- $E \in \mathcal{A}$,
- $\mathcal{A}$ est stable par intersection dénombrable,
- $\mathcal{A}$ est stable par union et intersection finies (prendre $A_n = \emptyset$ ou $A_n = E$ pour $n \geq n_0$),
- $\mathcal{A}$ est stable par différence : $A, B \in \mathcal{A} \Rightarrow A \setminus B \in \mathcal{A}$ (car $A \setminus B = A \cap B^c$).

Exemple. La plus petite tribu de E est $\mathcal{A} = \{\emptyset, E\}$ et la plus grande est $\mathcal{A} = \mathcal{P}(E)$ (au sens de l'inclusion).

Proposition 2.1.2. L'intersection de tribus de E est encore une tribu de E .

Preuve. Soit $(\mathcal{A}_i)_{i \in I}$ une famille de tribus sur E . On pose $\mathcal{A} = \bigcap_{i \in I} \mathcal{A}_i$.

Comme $\emptyset \in \mathcal{A}_i$ pour tout $i \in I$, on a $\emptyset \in \mathcal{A}$.

Soit $A \in \mathcal{A}$, alors $A \in \mathcal{A}_i$ et donc $A^c \in \mathcal{A}_i$ pour tout $i \in I$, d'où $A^c \in \mathcal{A}$. On établit de même la stabilité par union finie. $\square$

Remarque. *Attention ! L'union de tribus n'est en général pas une tribu.*

2.1.a) Tribu engendrée

Proposition 2.1.3. *Soit $\mathcal{A}$ un sous-ensemble de $\mathcal{P}(E)$ ($\mathcal{A} \subset \mathcal{P}(E)$ est un ensemble de parties de E). Il existe une plus petite tribu (au sens de l'inclusion) contenant tous les éléments de $\mathcal{A}$. Elle est appelée tribu engendrée par $\mathcal{A}$ et notée $\sigma(\mathcal{A})$.*

Preuve. On définit

$$\mathcal{M} = \{\mathcal{F} : \mathcal{F} \text{ tribu sur } E \text{ avec } \mathcal{A} \subset \mathcal{F}\}.$$

Cet ensemble est non vide car $\mathcal{P}(E) \in \mathcal{M}$. On définit

$$\mathcal{B} = \bigcap_{\mathcal{F} \in \mathcal{M}} \mathcal{F} = \{A \in \mathcal{P}(E) : \forall \mathcal{F} \in \mathcal{M}, A \in \mathcal{F}\}.$$

D'après la proposition précédente, $\mathcal{B}$ est une tribu. Par ailleurs, il est clair que $\mathcal{A} \subset \mathcal{B}$ et par définition, elle est contenue dans toute tribu contenant $\mathcal{A}$. $\square$

Remarque. • Soit $A \in \mathcal{P}(E)$, un sous-ensemble de E . On a alors

$$\sigma(\{A\}) = \{\emptyset, A, A^c, E\}$$

• Si $\mathcal{A}$ est une tribu sur E , alors $\sigma(\mathcal{A}) = \mathcal{A}$.

Remarque. Soit $A \subset E$ et $B \subset E$ avec $A \neq B$, $A \neq \overline{B}$, $A, B \neq \emptyset, E$. On remarque que

$$\sigma(\{A\}) \cup \sigma(\{B\}) = \{\emptyset, E, A, A^c, B, \overline{B}\}$$

Par conséquent, $\sigma(\{A\}) \cup \sigma(\{B\})$ n'est pas une tribu, car $A \cup B$ et/ou $A \cap B$ et/ou $A \cup \overline{B}$ et/ou $A^c \cup B$ ne sont pas dans $\sigma(\{A\}) \cup \sigma(\{B\})$. L'union de tribu n'est en général pas une tribu.

2.1.b) Image réciproque d'une tribu

Proposition 2.1.4. *Soit E et F deux ensembles et f une application de E dans F . Soit $\mathcal{B}$ une tribu sur F .*

Pour $B \in \mathcal{P}(F)$, on note $f^{-1}(B) = \{x \in E : f(x) \in B\}$. Alors $\mathcal{A} = f^{-1}(\mathcal{B})$ définie par

$$f^{-1}(\mathcal{B}) = \{f^{-1}(B) : B \in \mathcal{B}\}$$

est une tribu sur E , appelée tribu image réciproque de $\mathcal{B}$ par f .

Preuve. Preuve en exercice. On a besoin de vérifier que $f^{-1}(\emptyset) = \emptyset$, $\forall B \in \mathcal{P}(F)$, $f^{-1}(B^c) = f^{-1}(B)^c$ et $f^{-1}(\cup_{n \geq 0} B_n) = \cup_{n \geq 0} f^{-1}(B_n)$. $\square$

2.1.c) Tribu produit

Définition 2.1.5. Soit $\mathcal{A}$ une tribu sur E et $\mathcal{B}$ une tribu sur F . On appelle **tribu produit**, notée $\mathcal{A} \otimes \mathcal{B}$, la tribu sur $E \times F$, engendrée par l'ensemble des parties de la forme $A \times B$ avec $A \in \mathcal{A}$ et $B \in \mathcal{B}$:

$$\mathcal{A} \otimes \mathcal{B} = \sigma(\{A \times B : A \in \mathcal{A} \text{ et } B \in \mathcal{B}\}).$$

2.1.d) La tribu borélienne

On considère maintenant $E = \mathbb{R}$.

Rappelons que pour la topologie usuelle de $\mathbb{R}$, un ensemble O de $\mathbb{R}$ est un ouvert si

$$\forall x \in O, \exists r > 0 :]x - r, x + r[\subset O.$$

On note $\mathcal{O}$ l'ensemble des sous-ensembles ouverts de $\mathbb{R}$.

Définition 2.1.6. On appelle **tribu borélienne**, notée $\mathcal{B}(\mathbb{R})$, la tribu engendrée par $\mathcal{O}$ (engendrée par les ensembles ouverts). Ses éléments sont appelés **boréliens**.

Remarque. Comme une tribu est stable par passage au complémentaire, tout ensemble fermé est un borélien, toute union / intersection dénombrable d'ouverts et de fermés est un borélien, tout intervalle est un borélien.

Même si ce n'est pas évident, on peut montrer qu'il existe des parties de $\mathbb{R}$ qui ne sont pas boréliennes : l'inclusion $\mathcal{B}(\mathbb{R}) \subset \mathcal{P}(\mathbb{R})$ est stricte.

Remarque. Si E est un espace topologique, la tribu borélienne sur E est la tribu engendrée par les ouverts de E .

Proposition 2.1.7. La tribu borélienne est engendrée par l'un des sous-ensembles suivants

1. les intervalles ouverts bornés,
2. les intervalles de la forme $] - \infty, a[$ avec $a \in \mathbb{R}$,
3. les intervalles de la forme $] - \infty, a]$ avec $a \in \mathbb{R}$

Preuve. On rappelle que par définition $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{O})$.

1. Notons $\mathcal{A}_1$ l'ensemble des intervalles ouverts bornés. On a $\mathcal{A}_1 \subset \mathcal{O}$, donc $\sigma(\mathcal{A}_1) \subset \sigma(\mathcal{O})$.

De plus, tout ouvert peut s'écrire comme une union dénombrable d'intervalles ouverts bornés. En effet, si $O \in \mathcal{O}$, on peut écrire

$$O = \bigcup_{(x,r) \in I}]x - r, x + r[$$

avec $I = \{(x, r) \in \mathbb{Q} \times \mathbb{Q}_+^* :]x - r, x + r[\subset O\}$. Par conséquent, $O \in \sigma(\mathcal{A}_1)$ pour tout $O \in \mathcal{O}$, d'où $\mathcal{O} \subset \sigma(\mathcal{A}_1)$. On en déduit que $\sigma(\mathcal{O}) \subset \sigma(\mathcal{A}_1)$. D'où l'égalité.

2. Notons maintenant $\mathcal{A}_2$ l'ensemble des intervalles de la forme $] - \infty, a[$ avec $a \in \mathbb{R}$. On a de façon évidente $\sigma(\mathcal{A}_2) \subset \sigma(\mathcal{O})$.

Si on montre que $\mathcal{A}_1 \subset \sigma(\mathcal{A}_2)$, on aura alors $\sigma(\mathcal{A}_1) = \sigma(\mathcal{O}) \subset \sigma(\mathcal{A}_2)$ et donc l'égalité. On remarque que, pour $a < b$,

$$\begin{aligned}]a, b[&=] - \infty, b[\cap (] - \infty, a])^c \\ &=] - \infty, b[\cap \left(\bigcap_{n \geq 1}] - \infty, a + \frac{1}{n}[\right)^c \end{aligned}$$

Par conséquent, $]a, b[\in \sigma(\mathcal{A}_2)$ pour tout a, b , d'où $\mathcal{A}_1 \subset \sigma(\mathcal{A}_2)$. D'où le résultat.

3. La preuve est similaire pour le troisième cas.

□

Proposition 2.1.8. La tribu borélienne sur $\mathbb{R}^d$ est égale à la tribu engendrée par la classe des ouverts de la forme

$$\prod_{i=1}^d]a_i, b_i[\quad \text{avec} \quad -\infty < a_i < b_i < +\infty.$$

Preuve. En exercice.

□

Remarque. La tribu borélienne sur $\mathbb{R}^2$ correspond à la tribu produit $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$. De façon générale, pour $d \geq 1$, on a

$$\mathcal{B}(\mathbb{R}^d) = \underbrace{\mathcal{B}(\mathbb{R}) \otimes \dots \otimes \mathcal{B}(\mathbb{R})}_{d \text{ fois}}.$$

Remarque. Nous aurons par la suite à considérer la droite achevée $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. Sa topologie est définie par la base d'ouverts formés des intervalles ouverts de la forme $]a, b[$, $[-\infty, b[$, $]a, +\infty[$, avec $a, b \in \overline{\mathbb{R}}$. On démontre de façon analogue que la tribu borélienne sur $\overline{\mathbb{R}}$ est engendrée par les classes $\{[-\infty, a[, a \in \overline{\mathbb{R}}\}$ ou encore par $\{[-\infty, a], a \in \overline{\mathbb{R}}\}$.

2.1.e) Exercices sur les tribus

Exercice 2.1.1. Soit $m \geq 1$ un entier, $a, b \in \mathbb{R}$ avec $a \leq b$. Déterminer les ensembles suivants :

$$\begin{aligned} & \bigcup_{n=1}^m]-1/n, 1/n[, \quad \bigcup_{n=1}^{\infty }]-1/n, 1/n[, \quad \bigcap_{n=1}^m]a, b + 1/n[, \quad \bigcap_{n=1}^{\infty }]a, b + 1/n[, \\ & \bigcup_{n \in \mathbb{N}^*} \left[0, 1 - \frac{1}{n}\right], \quad \bigcap_{n \in \mathbb{N}^*} \left[0, \frac{1}{n}\right], \quad \bigcup_{n \in \mathbb{N}^*} \left[\frac{1}{n}, 1 + \frac{1}{n}\right]. \end{aligned}$$

Exercice 2.1.2. Donner un exemple de suite décroissante d'ensembles $(A_n)_{n \in \mathbb{N}}$ telle que pour tout n , A_n est (de cardinal) infini et $\bigcap_n A_n = \emptyset$.

Exercice 2.1.3. Soient E et F deux ensembles, f une application de E dans F .

1. Montrer que pour toute partie B de F , on a $f^{-1}(B^c) = (f^{-1}(B))^c$.
2. Soient $(A_i)_{i \in I}$ (resp. $(B_i)_{i \in I}$) une famille de partie de E (resp. F). Montrer que

$$\begin{aligned} f^{-1}\left(\bigcup_{i \in I} B_i\right) &= \bigcup_{i \in I} f^{-1}(B_i), & f^{-1}\left(\bigcap_{i \in I} B_i\right) &= \bigcap_{i \in I} f^{-1}(B_i), \\ f\left(\bigcup_{i \in I} A_i\right) &= \bigcup_{i \in I} f(A_i), & f\left(\bigcap_{i \in I} A_i\right) &\subset \bigcap_{i \in I} f(A_i). \end{aligned}$$

Montrer que la dernière inclusion est en général stricte et qu'il y a égalité si f est injective.

Exercice 2.1.4. [Vrai ou Faux?]

Si l'assertion est vraie, démontrez là. Si elle est fausse, trouvez un contre exemple.

1. Soit E un ensemble. Alors $A \subset E \Leftrightarrow A \in \mathcal{P}(E)$.
2. Soit $(E, \mathcal{A})$ un espace mesurable. Alors $E \in \mathcal{A}$.
3. $\mathcal{A}$ est une tribu sur E si et seulement si les conditions suivantes sont vérifiées :
 - $\emptyset \in \mathcal{A}$
 - $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
 - $(\forall n \in \mathbb{N} A_n \in \mathcal{A}) \Rightarrow \bigcap A_n \in \mathcal{A}$.
4. Si E est dénombrable et $\mathcal{A}$ une tribu sur E , alors $\mathcal{A}$ est dénombrable.
5. La tribu borélienne $\mathcal{B}(\mathbb{R})$ est la tribu engendrée par les fermés de $\mathbb{R}$.
6. Un borélien est soit un ouvert, soit son complémentaire (c'est à dire un fermé).

Exercice 2.1.5. Montrer que si $\mathcal{B}$ est une tribu sur F , f une application de E dans F , alors $f^{-1}(\mathcal{B})$ est une tribu sur E .

Exercice 2.1.6. (*) On considère une fonction $f : E \rightarrow F$ et $\mathcal{A} \subset \mathcal{P}(F)$. Pour $\mathcal{B} \subset \mathcal{P}(F)$, $f^{-1}(\mathcal{B})$ l'ensemble défini par

$$f^{-1}(\mathcal{B}) = \{f^{-1}(B) : B \in \mathcal{B}\}.$$

Montrer que $\sigma(f^{-1}(\mathcal{A})) = f^{-1}(\sigma(\mathcal{A}))$.

Exercice 2.1.7. Soit E un ensemble, $A, B, (A_i)_{i \in I}$ des parties de E .

1. Déterminer $\mathbb{1}_{\emptyset}$, $\mathbb{1}_E$ et calculer $\mathbb{1}_A^{-1}(J)$ pour $J \subset \mathbb{R}$.
2. Montrer que

$$(i) A \subset B \Leftrightarrow \mathbb{1}_A \leq \mathbb{1}_B \text{ et } A = B \Leftrightarrow \mathbb{1}_A = \mathbb{1}_B, \quad (ii) \mathbb{1}_{A \cap B} = \mathbb{1}_A \mathbb{1}_B \text{ et } \mathbb{1}_{A^c} = 1 - \mathbb{1}_A,$$

$$(iii) \mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A \mathbb{1}_B, \quad (iv) \mathbb{1}_{A \Delta B} = |\mathbb{1}_A - \mathbb{1}_B|, \quad (v) \mathbb{1}_{A \times B}(x, y) = \mathbb{1}_A(x) \mathbb{1}_B(y)$$

où $A \Delta B = A \cup B \setminus (A \cap B)$.

3. Montrer que $\mathbb{1}_{\bigcup_{i \in I} A_i} = \sup_{i \in I} \mathbb{1}_{A_i}$ et $\mathbb{1}_{\bigcap_{i \in I} A_i} = \inf_{i \in I} \mathbb{1}_{A_i}$.

Exercice 2.1.8. Soient $\mathcal{A}$ et $\mathcal{B}$ deux tribus sur E . Montrer que

$$\sigma(\mathcal{A} \cup \mathcal{B}) = \sigma(\{A \cup B : A \in \mathcal{A} \text{ et } B \in \mathcal{B}\}) = \sigma(\{A \cap B : A \in \mathcal{A} \text{ et } B \in \mathcal{B}\}).$$

Exercice 2.1.9. Chercher dans la littérature un exemple de sous-ensemble de $\mathbb{R}$ qui n'est pas un borélien.

Exercice 2.1.10. Soit $E =]0, 1]$. Pour $A \subset E$. On définit $\mathcal{F}$ tel que

$$\begin{aligned} \mathcal{F} &= \{A \in \mathcal{P}(E) : A \text{ peut être exprimé comme une union finie d'intervalles semi-ouverts}\} \\ &= \{A \in \mathcal{P}(E) : \exists r \geq 1, 0 \leq a_1 \leq b_1 \leq a_2 \leq b_2 \leq \dots \leq a_r \leq b_r \leq 1 \text{ tel que } A =]a_1, b_1] \cup \dots \cup]a_r, b_r]\} \end{aligned}$$

Est-ce que $\mathcal{F}$ est une tribu? Justifier votre réponse.

Exercice 2.1.11. Soit $E = [0, \infty[$. Pour chaque $n \geq 1$, on considère $\mathcal{F}_n$ la tribu engendrée par les intervalles

$$[0, 1[, [1, 2[, \dots, [n-1, n[.$$

1. Montrer que $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, pour tout $n \in \mathbb{N}$.
2. Montrer que $\cup_{n=1}^{100} \mathcal{F}_n$ est une tribu.
3. Montrer que $\cup_{n=1}^{\infty} \mathcal{F}_n$ n'est pas une tribu.

2.2 Applications mesurables

Définition 2.2.1. On appelle *espace mesurable* tout couple $(E, \mathcal{A})$ formé d'un ensemble E et d'une tribu $\mathcal{A}$ sur E .

Par exemple, on peut considérer les espace mesurables suivants $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ou $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$.

Définition 2.2.2. Soit $(E, \mathcal{A})$ et $(F, \mathcal{B})$ deux espaces mesurables et $f : E \rightarrow F$ une application. On dit que f est une **application mesurable** de $(E, \mathcal{A})$ à $(F, \mathcal{B})$ si l'image réciproque de tout élément de $\mathcal{B}$ est un élément de $\mathcal{A}$: $\forall B \in \mathcal{B}, f^{-1}(B) \in \mathcal{A}$.

Exemple. 1. Si on considère la tribu $\mathcal{P}(E)$ sur l'espace E alors toute application $f : (E, \mathcal{P}(E)) \rightarrow (F, \mathcal{B})$ est mesurable car $\forall B \in \mathcal{B}, f^{-1}(B) = \{x \in E : f(x) \in B\}$ est bien une partie de E .

2. Soit $(E, \mathcal{A})$ un espace mesurable et $A \in \mathcal{P}(E)$. On munit $\mathbb{R}$ de la tribu borélienne $\mathcal{B}(\mathbb{R})$. Alors $f : E \rightarrow \mathbb{R}$ définie par $f = \mathbb{1}_A$ est une application mesurable si et seulement si $A \in \mathcal{A}$.
En effet, pour tout $B \in \mathcal{B}(\mathbb{R})$, on a

$$f^{-1}(B) = \begin{cases} E & \text{si } 1 \in B \text{ et } 0 \in B \\ A & \text{si } 1 \in B \text{ et } 0 \notin B \\ A^c & \text{si } 0 \in B \text{ et } 1 \notin B \\ \emptyset & \text{si } 1 \notin B \text{ et } 0 \notin B \end{cases}$$

Proposition 2.2.3. Soient $(E, \mathcal{A})$ et $(F, \mathcal{B})$ deux espaces mesurables avec $\mathcal{B} = \sigma(\mathcal{F})$. Une application $f : (E, \mathcal{A}) \rightarrow (F, \mathcal{B})$ est mesurable si et seulement si $f^{-1}(B) \in \mathcal{A}$ pour tout $B \in \mathcal{F}$.

Preuve. En exercice. Il faut vérifier que $f^{-1}(\sigma(\mathcal{F})) = \sigma(f^{-1}(\mathcal{F}))$. □

Corollaire 2.2.4. On considère $(E, \mathcal{A})$ un espace mesurable et $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Une application $f : E \rightarrow \mathbb{R}$ est mesurable si et seulement si l'une des conditions suivantes est vérifiée

1. $\forall a \in \mathbb{R}, f^{-1}(]-\infty, a]) = \{x \in \mathbb{R} : f(x) < a\} \in \mathcal{A}$,
2. $\forall a \in \mathbb{R}, f^{-1}(]-\infty, a]) = \{x \in \mathbb{R} : f(x) \leq a\} \in \mathcal{A}$,
3. $\forall a \in \mathbb{R}, f^{-1}(]a, +\infty]) = \{x \in \mathbb{R} : f(x) > a\} \in \mathcal{A}$,
4. $\forall a \in \mathbb{R}, f^{-1}(]a, +\infty]) = \{x \in \mathbb{R} : f(x) \geq a\} \in \mathcal{A}$.

Preuve. En effet, l'une quelconque des classes suivantes de parties de $\mathbb{R}$

$$\{]-\infty, a[; a \in \mathbb{R}\}; \quad \{]-\infty, a]; a \in \mathbb{R}\}; \quad \{]a, +\infty[; a \in \mathbb{R}\}; \quad \{]a, +\infty]; a \in \mathbb{R}\};$$

engendre la tribu borélienne de $\mathbb{R}$. □

Exemple. La fonction $f : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ définie par $f(x) = x^2$ est mesurable. En effet, on a pour $a \in \mathbb{R}$

$$f^{-1}(]-\infty, a]) = \begin{cases} \emptyset & \text{si } a \leq 0 \\]-\sqrt{a}, \sqrt{a}[& \text{si } a > 0. \end{cases}$$

On a bien $f^{-1}(]-\infty, a]) \in \mathcal{B}(\mathbb{R})$.

2.2.a) Les fonctions boréliennes

Définition 2.2.5. Soient $d, n \geq 1$. Une application $f : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ mesurable est appelée **application borélienne**.

Exemple. La fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ définie par $f(x) = \mathbb{1}_{\mathbb{Q}}(x)$ où $\mathbb{Q}$ est l'ensemble des rationnels est borélienne.

En effet, il suffit de vérifier que $\mathbb{Q}$ est un ensemble borélien. Pour cela on remarque que pour tout $q \in \mathbb{Q}$, l'ensemble $\{q\}$ est fermé donc c'est un borélien. Par conséquent, comme $\mathbb{Q} = \bigcup_{q \in \mathbb{Q}} \{q\}$, $\mathbb{Q}$ est l'union dénombrable de boréliens ($\mathbb{Q}$ étant dénombrable), c'est donc un ensemble borélien.

Proposition 2.2.6. Soit $f : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ continue, alors f est borélienne. De façon générale, une fonction $f : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ continue est borélienne.

Preuve. Si f est continue alors l'image réciproque de tout ouvert est un ouvert. On conclut en utilisant la proposition 2.2.3, car $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{O})$ où $\mathcal{O}$ est l'ensemble des ouverts. $\square$

2.2.b) Propriétés de stabilité

La mesurabilité est stable par composition :

Proposition 2.2.7. On considère trois espaces mesurables $(E, \mathcal{A})$, $(F, \mathcal{B})$ et $(G, \mathcal{C})$. Soit $f : E \rightarrow F$ et $g : F \rightarrow G$ des applications mesurables. Alors la composée $g \circ f : (E, \mathcal{A}) \rightarrow (G, \mathcal{C})$ est mesurable.

Preuve. Soit $C \in \mathcal{C}$. Comme g est mesurable on a $g^{-1}(C) \in \mathcal{B}$ et comme f est mesurable on a $f^{-1}(g^{-1}(C)) \in \mathcal{A}$. Par ailleurs, $(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C))$, donc $f \circ g$ est mesurable. $\square$

Proposition 2.2.8. Soit $(E, \mathcal{A})$, $(F_1, \mathcal{B}_1)$ et $(F_2, \mathcal{B}_2)$ trois espaces mesurables. On munit $F_1 \times F_2$ de la tribu produit $\mathcal{B}_1 \otimes \mathcal{B}_2$ et on considère $p_1 : (F_1 \times F_2, \mathcal{B}_1 \otimes \mathcal{B}_2) \rightarrow (F_1, \mathcal{B}_1)$ et $p_2 : (F_1 \times F_2, \mathcal{B}_1 \otimes \mathcal{B}_2) \rightarrow (F_2, \mathcal{B}_2)$ les projections (i.e., $p_1(x_1, x_2) = x_1$ et $p_2(x_1, x_2) = x_2$). Alors

1. Les projections p_1 et p_2 sont mesurables ;
2. $f : (E, \mathcal{A}) \rightarrow (F_1 \times F_2, \mathcal{B}_1 \otimes \mathcal{B}_2)$ est mesurable si et seulement si $p_1 \circ f$ et $p_2 \circ f$ sont mesurables.

Conséquence. Si $f : \mathbb{R} \rightarrow \mathbb{R}^d$ avec $\mathbb{R}$ et $\mathbb{R}^d$ muni de leur tribu borélienne. On a $f = (f_1, f_2, \dots, f_d)$ avec $f_i : \mathbb{R} \rightarrow \mathbb{R}$. La fonction f est borélienne si et seulement si pour tout $i \in \{1, \dots, d\}$, f_i est borélienne.

Preuve. [Non fait en cours]

1. Soit $A \in \mathcal{B}_1$, on a $p_1^{-1}(A) = A \times F_2 \in \mathcal{B}_1 \times \mathcal{B}_2$. La fonction p_1 est bien mesurable et on montre de même que p_2 est mesurable.

2. Si f est mesurable, d'après la proposition 2.2.7, $p_1 \circ f$ et $p_2 \circ f$ sont mesurables.

Réciproquement, si $p_1 \circ f$ et $p_2 \circ f$ sont mesurables. Soit $A_1 \in \mathcal{B}_1$ et $A_2 \in \mathcal{B}_2$. On remarque que

$$f^{-1}(A_1 \times F_2) = (p_1 \circ f)^{-1}(A_1) \text{ et } f^{-1}(F_1 \times A_2) = (p_2 \circ f)^{-1}(A_2).$$

Par hypothèse, on a donc $f^{-1}(A_1 \times F_2) \in \mathcal{A}$ et $f^{-1}(F_1 \times A_2) \in \mathcal{A}$.

Par ailleurs,

$$f^{-1}(A_1 \times A_2) = f^{-1}((A_1 \times F_2) \cap (F_1 \times A_2)) = f^{-1}((A_1 \times F_2)) \cap f^{-1}((F_1 \times A_2)) \in \mathcal{A}.$$

Comme la tribu produit $\mathcal{B}_1 \otimes \mathcal{B}_2$ est engendrée par les parties de la forme $A_1 \times A_2$ avec $A_1 \in \mathcal{B}_1$ et $A_2 \in \mathcal{B}_2$, la fonction f est bien mesurable. □

Propriété 2.2.9. *Si f et g sont deux fonctions mesurables de $(E, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, alors*

1. *pour tout $\alpha, \beta \in \mathbb{R}$, $\alpha f + \beta g$ est une fonction mesurable,*
2. *le produit fg est une fonction mesurable,*
3. *$|f|$, $\min(f, g)$, $\sup(f, g)$ sont mesurables.*

Preuve. En effet, comme f et g sont mesurables, par la proposition 2.2.8, la fonction $(f, g) : E \times E \rightarrow \mathbb{R}^2$ est mesurable. Comme $(x, y) \in \mathbb{R}^2 \mapsto \alpha x + \beta y$ est continue donc mesurable, par composition $\alpha f + \beta g$ est une fonction mesurable. Le raisonnement est similaire pour les autres points. □

Si $(f_n)_{n \geq 0}$ est une suite de fonctions de E dans $\overline{\mathbb{R}}$, on note $\limsup f_n$ (resp. $\liminf f_n$) la fonction de E dans $\overline{\mathbb{R}}$ définie par $x \mapsto \limsup f_n(x)$ (resp. $x \mapsto \liminf f_n(x)$).

Proposition 2.2.10 (Stabilité par passage à la limite). *Soit (f_n) une suite de fonctions mesurables de $(E, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors*

1. *$\sup f_n$, $\inf f_n$, $\limsup f_n$ et $\liminf f_n$ sont mesurables.*
2. *Si la suite (f_n) converge ponctuellement vers f , i.e. pour tout $x \in E$ $\lim_{n \rightarrow +\infty} f_n(x) = f(x)$ existe, alors la fonction $f : E \rightarrow \overline{\mathbb{R}}$ est mesurable.*

Preuve. Comme f_n est mesurable, pour tout $a \in \mathbb{R}$, l'ensemble $\{f_n \leq a\} \in \mathcal{A}$ et comme

$$\begin{aligned} \left\{ \sup_{n \geq 0} f_n \leq a \right\} &= \left\{ x \in E : \sup_{n \geq 0} f_n(x) \leq a \right\} \\ &= \{x \in E : \forall n \geq 0, f_n(x) \leq a\} \\ &= \bigcap_{n \geq 0} \{f_n \leq a\} \end{aligned}$$

on déduit que $\left\{ \sup_{n \geq 0} f_n \leq a \right\} \in \mathcal{A}$ pour tout $a \in \mathbb{R}$ et donc $\sup_{n \geq 0} f_n$ est mesurable.

Comme $\inf_{n \geq 0} f_n = -\sup_{n \geq 0} (-f_n)$, on déduit que $\inf_{n \geq 0} f_n$ est mesurable. De même, comme $\limsup_{n \rightarrow +\infty} f_n = \inf_{n \geq 0} \sup_{k \geq n} f_k$ et $\liminf_{n \rightarrow +\infty} f_n = \sup_{n \geq 0} \inf_{k \geq n} f_k$, ces fonctions sont mesurables.

Si (f_n) admet une limite, on a $f = \limsup_{n \rightarrow +\infty} f_n = \liminf_{n \rightarrow +\infty} f_n$, la limite f est bien mesurable. □

Proposition 2.2.11. *Si f et g sont deux fonctions mesurables de $(E, \mathcal{A})$ dans $\overline{\mathbb{R}}$ muni de sa tribu borélienne. Alors les ensembles $\{f < g\}$ et $\{f \leq g\}$ sont des éléments de $\mathcal{A}$.*

Preuve. On remarque que

$$\begin{aligned}\{f < g\} &= \{x \in E : f(x) - g(x) < 0\} \\ &= h^{-1}([-\infty, 0[),\end{aligned}$$

avec $h = f - g$ fonction mesurable, donc $\{f < g\} \in \mathcal{A}$. Autre preuve :

$$\begin{aligned}\{f < g\} &= \{x \in E : f(x) < g(x)\} \\ &= \bigcup_{q \in \mathbb{Q}} (\{f < q\} \cap \{g > q\}),\end{aligned}$$

car $f(x) < g(x)$ si et seulement si $\exists q \in \mathbb{Q}$ tel que $f(x) < q < g(x)$ (par densité de $\mathbb{Q}$ dans $\mathbb{R}$).

De plus, on a

$$\{f \leq g\} = \bigcap_{n \geq 1} \left\{ f < g + \frac{1}{n} \right\}.$$

Par conséquent, ces ensembles appartiennent à la tribu $\mathcal{A}$. □

Remarque. Soit $f : E \times F \rightarrow G$ une fonction et $\mathcal{A}, \mathcal{B}, \mathcal{C}$ des tribus respectivement sur E, F, G . Comme pour la continuité des fonctions à plusieurs variables, avoir pour tout $y \in F$, $x \mapsto f(x, y)$ mesurable de $(E, \mathcal{A})$ and $(G, \mathcal{C})$ et pour tout $x \in E$, $y \mapsto f(x, y)$ mesurable de $(F, \mathcal{B})$ and $(G, \mathcal{C})$, ne suffit pas pour conclure que $(x, y) \mapsto f(x, y)$ mesurable de $(E \times F, \mathcal{A} \otimes \mathcal{B})$ and $(G, \mathcal{C})$.

Par exemple, si $B \subset \mathbb{R}$ avec $B \notin \mathcal{B}(\mathbb{R})$ et $f(x, y) = \begin{cases} 1 & \text{si } x = y \text{ et } x \in B \\ 0 & \text{sinon} \end{cases}$. Si $x \in B$, alors $y \mapsto f(x, y) = \mathbb{1}_x(y)$ qui est borélienne. Si $x \notin B$, alors $y \mapsto f(x, y) = 0$ qui est borélienne. Pour autant, si f était borélienne, on aurait $f \circ g = \mathbb{1}_B(x)$ borélienne, où $g(x, y) = (x, x)$.

Par contre, il existe certains cas où la mesurabilité par coordonnées est suffisante.

Propriété 2.2.12. Soit $(f_n)_{n \geq 0}$ une suite de fonctions mesurables de $(E, \mathcal{A})$ dans $(F, \mathcal{B})$. Alors la fonction $f : (x, n) \mapsto f_n(x)$ est mesurable de $(E \times \mathbb{N}, \mathcal{A} \otimes \mathcal{P}(\mathbb{N}))$ dans $(F, \mathcal{B})$.

Preuve. En effet, si $B \in \mathcal{B}$, on a

$$\begin{aligned}f^{-1}(B) &= \{(x, n) \in E \times \mathbb{N} : f_n(x) \in B\} \\ &= \{(x, n) \in E \times \mathbb{N} : x \in f_n^{-1}(B)\} \\ &= \bigcup_{n \in \mathbb{N}} (f_n^{-1}(B) \times \{n\}).\end{aligned}$$

En effet $(x, n) \in f^{-1}(B)$, implique $x \in f_n^{-1}(B)$. Réciproquement, $(x, m) \in \bigcup_{n \in \mathbb{N}} (f_n^{-1}(B) \times \{n\})$ implique qu'il existe $n \in \mathbb{N}$ tel que $(x, m) \in f_n^{-1}(B) \times \{n\}$. Donc $m = n$ et $x \in f_n^{-1}(B)$.

Comme $\forall n \in \mathbb{N}$, f_n est mesurable, on a $f_n^{-1}(B) \in \mathcal{A}$, donc $(f_n^{-1}(B) \times \{n\}) \in \mathcal{A} \otimes \mathcal{P}(\mathbb{N})$. Comme l'union est dénombrable, on déduit que $f^{-1}(B) \in \mathcal{A} \otimes \mathcal{P}(\mathbb{N})$. La fonction f est bien mesurable. □

2.2.c) Exercices sur les fonctions mesurables

Exercice 2.2.1. Soit $E = \{a, b, c, d\}$ un ensemble à 4 éléments.

- Déterminer la tribu $\mathcal{A} = \sigma(\{A, B, C\})$ où $A = \{a\}$, $B = \{b\}$ et $C = \{c, d\}$.

2. Donner un exemple de fonction $f : (E, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ qui n'est pas mesurable.

Exercice 2.2.2. On considère un ensemble E muni d'une tribu $\mathcal{A}$, avec $E \neq \emptyset$. Décrire les fonctions mesurables de $(E, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ dans les cas suivants :

- (i) $\mathcal{A} = \mathcal{P}(E)$ (ii) $\mathcal{A} = \{\emptyset, E\}$.

Exercice 2.2.3. On considère deux ensembles mesurables $(E, \mathcal{A})$ et $(F, \mathcal{B})$, et $f, g : E \rightarrow F$ deux fonctions mesurables. Soit $A \in \mathcal{A}$. On introduit la fonction h définie par

$$h(x) = \begin{cases} f(x) & \text{si } x \in A \\ g(x) & \text{si } x \notin A \end{cases}$$

Montrer que h est une fonction mesurable.

Exercice 2.2.4. Montrer la Proposition 2.2.3 :

Si f et g sont deux fonctions mesurables de $(E, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, alors

1. pour tout $\alpha, \beta \in \mathbb{R}$, $\alpha f + \beta g$ est une fonction mesurable,
2. le produit fg est une fonction mesurable,
3. $|f|$, $\min(f, g)$, $\sup(f, g)$ sont mesurables.

Exercice 2.2.5. Soit E et F deux ensembles.

1. Soit $\mathcal{A}$ une tribu sur E et une fonction $f : (E, \mathcal{A}) \rightarrow F$. Quelle est la plus grande tribu $\mathcal{B}$ (au sens de l'inclusion) sur F qui rend f mesurable ?
2. Soit $\mathcal{B}$ une tribu sur F et une fonction $f : E \rightarrow (F, \mathcal{B})$. Quelle est la plus petite tribu $\mathcal{A}$ (au sens de l'inclusion) sur E qui rend f mesurable ?

Exercice 2.2.6. On considère l'espace $E = \mathbb{R}$. Pour toute partie $A \in \mathcal{P}(E)$, on note $-A = \{-a : a \in A\}$. On considère la fonction $f : \mathbb{R} \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ définie par $f(x) = x^2$.

1. Montrer que $\mathcal{A} = \{A \in \mathcal{B}(\mathbb{R}) : A = -A\}$ est une tribu sur $\mathbb{R}$.
2. Montrer que la tribu image réciproque $\mathcal{A}_f := f^{-1}(\mathcal{B}(\mathbb{R}))$ de $\mathcal{B}(\mathbb{R})$ par f est donnée par

$$\mathcal{A}_f = \{A \in \mathcal{B}(\mathbb{R}) : A = -A\}.$$

3. Décrire toutes les applications mesurables $h : (\mathbb{R}, \mathcal{A}_f) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Exercice 2.2.7. (*) On rappelle qu'une fonction f définie sur un intervalle borné $[a, b]$ est continue par morceaux s'il existe un nombre fini $n \in \mathbb{N}$ et des points $(x_0, \dots, x_n)$ de I avec $a < x_0 < \dots < x_n < b$ tels que

- f est continue en tout point $x \in [a, b] \setminus \{x_0, \dots, x_n\}$
- f admet en tous les x_k une limite à droite et une limite à gauche (i.e., $\lim_{\substack{x \rightarrow x_k \\ x < x_k}} f(x)$ et $\lim_{\substack{x \rightarrow x_k \\ x > x_k}} f(x)$ existent).

Montrer qu'une fonction $f : [a, b] \rightarrow \mathbb{R}$ continue par morceaux est borélienne.

2.3 Approximation des fonctions mesurables

L'objet de ce paragraphe est d'établir un résultat d'approximation relativement élémentaire mais fondamental pour la construction de l'intégrale de Lebesgue : toute fonction mesurable à valeurs dans $\mathbb{R}_+$ est limite croissante de fonctions élémentaires, appelées fonctions étagées.

Définition 2.3.1. On note $\mathcal{M}$ (resp. $\mathcal{M}_+$) l'ensemble des fonctions mesurables (resp. mesurables positives) sur $(E, \mathcal{A})$ à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (resp. $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$).

Définition 2.3.2. Une fonction f mesurable de $(E, \mathcal{A})$ à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est dite **étagée** s'il existe $n \geq 1$, $\alpha_1, \dots, \alpha_n$ des réels et $A_1, \dots, A_n$ des éléments de $\mathcal{A}$ tels que

$$f(x) = \sum_{k=1}^n \alpha_k \mathbb{1}_{A_k}(x).$$

On note $\mathcal{E}_+$ (resp. $\mathcal{E}$) l'ensemble des fonctions étagées sur $(E, \mathcal{A})$ à valeurs dans $\mathbb{R}_+$ (resp. $\mathbb{R}$).

Une fonction étagée est mesurable et ne prend qu'un nombre fini de valeurs distinctes. On remarque facilement que toute combinaison linéaire de fonction étagée est étagée, de même tout produit de fonctions étagées est étagée.

Exemple. La fonction $f(x) = 2\mathbb{1}_{\mathbb{Q}}(x) - \mathbb{1}_{\mathbb{R} \setminus \mathbb{Q}}(x) + 3\mathbb{1}_{[2,6[}$ est une fonction étagée de $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Le théorème suivant est fondamental dans la construction de l'intégrale.

Théorème 2.3.3. Soit f une fonction mesurable de $(E, \mathcal{A})$ à valeurs dans $\overline{\mathbb{R}}_+$ (positive). Il existe une suite croissante $(f_n)_{n \in \mathbb{N}}$ de fonctions étagées qui converge ponctuellement vers f .

Preuve. [Idées de la preuve en cours]

Soit $n \in \mathbb{N}$ et $k \in \{0, 1, 2, \dots, n2^n - 1\}$. On introduit

$$A_n = \{f \geq n\} \quad \text{et} \quad A_{n,k} = \left\{ \frac{k}{2^n} \leq f < \frac{k+1}{2^n} \right\}$$

On définit f_n par

$$f_n(x) = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbb{1}_{A_{n,k}}(x) + n \mathbb{1}_{A_n}(x).$$

On remarque que $E = \bigcup_{k=0}^{n2^n-1} A_{n,k} \cup A_n$ (partition de E) et par construction, on a $f_n \leq f$.

De plus, comme f est mesurable, on a $A_n \in \mathcal{A}$ et $A_{n,k} \in \mathcal{A}$, donc f_n est une fonction étagée.

1. Montrons que la suite est croissante : pour tout $x \in E$, $f_n(x) \leq f_{n+1}(x)$.

Soit $x \in E$ et $n \geq 1$. Comme $E = \bigcup_{k=0}^{n2^n-1} A_{n,k} \cup A_n$ (partition de E), soit il existe k tel que $x \in A_{n,k}$, soit on a $x \in A_n$.

On remarque que si $x \in A_{n,k}$, on a $\frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n}$ et

$$f_{n+1}(x) = \begin{cases} \frac{2k}{2^{n+1}} = f_n(x) & \text{si } \frac{2k}{2^{n+1}} \leq f < \frac{2k+1}{2^{n+1}} \\ \frac{2k+1}{2^{n+1}} = f_n(x) + \frac{1}{2^{n+1}} & \text{si } \frac{2k+1}{2^{n+1}} \leq f < \frac{2k+2}{2^{n+1}} \end{cases}$$

et si $x \in A_n$,

$$f_{n+1}(x) = \begin{cases} n + \frac{k}{2^{n+1}} = f_n(x) + \frac{k}{2^{n+1}} & \text{si } n + \frac{k}{2^{n+1}} \leq f < n + \frac{k+1}{2^{n+1}}, \text{ pour } 0 \leq k \leq 2^{n+1} - 1 \\ n + 1 = f_n(x) + 1 & \text{si } f \geq n + 1 \end{cases}$$

On observe que pour tout $x \in E$, $f_n(x) \leq f_{n+1}(x)$, la suite est donc croissante.

2. Montrons que la suite (f_n) converge ponctuellement vers f sur E .

Soit $x \in E$.

- si $f(x) < +\infty$: il existe $n_0 \geq 1$ tel que $x \notin A_{n_0}$ (il suffit de prendre $n_0 = [f(x)] + 1$). Alors pour tout $n \geq n_0$, $x \notin A_n$ et donc

$$0 \leq f(x) - f_n(x) \leq \frac{1}{2^n}.$$

Par conséquent, la suite (f_n) converge ponctuellement vers f sur $\{x \in E : f(x) < +\infty\}$.

- si $f(x) = +\infty$, alors pour tout $n \geq 1$ $x \in A_n$ et donc $f_n(x) = n$ dont la limite est bien $+\infty$ quand $n \rightarrow +\infty$.

Par conséquent, la suite (f_n) converge ponctuellement vers f sur E .

□

Corollaire 2.3.4. *Toute fonction mesurable sur $(E, \mathcal{A})$ à valeurs dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ est limite simple d'une suite (f_n) de fonctions étagées à valeurs dans $\mathbb{R}$.*

Preuve. Soit f une fonction mesurable sur $(E, \mathcal{A})$ à valeurs dans $\overline{\mathbb{R}}$. On note $f^+ : x \mapsto \sup(f(x), 0)$ et $f^- : x \mapsto -\inf(f(x), 0)$. Ces fonctions sont positives et mesurables d'après la propriété 2.2.9. Il existe donc des suites croissantes de fonctions étagées (h_n) et (g_n) qui convergent vers f^+ et f^- respectivement. On remarque que $f(x) = f^+(x) - f^-(x)$ pour tout $x \in E$. Par conséquent, la fonction étagée $f_n = h_n - g_n$ converge simplement vers f . □

2.4 Mesures positives et mesure de Lebesgue

Définition 2.4.1. *Soit $(E, \mathcal{A})$ un espace mesurable (par exemple, $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ou $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$).*

*On appelle **mesure positive** sur $(E, \mathcal{A})$ une application $\mu : \mathcal{A} \rightarrow \overline{\mathbb{R}}_+$ telle que*

1. $\mu(\emptyset) = 0$,
2. Si $(A_n)_{n \geq 0}$ est une suite d'éléments de $\mathcal{A}$ deux à deux disjoints, alors

$$\mu\left(\bigcup_{n \geq 0} A_n\right) = \sum_{n=0}^{+\infty} \mu(A_n)$$

*Si $\mu(E) < +\infty$, on dit que la mesure μ est **finie**.*

*Si $\mu(E) = 1$, on dit que la mesure μ est une **mesure de probabilité**.*

*S'il existe une suite $(E_n)_{n \geq 0}$ d'éléments de $\mathcal{A}$ telle que $\mu(E_n) < +\infty$ pour tout $n \geq 0$ et $\bigcup_{n \geq 0} E_n = E$, on dit que la mesure μ est une **mesure σ -finie**.*

Définition 2.4.2. *On appelle **espace mesuré** tout triplet $(E, \mathcal{A}, \mu)$ où $(E, \mathcal{A})$ est un espace mesurable et μ une mesure positive sur $(E, \mathcal{A})$.*

Définition 2.4.3. On appelle **espace de probabilité** tout triplet $(E, \mathcal{A}, \mu)$ où $(E, \mathcal{A})$ est un espace mesurable et μ une mesure de probabilité sur $(E, \mathcal{A})$.

Remarque. La somme finie ou dénombrable de mesures positives est une mesure positive.

Si μ est une mesure positive alors $c\mu$, avec $c > 0$, est une mesure positive.

Tout combinaison convexe de mesures de probabilités est une mesure de probabilité :

si $(\mu_i)_{1 \leq i \leq n}$ probabilités, $(\alpha_i)_{1 \leq i \leq n}$, avec $\alpha_i \in [0, 1]$ et $\sum_{i=1}^n \alpha_i = 1$, alors $\mu = \sum_{i=1}^n \alpha_i \mu_i$ est une mesure de probabilité.

En particulier, si μ et ν sont deux mesures de probabilités et $\alpha \in [0, 1]$, alors $\alpha\mu + (1 - \alpha)\nu$ est une mesure de probabilité.

Proposition 2.4.4. Soit $(E, \mathcal{A}, \mu)$ un espace mesuré.

1. Si $A_1, \dots, A_n$ sont des éléments de $\mathcal{A}$ deux à deux disjoints alors,

$$\mu(A_1 \cup \dots \cup A_n) = \mu(A_1) + \dots + \mu(A_n).$$

2. Si A et B sont des éléments de $\mathcal{A}$, tels que $B \subset A$, alors $\mu(B) \leq \mu(A)$.

Si de plus, $\mu(B) < +\infty$ on a $\mu(A \setminus B) = \mu(A) - \mu(B)$.

3. Si A et B sont des éléments de $\mathcal{A}$, alors $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$.

4. Si $(A_n)_{n \geq 0}$ est une suite d'éléments de $\mathcal{A}$, alors

$$\mu\left(\bigcup_{n \geq 0} A_n\right) \leq \sum_{n=0}^{+\infty} \mu(A_n)$$

5. Si μ et ν sont deux mesures positives sur $(E, \mathcal{A})$ et $(\alpha, \beta) \in \mathbb{R}_+^2$, alors $\alpha\mu + \beta\nu$ est une mesure positive sur $(E, \mathcal{A})$.

Preuve.

1. On introduit $A_k = \emptyset$ pour $k \geq n + 1$. La suite $(A_k)_{k \geq 1}$ est une suite d'éléments de $\mathcal{A}$ deux à deux disjoints, donc par définition de μ :

$$\mu\left(\bigcup_{k \geq 1} A_k\right) = \sum_{k=1}^{+\infty} \mu(A_k) = \sum_{k=1}^n \mu(A_k),$$

car $\mu(A_k) = \mu(\emptyset) = 0$ pour $k \geq n + 1$.

2. Si $B \subset A$, on a $A = B \cup (A \setminus B)$ avec B et $A \setminus B$ ensembles disjoints. Donc d'après ce qui précède, on a

$$\mu(A) = \mu(B) + \mu(A \setminus B).$$

Comme $\mu(A \setminus B) \geq 0$, on a $\mu(B) \leq \mu(A)$ et si $\mu(B) < +\infty$, on a $\mu(A \setminus B) = \mu(A) - \mu(B)$.

3. En exercice.

4. En exercice.

□

On donne maintenant une définition équivalente à la notion de mesure positive.

Proposition 2.4.5. Une application μ de $\mathcal{A}$ dans $\overline{\mathbb{R}}_+$ est une mesure si et seulement si

1. $\mu(\emptyset) = 0$,
2. si A et B sont deux éléments disjoints de $\mathcal{A}$, $\mu(A \cup B) = \mu(A) + \mu(B)$,
3. pour toute suite croissante $(B_n)_{n \geq 0}$ d'éléments de $\mathcal{A}$, $\mu(\bigcup_{n \geq 0} B_n) = \lim_{n \rightarrow +\infty} \mu(B_n)$.

Preuve.

$\Rightarrow$] Supposons que μ soit une mesure positive.

Les premières propriétés sont vérifiées de façon évidente.

On considère $(B_n)_{n \geq 0}$ une suite croissante d'éléments de $\mathcal{A}$.

S'il existe k tel que $\mu(B_k) = +\infty$, alors par croissance de la suite $(B_n)_{n \geq 0}$ et de la mesure, $\forall n \geq k$, $\mu(B_n) = +\infty$. De même $\mu(\bigcup_{n \geq 0} B_n) = +\infty$. On a donc bien

$$\mu\left(\bigcup_{n \geq 0} B_n\right) = \lim_{n \rightarrow +\infty} \mu(B_n).$$

Supposons maintenant que pour tout $n \geq 0$, $\mu(B_n) < +\infty$.

On définit

$$A_0 = B_0, \quad A_1 = B_1 \setminus B_0, \dots, A_n = B_n \setminus B_{n-1}.$$

La suite $(A_n)_{n \geq 0}$ est une suite d'éléments de $\mathcal{A}$ deux à deux disjoints. De plus,

$$\bigcup_{n \geq 0} A_n = \bigcup_{n \geq 0} B_n.$$

par conséquent

$$\mu\left(\bigcup_{n \geq 0} B_n\right) = \mu\left(\bigcup_{n \geq 0} A_n\right) = \sum_{n=0}^{+\infty} \mu(A_n)$$

On remarque cependant que

$$\mu(A_n) = \mu(B_n) - \mu(B_{n-1})$$

et donc

$$\sum_{k=0}^n \mu(A_k) = \mu(B_n)$$

D'où,

$$\mu\left(\bigcup_{n \geq 0} B_n\right) = \lim_{n \rightarrow +\infty} \mu(B_n).$$

Ce qui conclut la première implication.

[$\Leftarrow$] Supposons que μ vérifie les hypothèses de la proposition.

On montre facilement par récurrence, à partir du point 2., que pour tout $n \geq 1$, si $A_1, \dots, A_n$ sont des ensembles deux à deux disjoints, alors

$$\mu(A_1 \cup \dots \cup A_n) = \mu(A_1) + \dots + \mu(A_n).$$

Soit (A_n) une suite d'éléments deux à deux disjoints de $\mathcal{A}$. On définit, pour $n \geq 0$

$$B_0 = A_0, \quad B_1 = A_0 \cup A_1, \dots, B_n = \bigcup_{k=0}^n A_k.$$

La suite $(B_n)_{n \geq 0}$ est une suite croissante d'éléments de $\mathcal{A}$. Par ailleurs,

$$\bigcup_{n \geq 0} B_n = \bigcup_{n \geq 0} A_n.$$

On a donc

$$\begin{aligned} \mu\left(\bigcup_{n \geq 0} A_n\right) &= \mu\left(\bigcup_{n \geq 0} B_n\right) \\ &= \lim_{n \rightarrow +\infty} \mu(B_n) \end{aligned}$$

Par ailleurs, les éléments $(A_n)_{n \geq 0}$ étant disjoints deux à deux

$$\mu(B_n) = \sum_{k=0}^n \mu(A_k).$$

D'où,

$$\mu\left(\bigcup_{n \geq 0} A_n\right) = \sum_{k=0}^{+\infty} \mu(A_k).$$

□

2.4.a) Mesures discrètes

Définition 2.4.6. Soit $(E, \mathcal{A})$ un espace mesurable et $a \in E$. On définit $\delta_a : \mathcal{A} \rightarrow \overline{\mathbb{R}}_+$ la fonction définie par

$$\delta_a(A) = \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{si } a \notin A. \end{cases}$$

Cette application est appelée **mesure (ou masse) de Dirac** au point a .

On remarque que $\delta_a(A) = \mathbb{1}_A(a)$.

Montrons qu'il s'agit bien d'une mesure en utilisant la définition alternative (voir proposition 2.4.5). On a bien

1. $\delta_a(\emptyset) = 0$,
2. si A et B sont deux éléments disjoints de $\mathcal{A}$. On a $\delta_a(A \cup B) = \mathbb{1}_{A \cup B}(a) = \mathbb{1}_A(a) + \mathbb{1}_B(a) = \delta_a(A) + \delta_a(B)$.
3. Soit $(B_n)_{n \geq 0}$ une suite croissante d'éléments de $\mathcal{A}$. Alors

$$a \in \bigcup_{n \geq 0} B_n \Leftrightarrow \exists n_0 \geq 0 : a \in B_{n_0} \Leftrightarrow \exists n_0 \geq 0 : \forall n \geq n_0, a \in B_n.$$

Par conséquent, si $a \in \bigcup_{n \geq 0} B_n$, on a $\delta_a(\bigcup_{n \geq 0} B_n) = 1 = \lim_{n \rightarrow +\infty} \delta_a(B_n)$ et si $a \notin \bigcup_{n \geq 0} B_n$, on a $\delta_a(\bigcup_{n \geq 0} B_n) = 0 = \lim_{n \rightarrow +\infty} \delta_a(B_n)$.

Exemple. On se place sur l'espace mesurable $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. Soit $p \in]0, 1[$. On définit

- la **mesure de Bernoulli** de paramètre p par $\mu = p\delta_1 + (1 - p)\delta_0$.

- la **mesure de Rademacher** de paramètre p par $\mu = p\delta_1 + (1-p)\delta_{-1}$.

On a $\mu(A) = p\mathbb{1}_A(1) + (1-p)\mathbb{1}_A(0)$ pour $A \in \mathcal{P}(\mathbb{N})$.

Il s'agit d'une mesure de probabilité.

Définition 2.4.7.

Soit $(E, \mathcal{A})$ un espace mesurable. Soit $(a_n)_{n \geq 0}$ une suite de points de E et $(\alpha_n)_{n \in \mathbb{N}}$ une suite de réels positifs. On définit $\mu : \mathcal{A} \rightarrow \overline{\mathbb{R}}_+$ par $\forall A \in \mathcal{A}$,

$$\mu(A) = \sum_{n=0}^{+\infty} \alpha_n \delta_{a_n}(A).$$

Alors μ est une mesure positive, appelée **mesure discrète** sur $(E, \mathcal{A})$.

Exemple. 1. La mesure de comptage sur un ensemble dénombrable $\mathbb{D}$ muni de la tribu $\mathcal{P}(\mathbb{D})$ est définie par

$$\nu = \sum_{x \in \mathbb{D}} \delta_x.$$

On l'appelle ainsi car si $A \in \mathcal{P}(\mathbb{D})$, $\nu(A) = \sum_{x \in \mathbb{D}} \mathbb{1}_A(x)$ est le nombre d'éléments dans A , i.e. $\nu(A) = \text{card}(A)$.

La mesure de comptage du $\mathbb{N}$ est définie par

$$\nu = \sum_{k=0}^{+\infty} \delta_k.$$

Il s'agit d'une mesure σ -finie, car en définissant $E_n = \{0, 1, \dots, n\}$, on a $\mathbb{N} = \cup_n E_n$ et $\nu(E_n) = n + 1 < +\infty$.

2. Mesure de probabilité discrète sur $\mathbb{N}$: On se place sur l'espace mesurable $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, alors $\mu = \sum_{k=0}^{+\infty} \alpha_k \delta_k$,

avec $\alpha_k \geq 0$ et $\sum_{k=0}^{+\infty} \alpha_k = 1$, est une mesure de probabilité.

Par exemple, la mesure de Poisson de paramètre $\lambda > 0$ est la mesure (de probabilité) discrète :

$$\mu = \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} e^{-\lambda} \delta_k.$$

2.4.b) Mesure de Lebesgue

Théorème 2.4.8. Il existe une unique mesure λ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que

(i) $\lambda([0, 1]) = 1$,

(ii) Invariance par translation : pour tout $a \in \mathbb{R}$, pour tout $B \in \mathcal{B}(\mathbb{R})$, $\lambda(a + B) = \lambda(B)$, où $a + B = \{a + x : x \in B\}$.

Elle est appelée mesure de Lebesgue sur $\mathbb{R}$.

Preuve. La preuve de ce théorème est compliquée, voir par exemple le Théorème 6.7 dans le livre de Briane et Pagès, *Théorie de l'intégration*, [BP06]. $\square$

Remarque. La mesure de Lebesgue est la seule mesure invariante par translation dont la mesure de l'ensemble $[0, 1]$ est 1.

La mesure de Lebesgue coïncide avec la notion intuitive de longueur.

Proposition 2.4.9. 1. La mesure de Lebesgue ne charge aucun point : $\lambda(\{a\}) = 0$ pour tout $a \in \mathbb{R}$.
2. Pour tous réels $a < b$, on a

$$\lambda([a, b]) = \lambda([a, b[) = \lambda(]a, b]) = \lambda(]a, b[) = b - a.$$

3. Si I est un intervalle non borné alors $\lambda(I) = +\infty$.

Preuve.

1. Montrons le premier point. Soit $\alpha = \lambda(\{0\})$. Alors par invariance par translation, pour tout $a \in \mathbb{R}$,

$$\lambda(\{a\}) = \lambda(a + \{0\}) = \lambda(\{0\}) = \alpha.$$

Soit $n \geq 1$, on a alors

$$\lambda\left(\bigcup_{k=1}^n \left\{\frac{1}{k}\right\}\right) = \sum_{k=1}^n \lambda\left(\left\{\frac{1}{k}\right\}\right) = n\alpha$$

et comme $\bigcup_{k=1}^n \left\{\frac{1}{k}\right\} \subset [0, 1]$, on a

$$n\alpha = \lambda\left(\bigcup_{k=1}^n \left\{\frac{1}{k}\right\}\right) \leq \lambda([0, 1]) = 1.$$

Comme ceci est vrai pour tout $n \geq 0$, on déduit $\alpha = 0$.

2. Comme λ ne charge aucun point, par conséquent pour tout intervalle on a

$$\lambda([a, b]) = \lambda([a, b[) = \lambda(]a, b]) = \lambda(]a, b[)$$

car on peut écrire les unions disjointes suivantes

$$[a, b] = [a, b[\cup \{b\} =]a, b[\cup \{a\} =]a, b[\cup \{a\} \cup \{b\}.$$

Par ailleurs, soit $n \geq 1$. On découpe l'intervalle $]0, 1]$ en n intervalles disjoints

$$]0, 1] = \bigcup_{k=1}^n \left] \frac{k-1}{n}, \frac{k}{n} \right].$$

Donc

$$\begin{aligned} 1 &= \lambda(]0, 1]) = \sum_{k=1}^n \lambda\left(\left] \frac{k-1}{n}, \frac{k}{n} \right]\right) \\ &= \sum_{k=1}^n \lambda\left(\left[\frac{k-1}{n}, \frac{k}{n} \right[\right) = \sum_{k=1}^n \lambda\left(\left[0, \frac{1}{n} \right[\right) \\ &= n\lambda\left(\left[0, \frac{1}{n} \right[\right). \end{aligned}$$

On a donc

$$\lambda\left(\left[0, \frac{1}{n}\right]\right) = \frac{1}{n}.$$

Maintenant, si $i, j \in \mathbb{Z}$ avec $i \leq j$, on a

$$\lambda\left(\left[\frac{i}{n}, \frac{j}{n}\right]\right) = \lambda\left(\bigcup_{k=i}^{j-1} \left[\frac{k}{n}, \frac{k+1}{n}\right]\right) = \sum_{k=i}^{j-1} \lambda\left(\left[0, \frac{1}{n}\right]\right) = \frac{j}{n} - \frac{i}{n}.$$

Ainsi pour tous rationnels, $q, q' \in \mathbb{Q}$ avec $q \leq q'$, on a

$$\lambda([q, q']) = q' - q.$$

Soit a, b deux réels avec $a \leq b$. Par densité de $\mathbb{Q}$, il existe une suite (u_n) strictement décroissante convergent vers a , une suite (v_n) strictement croissante convergent vers b , avec $u_n \leq v_n$. On a alors, d'après la proposition 2.4.5,

$$\lambda([a, b]) = \lambda\left(\bigcup_n]u_n, v_n[\right) = \lim_{n \rightarrow +\infty} \lambda([u_n, v_n]) = \lim_{n \rightarrow +\infty} (v_n - u_n) = b - a.$$

Soit I un intervalle non borné, par exemple de la forme $[a, +\infty[$. Alors pour tout $n \geq 1$, $[a, a+n] \subset I$, ce qui implique $\lambda(I) \geq n$. Ceci assure que $\lambda(I) = +\infty$.

□

Remarque. La mesure de Lebesgue λ est une mesure σ -finie. Il suffit d'introduire pour $n \geq 1$, $E_n = [-n, n] \in \mathcal{B}(\mathbb{R})$. On a $\mathbb{R} = \bigcup_n E_n$ et $\lambda(E_n) = 2n < +\infty$.

Remarque. La mesure de comptage sur un groupe discret $(\mathbb{D}, \mathcal{P}(\mathbb{D}))$ et la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ sont des cas particuliers de mesures de Haar.

Soit $(G, \star)$ un groupe, avec G localement compact. On note $\mathcal{B}(G)$ la tribu borélienne sur G . On dit que μ est une mesure de Haar si μ est une mesure non nulle, quasi régulière, et invariante par translation à gauche :

$$\lambda(gA) = \lambda(A) \quad \text{pour tout } A \in \mathcal{B}(G) \text{ et } g \in G,$$

où $gA = \{g \star x : x \in A\}$.

Une mesure μ est dite quasi régulière si elle vérifie les deux conditions

$$\begin{aligned} \forall A \in \mathcal{B}(G), \mu(A) &= \inf \{ \mu(O) : O \text{ ouvert avec } A \subset O \}, \\ \forall O \text{ ouvert}, \mu(O) &= \sup \{ \mu(K) : K \text{ compact avec } K \subset O \}. \end{aligned}$$

2.4.c) Exercices sur les mesures

Exercice 2.4.1. Soit E un ensemble non vide. On dit qu'un ensemble est au plus dénombrable s'il est vide, fini, ou dénombrable (en bijection avec $\mathbb{N}$).

On considère l'espace $(E, \mathcal{A})$ avec $\mathcal{A} = \{A \in \mathcal{P}(E) : A \text{ ou } A^c \text{ est au plus dénombrable}\}$.

1. Montrer que $(E, \mathcal{A})$ est un espace mesurable.
2. On introduit la fonction μ sur $\mathcal{A}$ défini par $\mu(A) = 0$ si A dénombrable et $\mu(A) = 1$ si A^c dénombrable.

Montrer que μ est une mesure de probabilité sur $(E, \mathcal{A})$.

Exercice 2.4.2. [Vrai ou Faux?] (Justifier votre réponse)

Soit $(E, \mathcal{A}, \mu)$ un espace mesuré.

1. Si $A \in \mathcal{A}$, alors $\mu(E) = \mu(A) + \mu(A^c)$.
2. Une réunion d'ensembles de mesure nulle est de mesure nulle.
3. Si $A, B \in \mathcal{A}$ tels que $\mu(A \cup B) = \mu(A) + \mu(B)$, alors A et B sont disjoints.
4. On considère $(E, \mathcal{A}) = (\mathbb{R}, \mathcal{P}(\mathbb{R}))$ et on introduit $\mu(A) = \sup \{b - a : (a, b) \in A^2\}$.
 - (a) On a $\mu(A) \leq \mu(B)$ si $A \subset B$;
 - (b) μ est une mesure sur $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$.

Exercice 2.4.3. 1. Montrer que si μ et ν sont deux mesures positives sur un espace mesurable $(E, \mathcal{A})$ et $(\alpha, \beta) \in \mathbb{R}_+^2$, alors $\alpha\mu + \beta\nu$ est une mesure positive sur $(E, \mathcal{A})$.

2. Si on suppose que μ et ν sont des mesures de probabilité, à quelle condition sur α et β , $\alpha\mu + \beta\nu$ est une mesure de probabilité?

3. (*) Soit $(a_n)_{n \geq 0}$ et $(\alpha_n)_{n \geq 0}$ deux suites de réels, avec $\alpha_n \geq 0$. Pour $n \geq 0$, on considère δ_{a_n} la mesure de Dirac au point a_n . Montrer que $\sum_{n=0}^{+\infty} \alpha_n \delta_{a_n}$ est une mesure positive sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Exercice 2.4.4. On considère l'espace mesurable $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On note λ la mesure de Lebesgue et pour $a \in \mathbb{R}$, δ_a la mesure de Dirac au point a .

Calculer la mesure des boréliens suivants pour la mesure de Lebesgue λ et pour la mesure $\mu = \lambda + \delta_0 + 2\delta_1$:

$$\begin{aligned} A_1 &= [-1, 1], & A_2 &= [-1, 1[\cup]3, 5], & A_3 &= \mathbb{Q}, \\ A_4 &= (\mathbb{R} \setminus \mathbb{Q}) \cap [0, 1], & A_5 &= \mathbb{R} \setminus \mathbb{Q}, & A_6 &= \bigcup_{n \geq 1} \left[0, 1 - \frac{1}{n}\right]. \end{aligned}$$

Exercice 2.4.5. Soit E un ensemble et μ la mesure définie sur $\mathcal{P}(E)$ par

$$\mu(A) = \begin{cases} \text{Card}(A) & \text{si } A \text{ est fini,} \\ +\infty & \text{sinon.} \end{cases}$$

Montrer que $(E, \mathcal{P}(E), \mu)$ est un espace mesuré.

La mesure μ est appelée mesure de comptage sur E .

Exercice 2.4.6. [Mesure conditionnelle]

Soit $(E, \mathcal{A}, \mu)$ un espace mesuré. Soit $B \in \mathcal{A}$ tel que $0 < \mu(B) < +\infty$. On définit $\mu(\cdot|B) : \mathcal{A} \rightarrow \overline{\mathbb{R}}_+$ par

$$\forall A \in \mathcal{A}, \quad \mu(A|B) = \frac{\mu(A \cap B)}{\mu(B)}.$$

1. montrer que $\mu(\cdot|B)$ est une mesure de probabilité sur $(E, \mathcal{A})$.
2. Supposons maintenant que μ est la mesure géométrique de paramètre $1/4$ sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$
 - (a) Donner l'expression de la mesure μ .
 - (b) Calculer $\mu(\{1, \dots, n\})$ et $\mu(\{n, n+1, \dots\})$.

- (c) Soit $B = 3\mathbb{N}$ l'ensemble des entiers multiples de 3. Calculer $\mu(B)$. Trouver l'expression de la mesure $\mu(\cdot|B)$, c'est à dire trouvez la valeur des p_k tels que

$$\mu(\cdot|B) = \sum_{k=1}^{\infty} p_k \delta_k.$$

Exercice 2.4.7. [Mesure image]

Soit $(E, \mathcal{A}, \mu)$ un espace mesuré, $(F, \mathcal{B})$ un espace mesurable et $\varphi : (E, \mathcal{A}) \rightarrow (F, \mathcal{B})$ une application mesurable. On définit l'application $\mu_\varphi : \mathcal{B} \rightarrow \overline{\mathbb{R}}_+$ par

$$\forall B \in \mathcal{B}, \quad \mu_\varphi(B) = \mu(\varphi^{-1}(B))$$

1. Montrer que μ_φ est une mesure sur $(F, \mathcal{B})$.
2. Montrer que si de plus μ est une mesure de probabilité, il en est de même pour μ_φ .
3. Application : $(E, \mathcal{A}, \mu) = (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), \lambda)$ et $(F, \mathcal{B}) = (\mathbb{N}, \mathcal{P}(\mathbb{N}))$, où λ est la mesure de Lebesgue et φ est la fonction partie entière. Décrire la mesure μ_φ .

2.5 Construction de l'intégrale

On se place sur un espace mesuré $(E, \mathcal{A}, \mu)$. On munit $\mathbb{R}$ de sa tribu borélienne $\mathcal{B}(\mathbb{R})$.

On va construire l'intégrale de fonctions définies sur E à valeurs dans $\mathbb{R}$ par rapport à la mesure μ de façon progressive, en généralisant au fur et à mesure le type de fonction que l'on peut intégrer.

2.5.a) Intégration des fonctions étagées positives

Définition 2.5.1. Soit f une fonction étagée positive, prenant les valeurs distinctes $\alpha_1, \dots, \alpha_n$ ($\alpha_i \geq 0$). On note $A_i = f^{-1}(\alpha_i)$, pour $i \in \{1, \dots, n\} = \{x \in E : f(x) = \alpha_i\}$. On a alors

$$f = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}.$$

On appelle **intégrale de f par rapport à μ** , notée $\int f d\mu$, la quantité de $\overline{\mathbb{R}}_+$ définie par

$$\int f d\mu = \sum_{i=1}^n \alpha_i \mu(A_i),$$

avec la convention usuelle en théorie de la mesure : $0 \times \infty = 0$.

On remarque qu'ainsi défini les ensembles $A_i \in \mathcal{A}$ sont deux à deux disjoints et $\bigcup_{i=1}^n A_i = E$.

Proposition 2.5.2. L'intégrale des fonctions étagées vérifie les propriétés suivantes :

1. **Linéarité** : si f et g sont deux fonctions étagées positives et $a \in \mathbb{R}_+$, alors

$$\int (af + g) d\mu = a \int f d\mu + \int g d\mu.$$

2. *Croissance* : si f et g sont deux fonctions étagées positives telles que $f \leq g$, alors

$$\int f d\mu \leq \int g d\mu.$$

Preuve. En exercice. □

Exemple. On considère la fonction étagée $f(x) = \mathbb{1}_{[0,2]} + 2\mathbb{1}_{[1,3]}$ sur l'espace mesuré $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, où λ est la mesure de Lebesgue. On remarque que

$$f(x) = \mathbb{1}_{[0,1[} + 3\mathbb{1}_{[1,2]} + 2\mathbb{1}_{[2,3]}.$$

D'après la définition de l'intégrale pour les fonctions étagées, on a donc

$$\begin{aligned} \int f d\lambda &= \lambda([0, 1[) + 3\lambda([1, 2]) + 2\lambda([2, 3]) \\ &= 1 + 3 + 2 = 6. \end{aligned}$$

On remarque que, d'après la définition de la mesure de Lebesgue, $\int \mathbb{1}_{[0,2]} d\lambda = 2$ et $\int 2\mathbb{1}_{[1,3]} d\lambda = 2 \times 2 = 4$. On a bien

$$\int f d\lambda = \int \mathbb{1}_{[0,2]} d\lambda + \int 2\mathbb{1}_{[1,3]} d\lambda.$$

Remarque. Si f est étagée positive avec $f(x) = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}$ avec α_i pas forcément distincts et A_i pas forcément disjoints, on a encore

$$\int f d\mu = \sum_{i=1}^n \alpha_i \mu(A_i).$$

(conséquence de la linéarité de l'intégrale).

2.5.b) Intégration des fonctions mesurables positives

Définition 2.5.3. Soit $f \in \mathcal{M}_+$, i.e. une fonction mesurable positive à valeurs dans $\overline{\mathbb{R}}_+$. On appelle intégrale de f par rapport à μ , notée $\int f d\mu$, la quantité de $\overline{\mathbb{R}}_+$ définie par

$$\int f d\mu = \sup \left\{ \int \psi d\mu : \psi \in \mathcal{E}_+ \text{ telle que } \psi \leq f \right\}.$$

On remarque que si f est étagée positive les deux définitions coïncident car le supremum est atteint en $\psi = f$.

Proposition 2.5.4 (Croissance de l'intégrale). Pour toutes fonctions f et g mesurables positives avec $f \leq g$, on a

$$\int f d\mu \leq \int g d\mu.$$

Preuve. C'est une conséquence de l'inclusion

$$\left\{ \int \psi d\mu : \psi \in \mathcal{E}_+ \text{ telle que } \psi \leq f \right\} \subset \left\{ \int \psi d\mu : \psi \in \mathcal{E}_+ \text{ telle que } \psi \leq g \right\}$$

et de la définition de l'intégrale. □

On va maintenant donner un des théorèmes limites important de la théorie de la mesure.

Théorème 2.5.5 (Théorème de Beppo-Levi, aussi appelé Théorème de convergence monotone).

Soit $(f_n)_{n \in \mathbb{N}}$ une suite **croissante** de fonctions **mesurables positives** (dans $\mathcal{M}_+$).

Alors $f = \lim_{n \rightarrow +\infty} f_n (= \sup_n f_n)$ est aussi dans $\mathcal{M}_+$ et

$$\int f d\mu = \lim_{n \rightarrow +\infty} \int f_n d\mu.$$

Il s'agit d'un théorème d'intervention de limite et intégrale : sous les conditions du théorème, on a

$$\int \lim_{n \rightarrow +\infty} f_n d\mu = \lim_{n \rightarrow +\infty} \int f_n d\mu.$$

Preuve. On sait déjà que le supremum d'éléments de $\mathcal{M}_+$ est dans $\mathcal{M}_+$. Comme $f_n \leq f_{n+1}$, la suite $(\int f_n d\mu)_{n \in \mathbb{N}}$ est croissante (par croissance de l'intégrale) et donc elle admet une limite dans $\overline{\mathbb{R}}_+$. Par ailleurs, pour tout $n \geq 0$, on a $f_n \leq f$, on en déduit que

$$\int f_n d\mu \leq \int f d\mu,$$

et donc

$$\lim_{n \rightarrow +\infty} \int f_n d\mu \leq \int f d\mu. \quad (2.1)$$

Montrons l'inégalité inverse.

Soit ψ une fonction étagée positive telle que $\psi \leq f$ et $a \in]0, 1[$. On introduit

$$E_n = \{x \in E : a\psi(x) \leq f_n(x)\}$$

La suite $(E_n)_{n \in \mathbb{N}}$ est croissante (au sens de l'inclusion) par croissance de la suite (f_n) . On souhaite montrer que $E = \bigcup_{n \in \mathbb{N}} E_n$.

Soit $x \in E$.

Si $\psi(x) = 0$, alors pour tout $n \in \mathbb{N}$, $x \in E_n$.

Si $\psi(x) > 0$, alors

$$a\psi(x) < \psi(x) \leq f(x) = \lim_{n \rightarrow +\infty} f_n(x).$$

Par conséquent, pour n assez grand $x \in E_n$.

On a donc $E = \bigcup_{n \in \mathbb{N}} E_n$.

D'autre part, par définition de E_n , on a pour tout $x \in E$, $a\psi \mathbb{1}_{E_n} \leq f_n$ et donc par croissance de l'intégrale, pour tout $n \in \mathbb{N}$,

$$a \int \psi \mathbb{1}_{E_n} d\mu = \int a\psi \mathbb{1}_{E_n} d\mu \leq \int f_n d\mu. \quad (2.2)$$

La fonction $a\psi \mathbb{1}_{E_n}$ est étagée positive. Donc si $\psi = \sum_{i=1}^p \alpha_i \mathbb{1}_{A_i}$, on a

$$\int \psi \mathbb{1}_{E_n} d\mu = \sum_{i=1}^p \alpha_i \mu(A_i \cap E_n).$$

Pour tout $i \in \{1, \dots, p\}$, la suite $(\mu(A_i \cap E_n))_{n \in \mathbb{N}}$ est croissante et converge vers $\mu(A_i)$ (voir proposition 2.4.5). On en déduit que

$$\lim_{n \rightarrow +\infty} \int \psi \mathbb{1}_{E_n} d\mu = \sum_{i=1}^p \alpha_i \mu(A_i) = \int \psi d\mu.$$

D'après l'équation (2.2), on a alors par passage à la limite,

$$\int a\psi d\mu \leq \lim_{n \rightarrow +\infty} \int f_n d\mu.$$

L'inégalité est vraie pour tout $a \in]0, 1[$, d'où, pour tout $\psi \in \mathcal{E}_+$ avec $\psi \leq f$,

$$\int \psi d\mu \leq \lim_{n \rightarrow +\infty} \int f_n d\mu.$$

Par conséquent,

$$\int f d\mu = \sup \left\{ \int \psi d\mu : \psi \in \mathcal{E}_+ \text{ avec } \psi \leq f \right\} \leq \lim_{n \rightarrow +\infty} \int f_n d\mu,$$

et donc l'égalité est établie. $\square$

Corollaire 2.5.6 (Linéarité de l'intégrale). *Si f et g sont deux fonctions mesurables positives, alors*

$$\int (f + g) d\mu = \int f d\mu + \int g d\mu.$$

Preuve. D'après le théorème 2.3.3 d'approximation des fonctions mesurables positives, il existe deux suites croissantes (f_n) et (g_n) de fonctions étagées positives telles que

$$\lim_{n \rightarrow +\infty} f_n = f \quad \text{et} \quad \lim_{n \rightarrow +\infty} g_n = g.$$

Alors $(f_n + g_n)$ est une suite croissante de fonctions étagées positives et

$$\lim_{n \rightarrow +\infty} (f_n + g_n) = f + g.$$

Par linéarité de l'intégrale pour les fonctions étagées positives, on a

$$\int (f_n + g_n) d\mu = \int f_n d\mu + \int g_n d\mu.$$

On conclut par passage à la limite en utilisant le théorème de Beppo-Levi. $\square$

Corollaire 2.5.7 (Intervention des signes somme et intégrale). *Si $(f_n)_{n \in \mathbb{N}}$ est une suite de fonctions mesurables positives, on a*

$$\int \left(\sum_{n=0}^{\infty} f_n \right) d\mu = \sum_{n=0}^{\infty} \left(\int f_n d\mu \right)$$

(égalité dans $\overline{\mathbb{R}}_+$).

Preuve. La suite de fonctions $g_n = \sum_{k=0}^n f_k$ est une suite croissante de fonctions mesurables positives. Donc par le théorème de Beppo-Levi

$$\int \lim_{n \rightarrow +\infty} g_n d\mu = \lim_{n \rightarrow +\infty} \int g_n d\mu,$$

d'où le résultat en utilisant la linéarité de l'intégrale,

$$\int g_n d\mu = \sum_{k=0}^n \int f_k d\mu.$$

$\square$

2.5.c) Intégration des fonctions mesurables

Soit $f : (E, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une fonction mesurable. On peut écrire

$$f = f^+ - f^-$$

avec $f^+ = \sup(f, 0)$ et $f^- = -\inf(f, 0)$. Les fonctions f^+ et f^- sont dans $\mathcal{M}^+$, dont l'intégrale a été définie dans la section précédente. Cependant, on ne peut pas définir directement l'intégrale de f car il se peut que $\int f^+ d\mu = \int f^- d\mu = +\infty$ et la différence est alors une forme indéterminée.

Définition 2.5.8. Une fonction f définie sur E à valeurs dans $\mathbb{R}$ est dite **intégrable** par rapport à μ si

1. $f : (E, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est mesurable,
2. $\int |f| d\mu < +\infty$.

On note $\mathcal{L}^1(\mu) = \mathcal{L}^1(E, \mathcal{A}, \mu)$ l'ensemble des fonctions intégrales à valeurs réelles.

Proposition 2.5.9. Soit f une fonction mesurable à valeurs dans $\mathbb{R}$. Alors f est intégrable si et seulement si f^+ et f^- le sont.

Preuve. En effet, $|f| = f^+ + f^-$. □

Définition 2.5.10. Soit $f \in \mathcal{L}^1(\mu)$. On appelle *intégrale* de f , notée $\int f d\mu$, le nombre réel

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

Remarque. • On utilisera aussi la notation suivante pour l'intégrale

$$\int f d\mu = \int f(x) \mu(dx).$$

- On remarque que l'intégrale est définie sur tout l'espace E . On peut aussi considérer l'intégrale de f sur un ensemble mesurable quelconque A est définissant

$$\int_A f d\mu = \int f \mathbb{1}_A d\mu.$$

Proposition 2.5.11. L'ensemble $\mathcal{L}^1(\mu)$ est un espace vectoriel sur $\mathbb{R}$. De plus, l'intégrale vérifie les propriétés suivantes :

1. *Linéarité* : si $a \in \mathbb{R}$, $f, g \in \mathcal{L}^1(\mu)$ alors $\int (af + g) d\mu = a \int f d\mu + \int g d\mu$,
2. *Positivité* : si $f \in \mathcal{L}^1(\mu)$ avec $f \geq 0$, alors $\int f d\mu \geq 0$,
3. *Croissance* : si $f, g \in \mathcal{L}^1(\mu)$ avec $f \leq g$, alors $\int f d\mu \leq \int g d\mu$,
4. si $f \in \mathcal{L}^1(\mu)$, alors $\left| \int f d\mu \right| \leq \int |f| d\mu$.

Preuve. On sait déjà que l'ensemble des fonctions mesurables est un espace vectoriel sur $\mathbb{R}$. Si $a \in \mathbb{R}$ et $f, g \in \mathcal{L}^1(\mu)$, alors $|af + g| \leq |a||f| + |g|$, donc

$$\int |af + g| d\mu \leq |a| \int |f| d\mu + \int |g| d\mu < +\infty,$$

c'est à dire $af + g \in \mathcal{L}^1(\mu)$. $\mathcal{L}^1(\mu)$ est bien un espace vectoriel sur $\mathbb{R}$. Pour obtenir les autres propriétés, il suffit de décomposer les fonctions f et g en partie positive et partie négative et d'utiliser les propriétés de l'intégrale des fonctions mesurables positives. $\square$

2.5.d) Intégrale par rapport à une mesure discrète

Soit $(E, \mathcal{A})$ un espace mesurable et $(a_k)_{k \geq 1}$ une suite de points de E et $(\alpha_k)_{k \geq 1}$ une suite de réels positifs. On considère la mesure discrète

$$\mu = \sum_{k=1}^{+\infty} \alpha_k \delta_{a_k}.$$

Rappel : pour $A \in \mathcal{A}$, on a $\mu(A) = \sum_{k=1}^{+\infty} \alpha_k \mathbf{1}_A(a_k)$.

On souhaite étudier l'ensemble $\mathcal{L}^1(\mu)$.

Si $f \in \mathcal{M}_+$, alors $\int f d\mu = \sum_{k=1}^{+\infty} \alpha_k f(a_k)$ (cette valeur est dans $\overline{\mathbb{R}}_+$).

On a $f \in \mathcal{L}^1(\mu)$ si f est mesurable et $\sum_{k=1}^{+\infty} \alpha_k |f(a_k)| < +\infty$. On a alors,

$$\int f d\mu = \sum_{k=1}^{+\infty} \alpha_k f(a_k).$$

Exemple. On considère $(E, \mathcal{A}) = (\mathbb{N}, \mathcal{P}(\mathbb{N}))$.

Comme $\mathcal{A} = \mathcal{P}(\mathbb{N})$, toute fonction $f : \mathbb{N} \rightarrow \mathbb{R}$ est mesurable. On considère la mesure de Bernoulli $\mu = p\delta_1 + (1-p)\delta_0$, avec $p \in]0, 1[$, qui est une mesure de probabilité. Alors si f est mesurable positive on a

$$\int f d\mu = pf(1) + (1-p)f(0).$$

Comme μ est une somme finie de mesure de Dirac, toute fonction $f : \mathbb{N} \rightarrow \mathbb{R}$ est intégrable. Par exemple,

$$\int x\mu(dx) = p \times 1 + (1-p) \times 0 = p \quad \text{et} \quad \int \cos\left(\frac{\pi x}{4}\right) \mu(dx) = p \frac{\sqrt{2}}{2} + (1-p).$$

2.5.e) Intégrale de Lebesgue et comparaison avec l'intégrale de Riemann

On se place sur l'espace mesuré $(E, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ où λ représente la mesure de Lebesgue sur $\mathbb{R}$.

Proposition 2.5.12. Soit $a, b \in \mathbb{R}$ avec $a < b$ et f une fonction continue sur $[a, b]$. Alors $f\mathbf{1}_{[a,b]} \in \mathcal{L}^1(\lambda)$ et

$$\int \mathbf{1}_{[a,b]} f d\lambda = \int_a^b f(x) dx,$$

où $\int_a^b f(x) dx$ représente l'intégrale de Riemann de la fonction f .

Remarque. La proposition s'étend facilement au cas d'une fonction f continue par morceaux.

Notation. Du fait de la proposition, on a par conséquent, tendance à écrire, $\int f \mathbb{1}_{[a,b]} d\lambda$ sous la forme $\int_{[a,b]} f(x) dx$, et même $\int_a^b f(x) dx$, même pour des fonctions f intégrables au sens de Lebesgue sans hypothèse de continuité.

Preuve. La fonction f étant continue sur $[a, b]$, il existe $M > 0$ tel que $\forall x \in [a, b]$, $|f(x)| \leq M$. La fonction $f \mathbb{1}_{[a,b]}$ est borélienne et

$$\int \mathbb{1}_{[a,b]} |f| d\lambda \leq M \lambda([a, b]) = M(b - a) < +\infty.$$

Par conséquent, $f \mathbb{1}_{[a,b]} \in \mathcal{L}^1(\lambda)$. De même, pour tout $z \in [a, b]$, $f \mathbb{1}_{[a,z]}$ est Lebesgue-intégrable.

Pour $z \in [a, b]$, on définit $F(z) = \int f \mathbb{1}_{[a,z]} d\lambda$. On va montrer que F est une primitive de f . Pour cela on étudie la dérivabilité de F .

Soit $z \in [a, b]$. Soit $h > 0$. On a

$$f \mathbb{1}_{[a,z+h]} = f \mathbb{1}_{[a,z]} + f \mathbb{1}_{]z,z+h]}.$$

Par conséquent,

$$\begin{aligned} \frac{F(z+h) - F(z)}{h} - f(z) &= \frac{1}{h} \int f \mathbb{1}_{]z,z+h]} d\lambda - f(z) \\ &= \frac{1}{h} \int (f - f(z)) \mathbb{1}_{]z,z+h]} d\lambda. \end{aligned}$$

Soit $\varepsilon > 0$. Comme f est continue en z , il existe $\eta > 0$ tel que si $y \in [a, b]$ avec $|y - z| \leq \eta$, on a $|f(z) - f(y)| \leq \varepsilon$.

Supposons que $h \in]0, \eta[$, alors

$$\left| \frac{F(z+h) - F(z)}{h} - f(z) \right| \leq \frac{1}{h} \int \varepsilon \mathbb{1}_{]z,z+h]} d\lambda = \varepsilon.$$

On vient montrer que F est dérivable à droite en z et sa dérivée est $f(z)$. On montre de même, en prenant $h < 0$, F est dérivable à gauche en z de dérivée est $f(z)$. Donc F est dérivable en z , pour tout $z \in [a, b]$. F est donc une primitive de f . D'après le théorème fondamental de l'analyse, il existe donc une constante $k \in \mathbb{R}$ telle que, $\forall z \in [a, b]$,

$$F(z) = \int_a^z f(t) dt + k.$$

On remarque que $F(a) = \int \mathbb{1}_a |f| d\lambda = 0$ car $\lambda(\{a\}) = 0$. Donc $k = 0$ et la proposition est démontrée. $\square$

Définition 2.5.13. Soit $a < b$, avec $(a, b) \in \overline{\mathbb{R}}$ et f une fonction borélienne.

On dit qu'une fonction f est **Lebesgue intégrable sur un intervalle** $[a, b[$ si $f \mathbb{1}_{[a,b[} \in \mathcal{L}^1(\lambda)$. La définition est équivalente pour des intervalles de la forme $]a, b[$, $]a, b]$ et $[a, b]$.

Plus généralement, si $A \in \mathcal{B}(\mathbb{R})$, f est Lebesgue intégrable sur A si $f \mathbb{1}_A \in \mathcal{L}^1(\lambda)$.

On rappelle que l'intégrale de Riemann d'une fonction f sur $[a, b[$ est convergente si f est localement Riemann intégrable (i.e. f Riemann intégrable sur $[a, c]$ pour tout $c \in [a, b[$) et si $\lim_{x \rightarrow b^-} \int_a^x f(t) dt$ existe et

est finie. On pose alors $\int_a^b f(t) dt = \lim_{x \rightarrow b^-} \int_a^x f(t) dt$.

On dit que l'intégrale de Riemann d'une fonction f sur $[a, b[$ est absolument convergente si l'intégrale $\int_a^b |f(t)| dt$ est convergente.

Proposition 2.5.14. *Soit $a < b$ avec $b \in \mathbb{R} \cup \{+\infty\}$. Soit $f : [a, b[\rightarrow \mathbb{R}$ une fonction continue. Alors $f\mathbb{1}_{[a, b[} \in \mathcal{L}^1(\lambda)$ si et seulement si $\int_a^b f(t) dt$ est absolument convergente au sens de Riemann et dans ce cas*

$$\int \mathbb{1}_{[a, b[} f d\lambda = \int_a^b f(x) dx.$$

Preuve. Si f est positive. Soit $(b_n)_{n \geq 0}$ une suite croissante de points de $[a, b[$ qui converge vers b . Alors, comme f est continue sur $[a, b_n]$, on a pour tout $n \geq 0$,

$$\int \mathbb{1}_{[a, b_n]} f d\lambda = \int_a^{b_n} f(x) dx.$$

La suite $(f\mathbb{1}_{[a, b_n]})$ est une suite croissante de fonction, donc d'après le théorème de Beppo-Levi, on a

$$\begin{aligned} \int \mathbb{1}_{[a, b[} f d\lambda &= \lim_{n \rightarrow +\infty} \int \mathbb{1}_{[a, b_n]} f d\lambda \\ &= \lim_{n \rightarrow +\infty} \int_a^{b_n} f(x) dx \in \overline{\mathbb{R}}_+. \end{aligned}$$

Par définition, f est intégrable au sens de Lebesgue si et seulement si cette limite est finie, donc si et seulement si l'intégrale de Riemann sur $[a, b[$ est convergente. De plus, les intégrales sont les mêmes.

Si on ne suppose plus f positive. On a f est intégrable au sens de Lebesgue si et seulement si $|f|$ est intégrable au sens de Lebesgue, donc si et seulement si l'intégrale de Riemann sur $[a, b[$ est absolument convergente. Si c'est le cas, on écrit $f = f^+ - f^-$. On a $f^+ \leq |f|$ et $f^- \leq |f|$, donc f^+ et f^- sont intégrables (au sens de Lebesgue et de Riemann) et

$$\int \mathbb{1}_{[a, b[} f^+ d\lambda = \int_a^b f^+(x) dx \text{ et } \int \mathbb{1}_{[a, b[} f^- d\lambda = \int_a^b f^-(x) dx,$$

d'où le résultat par linéarité. □

Remarque. 1. Si f est continue par morceaux et positive sur $[a, b]$, alors

$$\int \mathbb{1}_{[a, b]} f d\lambda = \int_a^b f(x) dx.$$

2. On a des résultats similaires pour des fonctions continues sur $]a, b]$ ou $]a, b[$ avec $a, b \in \overline{\mathbb{R}}$, $a < b$.

Lorsqu'on sort du cadre des fonctions continues par morceaux, les liens entre intégrales de Lebesgue et intégrale de Riemann sont plus subtils :

- Il existe des fonctions intégrables au sens de Lebesgue et qui ne sont pas intégrables au sens de Riemann. Par exemple, $f(x) = \mathbb{1}_{\mathbb{Q} \cap [0, 1]}(x)$. Cette fonction est mesurable et intégrable au sens de Lebesgue. On a $\int f d\lambda = 0$ car $\mathbb{Q}$ est de mesure nulle pour Lebesgue. Par contre, cette fonction n'est pas intégrable au sens de Riemann (les sommes de Darboux ne coïncident pas).

- Il existe des fonctions Riemann intégrables qui ne sont pas intégrables au sens de Lebesgue. Par exemple, l'intégrale de Riemann de $x \mapsto \frac{\sin x}{x}$ sur $[1, +\infty[$ est convergente, mais comme elle n'est pas absolument convergente, i.e. $\int_1^{+\infty} \left| \frac{\sin x}{x} \right| dx = +\infty$, elle n'est donc pas intégrable au sens de Lebesgue.

Théorème 2.5.15 (Théorème de Lebesgue, admis). *Une fonction $f : [a, b] \rightarrow \mathbb{R}$ bornée est intégrable au sens de Riemann si et seulement si il existe $N \subset [a, b]$ de mesure de Lebesgue nulle tel que f est continue en tout point $x \in [a, b] \setminus N$.*

La fonction $\mathbf{1}_{\mathbb{Q} \cap [0,1]}$ n'est pas Riemann intégrable, elle est en fait continue en aucun point de $[0, 1]$.

Exemple. Soit $f(x) = x^2$ sur l'intervalle $[0, 4]$, alors, comme la fonction est continue sur $[0, 4]$, on a

$$\int f \mathbf{1}_{[0,4]} d\lambda = \int_0^4 x^2 dx = \left[\frac{1}{3} x^3 \right]_0^4 = \frac{64}{3}.$$

Soit

$$g(x) = \begin{cases} -x & \text{si } x \in]-1, 0[\\ 1 & \text{si } x = 0 \\ x^2 & \text{si } x \in]0, 1] \end{cases}$$

Comme g est continue par morceaux, on a

$$\int g \mathbf{1}_{]-1,1]} d\lambda = \int_{-1}^0 x dx + \int_0^1 x^2 dx = -\frac{1}{2} + \frac{1}{3} = -\frac{1}{6}.$$

2.5.f) Exercices sur l'intégrale

Exercice 2.5.1. Vrai ou Faux ?

Soit $(E, \mathcal{A}, \mu)$ un espace mesuré.

1. Si $f = \mathbb{I}_A$ avec $A \in \mathcal{A}$, alors $\int f d\mu = \mu(A)$.
2. Si $f : E \rightarrow [0, +\infty]$ est mesurable et vérifie $\mu(f^{-1}(\{+\infty\})) = 0$, alors f est intégrable.
3. La somme de deux fonctions intégrables est intégrable.
4. Le produit de deux fonctions intégrables est intégrable.

Exercice 2.5.2. Calculer $\int x \mu(dx)$ et $\int x^2 \mu(dx)$ pour les mesures sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ suivantes :

$$\mu_1 = \frac{1}{2} \delta_0 + \frac{1}{2} \delta_1, \quad \mu_2 = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \delta_k, \quad \mu_3 = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \delta_k.$$

Exercice 2.5.3. Soit μ la mesure définie par $\mu = d\lambda + \delta_0 + \delta_1$ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, où λ est la mesure de Lebesgue et δ_a est la mesure de Dirac en a . Pour chacune des fonctions suivantes dire si elle est intégrable et si oui, calculer son intégrale par rapport à μ .

$$f_1(x) = x, \quad f_2(x) = x \mathbf{1}_{[-1,1]}(x), \quad f_3(x) = \frac{1}{\sqrt{x}} \mathbf{1}_{[0,1]}(x), \quad f_4(x) = \frac{1}{\sqrt{x}} \mathbf{1}_{]0,1]}(x).$$

Exercice 2.5.4. Soit $(E, \mathcal{A}, \mu)$ un espace mesuré. Soit $f : E \rightarrow \overline{\mathbb{R}}_+$ une fonction mesurable. Pour tout entier n , on définit $f_n = f \mathbb{1}_{\{f \leq n\}}$. Montrer que, pour tout n , f_n est mesurable et déterminer $\lim_{n \rightarrow \infty} \int f_n d\mu$.

Exercice 2.5.5. (*) [Théorème de transfert]

Soit $(E, \mathcal{A}, \mu)$ un espace mesuré, $(F, \mathcal{B})$ un espace mesurable et $\varphi : E \rightarrow F$ une application mesurable. Pour tout $B \in \mathcal{B}$, on définit la mesure image $\nu(B) = \mu_\varphi(B) := \mu(\varphi^{-1}(B))$ (voir Exercice 2.4.7).

Soit $h : F \rightarrow \mathbb{R}$ mesurable. Montrer que h est ν -intégrable si et seulement si $h \circ \varphi$ est μ -intégrable et dans ce cas, $\int_F h d\nu = \int_E h \circ \varphi d\mu$.

Exercice 2.5.6. [Convergence monotone décroissante]

Soit $(E, \mathcal{A}, \mu)$ un espace mesuré et $(f_n)_{n \in \mathbb{N}}$ une suite décroissante de fonctions mesurables positives.

1. On suppose que $\int f_0 d\mu < +\infty$. Montrer que

$$\lim_n \int f_n d\mu = \int \lim_n f_n d\mu.$$

2. Montrer que ce résultat peut être faux si $\int f_0 d\mu = +\infty$.
3. Quel résultat obtient-on si on choisit $f_n = \mathbb{1}_{A_n}$ avec $(A_n)_n$ suite décroissante d'éléments de $\mathcal{A}$, avec $\mu(A_0) < +\infty$?

Exercice 2.5.7. Étudier la convergence des séries suivantes :

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{1}{(1+n)^m} \quad \text{et} \quad \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{1}{(4n-1)^{2m}}.$$

(Indice : pensez au théorème de Beppo-Levi).

Exercice 2.5.8. Soit μ la mesure définie par $\mu = \sum_{k=1}^{\infty} p(1-p)^{k-1} \delta_k$. On définit la suite de fonctions $(f_n)_{n \geq 1}$ par

$$f_n(x) = \left(1 + \frac{x}{n}\right)^n \mathbb{1}_{[0, n]}(x).$$

1. Pour tout $x \geq 0$, déterminer la limite de la suite $(f_n(x))_n$.
2. On note g_n la fonction définie sur $[0, n[$ par $g_n(x) = \ln f_{n+1}(x) - \ln f_n(x)$. Montrer g_n est positive. En déduire que $(f_n)_{n \in \mathbb{N}}$ est une suite croissante.
3. Montrer que la suite $\left(\int f_n d\mu\right)_n$ converge vers une limite à déterminer.

2.6 Un aperçu de l'intégrale multiple

On considère deux espaces mesurés $(E, \mathcal{A}, \mu)$ et $(F, \mathcal{B}, \nu)$, avec μ et ν mesures σ -finies. On introduit la tribu produit, notée $\mathcal{A} \otimes \mathcal{B}$ sur $E \times F$, qui est la tribu engendrée par l'ensemble $\{A \times B : A \in \mathcal{A} \text{ et } B \in \mathcal{B}\}$.

Théorème 2.6.1 (Admis). Soit $(E, \mathcal{A}, \mu)$ et $(F, \mathcal{B}, \nu)$ deux espaces mesurés σ -finis.

1. Il existe une unique mesure m sur $(E \times F, \mathcal{A} \otimes \mathcal{B})$ telle que pour tout $A \in \mathcal{A}$ et $B \in \mathcal{B}$,

$$m(A \times B) = \mu(A)\nu(B),$$

avec la convention $0 \times +\infty = 0$. Cette mesure est σ -finie. Elle est en général notée $\mu \otimes \nu$ et on l'appelle **mesure produit** de μ et ν .

2. Soit $C \in \mathcal{A} \otimes \mathcal{B}$. Les fonctions $x \mapsto \int_F \mathbf{1}_C(x, y) \nu(dy)$ et $y \mapsto \int_E \mathbf{1}_C(x, y) \mu(dx)$ sont respectivement $\mathcal{A}$ -mesurable et $\mathcal{B}$ -mesurable et

$$\mu \otimes \nu(C) = \int_{E \times F} \mathbf{1}_C d\mu \otimes \nu = \int_E \left(\int_F \mathbf{1}_C(x, y) \nu(dy) \right) \mu(dx) = \int_F \left(\int_E \mathbf{1}_C(x, y) \mu(dx) \right) \nu(dy)$$

Exemple. [Mesure de Lebesgue sur $\mathbb{R}^d$]

Il existe une unique mesure λ_d sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ telle que pour tout produit d'intervalles $I_1 \times \dots \times I_d$, $\lambda_d(I_1 \times \dots \times I_d)$ soit égal au produit des longueurs des intervalles, i.e.

$$\lambda_d(I_1 \times \dots \times I_d) = \lambda(I_1) \dots \lambda(I_d).$$

où λ est la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

On a

$$\lambda_d = \underbrace{\lambda \otimes \dots \otimes \lambda}_{d \text{ fois}}.$$

λ_d est la seule mesure sur $\mathcal{B}(\mathbb{R}^d)$ telle que

1. $\lambda_d([0, 1]^d) = 1$,
2. $\forall a \in \mathbb{R}^d, \forall A \in \mathcal{B}(\mathbb{R}^d), \lambda_d(a + A) = \lambda_d(A)$.

On remarque que pour $A_i \in \mathcal{B}(\mathbb{R})$, on a $\lambda_d(A_1 \times \dots \times A_d) = 0$ dès que l'un des ensembles A_i est de mesure de Lebesgue nul.

Pour $C \in \mathcal{A} \otimes \mathcal{B}$, on a

$$\begin{aligned} \int_{E \times F} \mathbf{1}_C d\mu \otimes \nu &= \int_{E \times F} \mathbf{1}_C(x, y) \mu \otimes \nu(dx, dy) \\ &= \int_E \left(\int_F \mathbf{1}_C(x, y) \nu(dy) \right) \mu(dx) \\ &= \int_F \left(\int_E \mathbf{1}_C(x, y) \mu(dx) \right) \nu(dy), \end{aligned}$$

ce qui signifie que l'ordre dans lequel on intègre par rapport à chacune des mesures importe peu quand on calcule l'intégrale d'une fonction indicatrice. Cette propriété reste vraie pour les fonctions mesurables positives.

Exemple. On se place sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), \lambda_2)$, où λ_2 est la mesure de Lebesgue sur $\mathbb{R}^2$.

On a $\lambda_2([0, 1] \times \mathbb{Q}) = 0$, $\lambda_2(\mathbb{R} \times \mathbb{Q}) = 0$.

On considère les ensembles $C = [2, 13] \times [-4, -2]$ et $D = \{(x, y) \in \mathbb{R} : 0 \leq x \leq 1 \text{ et } 0 \leq y \leq x\}$. Ce sont bien des boréliens car ce sont des ensembles fermés. On a

$$\begin{aligned}\lambda_2(C) &= \lambda([2, 13])\lambda([-4, -2]) = 11 \times 2 = 22, \\ \lambda_2(D) &= \int_0^1 \left(\int_0^x \lambda(dy) \right) \lambda(dx) = \int_0^1 x \lambda(dx) = \frac{1}{2} \\ &= \int_0^1 \left(\int_y^1 \lambda(dx) \right) \lambda(dy) = \int_0^1 (1 - y) \lambda(dy) = \frac{1}{2}.\end{aligned}$$

Théorèmes de Tonelli et de Fubini

Théorème 2.6.2 (Théorème de Fubini-Tonelli).

Soit $(E, \mathcal{A}, \mu)$ et $(F, \mathcal{B}, \nu)$ deux espaces mesurés σ -finies. Soit f une fonction mesurable sur $(E \times F, \mathcal{A} \otimes \mathcal{B})$ à valeurs dans $\overline{\mathbb{R}}_+$.

1. Les fonctions

$$\begin{aligned}(E, \mathcal{A}) &\rightarrow (\overline{\mathbb{R}}_+, \mathcal{B}(\overline{\mathbb{R}}_+)) & \text{et} & & (F, \mathcal{B}) &\rightarrow (\overline{\mathbb{R}}_+, \mathcal{B}(\overline{\mathbb{R}}_+)) \\ x &\mapsto \int_F f(x, y) \nu(dy) & & & y &\mapsto \int_E f(x, y) \mu(dx)\end{aligned}$$

sont respectivement $\mathcal{A}$ -mesurable et $\mathcal{B}$ -mesurable.

2. On a

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

Preuve. Le raisonnement se fait en trois étapes : on montre le théorème

- (i) pour les fonctions indicatrices, puis
- (ii) pour les fonctions étagées positives, et enfin
- (iii) pour les fonctions mesurables positives.

Le point (i) est satisfait d'après le théorème précédent. On obtient le point (ii) par stabilité de la mesurabilité par combinaison linéaire et par linéarité de l'intégrale. Le dernier point est obtenu en utilisant l'approximation des fonctions mesurables positives par des suites croissantes de fonctions étagées positives, la stabilité par passage à la limite de la mesurabilité et par le théorème de Beppo-Lévi. $\square$

Corollaire 2.6.3. Soit f une fonction mesurable sur $(E \times F, \mathcal{A} \otimes \mathcal{B})$ à valeurs dans $\overline{\mathbb{R}}$. Alors f est intégrable pour la mesure $\nu \otimes \mu$ si et seulement si l'une des conditions suivantes est vérifiée :

$$\int_E \left(\int_F |f(x, y)| \nu(dy) \right) \mu(dx) < +\infty \quad \text{ou} \quad \int_F \left(\int_E |f(x, y)| \mu(dx) \right) \nu(dy) < +\infty.$$

Définition 2.6.4. Soit $(E, \mathcal{A}, \mu)$ un espace mesuré.

Un sous-ensemble $N \subset E$ est dit **négligeable** pour la mesure μ , s'il existe $A \in \mathcal{A}$ tel que $N \subset A$ et $\mu(A) = 0$.

Une propriété P est dite vraie pour presque tout $x \in E$ si l'ensemble $\{x \in E : \text{la propriété } P \text{ ne soit pas vraie}\}$ est un ensemble négligeable.

On remarque qu'un ensemble négligeable n'est pas forcément mesurable, mais il est inclus dans un ensemble mesurable de mesure nulle.

Proposition 2.6.5. Soit f une fonction mesurable de $(E, \mathcal{A}, \mu)$ à valeurs dans $\overline{\mathbb{R}}_+$. Si $f \in \mathcal{L}^1(\mu)$, i.e $\int f d\mu < +\infty$, alors $\{f = +\infty\}$ est un ensemble de mesure nulle :

$$\mu(\{f = +\infty\}) = 0.$$

Lemme 2.6.6 (Inégalité de Markov). Soit f une fonction mesurable de $(E, \mathcal{A}, \mu)$ à valeurs dans $\overline{\mathbb{R}}_+$. Alors $\forall a > 0$,

$$\mu(\{f \geq a\}) \leq \frac{1}{a} \int f d\mu.$$

Preuve. En effet, on a

$$\int f d\mu \geq \int_{\{f \geq a\}} f d\mu \geq a\mu(\{f \geq a\}).$$

□

Preuve. (Proposition 2.6.5) On remarque que $\{f = +\infty\} = \bigcap_{n \geq 1} \{f \geq n\}$, avec $\{f \geq n\}$ suite décroissante de d'ensembles mesurables et, comme $f \in \mathcal{L}^1(\mu)$, d'après l'inégalité de Markov

$$\mu(\{f \geq n\}) \leq \frac{1}{n} \int f d\mu.$$

On déduit ainsi que $\mu(\{f = +\infty\}) = \lim_{n \rightarrow +\infty} \mu(\{f \geq n\}) = 0$.

□

D'après les résultats précédents, on a le résultat suivant :

Théorème 2.6.7 (Théorème de Fubini).

Soit $(E, \mathcal{A}, \mu)$ et $(F, \mathcal{B}, \nu)$ deux espaces mesurés σ -finies. Soit f une fonction intégrable sur $(E \times F, \mathcal{A} \otimes \mathcal{B}, \mu \otimes \nu)$ à valeurs réelles. Alors

1. pour presque tout $x \in E$, la fonction $y \mapsto f(x, y)$ est dans $\mathcal{L}^1(\nu)$; de plus la fonction $x \mapsto \int_F f(x, y) \nu(dy)$ est μ -intégrable,
2. pour presque tout $y \in F$, la fonction $x \mapsto f(x, y)$ est dans $\mathcal{L}^1(\mu)$; de plus la fonction $y \mapsto \int_E f(x, y) \mu(dx)$ est ν -intégrable,
3. on a

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

Preuve. D'après le théorème de Tonelli, on a

$$\int_{E \times F} |f| d\mu \otimes \nu = \int_E \left[\int_F |f(x, y)| \nu(dy) \right] \mu(dx)$$

Par conséquent, d'après la proposition 2.6.5, $x \mapsto \int_F |f(x, y)| \nu(dy)$ est finie μ -p.p. Il existe donc un ensemble μ -négligeable $N \in \mathcal{A}$ tel que $\forall x \notin N$, $y \mapsto f(x, y)$ est ν -intégrable. On décompose f en la différence de ses parties positive et négative : $f = f^+ - f^-$. Comme f intégrable, f^+ et f^- sont intégrables (en effet, $0 \leq f^+ \leq |f|$ et $0 \leq f^- \leq |f|$).

Si $x \notin N$, les fonctions $y \mapsto f^+(x, y)$ et $y \mapsto f^-(x, y)$ sont ν -intégrables et on a

$$\forall x \notin N, \int f(x, y) \nu(dy) = \int f^+(x, y) \nu(dy) - \int f^-(x, y) \nu(dy)$$

D'après le théorème de Tonelli, les fonctions $x \mapsto \int f^\pm(x, y) \nu(dy) \mathbf{1}_{E \setminus N}$ sont mesurables et on a

$$\int_{E \setminus N} \left[\int_F f^\pm(x, y) \nu(dy) \right] \mu(dx) = \int_E \left[\int_F f^\pm(x, y) \nu(dy) \right] \mu(dx) = \int_{E \times F} f^\pm d\mu \otimes \nu < +\infty.$$

Par conséquent, $x \mapsto \int_F f(x, y) \nu(dy)$ définie sur $E \setminus N$ est μ -intégrable comme combinaison linéaire de deux fonctions intégrables. On a enfin, comme N est négligeable,

$$\begin{aligned} \int_{E \times F} f d\mu \otimes \nu &= \int_{E \times F} f^+ d\mu \otimes \nu - \int_{E \times F} f^- d\mu \otimes \nu \\ &= \int_{E \setminus N} \left[\int_F f^+(x, y) \nu(dy) \right] \mu(dx) - \int_{E \setminus N} \left[\int_F f^-(x, y) \nu(dy) \right] \mu(dx) \\ &= \int_{E \setminus N} \left[\int_F (f^+(x, y) - f^-(x, y)) \nu(dy) \right] \mu(dx) \\ &= \int_{E \setminus N} \left[\int_F f(x, y) \nu(dy) \right] \mu(dx) = \int_E \left[\int_F f(x, y) \nu(dy) \right] \mu(dx), \end{aligned}$$

ce qui achève la preuve. $\square$

Notation. Lorsque l'on intègre par rapport à des mesures produits $\mu \otimes \nu$, on utilise souvent les notations $d\mu \otimes \nu$, ou $\mu \otimes \nu(dx, dy)$, ou $\mu(dx) \nu(dy)$ (ce qui est justifié par le théorème de Fubini).

Exemple. 1. On se place sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), \lambda_2)$. On considère la fonction

$$f(x, y) = xy^2 \mathbf{1}_C(x, y) \quad \text{avec } C = \{(x, y) \in \mathbb{R}^2 : -1 \leq x \leq 1 \text{ et } 0 < y \leq x^2\}.$$

On a $f \in \mathcal{L}^1(\mathbb{R}^2, \lambda_2, \mathcal{B}(\mathbb{R}^2))$.

Cette fonction est mesurable comme produit d'une fonction continue et d'une fonction indicatrice d'un borélien (en effet, C est un borélien car $C = h^{-1}([0, +\infty[) \cap [-1, 1] \times]0, +\infty[$ avec $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ avec $h(x, y) = x^2 - y$ fonction continue). On remarque que $C \subset [-1, 1] \times [0, 1]$ et $\|f\|_\infty \leq 1$, donc $f \in \mathcal{L}^1(\mathbb{R}^2, \lambda_2, \mathcal{B}(\mathbb{R}^2))$.

2. Soit $\alpha, \beta > 0$ et λ_2 la mesure de Lebesgue sur $\mathbb{R}^2$. On considère la fonction définie sur $\mathbb{R}^2$ par

$$f(x, y) = e^{-(\alpha x + \beta y)} \mathbf{1}_{0 < x < y}.$$

La fonction f est borélienne sur $\mathbb{R}^2$ comme produit d'une fonction continue et d'une fonction indicatrice d'un borélien (en effet, $\{(x, y) \in \mathbb{R}^2 : 0 < x < y\} = h^{-1}([0, +\infty[) \cap]0, +\infty[^2$ avec $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ $h(x, y) = y - x$ fonction continue). Elle est de plus positive, donc d'après le théorème de Tonelli

$$\begin{aligned} \int f d\lambda_2 &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} f(x, y) dx \right] dy \\ &= \int_0^\infty \left[\int_x^\infty e^{-(\alpha x + \beta y)} dy \right] dx \\ &= \frac{1}{\beta} \left[\int_0^\infty e^{-(\alpha + \beta)x} dx \right] \\ &= \frac{1}{\beta(\alpha + \beta)} \end{aligned}$$

3. On se place sur $(\mathbb{R} \times \mathbb{N}, \mathcal{B}(\mathbb{R}) \otimes \mathcal{P}(\mathbb{N}), \lambda \otimes \nu)$ où λ mesure de Lebesgue sur $\mathbb{R}$ et ν mesure de comptage sur $\mathbb{N}$. On considère la fonction

$$f(x, n) = \frac{(-1)^n}{x^n} \mathbb{1}_{x \geq 2} \mathbb{1}_{n \geq 2}.$$

f est une fonction mesurable car pour tout $n \geq 2$, $x \mapsto f(x, n)$ est continue sur $]1, +\infty[$ et pour tout $x > 1$ $n \mapsto f(x, n)$ est mesurable par rapport à $\mathcal{P}(\mathbb{N})$. Étudions l'intégrabilité de f . D'après le théorème de Tonelli, on a

$$\begin{aligned} \int |f| d\lambda \otimes d\nu &= \int \left(\sum_{n=2}^{+\infty} \frac{1}{x^n} \right) \mathbb{1}_{x \geq 2} dx \\ &= \int_2^{+\infty} \frac{1}{x^2} \frac{1}{1 - 1/x} dx \end{aligned}$$

On a $\frac{1}{x(x-1)} = \frac{1}{x-1} - \frac{1}{x}$, qui est une fonction continue sur $[2, +\infty[$ et dont une primitive est $x \mapsto \ln\left(\frac{x-1}{x}\right)$. La primitive tend vers 0 quand $x \rightarrow +\infty$, donc $\int |f| d\lambda \otimes d\nu = \ln 2$ et f est intégrable par rapport à $\lambda \otimes \nu$.

De plus, on a par le théorème de Fubini,

$$\begin{aligned} \int f d\lambda \otimes d\nu &= \int \left(\sum_{n=2}^{+\infty} \frac{(-1)^n}{x^n} \right) \mathbb{1}_{x \geq 2} dx = \sum_{n=2}^{+\infty} \left(\int \frac{(-1)^n}{x^n} \mathbb{1}_{x \geq 2} dx \right) \\ &= \int_2^{+\infty} \frac{1}{x(x+1)} dx = \left[\ln\left(\frac{x}{x+1}\right) \right]_2^{+\infty} \\ &= \ln \frac{2}{3} \end{aligned}$$

On remarque que $g(x, n) = \frac{(-1)^n}{x^n} \mathbb{1}_{x \geq 1} \mathbb{1}_{n \geq 2}$ n'est pas intégrable par rapport à $\lambda \otimes \nu$.

Théorème de changement de variables

Soit $d \geq 1$. On se place sur l'espace mesuré $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$, où λ_d est la mesure de Lebesgue sur $\mathbb{R}^d$.

On notera dans la suite $\lambda_d(dx) = dx$.

Soit U et D deux ouverts de $\mathbb{R}^d$.

Soit $\varphi : U \rightarrow D$ un $\mathcal{C}^1$ -difféomorphisme (fonction de classe $\mathcal{C}^1$, inversible et d'inverse de classe $\mathcal{C}^1$). On a $\varphi = (\varphi_1, \dots, \varphi_d)$. On note J_φ le Jacobien, i.e. le déterminant de sa matrice jacobienne :

$$J_\varphi(y) = \det \begin{pmatrix} \frac{\partial \varphi_1}{\partial y_1}(y) & \dots & \frac{\partial \varphi_1}{\partial y_d}(y) \\ \vdots & \ddots & \vdots \\ \frac{\partial \varphi_d}{\partial y_1}(y) & \dots & \frac{\partial \varphi_d}{\partial y_d}(y) \end{pmatrix}$$

D'après le théorème d'inversion globale, si U est un ouvert, φ est de classe $\mathcal{C}^1$ inversible, et J_φ ne s'annule pas sur U alors $\varphi(U)$ est un ouvert et φ est un $\mathcal{C}^1$ -difféomorphisme de U dans $\varphi(U)$.

Théorème 2.6.8 (Théorème de changement de variables, voir Théorème 12.4 dans le livre de Briane et Pagès, [BP06]).

1. Soit f une fonction borélienne positive $f : D \rightarrow \overline{\mathbb{R}}^+$. Alors

$$\int_D f(x) dx = \int_U f(\varphi(y)) |J_\varphi(y)| dy.$$

2. Soit f une fonction borélienne $f : D \rightarrow \mathbb{R}$ (ou $f : D \rightarrow \mathbb{C}$). Alors f est intégrable sur D si et seulement si $\int_U |f(\varphi(y))| |J_\varphi(y)| dy < +\infty$ et dans ce cas,

$$\int_D f(x) dx = \int_U f(\varphi(y)) |J_\varphi(y)| dy.$$

Exemple. Passage en coordonnées polaires. On considère le difféomorphisme

$$\begin{aligned} \varphi : \mathbb{R}_+^* \times]-\pi, \pi[&\rightarrow \mathbb{R}^2 \setminus (\mathbb{R}_- \times \{0\}) \\ (\rho, \theta) &\mapsto \varphi(\rho, \theta) = (\rho \cos \theta, \rho \sin \theta) \end{aligned}$$

On a $J_\varphi(\rho, \theta) = \rho \cos^2 \theta + \rho \sin^2 \theta = \rho$.

En effet, $U = \mathbb{R}_+^* \times]-\pi, \pi[$ et $D = \mathbb{R}^2 \setminus (\mathbb{R}_- \times \{0\})$ sont bien des ensembles ouverts. La fonction φ est de classe $\mathcal{C}^1$ et son Jacobien ne s'annule pas. De plus, on a $\varphi(U) \subset D$ car si $\rho \sin \theta = 0$ (car $\theta \in]-\pi, \pi[$), comme $\rho > 0$ on a forcément $\theta = 0$ et donc $\rho \cos \theta = \rho > 0$. φ est surjective, car si $(x, y) \in D$, on a $(x, y) = \varphi(\rho, \theta)$ avec $\rho = \sqrt{x^2 + y^2}$ et $\theta \in]-\pi, \pi[$ tel que $\cos \theta = x/\rho$ et $\sin \theta = y/\rho$ (on distingue selon le signe de x et y pour identifier θ).

Comme l'ensemble $\mathbb{R}_- \times \{0\}$ est de mesure de Lebesgue nulle, pour toute fonction $f \in \mathcal{L}^1(\mathbb{R}^2)$,

$$\int_{\mathbb{R}^2} f(x, y) \lambda_2(dx, dy) = \int_{\mathbb{R}^2 \setminus (\mathbb{R}_- \times \{0\})} f(x, y) \lambda_2(dx, dy) = \int_0^{2\pi} \int_0^{+\infty} f(\rho \cos \theta, \rho \sin \theta) \rho d\rho d\theta.$$

En remarquant que $\left(\int_0^{+\infty} e^{-x^2} dx \right)^2 = \int_0^{+\infty} \int_0^{+\infty} e^{-(x+y)^2} dx dy$ et en utilisant le changement de variables en coordonnées polaires, on déduit que $\int_0^{+\infty} e^{-x^2} dx = \sqrt{\pi}/2$.

2.6.a) Exercices sur l'intégrale multiple

Exercice 2.6.1. 1. Montrer que la fonction $f : x \mapsto \frac{\sin x}{e^x - 1}$ est Lebesgue-intégrable sur $[0, +\infty[$.

2. Montrer que, pour tout $x > 0$, on peut encore écrire $f(x)$ sous la forme :

$$f(x) = \sum_{n=1}^{+\infty} e^{-nx} \sin x.$$

3. En déduire que $\int_0^{+\infty} \frac{\sin x}{e^x - 1} dx = \sum_{n=1}^{\infty} \frac{1}{n^2 + 1}$.

Exercice 2.6.2. On se place sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), \lambda_2)$ où λ_2 est la mesure de Lebesgue sur $\mathbb{R}^2$. Soit $D = \{(x, y) \in \mathbb{R}^2 : 0 \leq 2y < x \leq 2\}$ et $f : (x, y) \mapsto e^{x^2}$.

1. Justifier le fait que D soit un borélien.

2. Dessiner D et donner son aire.
3. Calculer l'intégrale de $\int_D f d\lambda_2$.

Exercice 2.6.3. On se place sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), \lambda_2)$ où λ_2 est la mesure de Lebesgue sur $\mathbb{R}^2$. Soit f la fonction définie par $f(x, y) = (2e^{-2xy} - e^{-xy})\mathbf{1}_{x>0, y \in [0,1]}$.

1. Vérifier que f est borélienne.
2. Calculer $\int_{[0,1]} \left(\int_{\mathbb{R}_+} f(x, y) dx \right) dy$ et $\int_{\mathbb{R}_+} \left(\int_{[0,1]} f(x, y) dy \right) dx$. Conclure.

Exercice 2.6.4. (*) On considère le domaine $D = \{(x, y) \in \mathbb{R}^2 : x + y > 0 \text{ et } x^2 + y^2 \leq 1\}$ et la fonction $f(x, y) = xy^2$. Étudier l'intégrabilité de f par rapport à la mesure de Lebesgue et si $f \in \mathcal{L}^1(\lambda_2)$, calculer son intégrale.

Exercice 2.6.5. Soit f une fonction de $\mathbb{R}^2$ dans $\mathbb{R}$. Soit I un intervalle de $\mathbb{R}$. Dans chacun des cas suivants, déterminer si f est Lebesgue intégrable sur I^2 et calculer, si elles existent, les intégrales $\int_I \left(\int_I f(x, y) \lambda(dx) \right) \lambda(dy)$ et $\int_I \left(\int_I f(x, y) \lambda(dy) \right) \lambda(dx)$.

$$f(x, y) = \begin{cases} -1 & \text{si } x > 0 \text{ et } 0 < y - x \leq 1, \\ 2 & \text{si } x > 0 \text{ et } 1 < y - x \leq 2, \\ -1 & \text{si } x > 0 \text{ et } 2 < y - x \leq 3, \\ 0 & \text{sinon.} \end{cases} \quad \text{avec } I = \mathbb{R}$$

et

$$(*) \quad f(x, y) = \frac{x^2 - y^2}{(x^2 + y^2)^2} \quad \text{avec } I = [0, 1].$$

Exercice 2.6.6. Soit B_2 le disque centré en 0 de rayon 1 dans $\mathbb{R}^2$. Calculer les intégrales suivantes

$$(a) \int_{B_2} \frac{1}{1 + x^2 + y^2} dx dy \quad (b) \int_{B_2} (x^2 + y^2)^{-1/4} dx dy \quad (c) \int_{B_2} x^2 dx dy.$$

Exercice 2.6.7. On note D le domaine du plan défini par

$$D = \left\{ (x, y) \in \mathbb{R}^2 : 0 < x < \frac{1}{1+2y}, 0 < y < \frac{1}{1+2x} \right\}.$$

On considère l'application $\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ telle que $\Phi(x, y) = (u, v)$ avec

$$u = \frac{x(1+y)}{1-xy} \quad v = \frac{y(1+x)}{1-xy}.$$

1. Déterminer le domaine $\Delta = \Phi(D)$.
2. Montrer que $\Phi : D \rightarrow \Delta$ est une bijection et calculer son inverse.

3. En utilisant le changement de variable Φ , Calculer les intégrales suivantes

$$(a) \int_D \frac{(1+x)(1+y)}{(1-xy)^3} dx dy, \quad (b) \int_D \frac{1-xy}{(1+x)(1+y)} dx dy, \quad (c) \int_D \frac{1}{1-xy} dx dy, .$$

Exercice 2.6.8. (*)[Calcul de volume de la boule unité]

La boule euclidienne unité de $\mathbb{R}^d$ est définie par $B_d = \{(x_1, \dots, x_d) \in \mathbb{R}^d : x_1^2 + \dots + x_d^2 \leq 1\}$. On note V_d son volume. Pour $d \geq 1$, on a

$$V_d = \int_{\mathbb{R}^d} \mathbb{1}_{x_1^2 + \dots + x_d^2 \leq 1} dx_1 \dots dx_d.$$

En utilisant le théorème de changement de variables, montrer que, pour $d \geq 2$, V_d satisfait la relation suivante : $V_d = \frac{2\pi}{d} V_{d-1}$. En déduire la valeur de V_d

Exercice 2.6.9. On souhaite calculer $\sum_{n \geq 1} \frac{1}{n^2}$.

1. Montrer que $\sum_{n \geq 1} \frac{1}{n^2} = \int_{[0,1]^2} \frac{dx dy}{1-xy}$.
2. En justifiant l'utilisation du théorème de changement de variables avec $(x, y) = \varphi(\theta, t)$ avec $\varphi(\theta, t) = (\cos \theta - t, \cos \theta + t)$ pour le calcul de l'intégrale, montrer que $\sum_{n \geq 1} \frac{1}{n^2} = \frac{\pi^2}{6}$.

Indication. On pourra utiliser les identités trigonométriques

$$\forall x > 0, \quad \arctan(x) + \arctan(1/x) = \pi/2 \quad \text{et} \quad \frac{1 - \cos(\theta)}{\sin \theta} = \tan \frac{\theta}{2}.$$

Exercice 2.6.10. Soient $a, b > 0$. On note

$$\Gamma(a) = \int_0^{+\infty} e^{-t} t^{a-1} dt \quad \text{et} \quad B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt.$$

1. Vérifier que $\Gamma(a)$ et $B(a, b)$ sont bien définies et finies.
2. Montrer que $\Gamma(a)\Gamma(b) = 4 \int_{\mathbb{R}_+^2} e^{-(u^2+v^2)} u^{2a-1} v^{2b-1} du dv$.
3. En utilisant les théorème de changement de variables (dont celui en coordonnées polaires), déduire que

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

4. Soit $\alpha \in]0, 1[$. En justifiant que l'on peut utiliser le changement de variables $(x, y) \rightarrow (x + y, \frac{y}{x})$, montrer que

$$\Gamma(\alpha)\Gamma(1-\alpha) = \int_0^{+\infty} \frac{x^{-\alpha}}{1+x} dx.$$

Chapitre 3

Variables aléatoires, Espérance et Indépendance

Les variables aléatoires sont l'objet de base en probabilité. Il s'agit de fonction du hasard dont nous présenterons dans ce chapitre les outils clefs pour leur étude.

3.1 Espaces de probabilité

3.1.a) Définitions et propriétés

Quand, on ne peut pas toujours prédire avec certitude le résultat d'une expérience, on parle alors d'expérience aléatoire. Par exemple, lancer un dé ou une pièce de monnaie sont des expériences aléatoires. D'autres phénomènes ne sont peut-être pas aléatoires (comme la succession des nucléotides sur un brin d'ADN), mais n'en connaissant pas tous les mécanismes, il peut être utile pour pouvoir les étudier de les supposer aléatoires.

Définition 3.1.1. *L'espace fondamental est un triplet généralement noté $(\Omega, \mathcal{F}, \mathbb{P})$, appelé **espace de probabilité** (ou **espace probabilisé**), où $(\Omega, \mathcal{F})$ est un espace mesurable muni d'une mesure de probabilité $\mathbb{P}$.*

Rappel :

Une *mesure probabilité*, ou plus simplement une probabilité, sur un espace mesurable $(\Omega, \mathcal{F})$ est une application $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ telle que

1. $\mathbb{P}(\emptyset) = 0$,
2. si $A_n \in \mathcal{F}$ pour tout $n \in \mathbb{N}$, tel que, pour $n \neq m$, $A_n \cap A_m = \emptyset$, alors

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n) \quad (\sigma\text{-additivité}),$$

3. $\mathbb{P}(\Omega) = 1$.

L'espace Ω permet en quelque sorte de décrire toutes des réalisations possibles d'une expérience. Un point $\omega \in \Omega$ est une réalisation particulière de cette expérience.

Exemple. On lance un dé, on s'intéresse au chiffre obtenu. L'ensemble fondamentale est alors fini, plus précisément

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Dans cet exemple, on munit Ω de la tribu $\mathcal{F} = \mathcal{P}(\Omega)$ et de la mesure d'équiprobabilité $\mathbb{P} = \frac{1}{6} \sum_{k=1}^6 \delta_k$.

En général l'espace $(\Omega, \mathcal{F}, \mathbb{P})$ reste abstrait, il nous permet de poser un cadre théorique sans que l'on ait besoin de l'expliciter.

Définition 3.1.2. *Un événement $A \subset \Omega$ est un élément de la tribu $\mathcal{F}$ (i.e., un ensemble mesurable).*

Exemple. On lance deux dés. L'ensemble fondamental est

$$\Omega = \{(1, 1), (1, 2), (1, 3), \dots, (2, 1), (2, 2), \dots, (6, 5), (6, 6)\}.$$

On peut considérer les événements suivants :

- la somme des points est 6 : $A = \{(1, 5), (5, 1), (2, 4), (4, 2), (3, 3)\}$,

- la somme des points est un multiple de 3 :

$$B = \{(1, 2), (2, 1), (1, 5), (5, 1), (2, 4), (4, 2), (3, 3), (3, 6), (6, 3), (4, 5), (5, 4), (6, 6)\}.$$

La mesure de probabilité $\mathbb{P}$ permet de quantifier les "chances" de réalisation des événements.

Exemple. 1. $([0, 1], \mathcal{B}([0, 1]), \lambda)$ où λ est la mesure de Lebesgue sur $[0, 1]$.

2. Mesure de Dirac sur $(\Omega, \mathcal{P}(\Omega))$: Soit $a \in \Omega$, pour tout $A \in \mathcal{F}$, on définit la mesure de Dirac au point a par

$$\delta_a(A) = \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{si } a \notin A \end{cases}$$

3. Pour définir une probabilité sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, il suffit de se donner une suite de réels positifs p_n tel que $\sum_{n \in \mathbb{N}} p_n = 1$. Alors $\mathbb{P} = \sum_{n \in \mathbb{N}} p_n \delta_n$, où δ_n est la masse de Dirac en n , est une mesure de probabilité sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. On a alors $\mathbb{P}(A) = \sum_{n \in \mathbb{N}} p_n \mathbb{1}_A(n)$.

Propriété 3.1.3. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et A, B et $(A_n)_{n \in \mathbb{N}}$ des événements, alors*

1. Si $A \cap B = \emptyset$, alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ (additivité)
2. De manière générale, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
3. Si $B \subset A$, alors $\mathbb{P}(B) \leq \mathbb{P}(A)$ (croissance) et $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(B)$.
4. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
5. $\mathbb{P}(\cup_{n \in \mathbb{N}} A_n) \leq \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$
6. si $(A_n)_{n \in \mathbb{N}}$ suite croissante (i.e., $A_n \subset A_{n+1}$), alors

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \sup_{n \in \mathbb{N}} \mathbb{P}(A_n)$$

7. si $(A_n)_{n \in \mathbb{N}}$ suite décroissante (i.e., $A_{n+1} \subset A_n$), alors

$$\mathbb{P}\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \inf_{n \in \mathbb{N}} \mathbb{P}(A_n)$$

3.1.b) Exercices sur les espaces de probabilités

Exercice 3.1.1. [Convergence monotone pour les probabilités] Soit l'espace mesurable $(\Omega, \mathcal{F})$ et $\mathbb{P} : \mathcal{F} \rightarrow [0, +\infty[$ une application additive (c'est-à-dire $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ lorsque $A, B \in \mathcal{F}$ et $A \cap B = \emptyset$), telle que $\mathbb{P}(\Omega) = 1$. Montrer que les quatre affirmations suivantes sont équivalentes :

1. $\mathbb{P}$ est une probabilité (c'est-à-dire elle est σ -additive).
2. $\mathbb{P}$ est continue sur des suites croissantes :

$$(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}, A_n \subset A_{n+1} \Rightarrow \mathbb{P}(\cup_{n \in \mathbb{N}} A_n) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n).$$

(On notera pour des telles suites croissantes $\lim_n A_n = \cup_{n \in \mathbb{N}} A_n$.)

3. $\mathbb{P}$ est continue sur des suites décroissantes :

$$(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}, A_n \supset A_{n+1} \Rightarrow \mathbb{P}(\cap_{n \in \mathbb{N}} A_n) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n).$$

(On notera pour des telles suites décroissantes $\lim_n A_n = \cap_{n \in \mathbb{N}} A_n$.)

4. $\mathbb{P}$ est continue sur des suites décroissantes vers $\emptyset$:

$$(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}, A_n \supset A_{n+1} \text{ et } \cap_{n \in \mathbb{N}} A_n = \emptyset \Rightarrow \lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = 0.$$

Exercice 3.1.2. [Limites inférieure et supérieure] Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. On considère une suite d'ensembles mesurables $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$ et on note

$$\liminf_n A_n = \bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} A_m, \quad \limsup_n A_n = \bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} A_m.$$

1. Montrer que $\omega \in \liminf_n A_n$ ssi à partir d'un certain rang, ω est dans tous les A_n .
2. Montrer que $\omega \in \limsup_n A_n$ ssi ω est dans une infinité de A_n .
3. Montrer que $\mathbb{P}(\liminf_n A_n) \leq \liminf_{n \rightarrow +\infty} \mathbb{P}(A_n) \leq \limsup_{n \rightarrow +\infty} \mathbb{P}(A_n) \leq \mathbb{P}(\limsup_n A_n)$.
4. On dit que la suite $(A_n)_{n \in \mathbb{N}}$ est convergente si $\liminf_n A_n = \limsup_n A_n$. Montrer que si la suite est croissante (respectivement décroissante) alors elle est convergente et

$$\lim_n A_n = \bigcup_{n \in \mathbb{N}} A_n \quad (\text{respectivement } \lim_n A_n = \bigcap_{n \in \mathbb{N}} A_n).$$

5. Montrer que si la suite $(A_n)_{n \in \mathbb{N}}$ est convergente, on a la propriété de continuité de la mesure : $\mathbb{P}(\lim_n A_n) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n)$.

Exercice 3.1.3. [Presque sûr] On dit qu'un événement $A \in \mathcal{F}$ est presque sûr si A est presque sûrement égal à Ω , c'est-à-dire $\Omega = A \cup N$ avec N un ensemble négligeable (ie. $\exists B \in \mathcal{F} : N \subset B, \mathbb{P}(B) = 0$). Soit $(A_j)_{j \in J}$, $J \subset \mathbb{N}$, une famille d'événements presque sûrs. Montrer que $\bigcap_{j \in J} A_j$ est presque sûr.

Exercice 3.1.4. [Tribu complète] Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Soit la famille d'ensembles suivante :

$$\mathcal{F}_{\mathbb{P}} = \{C \in \mathcal{P}(\Omega) : \exists A_1, A_2 \in \mathcal{F}, \text{ tels que } A_1 \subset C \subset A_2 \text{ et } \mathbb{P}(A_2 \setminus A_1) = 0\}.$$

1. Montrer que $\mathcal{F}_{\mathbb{P}} = \sigma(\mathcal{F} \cup \mathcal{N})$ où $\mathcal{N}$ est la classe des ensembles $\mathbb{P}$ -négligeables, ie. $\mathcal{N} = \{N \subset \Omega : \exists A \in \mathcal{F} : N \subset A, \mathbb{P}(A) = 0\}$.
2. On définit $\bar{\mathbb{P}}$ sur $\mathcal{F}_{\mathbb{P}}$ par $\bar{\mathbb{P}}(C) = \mathbb{P}(A_1) = \mathbb{P}(A_2)$. Montrer que $\bar{\mathbb{P}}$ est bien définie (c'est-à-dire que sa valeur ne dépend pas du choix de A_1 et de A_2). Montrer que $\bar{\mathbb{P}}$ est la seule mesure sur $\mathcal{F}_{\mathbb{P}} = \sigma(\mathcal{F} \cup \mathcal{N})$ qui prolonge $\mathbb{P}$ (ie. qui coïncide avec $\mathbb{P}$ sur $\mathcal{F}$).
3. Montrer que pour toute fonction X réelle $\mathcal{F}_{\mathbb{P}}$ -mesurable, il existe deux variables aléatoires (ie. $\mathcal{F}$ -mesurables) U, V réelles telles que $U \leq X \leq V$ et $V - U = 0$ $\mathbb{P}$ -p.s.

3.2 Variables aléatoires

3.2.a) Définition

Imaginons une personne jouant une partie de dés en 10 coups : à chaque lancer, il gagne 2 euros s'il obtient un 5 ou un 6, un euro pour un quatre, rien pour un 3, perd un euros pour un 2 et trois euros s'il obtient 1. Si on cherche à modéliser l'expérience, on peut prendre $\Omega = \{1, \dots, 6\}^{10}$, $\mathcal{F} = \mathcal{P}(\Omega)$ et si le dé est bien équilibré chaque configuration a la même probabilité d'apparaître, il est donc naturel de choisir $\mathbb{P}(\{\omega\}) = 6^{-10}$ pour tout $\omega \in \Omega$. Pour le joueur, ce qui est intéressant, ce n'est pas la description de l'expérience, mais juste le gain final qu'il va avoir. Bien entendu le gain dépend du résultat de l'expérience. Le gain est une application sur Ω à valeurs dans $\{-30, \dots, 20\}$. On parle de variable aléatoire.

Soit $(\Omega, \mathcal{F})$ et $(E, \mathcal{A})$ deux espaces mesurables et X une application de Ω dans E . Pour $A \in \mathcal{A}$, on note $\{X \in A\}$ l'image réciproque de A par X :

$$\{X \in A\} := X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$$

Définition 3.2.1. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $(E, \mathcal{A})$ un espace mesurable.

Une variable aléatoire (v.a.) est une application X mesurable d'un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans un espace mesurable $(E, \mathcal{A})$, i.e. $X^{-1}(\mathcal{A}) = \{X^{-1}(A) : A \in \mathcal{A}\} \subset \mathcal{F}$.

Si $(E, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on parle alors de v.a. réelle. Si $(E, \mathcal{A}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, on parle de vecteur aléatoire. On peut alors écrire $X = (X_1, \dots, X_d)$ avec $X_i : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ variable aléatoire appelée $i^{\text{ème}}$ marginale de X . Si $(E, \mathcal{A}) = (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ alors X est dite v.a. positive et si $(E, \mathcal{A}) = (\mathbb{N}, \mathcal{P}(\mathbb{N}))$, X est dite v.a. entière.

Propriété 3.2.2. Si $\mathcal{A} = \sigma(\mathcal{C})$, X est une v.a. si et seulement si $X^{-1}(\mathcal{C}) \subset \mathcal{F}$.

Par conséquent si $(E, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, X est une v.a. si et seulement si $\{X \leq x\} \in \mathcal{F}$ pour tout $x \in \mathbb{R}$.

Exemple. Une fonction étagée est une fonction mesurable qui ne prend qu'un nombre fini de valeurs. Elle s'écrit sous la forme $X = \sum_{i=1}^n x_i \mathbb{1}_{A_i}$ avec $A_i \in \mathcal{F}$ et $x_1, \dots, x_n \in \mathbb{R}$ (il n'y a pas unicité de la représentation).

Rappel : toute v.a. est limite simple d'une suite de fonctions étagées. Et si de plus la v.a. est positive alors la suite peut-être choisie croissante. La mesurabilité est stable par composition. La somme, différence, quotient de v.a. réelles est encore une v.a. Le max et le min de v.a. réelles est encore une v.a., la partie imaginaire et la partie réelle d'un v.a complexe est une v.a. Si (X_n) est une suite de v.a., alors l'inf, le sup, les lim inf et lim sup sont des v.a.

3.2.b) Loi d'une variable aléatoire

Définition 3.2.3. Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{A})$ une v.a.. On appelle la loi de X la mesure image $\mathbb{P}_X$ de $\mathbb{P}$ sur $(E, \mathcal{A})$ par X : pour $A \in \mathcal{A}$

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$$

Dans le cas d'un vecteur aléatoire, $X = (X_1, \dots, X_n)$, la loi $\mathbb{P}_{(X_1, \dots, X_n)}$ est appelée la loi jointe des variables aléatoires $X_1, \dots, X_n$. Les lois des coordonnées de la variable s'appellent les lois marginales.

Propriété 3.2.4. La loi image $\mathbb{P}_X$ est une mesure de probabilité sur $(E, \mathcal{A})$.

Preuve. Voir Exercice 2 du devoir 2. □

Par abus de notation, si $A = \{x\}$, on notera $\mathbb{P}(X = x)$ au lieu de $\mathbb{P}(X \in \{x\})$.

Exemple. On lance deux dés et on regarde X le minimum des deux chiffres obtenu. On a $\Omega = \{1, 2, 3, 4, 5, 6\}^2$, $\mathcal{F} = \mathbb{P}(\Omega)$ et $X : \Omega \rightarrow \{1, 2, 3, 4, 5, 6\}$, $(\omega_1, \omega_2) \mapsto \min(\omega_1, \omega_2)$ est bien une v.a. Les dés n'étant pas truqués il est naturel de considérer la loi produit $\mathbb{P} = \nu \otimes \nu$ avec $\nu = \frac{1}{6} \sum_{k=1}^6 \delta_k$. Par conséquent, $\mathbb{P}(\{\omega_1, \omega_2\}) = 1/36$. Donc

k	1	2	3	4	5	6
$\mathbb{P}(X = k)$	11/36	9/36	7/36	5/36	3/36	1/36

On vérifie rapidement qu'il s'agit d'une mesure de probabilité, car la somme vaut bien 1.

Définition 3.2.5. On dit que deux variables X et Y à valeurs dans le même espace $(E, \mathcal{A})$ sont de même loi si $\mathbb{P}_X = \mathbb{P}_Y$. On notera alors $\mathcal{L}(X) = \mathcal{L}(Y)$ ou $X \stackrel{\mathcal{L}}{=} Y$.

On remarque que deux variables peuvent avoir la même loi sans pour autant être définie sur le même espace de probabilité (on identifie seulement les mesures images).

Définition 3.2.6. Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{A})$ une variable aléatoire. On appelle atome de (la loi de) X tout point $x \in E$ tel que $\mathbb{P}(X = x) \neq 0$.

On appellera support d'une variable aléatoire $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{A})$ « le plus petit » événement S de E tel que $\mathbb{P}(X \in S) = 1$ (cet objet est mal défini dans le cas général, il faut juste en retenir le sens intuitif).

3.2.c) Classe monotone

Définition 3.2.7. Une famille $\mathcal{M}$ de parties de E est appelée **classe monotone** (ou λ -système) si

1. $E \in \mathcal{M}$,
2. $\mathcal{M}$ est stable par différence : lorsque $A, B \in \mathcal{M}$ et $B \subset A$ alors $A \setminus B \in \mathcal{M}$,
3. $\mathcal{M}$ est stable par union croissante (si $A_n \in \mathcal{M}$ et $A_n \subset A_{n+1}$ alors $\cup_{n \geq 0} A_n \in \mathcal{M}$).

Remarque. — Une intersection quelconque de classes monotones est encore une classe monotone.

— Une tribu est une classe monotone, car $A \setminus B = A \cap B^c$.

Propriété 3.2.8. Une classe monotone stable par intersection finie est une tribu.

Preuve. À faire en exercice. □

Définition 3.2.9. Pour $\mathcal{E} \subset \mathcal{P}(E)$, on appelle classe monotone engendrée par $\mathcal{E}$ la plus petite classe monotone contenant $\mathcal{E}$, i.e. l'intersection de toutes les classes monotones contenant $\mathcal{E}$. On la note $\mathcal{M}(\mathcal{E})$.

Théorème 3.2.10 (Théorème des classes monotones, Théorème de Dynkin). Soit $\mathcal{E}$ une famille de parties de E stable par intersection finie (appelé π -système). Alors $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E})$.

Preuve. Il suffit de montrer que $\mathcal{M}(\mathcal{E})$ est stable par intersection finie, ce qui impliquera que $\mathcal{M}(\mathcal{E})$ est une tribu qui contient $\mathcal{E}$, car par définition $\mathcal{M}(\mathcal{E}) \subset \sigma(\mathcal{E})$.

Soit $A \in \mathcal{E}$, on note $\mathcal{M}_A = \{B \in \mathcal{M}(\mathcal{E}) : A \cap B \in \mathcal{M}(\mathcal{E})\}$. On remarque que $\mathcal{E} \subset \mathcal{M}_A$ et $\mathcal{M}_A$ est une classe monotone. Par conséquent, $\mathcal{M}(\mathcal{E}) \subset \mathcal{M}_A$ pour tout $A \in \mathcal{E}$.

On a donc

$$\forall A \in \mathcal{E}, \forall B \in \mathcal{M}(\mathcal{E}), \quad A \cap B \in \mathcal{M}(\mathcal{E}).$$

Soit $A \in \mathcal{M}(\mathcal{E})$ et $\mathcal{M}_A = \{B \in \mathcal{M}(\mathcal{E}) : A \cap B \in \mathcal{M}(E)\}$. De même, $\mathcal{M}_A$ est une classe monotone qui contient $\mathcal{E}$, donc $\mathcal{M}(\mathcal{E}) \subset \mathcal{M}_A$ pour tout $A \in \mathcal{M}(\mathcal{E})$.

On déduit que $\mathcal{M}(\mathcal{E})$ est stable par intersection finie et le théorème se déduit de la Propriété 3.2.8. $\square$

Théorème 3.2.11. *Soit $(E, \mathcal{A})$ un espace mesurable. Si deux probabilités $\mathbb{P}_1$ et $\mathbb{P}_2$ coïncident sur $\mathcal{E}$ stable par intersections finies, alors elles coïncident sur $\sigma(\mathcal{E})$.*

Preuve. Soit $\mathcal{M} = \{A \in \mathcal{A} : \mathbb{P}_1(A) = \mathbb{P}_2(A)\}$. Alors $E \in \mathcal{M}$, si $A, B \in \mathcal{M}$ avec $B \subset A$ alors $A \setminus B \in \mathcal{M}$ (car $\mathbb{P}_1(A \setminus B) = \mathbb{P}_1(A) - \mathbb{P}_1(B) = \mathbb{P}_2(A \setminus B)$) et si $(A_n)_{n \geq 0}$ suite croissante de $\mathcal{M}$, $\bigcup A_n \in \mathcal{M}$ (car $\mathbb{P}_1(\bigcup A_n) = \lim_{n \rightarrow \infty} \mathbb{P}_1(A_n) = \mathbb{P}_2(\bigcup A_n)$). Par conséquent, $\mathcal{M}$ est une classe monotone, avec $\mathcal{E} \subset \mathcal{M}$ et donc $\mathcal{M}(\mathcal{E}) \subset \mathcal{M}$. Comme $\mathcal{E}$ est un π -système, par le théorème de la classe monotone $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E})$. $\square$

Conséquence. D'après le théorème de la classe monotone, si $\mathcal{A} = \sigma(\mathcal{C})$ où $\mathcal{C}$ est stable par intersection finie (π -système), alors X et Y sont de même loi si et seulement si $\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$ pour tout $A \in \mathcal{C}$.

De même, si μ mesure de probabilité sur $(E, \mathcal{A})$, dans ce cas, X suit la loi μ si et seulement si $\mathbb{P}(X \in A) = \mu(A)$ pour tout $A \in \mathcal{C}$.



Dans la suite, on aura tendance à introduire des v.a. en omettant de définir l'espace $(\Omega, \mathcal{F}, \mathbb{P})$ sur lequel on se place. Cependant, il ne faut jamais oublier qu'une v.a. est une fonction.

3.2.d) Fonction de répartition

On considère des variables à valeurs réelles : $(E, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Définition 3.2.12. *On appelle fonction de répartition d'une variable aléatoire $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ la fonction F_X définie sur $\mathbb{R}$ à valeurs dans $[0, 1]$ par*

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}_X([-\infty, x]).$$

Proposition 3.2.13. *La fonction de répartition caractérise la loi d'un variable aléatoire, i.e. si X, Y deux v.a. telles que $F_X = F_Y$ alors $\mathbb{P}_X = \mathbb{P}_Y$.*

Preuve. La famille $\mathcal{C} = \{[-\infty, x], X \in \mathbb{R}\}$ est stable par intersection finie et engendre $\mathcal{B}(\mathbb{R})$, on conclut par le théorème de la classe monotone. $\square$

Propriété 3.2.14. *Soit F_X la fonction de répartition d'une variable aléatoire réelle X . Alors,*

1. F_X est une fonction croissante,
2. $\lim_{x \rightarrow -\infty} F_X(x) = 0$ et $\lim_{x \rightarrow +\infty} F_X(x) = 1$,
3. F_X est une fonction continue à droite et admet une limite à gauche,

$$F_X(x^-) = \lim_{\substack{y \rightarrow x \\ y < x}} F_X(y) = \mathbb{P}(X < x),$$

4. $\mathbb{P}(X \in [a, b]) = F_X(b) - F_X(a^-),$

5. si x n'est pas un atome de X alors F_X est continue à gauche en x (donc continue). L'ensemble des points de discontinuité de F_X est l'ensemble des atomes de X .

Preuve.

1. Si $x \leq y$, on a $\{X \leq x\} \subset \{X \leq y\}$ et donc par croissance de la mesure de probabilité $\mathbb{P}$, $F_X(x) \leq F_X(y)$.
La fonction F_X étant monotone, elle admet des limites à gauche et à droite en tout point et l'ensemble de ses points de discontinuité est au plus dénombrable.
2. F_X étant croissante, elle admet des limites en $-\infty$ et $+\infty$. On a donc $\lim_{x \rightarrow -\infty} F_X(x) = \lim_{n \rightarrow +\infty} F_X(-n)$. On pose $A_n =]-\infty, -n]$ qui est une suite décroissante de boréliens avec $\bigcap_{n \geq 1} A_n = \emptyset$. Par conséquent, $\lim_{n \rightarrow +\infty} F_X(-n) = \mathbb{P}_X(\emptyset) = 0$. De même, $\lim_{x \rightarrow +\infty} F_X(x) = \lim_{n \rightarrow +\infty} F_X(n)$. On pose $A_n =]-\infty, n]$ qui est une suite croissante de boréliens avec $\bigcup_{n \geq 1} A_n = \mathbb{R}$. Par conséquent, $\lim_{n \rightarrow +\infty} F_X(n) = \mathbb{P}_X(\mathbb{R}) = 1$.
3. On introduit la suite décroissante $A_n =]-\infty, x + 1/n]$. On a $\bigcap A_n =]-\infty, x]$ et donc $\lim_{y \rightarrow x^+} F_X(y) = \lim_{n \rightarrow \infty} F_X(x + 1/n) = \lim_{n \rightarrow \infty} \mathbb{P}_X(A_n) = \mathbb{P}_X(]-\infty, x]) = F_X(x)$. La fonction est continue à droite.
On introduit maintenant la suite croissante $A_n =]-\infty, x - 1/n]$. On a $\bigcup A_n =]-\infty, x[$ et donc $\lim_{y \rightarrow x^-} F_X(y) = \lim_{n \rightarrow \infty} \mathbb{P}_X(A_n) = \mathbb{P}_X(]-\infty, x[) = F_X(x^-)$. La fonction admet une limite à gauche qui vaut $\mathbb{P}(X < x)$.
4. $\mathbb{P}(X \in [a, b]) = \mathbb{P}(X \leq b) - \mathbb{P}(X < a) = F_X(b) - F_X(a^-)$ car $[a, b] =]-\infty, b] \setminus]-\infty, a[$.
5. En général, $\mathbb{P}(X < x) \neq \mathbb{P}(X \leq x)$. En fait, $\mathbb{P}(X \leq x) = \mathbb{P}(X < x) + \mathbb{P}(X = x)$ car ensembles disjoints et donc la fonction est continue en x si et seulement si x n'est pas un atome.

□

Proposition 3.2.15. *La fonction de répartition admet au plus un nombre dénombrable de points de discontinuité. Une variable aléatoire réelle a par conséquent un nombre au plus dénombrable d'atomes.*

Preuve. Ceci est lié à la monotonie de la fonction. On peut aussi le redémontrer dans le cas particulier des fonctions de répartition. Si on note $D_n = \{x : F_X(x) - F_X(x^-) \geq 1/n\}$ l'ensemble des points de discontinuité avec un saut d'amplitude supérieur à $1/n$, comme $F_X(x) \in [0, 1]$, on a $\text{card } D_n \leq n$. Donc l'ensemble des points de discontinuité $D = \bigcup_{n \geq 1} D_n$ est dénombrable. □

Théorème 3.2.16. *Une fonction $F : \mathbb{R} \rightarrow [0, 1]$ croissante càdlàg (continue à droite et limitée à gauche) avec $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$ est la fonction de répartition d'une v.a. X . De plus, l'ensemble des points où la fonction a un saut est l'ensemble des atomes de X .*

Preuve. La preuve est complexe si on souhaite construire X sur un espace de probabilité général $(\Omega, \mathcal{F}, \mathbb{P})$ et repose sur le théorème de Carathéodory.

On se place dans le cas particulier $(]0, 1[, \mathcal{B}(]0, 1[), \mathbb{P})$ où $\mathbb{P} = \lambda$ mesure de Lebesgue sur $]0, 1[$. Soit $\omega \in]0, 1[$. On note $X(\omega) = \inf \{t \in \mathbb{R} : F(t) \geq \omega\}$.

X est bien défini car F est croissante et $\overline{F(\mathbb{R})} = [0, 1]$. Si F est strictement croissante, alors elle est inversible et $X(\omega) = F^{-1}(\omega)$.

X est une variable aléatoire (X croissante donc mesurable, cf devoir 1).

Montrons que $\{t \in \mathbb{R} : F(t) \geq \omega\} = [X(\omega), +\infty[$. On a de façon évidente $\{t \in \mathbb{R} : F(t) \geq \omega\} \subset [X(\omega), +\infty[$. Par ailleurs, soit $s \geq X(\omega)$, alors par croissance $F(s) \geq F(X(\omega))$ et par continuité à droite de F on a

$F(X(\omega)) \geq \omega$, donc $F(s) \geq \omega$ et $s \in \{t \in \mathbb{R} : F(t) \geq \omega\}$.

Par ailleurs, vérifions que la fonction de répartition de X est bien F . On a, en effet,

$$\begin{aligned}\mathbb{P}(X \leq t) &= \mathbb{P}(\{\omega \in]0, 1[: X(\omega) \leq t\}) \\ &= \mathbb{P}(\{\omega \in]0, 1[: t \in [X(\omega), +\infty[)\}) \\ &= \mathbb{P}(\{\omega \in]0, 1[: F(t) \geq \omega\}) \\ &= \int_0^{F(t)} d\lambda(\omega) = F(t).\end{aligned}$$

□

3.2.e) Exercices sur les variables aléatoires

Exercice 3.2.1. [Tribu discrète] Soit $\Omega = \{1, 2, 3, 4, 5\}$. Déterminer la tribu $\sigma(\{A, B, C\})$ où $A = \{1, 2\}$, $B = \{2, 3\}$, $C = \emptyset$.

1. Quelles sont les variables aléatoires réelles définies sur Ω ?
2. Que se passe-t-il si on retire C ?
3. Que rajouter à $\{A, B, C\}$ pour engendrer $\mathcal{P}(\Omega)$?

Exercice 3.2.2. Soit X une variable aléatoire à valeurs dans $(E, \mathcal{E})$. Montrer que la loi image $\mathbb{P}_X$ est une mesure de probabilité sur $(E, \mathcal{E})$.

Exercice 3.2.3. [Variable aléatoire constante] Montrer qu'une application $\Omega \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est une variable aléatoire par rapport à la tribu triviale sur Ω si et seulement si elle est constante.

Exercice 3.2.4. [Approximation] Soit X une variable aléatoire positive. On considère

$$A_{n,k} = \left\{ \omega : \frac{k-1}{2^n} \leq X(\omega) < \frac{k}{2^n} \right\}, \quad B_n = \{\omega : X(\omega) > n\}.$$

1. Montrer que $X_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbb{1}_{A_{n,k}} + n \mathbb{1}_{B_n}$ est une variable aléatoire.
2. Montrer que $\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)$ pour tout $\omega \in \Omega$.
3. Montrer que $(X_n(\omega))_{n \geq 1}$ est une suite croissante.

Exercice 3.2.5. [Copies ordonnées]

1. Soient X, Y deux variables aléatoires réelles sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ telles que

$$\mathbb{P}(Y \leq t < X) = 0 \quad \forall t \in \mathbb{R}.$$

Montrer que $\mathbb{P}(Y < X) = 0$.

2. On suppose cette fois que X et Y ont même loi. Montrer que si $X \geq Y$ p.s., alors X et Y sont presque sûrement égales : c'est à dire

$$(\exists A \in \mathcal{F} \text{ avec } \mathbb{P}(A) = 1 : \forall \omega \in A, X(\omega) \geq Y(\omega)) \implies (\exists B \in \mathcal{F} \text{ avec } \mathbb{P}(B) = 1 : \forall \omega \in B, X(\omega) = Y(\omega)).$$

3.3 Exemples de variables aléatoires

3.3.a) Variables discrètes

Définition 3.3.1. Une v.a. $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est dite *discrète* si son support est au plus dénombrable.

Dans ce cas, son support est exactement l'ensemble $\mathfrak{A}$ de ses atomes.

Rappel : La mesure de dirac δ_a en $a \in \mathbb{R}$ est définie pour tout borélien $A \in \mathcal{B}(\mathbb{R})$

$$\delta_a(A) = \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{si } a \notin A. \end{cases}$$

Propriété 3.3.2. Si X est une variable aléatoire discrète alors sa loi est la somme pondérée des mesures de Dirac en ses atomes

$$\mathbb{P}_X = \sum_{x \in \mathcal{A}} p_x \delta_x,$$

avec $p_x = \mathbb{P}(X = x)$.

Preuve. Comme $\mathcal{A}$ est au plus dénombrable et $\mathcal{A} = \bigcup_{x \in \mathcal{A}} \{x\}$, pour tout $A \in \mathcal{B}(\mathbb{R})$, on a

$$\mathbb{P}_X(A) = \mathbb{P}_X\left(A \cap \bigcup_{x \in \mathcal{A}} \{x\}\right) = \sum_{x \in \mathcal{A}} \mathbb{P}_X(A \cap \{x\}) = \sum_{x \in \mathcal{A}} \mathbb{P}(X = x, x \in A) = \sum_{x \in \mathcal{A}} \mathbb{P}(X = x) \delta_x(A).$$

□

Remarque. Pour caractériser la loi d'une v.a. discrète, il suffit de donner l'ensemble de ses atomes et d'exprimer $\mathbb{P}(X = x)$ pour tout atome x .

Proposition 3.3.3. Soit X une v.a. discrète et $\mathcal{A} = \{x_i : i \in I\}$ l'ensemble de ses atomes (supposés ordonnés : $x_i < x_{i+1}$). La fonction de répartition F_X de X est croissante, constante sur chaque intervalle $[x_k, x_{k+1}[$ avec un saut $p_k = \mathbb{P}(X = x_k)$ en chaque atome x_k : $p_k = F_X(x_k) - F_X(x_{k-1})$

Preuve. Si $x < \min(x_i)$, alors $F_X(x) = 0$ et si $x > \sup(x_i)$, $F_X(x) = 1$ car en dehors du support. Soient $s, t \in [x_k, x_{k+1}[$ avec $s < t$, $F_X(t) - F_X(s) = \mathbb{P}(X \in]s, t]) = \sum_{i: s < x_i \leq t} \mathbb{P}(X = x_i) = 0$ car la somme est vide. La fonction est bien constante par intervalle. □

Exemple. • Une variable aléatoire constante est discrète, $X = c$ a pour loi $\mathbb{P}_X = \delta_c$.

- X une variable aléatoire qui prend les valeurs $x_1, \dots, x_n$ avec probabilité respectives $p_1, \dots, p_n$ a pour loi $\mathbb{P}_X = p_1 \delta_{x_1} + \dots, p_n \delta_{x_n}$.
- Lorsque tous les atomes ont la même probabilité de se réaliser on parle de loi équirépartie (équiprobable ou encore uniforme). Si X est équirépartie sur $\{x_1, \dots, x_n\}$ alors $\mathbb{P}_X = \frac{1}{n} \sum_{k=1}^n \delta_{x_k}$.
- Si $A \in \mathcal{F}$ et $X = \mathbb{1}_A$ alors X est discrète à valeurs dans $\{0, 1\}$ et $\mathbb{P}_X = \mathbb{P}(A) \delta_1 + (1 - \mathbb{P}(A)) \delta_0$. Cette loi est appelée loi de Bernoulli $\mathcal{B}(p)$ avec $p = \mathbb{P}(A)$.

De manière générale, une variable qui ne prend que deux valeurs suit une loi de Bernoulli.

- On appelle loi Binomiale $\mathcal{B}(n, p)$ avec $n \in \mathbb{N}^*$ et $p \in]0, 1[$ la loi d'une v.a. X à valeurs dans $\{0, \dots, n\}$ avec

$$\mathbb{P}_X = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \delta_k.$$

On trouve cette loi quand on considère le nombre de succès dans une suite de n épreuves identiques ayant chacune une probabilité de succès p .

- On appelle loi Géométrique $\mathcal{G}(p)$ avec $p \in]0, 1[$ la loi d'une v.a. à valeurs dans $\mathbb{N}^*$ avec

$$\mathbb{P}_X = \sum_{k=1}^{\infty} p(1-p)^{k-1} \delta_k.$$

On trouve cette loi quand on considère la position du premier succès dans une suite d'épreuves aléatoires identiques de probabilité de succès p .

- On appelle loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda > 0$ la loi d'une v.a. à valeurs dans $\mathbb{N}$ avec

$$\mathbb{P}_X = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \delta_k.$$

Cette loi est souvent utilisée pour modéliser des événements rares.

Remarque. Soit X une v.a. discrète à valeurs dans $\{x_i : i \in I\}$ et on note $p_i = \mathbb{P}(X = x_i)$. Soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une application mesurable. Alors $Y = g(X)$ est une v.a. discrète à valeurs dans $\{y_j : j \in J\} = g(\{x_i : i \in I\})$ et

$$\mathbb{P}(Y = y_j) = \sum_{i \in I : g(x_i) = y_j} p_i.$$

Preuve. Il est clair que Y est à valeurs dans $g(\{x_i : i \in I\})$. Par ailleurs, on a $g^{-1}(\{y_j\}) = \{x_i : g(x_i) = y_j\}$ et donc

$$\mathbb{P}(Y = y_j) = \mathbb{P}(g \circ X = y_j) = \mathbb{P}(X^{-1}(g^{-1}(\{y_j\}))) = \mathbb{P}_X(g^{-1}(\{y_j\})) = \sum_{x_i \in g^{-1}(\{y_j\})} \mathbb{P}(X = x_i).$$

□

3.3.b) Théorème de Radon-Nikodym

Définition 3.3.4. Soit μ et ν deux mesures définies sur un espace mesurable $(\Omega, \mathcal{F})$. On dit que μ est absolument continue par rapport à ν , noté $\mu \ll \nu$, si pour tout $A \in \mathcal{F}$,

$$\nu(A) = 0 \text{ implique } \mu(A) = 0.$$

Théorème 3.3.5 (Radon-Nikodym). Soient μ et ν deux mesures σ -finies. Si $\mu \ll \nu$ alors il existe une fonction $f : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ mesurable telle que

$$\forall A \in \mathcal{F}, \quad \mu(A) = \int_A f d\nu.$$

La fonction f s'appelle la densité de μ par rapport à ν . Elle est souvent notée $f = \frac{d\mu}{d\nu}$. De plus, on a le lien suivant pour tout fonction g telle que les intégrales soient bien définies

$$\int g d\mu = \int g f d\nu.$$

Si μ est une mesure finie alors $f \in L^1(\nu)$.

Voir Théorèmes 10.4 et 10.5 dans Briane-Pagès [BP06] pour la preuve de ce théorème.

Exemple. 1. soit $\mu = \sum_{k \geq 0} p_k \delta_k$ avec p_k des réels positifs, alors $\mu \ll \nu$ avec $\nu = \sum_{k \geq 0} \delta_k$ la mesure de comptage sur $\mathbb{N}$. La densité est donnée par $f(k) = p_k$.

2. Soit λ la mesure de Lebesgue. Aucune loi discrète n'est absolument continue par rapport à la mesure de Lebesgue car une loi discrète possède des atomes : $\mathbb{P}(X = x) > 0$ et $\lambda(\{x\}) = 0$. Il suffit d'avoir un atome pour qu'une loi ne soit pas absolument continue par rapport à la mesure de Lebesgue.

3.3.c) Variables à densité

On appelle variables à densité les variables dont la loi est absolument continue par rapport à la mesure de Lebesgue. On note λ la mesure de Lebesgue sur $\mathbb{R}$.

Définition 3.3.6. Une variable aléatoire X est dite **à densité** f si $\mathbb{P}_X \ll \lambda$, i.e.

$$\forall A \in \mathcal{B}(\mathbb{R}) \quad \mathbb{P}(X \in A) = \int_A f d\lambda.$$

On a alors $\mathbb{P}(X \in [a, b]) = \int_a^b f(x) dx$ pour tout $a, b \in \mathbb{R}$. On note dans ce cas $\mathbb{P}_X = f(x) dx$.

Remarque. La fonction f est mesurable positive et $\int_{\mathbb{R}} f(x) dx = 1$. De façon réciproque, une telle fonction est une densité de probabilité.

Exemple. 1. On appelle loi uniforme sur $[a, b]$, notée $\mathcal{U}([a, b])$ la loi de densité $\frac{1}{b-a} \mathbb{1}_{[a, b]}$.

2. On appelle loi exponentielle de paramètre $\alpha > 0$, notée $\mathcal{E}(\alpha)$, la loi de densité $\alpha e^{-\alpha x} \mathbb{1}_{x > 0}$. Cette loi est souvent utilisée pour modéliser des durées de vie.

3. On appelle loi normale centrée réduite, notée $\mathcal{N}(0, 1)$, la loi de densité $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

4. On appelle loi normale de moyenne m et de variance σ^2 , notée $\mathcal{N}(m, \sigma^2)$, la loi de densité $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$.

5. On appelle loi de Cauchy $\mathcal{C}$ de paramètre $a > 0$ la loi de densité $\frac{a}{\pi(a^2 + x^2)}$.

Proposition 3.3.7 (Lien entre fonction de répartition et densité). Soit X une variable aléatoire à densité f . Alors sa fonction de répartition vérifie

1. $\forall x \in \mathbb{R}, F_X(x) = \int_{-\infty}^x f(t) dt$.
2. F_X est continue sur $\mathbb{R}$.
3. Si f est continue en x_0 , alors F_X est dérivable en x_0 de dérivée $F'_X(x_0) = f(x_0)$.
4. Si X v.a. ayant une fonction de répartition F de la forme $F(x) = \int_{-\infty}^x f(t) dt$ où f fonction mesurable positive, alors X admet nécessairement f pour densité.

Preuve.

1. $\forall x \in \mathbb{R}, F_X(x) = \mathbb{P}_X([-\infty, x]) = \int_{-\infty}^x f(t) dt$.
2. Ok, car pas d'atome.
3. Si f est continue en x_0 , donc définie sur un voisinage en x_0 . Soit $\varepsilon > 0$, il existe $\delta > 0$ tel que $|t - x_0| < \delta \Rightarrow |f(t) - f(x_0)| < \varepsilon$. Soit h tel que $|h| < \delta$. On a alors

$$\begin{aligned} |F_X(x_0 + h) - F_X(x_0) - hf(x_0)| &= \left| \int_{x_0}^{x_0+h} [f(t) - f(x_0)] dt \right| \\ &\leq \varepsilon |h| \end{aligned}$$

Par conséquent, le taux d'accroissement $(F_X(x_0 + h) - F_X(x_0))/h$ converge vers $f(x_0)$ quand $|h| \rightarrow 0$.

4. F étant une fonction de répartition, f est une densité car $\int_{-\infty}^{+\infty} f(t)dt = \lim_{x \rightarrow \infty} F(x) = 1$. On introduit μ la mesure de probabilité définie par $\mu(A) = \int_A f(t)dt$ pour $A \in \mathcal{B}(\mathbb{R})$. Cette mesure est absolument continue par rapport à Lebesgue de densité f . $\mathbb{P}_X$ et μ coïncident sur $\mathcal{C} = \{] - \infty, x], x \in \mathbb{R} \}$ qui engendre $\mathcal{B}(\mathbb{R})$ et stable par intersection finie, donc par le théorème de classe monotone $\mathbb{P}_X = \mu$ et $\mathbb{P}_X$ admet f pour densité.

□



Avoir une fonction de répartition continue n'implique pas que la v.a. soit à densité. Les lois à densité sont en fait absolument continues : on dit que la fonction F est absolument continue sur $[a, b]$ si, pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que, pour toute suite $([a_n, b_n])_{n \in \mathbb{N}}$ de sous-intervalles de $[a, b]$ d'intérieurs disjoints,

$$\sum_{n \geq 0} (b_n - a_n) < \delta \Rightarrow \sum_{n \geq 0} |F(a_n) - F(b_n)| < \varepsilon.$$

En fait, F est absolument continue sur $[a, b]$ si et seulement s'il existe une fonction f intégrable sur $[a, b]$ telle que pour tout $x \in [a, b]$,

$$F(x) - F(a) = \int_a^x f(t)dt.$$

Exemple. Une fonction continue, dérivable presque partout et de dérivée dans L^1 n'est pas forcément l'intégrale de sa dérivée. En effet, l'escalier de Cantor est une fonction $f : [0, 1] \rightarrow [0, 1]$ continue dérivable presque partout (croissante) et de dérivée nulle presque partout.

On note K_3 l'ensemble triadique de Cantor. K_3 est un sous-ensemble compact de $[0, 1]$ d'intérieur vide donc totalement discontinu. Il est non dénombrable mais négligeable pour la mesure de Lebesgue. $K_3 = \bigcap_{n \in \mathbb{N}} K_n$ avec $K_0 = [0, 1]$, $K_1 = K_0 \setminus]1/3, 2/3[$, ..., K_n est la réunion de 2^n ensemble disjoints de longueur $1/3^n$: $K_n = \bigcup I_{q_n}$ avec $I_{q_n} = [q_n/3^n, (q_n+1)/3^n]$ et $q_n = \sum_{i=1}^n a_i 3^{n-i}$ avec $a_i \in \{0, 2\}$. Par conséquent,

$$K_3 = \left\{ \sum_{n=1}^{\infty} \frac{x_n}{3^n} : x_n \in \{0, 2\} \right\}.$$

On considère la suite de fonctions affines par morceaux définies sur $[0, 1]$ de la façon suivante :

$$\begin{aligned} f_0(x) &= x \\ f_1(0) &= 0, f_1(1) = 1, f_1([1/3, 2/3]) = 1/2 \end{aligned}$$

On construit la fonction f_{n+1} à partir de f_n : sur chaque intervalle $[u, v]$ où f_n est non constante, on remplace f_n par la fonction linéaire par morceaux qui vaut $\frac{f_n(u)+f_n(v)}{2}$ sur $[(2u+v)/3, (2v+u)/3]$. On a par conséquent

$$|f_{n+1}(x) - f_n(x)| \leq 2^{-n}$$

et donc la série $\sum_{n \geq 1} (f_{n+1} - f_n)(x)$ converge uniformément, i.e. (f_n) converge uniformément. Sa limite f est par conséquent continue, croissante, avec $f(0) = 0$ et $f(1) = 1$. De plus $f' = 0$ sur K_3^c car celui-ci est une réunion d'intervalles sur lesquelles f est constante.

3.3.d) Lois marginales d'un vecteur aléatoire

À partir de la loi d'un vecteur aléatoire, on obtient facilement les lois des marginales.

Propriété 3.3.8. Si (X, Y) est un couple de variables à valeurs dans E^2 de loi $\mathbb{P}_{(X,Y)}$, alors les lois marginales $\mathbb{P}_X$ et $\mathbb{P}_Y$ de X et Y sont données par : pour $A \in \mathcal{A}$

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

Plus généralement, si on considère un n -uplet,

$$\mathbb{P}_{X_i}(A) = \mathbb{P}_{(X_1, \dots, X_n)}(E^{i-1} \times A \times E^{n-i}).$$

Preuve. La preuve est évidente car $\{X \in A\} = \{(X, Y) \in A \times E\}$. □

La loi d'un vecteur aléatoire détermine la loi de chacune des marginales, mais la réciproque est fausse. De la loi des marginales, on ne peut pas en déduire (en général) la loi du vecteur.

Exemple. Soit (X, Y) un couple aléatoire de loi

$$\mathbb{P}_{(X,Y)} = \frac{1}{6}\delta_{(0,0)} + \frac{1}{3}\delta_{(0,1)} + \frac{1}{12}\delta_{(1,0)} + \frac{5}{12}\delta_{(1,1)}$$

et (U, V) un autre couple de loi

$$\mathbb{P}_{(U,V)} = \frac{1}{4}\delta_{(0,0)} + \frac{1}{4}\delta_{(0,1)} + \frac{1}{2}\delta_{(1,1)}$$

On a $\mathbb{P}_{(X,Y)} \neq \mathbb{P}_{(U,V)}$ et pourtant

$$\begin{aligned} \mathbb{P}_X &= \mathbb{P}_U = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1 \\ \mathbb{P}_Y &= \mathbb{P}_V = \frac{1}{4}\delta_0 + \frac{3}{4}\delta_1. \end{aligned}$$

Définition 3.3.9. Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire de $\mathbb{R}^d$. On appelle fonction de répartition de X la fonction définie sur $\mathbb{R}^d$ par

$$F_X(t) = F_X(t_1, \dots, t_d) = \mathbb{P}([-\infty, t_1] \times \dots \times [-\infty, t_d]) = \mathbb{P}(X_1 \leq t_1, \dots, X_d \leq t_d).$$

On remarque que la fonction de répartition de la $i^{\text{ème}}$ marginale vérifie $F_{X_i}(t_i) = \lim_{\substack{t_j \rightarrow +\infty \\ j \neq i}} F_X(t)$.

La fonction de répartition caractérise la loi d'un vecteur aléatoire (même preuve qu'en dimension $d = 1$).

Définition 3.3.10 (Densité en dimension d). Une fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ est une densité de probabilité (en dimension d) si

- f est positive : $\forall t \in \mathbb{R}^d, f(t) \geq 0$,
- f est intégrable d'intégrale 1 : $\int_{\mathbb{R}^d} f(t) dt = 1$.

Un vecteur aléatoire $X = (X_1, \dots, X_d)$ suit la loi de densité f si pour tous les pavés $\Pi_{i=1}^d A_i$ de $\mathbb{R}^d$,

$$\mathbb{P}\left((X_1, \dots, X_d) \in \Pi_{i=1}^d A_i\right) = \int_{\Pi_{i=1}^d A_i} f(x) dx = \int_{A_1} \dots \int_{A_d} f(x_1, \dots, x_d) dx_d \dots dx_1.$$

Proposition 3.3.11 (Théorème de changement de variables). Soit X un vecteur aléatoire de $\mathbb{R}^d$ de densité f et $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$ un difféomorphisme (bijection continument différentiable, ainsi que son inverse). Alors $Y = g(X)$ est un vecteur aléatoire de densité

$$f_Y(y) = f_X(g^{-1}(y)) |Jg^{-1}(y)|$$

où $Jg = \det((\partial_j g_i)_{1 \leq i, j \leq d})$ est le jacobien de g .

Preuve. $\mathbb{P}(g(X) \in B) = \mathbb{P}(X \in g^{-1}(B)) = \int_{\mathbb{R}^d} \mathbb{1}_{x \in g^{-1}(B)} f_X(x) dx = \int_{\mathbb{R}^d} \mathbb{1}_{g(x) \in B} f_X(x) dx$. Puis appliquer le théorème de changement de variables. $\square$

Proposition 3.3.12 (Densités marginales). *Si (X, Y) est un couple de densité f alors X et Y sont des v.a. de à densité*

$$f_X(x) = \int_{\mathbb{R}} f(x, y) dy, \quad f_Y(y) = \int_{\mathbb{R}} f(x, y) dx.$$

La réciproque est fausse, si X et Y sont à densité, on ne sait rien sur le couple. Par exemple le couple (X, X) avec X variable à densité f_X n'est pas à densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$.

Preuve. Étudions la loi de X , la preuve étant identique pour Y . Soit $A \in \mathcal{B}(\mathbb{R})$. Par Fubini-Tonelli, on a

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A, Y \in \mathbb{R}) = \iint_{A \times \mathbb{R}} f(x, y) dx dy = \int_A \left[\int_{\mathbb{R}} f(x, y) dy \right] dx.$$

Par ailleurs, $\int_{\mathbb{R}} f(x, y) dy$ est bien une densité de probabilité car positive et d'intégrale 1. $\square$

Exemple. — Vérifier que $f(x, y) = \frac{1}{3} \mathbb{1}_{[0,1] \times [-1,2]}(x, y)$ est une densité de probabilité sur $\mathbb{R}^2$. Montrer que $f_X(x) = \mathbb{1}_{[0,1]}(x)$ et $f_Y(y) = \frac{1}{3} \mathbb{1}_{[-1,2]}(y)$.

— Les densités marginales de $f(x, y) = ye^{-xy} \mathbb{1}_{\mathbb{R}^+}(x) \mathbb{1}_{[0,1]}(y)$ sont $f_X(x) = (1 - e^{-x} - xe^{-x})/x^2 \mathbb{1}_{\mathbb{R}^+}(x)$ et $Y \sim \mathcal{U}[0, 1]$.

— Vérifier que les lois marginales de $f(x, y) = \frac{1}{3\pi} e^{-\frac{x^2 + 2xy + 5y^2}{6}}$ sont des lois gaussiennes.

3.4 Espérance d'une variable aléatoire

Définition 3.4.1. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $X : \begin{array}{ccc} (\Omega, \mathcal{F}, \mathbb{P}) & \rightarrow & (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ \omega & \mapsto & X(\omega) \end{array}$ une variable aléatoire. X est dite intégrable si

$$\int_{\Omega} |X(\omega)| d\mathbb{P}(\omega) < \infty.$$

On définit l'espérance d'une v.a. positive ou intégrable X , notée $\mathbb{E}[X]$, comme l'intégrale de X par rapport à la mesure $\mathbb{P}$:

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

L'espérance représente la valeur moyenne prise par la variable.

On remarque que comme les mesures de probabilités sont de masse finie, les constantes sont intégrables et de même les v.a. bornées sont intégrables.

— Si $X = \mathbb{1}_A$ avec $A \in \mathcal{F}$ une v.a. de Bernoulli, alors $\mathbb{E}[X] = \mathbb{P}(A)$.

— Si $X = (X_1, \dots, X_d)$ vecteur aléatoire de $\mathbb{R}^d$ avec $\mathbb{E}[|X_i|] < \infty \forall i \in \{1, \dots, d\}$, alors on définit

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d]).$$

Définition 3.4.2. Une variable aléatoire intégrable est dite centrée si $\mathbb{E}[X] = 0$.

L'espérance étant l'intégrale de X par rapport à la mesure $\mathbb{P}$, l'espérance vérifie les propriétés classiques de l'intégrale. De même, on peut utiliser les théorèmes de Fubini-Tonelli et de Fubini.

Propriété 3.4.3. Soit X, Y deux v.a. intégrables et $a, b \in \mathbb{R}$.

1. L'espérance est croissante : si $X \leq Y$ p.s., alors $\mathbb{E}[X] \leq \mathbb{E}[Y]$,
2. l'espérance est linéaire : $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$,
3. $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$,

Pour le calcul effectif de l'espérance d'une v.a., on se ramène au calcul d'une intégrale sur $\mathbb{R}$ grâce au théorème de transfert. En effet, Ω est en général un espace abstrait et $\mathbb{P}$ une mesure non explicite, par contre $\mathbb{P}_X$ est la mesure image de $\mathbb{P}$ par l'application X .

Théorème 3.4.4 (Théorème de transfert). Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une variable aléatoire et $h : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une fonction borélienne.

Alors l'espérance de $h(X) = h \circ X$ est bien définie si et seulement si l'intégrale de h par rapport à $\mathbb{P}_X$ est bien définie (i.e, soit mesurable positive, soit intégrable). Et dans ce cas, on a

$$\begin{aligned} \int_{\Omega} h \circ X d\mathbb{P} &= \int_{\mathbb{R}} h d\mathbb{P}_X, \text{ c'est à dire} \\ \int_{\Omega} h \circ X(\omega) \mathbb{P}(d\omega) &= \int_{\mathbb{R}} h(x) \mathbb{P}_X(dx) \end{aligned}$$

Preuve. Voir Exercice 2 du devoir 2, avec $\varphi = X$ et $\mu = \mathbb{P}$. □

Conséquence. Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une v.a. de loi $\mathbb{P}_X$. Alors X est intégrable si et seulement si

$$\int_{\mathbb{R}} |x| \mathbb{P}_X(dx) < \infty$$

et alors $\mathbb{E}[X] = \int_{\mathbb{R}} x \mathbb{P}_X(dx)$.

Généralisation : si $h : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une fonction borélienne. Alors $h(X)$ est une variable aléatoire. $h(X)$ est $\mathbb{P}$ -intégrable si et seulement si h est $\mathbb{P}_X$ -intégrable : $\int_{\mathbb{R}} |h(x)| \mathbb{P}_X(dx) < \infty$ et alors

$$\mathbb{E}[h(X)] = \int_{\Omega} h(X) \mathbb{P} = \int_{\mathbb{R}} h(x) \mathbb{P}_X(dx).$$

3.4.a) Exemples de calcul d'espérance

Espérance d'une v.a. discrète

Soit X une v.a. de loi $\mathbb{P}_X = \sum_{i \in I} p_i \delta_{x_i}$ où I ensemble au plus dénombrable et $x_i \in \mathbb{R}$. Alors X est intégrable si et seulement si $\sum_{i \in I} |x_i| p_i < \infty$ et dans ce cas

$$\mathbb{E}[X] = \sum_{i \in I} x_i p_i.$$

On remarque que si l'ensemble des atomes $\mathfrak{A}$ de X est fini alors X est toujours intégrable.

Soit $h : \mathbb{R} \rightarrow \mathbb{R}$ mesurable, alors $h(X)$ intégrable si et seulement si $\sum_{i \in I} |h(x_i)| p_i < \infty$ et dans ce cas

$$\mathbb{E}[h(X)] = \sum_{i \in I} h(x_i) p_i.$$

Exemple. 1. Si $X = c$ est constance, alors $\mathbb{E}[X] = c$.

2. Si X est équirépartie sur I de cardinal n , alors $\mathbb{E}[X] = \frac{1}{n} \sum_{i=1}^n x_i$.

3. Si $X \sim \mathcal{P}(\lambda)$ de loi de Poisson de paramètre $\lambda > 0$, alors $\mathbb{E}[X] = \lambda$. En effet,

$$\mathbb{E}[X] = e^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

Espérance d'une variable à densité

Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une v.a. de densité f , alors X est intégrable si et seulement si $\mathbb{E}[|X|] = \int_{\mathbb{R}} |x|f(x)dx < \infty$ et dans ce cas

$$\mathbb{E}[X] = \int_{\mathbb{R}} xf(x)dx.$$

De plus, si $h : \mathbb{R} \rightarrow \mathbb{R}$ mesurable, alors $h(X)$ intégrable si et seulement si $\int_{\mathbb{R}} |h(x)|f(x)dx < \infty$ et dans ce cas

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

Exemple. 1. Si $X \sim \mathcal{U}([a, b])$ de loi uniforme sur $[a, b]$, alors X intégrable car bornée et $\mathbb{E}[X] = \frac{a+b}{2}$.
 2. Si $X \sim \mathcal{E}(\alpha)$ de loi exponentielle de paramètre $\alpha > 0$, alors $\mathbb{E}[X] = \frac{1}{\alpha}$.
 3. Si $X \sim \mathcal{C}$ de loi de Cauchy de densité sur $\mathbb{R}$ donnée par $f(x) = \frac{1}{\pi(1+x^2)}$, on a $\mathbb{E}[|X|] = \frac{2}{\pi} \int_0^\infty \frac{x}{1+x^2}dx = \infty$. La variable X n'est pas intégrable.

Quelques autres lois

Supposons que la loi de X soit la combinaison convexe de deux mesures de probabilité μ et ν : il existe $\alpha \in [0, 1]$ tel que

$$\mathbb{P}_X = \alpha\mu + (1 - \alpha)\nu$$

alors

$$\mathbb{E}[X] = \alpha \int x\mu(dx) + (1 - \alpha) \int x\nu(dx).$$

Par exemple, $\mathbb{P}_X = \frac{1}{2}\delta_2 + \frac{1}{2}\mathbb{1}_{[0,1]}(x)dx$. X est une v.a. positive et

$$\mathbb{E}[X] = \int_{\mathbb{R}} x\mathbb{P}_X(dx) = \frac{1}{2}2 + \frac{1}{2} \int_{\mathbb{R}} x\mathbb{1}_{[0,1]}(x)dx = \frac{5}{4}.$$

De manière plus générale, si la loi de X est une combinaison convexe d'une loi discrète et d'une loi à densité,

$$\mathbb{P}_X = \sum_{i=1}^n p_i\delta_{x_i} + f(x)dx$$

où $p_i > 0$ et h fonction mesurable positive, tels que $\sum_{i=1}^n p_i + \int_{\mathbb{R}} f(x)dx = 1$, alors la variable est intégrable si $\sum_{i=1}^n p_i|x_i| + \int_{\mathbb{R}} |x|f(x)dx < +\infty$ et dans ce cas

$$\mathbb{E}[X] = \sum_{i=1}^n p_i x_i + \int_{\mathbb{R}} xf(x)dx.$$

3.4.b) Inégalités importantes

Proposition 3.4.5 (Inégalité de Markov). *si X v.a. positive, alors pour tout $t > 0$*

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

Preuve. Si X positive alors $a\mathbb{1}_{X \geq a} \leq X$ et on conclut en prenant l'espérance. □

Exemple. Supposons que X v.a. de loi exponentielle de paramètre 1. Alors d'après l'inégalité de Markov, pour tout $a > 0$,

$$\mathbb{P}(X > a) \leq \frac{1}{a}.$$

Cependant, on sait que dans ce cas $\mathbb{P}(X > a) = e^{-a}$. On voit que la borne $1/a$ est très grossière (loin de la vraie valeur).

La proposition suivante est une conséquence directe de l'inégalité de Markov.

Proposition 3.4.6 (Inégalité de Tchebychev). *Si X est une variable telle que $\mathbb{E}[X^2] < +\infty$, alors pour tout $a > 0$,*

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

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

Proposition 3.4.7 (Inégalité de Jensen). *Soient $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction convexe et X une v.a. réelle tels que X et $\varphi(X)$ soit intégrables, alors*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$$

Preuve. Une fonction convexe est le sup de fonctions affines :

$$\varphi(x) = \sup \{h(x) : h \text{ affine telle que } h \leq \varphi\}.$$

Par linéarité et croissance de l'espérance, si h est affine avec $h \leq \varphi$, alors $h(\mathbb{E}[X]) = \mathbb{E}[h(X)] \leq \mathbb{E}[\varphi(X)]$. On conclut en prenant le sup. $\square$

Conséquence. On a

$$\begin{aligned} \mathbb{E}[X]^+ &\leq \mathbb{E}[X^+], \quad \mathbb{E}[X]^2 \leq \mathbb{E}[X^2], \\ \exp(\lambda \mathbb{E}[X]) &\leq \mathbb{E}[e^{\lambda X}] \text{ pour } \lambda > 0, \\ \max(\mathbb{E}[X], c) &\leq \mathbb{E}[\max(X, c)], \\ \frac{1}{\mathbb{E}[X]} &\leq \mathbb{E}\left[\frac{1}{X}\right] \text{ si } X \text{ positive.} \end{aligned}$$

Comme $x \mapsto \ln(x)$ et $x \mapsto \min(X, c)$ est concave, on a alors

$$\mathbb{E}[\ln(X)] \leq \ln \mathbb{E}[X], \quad \text{et} \quad \mathbb{E}[\min(X, c)] \leq \min(\mathbb{E}[X], c).$$

On remarque que si φ est une fonction mesurable telle que $\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$ pour toute v.a. X , alors φ est convexe (il suffit de prendre X de loi $p\delta_x + (1-p)\delta_y$, avec $x, y \in \mathbb{R}$).

3.4.c) Identification de la loi d'une variable aléatoire par l'espérance

L'espérance d'une variable aléatoire, représentant juste la valeur moyenne de la variable, ne peut pas caractériser la loi de la variable (il est facile d'imaginer que deux variables aient la même moyenne sans pour autant avoir la même loi). On peut cependant utiliser l'espérance pour caractériser la loi d'une variable.

Théorème 3.4.8. Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ un vecteur aléatoire et μ une mesure de probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Alors la loi de X est μ si et seulement si l'un des conditions suivantes est réalisée

1. Pour tout $A \in \mathcal{B}(\mathbb{R}^d)$, $\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mu(A)$,
2. la fonction de répartition satisfait pour tout $x = (x_1, \dots, x_d) \in \mathbb{R}^d$, en notant $]-\infty, x] = \prod_{i=1}^d]-\infty, x_i]$, $F_X(x) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d) = \mu(]-\infty, x])$,
3. pour toute fonction $h : \mathbb{R}^d \rightarrow \mathbb{R}$ continue à support compact positive,

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h(x) \mu(dx).$$

Le théorème reste vrai si on se restreint aux fonctions h de classe C^∞ à support compact et positives.

Preuve. Il suffit de vérifier la dernière équivalence. Il est clair que si les mesures sont égales, on a égalité des intégrales.

Réciproquement, notons ν la loi de X . Supposons que

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h(x) \nu(dx) = \int_{\mathbb{R}^d} h(x) \mu(dx)$$

pour tout fonction h continue à support compact positive. Soit $K \in \mathcal{B}(\mathbb{R}^d)$ un ensemble compact de $\mathbb{R}^d$. On souhaite montrer que $\nu(K) = \mu(K)$. Malheureusement, la fonction $h = \mathbf{1}_K$ n'est pas continue et donc on ne peut pas utiliser l'hypothèse directement.

On introduit donc O un ouvert tel que $K \subset O$ et l'adhérence de O est compact, et on note h la fonction

$$h(x) = \frac{d(x, O^c)}{d(x, K) + d(x, O^c)}.$$

La fonction h est positive, bornée par 1, continue sur $\mathbb{R}^d$, à support compact car le complémentaire $O^c = \{x : h(x) = 0\}$ et qui vaut 1 si et seulement si $x \in K$. Par conséquent,

$$\mathbf{1}_K \leq h(x) \leq \mathbf{1}_O.$$

On construit une suite décroissante d'ouverts O_n telle que $\lim_{n \rightarrow \infty} O_n = K$. On peut prendre par exemple

$$O_n = \left\{ x \in \mathbb{R}^d : d(x, K) < 1/n \right\},$$

ce qui donne en dimension 1, pour $K = [a, b]$, $O_n =]a - 1/n, b + 1/n[$.

Comme $n \rightarrow d(., O_n^c)$ est décroissante et $u \rightarrow u/(a + u)$ est croissante, la suite (h_n) est décroissante vers $\mathbf{1}_K$. Par convergence monotone (avec la suite $(1 - h_n)_{n \geq 0}$ positive croissante),

$$\lim_{n \rightarrow \infty} \nu(h_n) = \nu(K) \quad \text{et} \quad \lim_{n \rightarrow \infty} \mu(h_n) = \mu(K).$$

Par hypothèse, $\forall n \geq 0$, $\nu(h_n) = \mu(h_n)$. Donc $\mu(K) = \nu(K)$ pour tout compact K , donc les mesures sont égales (l'ensemble des compacts engendre les boréliens et stable par intersection finie).

□

Le théorème précédent est très utile lorsqu'on cherche la loi d'une variable Y de la forme $Y = g(X)$ où X est une variable aléatoire de loi connue. En effet, on peut exprimer l'espérance d'une fonctionnelle de Y via la loi de X : si h est mesurable bornée (ou mesurable positive, peu importe), alors

$$\int h(y) \mathbb{P}_Y(dy) = \mathbb{E}[h(Y)] = \mathbb{E}[h(g(X))] = \int h(g(x)) \mathbb{P}_X(dx).$$

C'est notamment très utile pour trouver la loi de Y lorsque X est à densité sur $\mathbb{R}^d$ et que l'on peut utiliser le théorème de changement de variables.

Exemple. Soit X une variable de loi de Cauchy. Cherchons la loi de $Y = X^+$. Il s'agit bien d'une v.a. car transformation continue (et donc mesurable) d'une v.a.

Soit h une fonction mesurable positive, alors

$$\begin{aligned}\mathbb{E}[h(Y)] &= \mathbb{E}[h(X^+)] = \int_{\mathbb{R}} h(x^+) \frac{1}{\pi(1+x^2)} dx \\ &= h(0)\mathbb{P}(X \leq 0) + \int_0^\infty h(x) \frac{1}{\pi(1+x^2)} dx \\ &= h(0)\frac{1}{2} + \int_{\mathbb{R}} h(x) \frac{1}{\pi(1+x^2)} \mathbb{1}_{x>0} dx.\end{aligned}$$

Par conséquent, $\mathbb{P}_Y = \frac{1}{2}\delta_0 + \frac{1}{\pi(1+x^2)}\mathbb{1}_{x>0}dx$.

Cherchons maintenant la loi de $Z = \sqrt{1+X^2}$. Il s'agit bien d'une v.a. car transformation continue d'une v.a. Soit h une fonction mesurable positive, alors

$$\begin{aligned}\mathbb{E}[h(Z)] &= \mathbb{E}[h(\sqrt{1+X^2})] = \int_{\mathbb{R}} h(\sqrt{1+x^2}) \frac{1}{\pi(1+x^2)} dx \\ &= 2 \int_0^\infty h(\sqrt{1+x^2}) \frac{1}{\pi(1+x^2)} dx \quad \text{par parité} \\ &= 2 \int_1^\infty h(z) \frac{1}{\pi z^2} \frac{z}{\sqrt{z^2-1}} dz \quad \text{en utilisant le changement de variables } z = \sqrt{1+x^2}\end{aligned}$$

Par conséquent, $\mathbb{P}_Z = \frac{2}{\pi z \sqrt{z^2-1}} \mathbb{1}_{z>1} dz$.

3.4.d) Exercices sur l'espérance

Exercice 3.4.1. 1. Pour quelle valeur de a la loi de la variable aléatoire X à valeurs dans $\mathbb{N}$ donnée

$$\mathbb{P}(X = k) = \frac{e^{-2}}{4} \frac{2^k}{k!} (1 + ak) \quad \forall k \in \mathbb{N},$$

est bien définie ?

2. Soit Y une variable aléatoire de loi de Poisson de paramètre $\lambda > 0$. Trouver la loi de

$$Z := \begin{cases} Y/2 & \text{si } Y \text{ est pair,} \\ (1-Y)/2 & \text{si } Y \text{ est impair.} \end{cases}$$

3. Soit T une variable aléatoire de loi géométrique de paramètre $p \in]0, 1[$. On considère $U = 4[T/2] - 2T + 1$ où $[\cdot]$ désigne la partie entière. Trouver la loi de U .

Exercice 3.4.2. 1. Vérifier qu'il existe une variable aléatoire X dont la fonction de répartition vaut

$$F(t) = (1 + e^{-t})^{-1} \quad \forall t \in \mathbb{R}.$$

2. Calculer la densité de X .

3. On pose $U = e^X$, $V = \mathbb{1}_{\{0 < X < 1\}}$ et $W = X\mathbb{1}_{\{0 < X < 1\}}$. Trouver les lois de U, V et W .

Exercice 3.4.3. Soit $X \sim \mathcal{U}([-1, 3])$. On pose $Y = |X|$. Trouver la fonction de répartition et la densité de Y .

Exercice 3.4.4. [Variable aléatoire mixte] Soit X une variable aléatoire qui prend ses valeurs dans $[0, 1]$ et telle que $\mathbb{P}(X = 0) = 1/4$ et $\mathbb{P}(X = 1) = p$ et pour tout intervalle $[a, b] \subset]0, 1[$, on a

$$\mathbb{P}(X \in [a, b]) = \frac{b - a}{3}.$$

1. Déterminer p pour qu'on ait ainsi bien défini une loi de probabilité.
2. Exprimer la loi $\mathbb{P}_X$ sous la forme $\mathbb{P}_X = \mu_1 + \mu_2$ où μ_1 est une mesure discrète et μ_2 est une mesure à densité.
3. Calculer la fonction de répartition de X .
4. Calculer l'espérance et la variance de X .

Exercice 3.4.5. Soit X une variable de loi exponentielle de paramètre $\alpha > 0$. Exprimer la loi de $Y = X\mathbf{1}_{X>2}$. En déduire l'espérance de Y .

Exercice 3.4.6. Soit X une v.a. positive.

1. Montrer que

$$\mathbb{E}[X] = \int_0^{+\infty} \mathbb{P}(X > t) dt = \int_0^{+\infty} [1 - F_X(t)] dt.$$

2. Montrer que, $\mathbb{E}[X] < \infty$ si et seulement s'il existe $\varepsilon > 0$ (ou pour tout $\varepsilon > 0$) tel que

$$\sum_{n \geq 0} \mathbb{P}(X > \varepsilon n) < \infty.$$

3. Montrer de même que $\mathbb{E}[X] < \infty$ si et seulement s'il existe $\varepsilon > 0$ (ou pour tout $\varepsilon > 0$) tel que

$$\sum_{n \geq 0} 2^n \mathbb{P}(X > \varepsilon 2^n) < \infty.$$

Exercice 3.4.7. Soit U une variable aléatoire de loi $\mathcal{U}(]0, 1[)$. On considère $X = \lfloor \frac{1}{U} \rfloor$ où $\lfloor \cdot \rfloor$ désigne la fonction partie entière.

1. Montrer que X est une variable aléatoire discrète et donner sa loi.
2. Calculer la probabilité de l'évènement $\{X \geq 100\}$.
3. La variable aléatoire X est-elle intégrable ?

Exercice 3.4.8. Soit X une variable aléatoire de loi de Cauchy, i.e de densité

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

Calculer $\mathbb{E}[\min(|X|, 1)]$.

Exercice 3.4.9. En utilisant l'approximation de Tchebychev, combien de fois faut-il lancer une pièce équilibrée pour que la moyenne du nombre de pile soit dans $[1/2 - 0.1, 1/2 + 0.1]$ avec probabilité supérieure à 0.96 ?

Exercice 3.4.10. Soient $n \geq 1$ un entier fixé et p_1, p_2, p_3 trois réels positifs. On note

$$p_{i,j} = \begin{cases} \frac{n!}{i!j!(n-i-j)!} p_1^i p_2^j p_3^{n-i-j} & \text{si } i+j \leq n \\ 0 & \text{sinon.} \end{cases}$$

1. À quelle condition sur p_1, p_2, p_3 existe-t-il un couple (X, Y) tel que $\mathbb{P}(X = i, Y = j) = p_{i,j}$ (i.e, à quelle condition $(p_{i,j})_{i,j \geq 0}$ est une probabilité sur $\mathbb{N} \times \mathbb{N}$).
2. Déterminer les lois marginales de X et de Y .

Exercice 3.4.11. Le couple aléatoire (X, Y) a pour densité, par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$,

$$f(x, y) = cy \mathbb{1}_{[0,2]}(x) \mathbb{1}_{[0,1]}(y).$$

1. Calculer la constante c .
2. Trouver les densités marginales de X et de Y .
3. Mêmes questions pour

$$g(x, y) = c(x + 2y)e^{-2x-y} \mathbb{1}_{[0,+\infty[}(x) \mathbb{1}_{[0,+\infty[}(y).$$

4. Mêmes questions encore pour

$$h(x, y) = c \exp \left(-\frac{x^2 + 2xy + 4y^2}{8} \right).$$

Exercice 3.4.12. Soit $U = (X, Y)$ une v.a. dans $\mathbb{R}^2$ de densité $(x, y) \mapsto ce^{-x} \mathbb{1}_{0 < |y| < x}$.

1. Quelle est la valeur de c ?
2. Déterminer les lois marginales.
3. Quelle est la loi du vecteur $(\frac{X-Y}{2}, \frac{X+Y}{2})$?

3.5 Moments d'une variable aléatoire

3.5.a) Les espaces L^p

Définition 3.5.1. Une variable aléatoire $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{E})$, avec $E = \mathbb{R}, \mathbb{C}$ ou $\mathbb{R}^d$, admet un moment d'ordre $p > 0$ si

$$\mathbb{E}[|X|^p] = \int_{\Omega} |X(\omega)|^p \mathbb{P}(d\omega) = \int_E |x|^p \mathbb{P}_X(dx) < \infty.$$

On définit l'espace des variables ayant un moment d'ordre $p > 0$:

$$\mathcal{L}^p = \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P}) = \{X \text{ variables aléatoires sur } (\Omega, \mathcal{F}, \mathbb{P}) \text{ telle que } \mathbb{E}[|X|^p] < \infty\}.$$

Pour $p = \infty$, on définit l'ensemble des variables p.s. bornées,

$$\mathcal{L}^\infty = \mathcal{L}^\infty(\Omega, \mathcal{F}, \mathbb{P}) = \{X \text{ variable aléatoire sur } (\Omega, \mathcal{F}, \mathbb{P}) \text{ telle qu'il existe } c > 0 : \mathbb{P}(|X| > c) = 0\}.$$

Pour $X \in \mathcal{L}^p$, on introduit $\|X\|_p = \mathbb{E}[|X|^p]^{1/p}$ et pour $X \in \mathcal{L}^\infty$, $\|X\|_\infty = \inf \{c > 0 : \mathbb{P}(|X| > c) = 0\}$.

Du fait de la convexité de $x \mapsto x^p$ pour $p \geq 1$, on déduit facilement que pour $p \in [1, +\infty]$, $\mathcal{L}^p$ est un espace vectoriel (en effet, $(x + y)^p \leq 2^{p-1}(x^p + y^p)$).

Exemple. 1. Si $X = \mathbb{1}_A$ (variable de Bernoulli), alors X est à valeurs dans $\{0, 1\}$ et donc pour tout $c \geq 1$, $\mathbb{P}(|X| > c) = 0$ et pour tout $c < 1$ $\mathbb{P}(X > c) > 0$. Donc $X \in \mathcal{L}^\infty$ et $\|X\|_\infty = 1$.

2. Si X de loi exponentielle de paramètre $\alpha > 0$. Alors X est à valeurs dans $]0, +\infty[$ et pour tout $p > 0$, $X \in \mathcal{L}^p$, car

$$\mathbb{E}[|X|^p] = \mathbb{E}[X^p] = \int_0^\infty x^p \alpha e^{-\alpha x} dx = \frac{p!}{\alpha^p} < +\infty.$$

Par contre, $X \notin \mathcal{L}^\infty$, car $\forall c > 0$, $\mathbb{P}(X > c) = e^{-\alpha c} \neq 0$.

3. Si X est une v.a. de densité de Cauchy : pour $x \in \mathbb{R}$, $f_X(x) = \frac{1}{\pi(1+x^2)}$, alors $X \notin \mathcal{L}^1$, car

$$\mathbb{E}[|X|] = 2 \int_0^\infty \frac{x}{\pi(1+x^2)} dx = +\infty$$

On remarque cependant que pour $p \in]0, 1[$, $X \in \mathcal{L}^p$.

Proposition 3.5.2. Une variable X est négligeable ($X = 0$ p.s.) si et seulement si $\mathbb{E}[|X|] = 0$.

Revoir l'Exercice 3.1.3 sur la notion de propriété presque-sûr (p.s.).

Preuve. Supposons tout d'abord que X est négligeable : $X = 0$ presque sûrement. On a $\mathbb{P}(X \neq 0) = 0$ et comme $\min(|X|, n) \leq n \mathbb{1}_{\{X \neq 0\}}$, on a

$$\mathbb{E}[\min(|X|, n)] \leq n \mathbb{P}(X \neq 0) = 0$$

d'où $\mathbb{E}[\min(|X|, n)] = 0$ pour tout $n \geq 0$. La suite de v.a. $X_n = \min(|X|, n)$ est une suite croissante de v.a. positives qui converge vers $|X|$. Par le théorème de convergence monotone, on a alors

$$\mathbb{E}[|X|] = \lim_{n \rightarrow +\infty} \mathbb{E}[\min(|X|, n)] = 0.$$

Réciproquement, si $\mathbb{E}[|X|] = 0$, alors pour tout $n \geq 1$, on a par l'inégalité de Markov

$$\mathbb{P}\left(|X| \geq \frac{1}{n}\right) \leq n \mathbb{E}[|X|] = 0.$$

Par ailleurs,

$$\{|X| \neq 0\} = \bigcup_{n \geq 1} \left\{|X| \geq \frac{1}{n}\right\}$$

l'ensemble $\{|X| \neq 0\}$ est une union dénombrable d'ensembles négligeables, il est donc de mesure nulle. $\square$

Conséquence. Si X v.a. et $A \in \mathcal{F}$ avec $\mathbb{P}(A) = 0$, alors

$$\mathbb{E}[X \mathbb{1}_A] = 0.$$

En effet, $X \mathbb{1}_A$ est négligeable car $\{X \mathbb{1}_A \neq 0\} = \{\omega \in \Omega : X(\omega) \mathbb{1}_A(\omega) \neq 0\} \subset A$.

La relation $X \stackrel{p.s.}{=} Y$ est une relation d'équivalence. On note L^p (resp. L^∞) l'espace $\mathcal{L}^p$ (resp. $\mathcal{L}^\infty$) quotienté par cette relation d'équivalence.

Propriété 3.5.3. 1. **Inégalité de Cauchy-Schwartz :** Soient $X, Y \in L^2$ alors $XY \in L^1$ et

$$|\mathbb{E}[XY]| \leq \|XY\|_1 \leq \|X\|_2 \|Y\|_2.$$

Il y a égalité si et seulement si les variables X et Y sont liées, i.e. $\exists a, b \in \mathbb{R}$ tels que $aX + bY$ est nul p.s.

2. **Inégalité de Hölder** : Soient $p, q \in [1, \infty]$ deux réels conjugués, i.e. tels que $\frac{1}{p} + \frac{1}{q} = 1$. Si X, Y deux variables sur $(\Omega, \mathcal{F}, \mathbb{P})$ avec $X \in L^p$ et $Y \in L^q$, alors $XY \in L^1$ avec

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

3. **Inégalité de Minkowski** : Soient $p \in [1, +\infty]$ et X, Y deux v.a. dans L^p , alors

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

4. Pour $p \in [1, +\infty]$, $(L^p, \|\cdot\|_p)$ est un espace vectoriel normé.
5. **Théorème de Riesz-Fischer** : $(L^p, \|\cdot\|_p)$ est un espace vectoriel complet (i.e. espace de Banach) pour tout $p \geq 1$, et L^2 est un espace de Hilbert pour le produit scalaire $\langle X, Y \rangle = \mathbb{E}[XY]$
6. Si $1 \leq p \leq q \leq +\infty$, alors $L^\infty \subset L^q \subset L^p \subset L^1$. De plus $p \rightarrow \|\cdot\|_p$ est croissante et $\lim_{p \rightarrow \infty} \|\cdot\|_p = \|\cdot\|_\infty$.

Preuve.

1. Soit $a \in \mathbb{R}$, alors en développant le carré, on a

$$\mathbb{E}[(a|X| + |Y|)^2] = a^2\mathbb{E}[X^2] + 2a\mathbb{E}[|XY|] + \mathbb{E}[Y^2].$$

Il s'agit d'un polynôme en a de degré 2 positif pour tout $a \in \mathbb{R}$, alors son discriminant est négatif

$$\Delta' = \mathbb{E}[|XY|]^2 - \mathbb{E}[X^2]\mathbb{E}[Y^2] \leq 0$$

D'où $\mathbb{E}[|XY|] \leq \mathbb{E}[X^2]^{1/2}\mathbb{E}[Y^2]^{1/2}$. On savait déjà que $|\mathbb{E}[XY]| \leq \mathbb{E}[|XY|]$.

En utilisant, $\mathbb{E}[(aX + Y)^2] = a^2\mathbb{E}[X^2] + 2a\mathbb{E}[XY] + \mathbb{E}[Y^2]$, on obtient juste $|\mathbb{E}[XY]| \leq \mathbb{E}[X^2]^{1/2}\mathbb{E}[Y^2]^{1/2}$.

Il y a égalité lorsque le discriminant $\Delta' = \mathbb{E}[XY]^2 - \mathbb{E}[X^2]\mathbb{E}[Y^2]$ est nul, donc lorsque le polynôme en a a une racine double. Il existe par conséquent $a_0 \in \mathbb{R}$, $a_0X + Y = 0$ p.s.

2. Si $\|X\|_p\|Y\|_q = 0$, alors l'un des termes est nul, par exemple $\|X\|_p = 0$. Par conséquent $X = 0$ p.s., ce qui implique que $XY = 0$ p.s. et donc $XY \in L^1$ et $\|XY\|_1 = 0$.

Si $\|X\|_p\|Y\|_q > 0$.

Lemme 3.5.4. Pour tous $a, b > 0$, $p, q > 0$ avec $\frac{1}{p} + \frac{1}{q} = 1$,

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Preuve. En effet, il existe $s, t \in \mathbb{R}$ tels que $a = e^{s/p}$ et $b = e^{t/q}$ et par convexité de la fonction exponentielle et du fait que $1/p + 1/q = 1$,

$$\exp\left(\frac{s}{p} + \frac{t}{q}\right) \leq \frac{1}{p}e^s + \frac{1}{q}e^t = \frac{a^p}{p} + \frac{b^q}{q}.$$

En considérant $a = \frac{|X|}{\|X\|_p}$ et $b = \frac{|Y|}{\|Y\|_q}$ et en utilisant le résultat du lemme on déduit que

$$\frac{|XY|}{\|X\|_p\|Y\|_q} \leq \frac{1}{p} \frac{|X|^p}{\mathbb{E}[|X|^p]} + \frac{1}{q} \frac{|Y|^q}{\mathbb{E}[|Y|^q]}.$$

On conclut en prenant l'espérance.

□

3. Si $\|X + Y\|_p = 0$, l'inégalité est vérifiée.

Sinon, en utilisant l'inégalité de Hölder et l'inégalité triangulaire, il vient

$$\begin{aligned}\|X + Y\|_p^p &= \mathbb{E}[|X + Y|^p] \leq \mathbb{E}[|X + Y|^{p-1}(|X| + |Y|)] \\ &\leq \mathbb{E}[|X + Y|^{p-1}|X|] + \mathbb{E}[|X + Y|^{p-1}|Y|] \\ &\quad \text{en utilisant Hölder avec } q = \frac{p}{p-1} \\ &\leq \mathbb{E}[|X + Y|^{p-1}]^{1-1/p} \mathbb{E}[|X|^p]^{1/p} + \mathbb{E}[|X + Y|^{p-1}]^{1-1/p} \mathbb{E}[|Y|^p]^{1/p} \\ &\leq (\|X\|_p + \|Y\|_p) \frac{\|X + Y\|_p^p}{\|X + Y\|_p},\end{aligned}$$

D'où le résultat.

4. Conséquence de 3 et de la Proposition 3.5.2, dans le sens où on a égalité entre deux variables aléatoires si $X = Y$ p.s (i.e, $\mathbb{P}(X \neq Y) = 0$).
5. Admis (Théorème 9.3 et Corollaire 9.3 dans [BP06])
6. En exercice.

□

3.5.b) Variance, Covariance

Définition 3.5.5. Soit $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs réelles. On définit la variance de X par

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

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

L'espérance correspond à la valeur moyenne d'une variable aléatoire. L'écart-type correspond à la distance moyenne entre la variable et sa moyenne. C'est une mesure de dispersion. Par exemple, les mesures $\mathbb{P}_1 = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ et $\mathbb{P}_2 = \frac{1}{2}\delta_{-1000} + \frac{1}{2}\delta_{1000}$ ont la même moyenne égale à 0, mais la dispersion de la seconde est bien plus grande que celle de la première.

Proposition 3.5.6. 1. $\text{Var}(X) \geq 0$,

2. en développant le carré, on a $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ (Formule de Koenig-Huygens),

3. $\text{Var}(aX) = a^2\text{Var}(X)$,

4. $\text{Var}(X + b) = \text{Var}(X)$,

5. $\text{Var}(X) = 0$ si et seulement si X est constant p.s.

Définition 3.5.7. La covariance entre deux variables X et Y dans L^2 est définie par

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

On remarque que la covariance est une application bilinéaire, $\text{Cov}(X, X) = \text{Var}(X)$, $\text{Cov}(X, a) = 0$. Par ailleurs, en développant le carré de la variance d'une somme, on a

Propriété 3.5.8. Soient $X, Y \in L^2$, alors

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y),$$

et en utilisant l'inégalité de Cauchy-Schwartz,

$$|Cov(X, Y)| \leq \sqrt{\text{Var}(X)\text{Var}(Y)}.$$

Propriété 3.5.9 (Inégalité de Tchebychev). Soit X une variable dans L^2 , alors pour tout $t > 0$

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

Preuve. Utiliser l'inégalité de Markov. □

Définition 3.5.10. Soit $X = (X_1, \dots, X_d)^t$ un vecteur aléatoire colonne) dans L^2 . On appelle matrice de covariance la matrice $d \times d$

$$\Sigma_X = (Cov(X_i, X_j))_{1 \leq i, j \leq d}.$$

On remarque que $\Sigma_X = \mathbb{E}[XX^t] - \mathbb{E}[X]\mathbb{E}[X]^t = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^t]$

La matrice Σ_X est positive, symétrique et

$$a^t \Sigma_X a = \mathbb{E}[a^t (X - \mathbb{E}[X])(X - \mathbb{E}[X])^t a] = \mathbb{E}[(a^t (X - \mathbb{E}[X]))^2]$$

où a^t est le vecteur transposé. La matrice Σ_X est définie si aucune combinaison linéaire des marginales est p.s. constante.

3.5.c) Exercices sur les moments

Exercice 3.5.1. Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ une variable aléatoire. Calculer $\mathbb{E}[X]$, $\mathbb{E}[|X|]$, $\mathbb{E}[X^2]$ dans les cas suivants :

1. $\mathbb{P}_X(dx) = \frac{1}{2} \mathbb{1}_{[-1,1]}(x) dx$;
2. $\mathbb{P}_X(dx) = \exp(-(x^2 - 2x)/4) dx / (\sqrt[4]{e} \sqrt{4\pi}) dx$.

Exercice 3.5.2. Soit X une variable aléatoire réelle de densité

$$f_X(x) = c(x \mathbb{1}_{[0,1]}(x) + (2-x) \mathbb{1}_{[1,2]}(x)).$$

Déterminer la valeur de c et calculer les moments $\mu_n(X) = \mathbb{E}[X^n]$, $n \in \mathbb{N}^*$, de X .

Exercice 3.5.3. [Couple gaussien] Soit

$$f(x, y) = C \exp\left(-\frac{x^2 + 2\alpha xy + y^2}{2}\right).$$

1. Pour quels $\alpha \in \mathbb{R}$, la fonction f peut-elle définir la densité d'un couple (X, Y) ? Déterminer alors la constante C .
2. Déterminer les lois de X et de Y .
3. Calculer $\mathbb{E}[XY]$ et $\text{Var}(X + Y)$.

Exercice 3.5.4. [Tirages avec remise] Soit X de loi géométrique de paramètre $p \in]0, 1[$ et soit $n \in \mathbb{N}^*$.

1. On note $Y = \inf(X, n)$. Calculer $\mathbb{E}[Y]$.

2. Une urne contient 2 boules rouges et 3 boules noires. On tire successivement, et avec remise à chaque fois, une boule de l'urne et on cesse les tirages dès qu'une boule rouge est sortie. On désigne par Z le nombre de boules noires obtenues pendant l'expérience. Calculer $\mathbb{E}[Z]$ et $\text{Var}(Z)$.

Exercice 3.5.5. Soit X une v.a. positive de fonction de répartition F_X .

1. Soit ϕ une fonction positive strictement croissante de classe C^1 sur $[0, +\infty[$, nulle en 0. Montrer que

$$\mathbb{E}[\phi(X)] = \int_0^{+\infty} \phi'(t)(1 - F_X(t)) dt.$$

2. Soit ϕ une fonction positive croissante avec $\phi(X)$ est intégrable. Montrer que

$$\lim_{t \rightarrow +\infty} \phi(t)\mathbb{P}(X > t) = 0.$$

3. Expliciter le cas particulier $\phi(t) = t^n$, $n \in \mathbb{N}^*$.
4. On suppose maintenant que pour t assez grand, on a

$$\mathbb{P}(X \geq t) \leq \frac{c}{t^\alpha}$$

où $c > 0$. Montrer que X admet un moment d'ordre k pour tout $k \in \mathbb{N}^*$ avec $k < \alpha$.

Exercice 3.5.6. Soit X une v.a. positive, ayant un moment d'ordre 2 fini. Soit $\alpha \in]0, 1[$.

1. Montrer que $(1 - \alpha)\mathbb{E}[X] \leq \mathbb{E}[X\mathbf{1}_{[\alpha\mathbb{E}[X], +\infty[}(X)]$.
2. En déduire que

$$\mathbb{P}(X \geq \alpha\mathbb{E}[X]) \geq (1 - \alpha)^2 \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}.$$

3. Appliquer cette formule à une variable X de loi exponentielle de paramètre $\lambda > 0$.

Exercice 3.5.7. [Moments et loi] Soit X une v.a. de loi normale $\mathcal{N}(0, 1)$.

1. Déterminer tous les moments $\mathbb{E}[X^n]$, $n \geq 1$, de X .
2. Trouver la densité f_Y de $Y = e^X$ et calculer les moments $\mu_n(Y) = \mathbb{E}[Y^n]$, $n \in \mathbb{N}^*$.
3. On note Y_a , $|a| \leq 1$, la variable aléatoire de densité

$$f_{Y_a}(x) = f_Y(x)(1 + a \sin(2\pi \ln(x))).$$

Calculer les moments $\mu_n(Y_a) = \mathbb{E}[Y_a^n]$, $n \in \mathbb{N}^*$, de Y_a .

4. En déduire une conclusion intéressante.

3.6 Fonction caractéristique

Définition 3.6.1. Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ un vecteur aléatoire. On appelle **fonction caractéristique** $\varphi_X : \mathbb{R}^d \rightarrow \mathbb{C}$ la fonction définie pour $t = (t_1, \dots, t_d) \in \mathbb{R}^d$

$$\varphi_X(t) = \mathbb{E}\left[e^{i\langle t, X \rangle}\right] = \mathbb{E}\left[e^{i\sum_{k=1}^d t_k X_k}\right],$$

où $\langle \cdot, \cdot \rangle$ est le produit scalaire euclidien sur $\mathbb{R}^d$.

On remarque que φ_X est bien définie pour tout $t \in \mathbb{R}^d$ car $|e^{i\langle t, X \rangle}| = 1$ et donc intégrable.

Par ailleurs, si X est une v.a. réelle à densité, on retrouve la transformée de Fourier, généralement vue en cours d'analyse,

$$\mathbb{E}[e^{itX}] = \int_{\mathbb{R}} e^{itx} f_X(x) dx = \mathcal{F}(f)\left(\frac{t}{2\pi}\right).$$

3.6.a) Théorème de Fourier

Théorème 3.6.2 (Admis). *La fonction caractéristique caractérise la loi, i.e. si X, Y deux v.a., on a $\varphi_X = \varphi_Y$ si et seulement si $\mathbb{P}_X = \mathbb{P}_Y$.*

Preuve. (Non fait en cours, car utilise le théorème de convergence dominé qui n'a pas été introduit)

Preuve en dimension 1. Supposons que $\varphi_X = \varphi_Y$.

Soit $h \in \mathcal{C}_c(\mathbb{R}, \mathbb{R})$ continue à support compact. On remarque, en utilisant une approximation de l'unité, que par convergence dominée (pas encore introduit !)

$$h(x) = \frac{1}{2\pi} \lim_{\lambda \searrow 0} \int_{\mathbb{R}} h(x - \lambda t) \frac{2}{1 + t^2} dt.$$

En effet,

$$\left| h(x - \lambda t) \frac{2}{1 + t^2} \right| \leq \|h\|_{\infty} \frac{2}{1 + t^2} \in L^1(\mathbb{R}).$$

Exprimons maintenant $\mathbb{E}[h(X)]$ en terme de φ_X :

$$\begin{aligned} \mathbb{E}[h(X)] &= \int_{\mathbb{R}} h(x) \mathbb{P}_X(dx) \\ &= \int_{\mathbb{R}} \left(\frac{1}{2\pi} \lim_{\lambda \searrow 0} \int_{\mathbb{R}} h(x - \lambda t) \frac{2}{1 + t^2} dt \right) \mathbb{P}_X(dx) \\ &= \frac{1}{2\pi} \lim_{\lambda \searrow 0} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x - \lambda t) \frac{2}{1 + t^2} dt \right) \mathbb{P}_X(dx) \text{ par CV dominée par } \|h\|_{\infty} \frac{2}{1 + t^2} \in L^1(dt \otimes \mathbb{P}_X) \\ &= \frac{1}{2\pi} \lim_{\lambda \searrow 0} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(s) \frac{2\lambda}{\lambda^2 + (x - s)^2} ds \right) \mathbb{P}_X(dx) \text{ par chgt de variable } s = x - \lambda t \\ &= \frac{1}{2\pi} \lim_{\lambda \searrow 0} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \frac{2\lambda}{\lambda^2 + (x - s)^2} \mathbb{P}_X(dx) \right) h(s) ds \text{ par Fubini.} \end{aligned}$$

Par ailleurs, en décomposant en éléments simples $2\lambda/(\lambda^2 + (x - s)^2)$, on a

$$\begin{aligned} \frac{2\lambda}{\lambda^2 + (x - s)^2} &= \frac{1}{\lambda - i(x - s)} + \frac{1}{\lambda + i(x - s)} \\ &= \int_0^{\infty} e^{-(\lambda - i(x - s))t} dt + \int_0^{\infty} e^{-(\lambda + i(x - s))t} dt \\ &= \int_0^{\infty} e^{i(x - s)t} e^{-\lambda t} dt + \int_0^{\infty} e^{-i(x - s)t} e^{-\lambda t} dt \\ &= \int_{\mathbb{R}} e^{i(x - s)t} e^{-\lambda|t|} dt. \end{aligned}$$

Par conséquent, par Fubini (voir justification plus bas),

$$\begin{aligned} \mathbb{E}[h(X)] &= \frac{1}{2\pi} \lim_{\lambda \searrow 0} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{i(x - s)t} e^{-\lambda|t|} dt \right) \mathbb{P}_X(dx) \right) h(s) ds \\ &= \frac{1}{2\pi} \lim_{\lambda \searrow 0} \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{ixt} \mathbb{P}_X(dx) \right) e^{-ist} e^{-\lambda|t|} dt h(s) ds \\ \mathbb{E}[h(X)] &= \frac{1}{2\pi} \lim_{\lambda \searrow 0} \int_{\mathbb{R}} h(s) \int_{\mathbb{R}} \varphi_X(t) e^{-\lambda|t|} e^{-ist} dt ds. \end{aligned}$$

Par conséquent, $\mathbb{E}[h(X)]$ ne dépend que de h et de φ_X . Si $\varphi_X = \varphi_Y$, on en déduit que $\mathbb{E}[h(X)] = \mathbb{E}[h(Y)]$ pour tout $h \in \mathcal{C}_c(\mathbb{R}, \mathbb{R})$ et donc $\mathbb{P}_X = \mathbb{P}_Y$.

Justification de Fubini,

$$\int_{\mathbb{R}^3} \left| e^{i(x-s)t} e^{-\lambda|t|} h(s) \right| \mathbb{P}_X(dx) dt ds = \int_{\mathbb{R}} e^{-\lambda|t|} dt \int_{\mathbb{R}} |h(s)| ds < \infty$$

car h est à support compact.

La réciproque est évidente. □

Proposition 3.6.3. *Soit X un vecteur aléatoire de $\mathbb{R}^d$ alors sa fonction caractéristique vérifie*

1. $|\varphi_X(t)| \leq 1$,
2. $\varphi_X(-t) = \overline{\varphi_X(t)}$,
3. $\varphi_X(0) = 1$,

Preuve. Évident. □

Exemple. • Si X est discrète à valeurs dans $S = \{x_k, k \in K\}$, alors, en notant $p_k = \mathbb{P}(X = x_k)$,

$$\varphi_X(t) = \sum_{k \in K} e^{itx_k} p_k.$$

Si $X \sim \mathcal{B}(n, p)$, on a alors, par la formule du binôme de Newton,

$$\varphi_X(t) = \sum_{k=0}^n \binom{n}{k} e^{itk} p^k (1-p)^{n-k} = (pe^{it} + 1 - p)^n.$$

- Si X à densité alors

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx} f_X(x) dx.$$

Par exemple, si $X \sim \mathcal{N}(0, 1)$ alors

$$\varphi_X(t) = \int e^{\frac{-x^2}{2}} e^{ixt} \frac{dx}{\sqrt{2\pi}} = e^{\frac{-t^2}{2}} \int e^{\frac{-(x-it)^2}{2}} \frac{dx}{\sqrt{2\pi}}.$$

On voit apparaître l'intégrale de la densité de $Y = Z + it$ où $Z \sim \mathcal{N}(0, 1)$ et donc $\varphi_X(t) = e^{-\frac{t^2}{2}}$.

On en déduit que si $X \sim \mathcal{N}(m, \sigma^2)$,

$$\varphi_X(t) = e^{itm} e^{-\frac{t^2 \sigma^2}{2}}.$$

car X peut s'écrire sous la forme $X = \sigma Z + m$ avec $Z \sim \mathcal{N}(0, 1)$.

Si $X \sim \mathcal{E}(\alpha)$ avec $\alpha > 0$, alors

$$\begin{aligned} \varphi_X(t) &= \int_0^\infty \alpha e^{-\alpha x} e^{ixt} dx \\ &= \alpha \left[\frac{-1}{\alpha - it} e^{-(\alpha - it)x} \right]_0^\infty \\ &= \frac{\alpha}{\alpha - it}. \end{aligned}$$

3.6.b) Théorème d'inversion de Fourier

Théorème 3.6.4 (Inversion de Fourier, Admis). *Soit X une v.a. de fonction caractéristique φ_X . Si φ_X est intégrable sur $\mathbb{R}^d$ par rapport à la mesure de Lebesgue, alors X admet une densité continue bornée donnée par*

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

Preuve. (Non fait en cours, car utilise le théorème de convergence dominée qui n'a pas été introduit)

Preuve en dimension $d = 1$. Soit $h \in \mathcal{C}_c(\mathbb{R})$. Alors

$$\begin{aligned} \mathbb{E}[h(X)] &= \frac{1}{2\pi} \lim_{\lambda \searrow 0} \int_{\mathbb{R}} h(s) \int_{\mathbb{R}} \varphi_X(t) e^{-\lambda|t|} e^{-ist} dt ds \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} h(s) \left(\int_{\mathbb{R}} \varphi_X(t) e^{-ist} dt \right) ds, \end{aligned}$$

par convergence dominée et Fubini. En effet,

$$\left| h(s) \varphi_X(t) e^{-\lambda|t|} e^{-ist} \right| \leq |h(s)| |\varphi_X(t)| \in L^1(ds \otimes dt).$$

□

Exemple. Si X de loi de Cauchy de paramètre $a > 0$, on a

$$\varphi_X(t) = \int_{\mathbb{R}} \frac{a}{\pi(a^2 + x^2)} e^{ixt} dx.$$

En décomposant $a/(a^2 + x^2)$ en éléments simples on a

$$\frac{1}{\pi} \frac{a}{a^2 + x^2} = \frac{1}{2\pi} \left(\frac{1}{a + ix} + \frac{1}{a - ix} \right) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-a|s|} e^{-ixs} ds \in L^1(dx).$$

Par conséquent, par le théorème d'inversion de Fourier $\varphi_X(t) = e^{-a|t|}$. On remarque que cette fonction est continue en 0, mais n'est pas dérivable.

Propriété 3.6.5 (Admis). *Soit X une v.a. réelle de fonction caractéristique φ_X . Si X admet un moment d'ordre $p \in \mathbb{N}^*$ alors φ_X est de classe $\mathcal{C}^p$ et pour tout $k \leq p$,*

$$\varphi_X^{(k)}(0) = i^k \mathbb{E}[X^k]$$

Réciproquement, si φ_X est p fois dérivable en 0, alors X admet des moments d'ordre inférieur ou égal à $2[p/2]$.

La proposition est encore vraie en dimension quelconque.

Par conséquent, si φ_X est suffisamment régulière, on a

$$\mathbb{E}[X] = -i\varphi_X'(0), \quad \mathbb{E}[X^2] = -\varphi_X^{(2)}(0), \quad \mathbb{E}[X^3] = i\varphi_X^{(3)}(0), \dots$$

Exemple. La loi Binomiale $X \sim \mathcal{B}(n, p)$ admet des moments de tous ordres car à support fini. Par ailleurs, on a

$$\varphi_X(t) = (pe^{it} + (1-p))^n$$

donc en dérivant successivement, on obtient

$$\mathbb{E}[X] = np, \quad \mathbb{E}[X]^2 = np + n(n-1)p^2 \quad \Rightarrow \quad \text{Var}(X) = np(1-p).$$

3.6.c) Formulaire des lois classiques

• Lois discrètes usuelles

Loi	Fonction de masse	Moments	$\varphi_X(t)$
Bernoulli $\mathcal{B}(p)$	$p_X(1) = p, p_X(0) = 1 - p$	$\mathbb{E}[X] = p, \text{Var}(X) = p(1 - p)$	$pe^{it} + 1 - p$
Binomiale $\mathcal{B}(n, p)$	$\binom{n}{k} p^k (1 - p)^{n-k}, 0 \leq k \leq n$	$\mathbb{E}[X] = np, \text{Var}(X) = np(1 - p)$	$(pe^{it} + 1 - p)^n$
Géométrique $\mathcal{G}(p)$	$p(1 - p)^{k-1}, k \geq 1$	$\mathbb{E}[X] = \frac{1}{p}, \text{Var}(X) = \frac{1-p}{p^2}$	$\frac{pe^{it}}{1 - (1 - p)e^{it}}$
Poisson $\mathcal{P}(\lambda)$	$\frac{\lambda^k}{k!} e^{-\lambda}, k \geq 0$	$\mathbb{E}[X] = \lambda, \text{Var}(X) = \lambda$	$e^{\lambda(e^{it} - 1)}$
Binomiale Négative $BN(r, p)$	$\binom{k-1}{r-1} p^r (1 - p)^{k-r}, k \geq r$	$\mathbb{E}[X] = \frac{r}{p}, \text{Var}(X) = \frac{r(1-p)}{p^2}$	$\left(\frac{pe^{it}}{1 - (1 - p)e^{it}} \right)^r$

• Lois à densité usuelles

Loi	Densité	Moments	$\varphi_X(t)$
Uniforme $\mathcal{U}([a, b]), a < b$	$\frac{1}{b-a} \mathbb{1}_{x \in [a, b]}$	$\mathbb{E}[X] = \frac{a+b}{2}, \text{Var}(X) = \frac{(b-a)^2}{12}$	$\frac{e^{ibt} - e^{iat}}{it(b-a)}$
Exponentielle $\mathcal{E}(\lambda), \lambda > 0$	$\lambda e^{-\lambda x} \mathbb{1}_{x > 0}$	$\mathbb{E}[X] = \frac{1}{\lambda}, \text{Var}(X) = \frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda - it}$
Normale $\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma > 0$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\mathbb{E}[X] = \mu, \text{Var}(X) = \sigma^2$	$e^{i\mu t} e^{-\frac{\sigma^2 t^2}{2}}$
Gamma $G(\alpha, \lambda), \alpha, \lambda > 0$	$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathbb{1}_{x > 0}$ with $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$	$\mathbb{E}[X] = \frac{\alpha}{\lambda}, \text{Var}(X) = \frac{\alpha}{\lambda^2}$	$\left(\frac{\lambda}{\lambda - it} \right)^\alpha$

Rappel $X \sim \mathcal{N}(\mu, \sigma^2)$ si et seulement si $X = \sigma Z + \mu$ avec $Z \sim \mathcal{N}(0, 1)$.

3.6.d) Exercices sur les fonctions caractéristiques

Exercice 3.6.1. Calculer les fonctions caractéristiques des lois classiques : bernoulli, géométrique, Binomiale, Poisson, uniforme sur un intervalle $[a, b]$, Exponentielle, Gamma, Normale.

Exercice 3.6.2. Une variable aléatoire X suit une loi arithmétique s'il existe $a \geq 0$ et $b > 0$ tels que X prenne ses valeurs dans le réseau $a + b\mathbb{Z}$, c'est à dire :

$$\mathbb{P}(X \in \{a + nb, n = 0, \pm 1, \pm 2, \dots\}) = 1.$$

1. On suppose que X suit une loi arithmétique. Montrer qu'il existe $c \neq 0$ tel que $|\varphi_X(c)| = 1$.
2. Réciproquement, s'il existe $c \neq 0$ tel que $|\varphi_X(c)| = 1$, on montre que X suit une loi arithmétique.

- (a) Montrer que si $|\varphi_X(c)| = 1$ alors l'argument de e^{icX} est ps constant.
 - (b) En déduire que X suit une loi arithmétique.
3. S'il existe $c \neq 0$ et $c' \neq 0$ tels que $|\varphi_X(c)| = |\varphi_X(c')| = 1$ avec $c'/c \notin \mathbb{Q}$, montrer que X est ps constante.

Exercice 3.6.3. [Dérivabilité en 0 de la fonction caractéristique]

On considère une variable aléatoire X de support $\mathbb{Z} \setminus \{-1, 0, 1\}$ et loi donnée par

$$p_n = p_{-n} = \frac{c}{n^2 \ln(n)}, \quad n \in \mathbb{N} \setminus \{0, 1\}, \quad p_{-1} = p_0 = p_1 = 0$$

c'est-à-dire $\mathbb{P}_X = \sum_{n=2}^{+\infty} \frac{c}{n^2 \ln(n)} (\delta_n + \delta_{-n})$ (où c est une constante qu'on ne demande pas de préciser).

1. Justifier qu'on définit bien de cette façon une loi de probabilité.
2. Montrer que la fonction caractéristique φ_X de X s'exprime sous la forme

$$\varphi_X(t) = \sum_{n=2}^{+\infty} \frac{2c(1 - 2\sin^2(nt/2))}{n^2 \ln(n)}.$$

3. Soit, pour $N \geq 2$,

$$f_N(t) = \sum_{2 \leq n \leq N} \frac{\sin^2(nt/2)}{tn^2 \ln(n)}, \quad g_N(t) = \sum_{n > N} \frac{\sin^2(nt/2)}{tn^2 \ln(n)}.$$

- (a) Exprimer $\frac{\varphi_X(t)-t}{t}$ à l'aide de $f_N(t)$ et de $g_N(t)$.
 - (b) Montrer que $|f_N(t)| \leq \frac{|t|N}{4 \ln 2}$.
 - (c) Montrer que $|g_N(t)| \leq \frac{1}{|t|N \ln(N)}$.
4. Trouver une fonction $t \mapsto N(t) \in [0, +\infty[$ telle que $\lim_{t \rightarrow 0} f_{N(t)}(t) = 0$ et $\lim_{t \rightarrow 0} g_{N(t)}(t) = 0$.
 5. En déduire que φ est dérivable en 0 et formuler une remarque pertinente.

Exercice 3.6.4. Soit X une variable aléatoire réelle de fonction caractéristique φ .

On dit que X est de loi symétrique si la loi de $-X$ est la même que celle de X , i.e. $\mathbb{P}(X \in A) = \mathbb{P}(-X \in A)$ pour tout $A \in \mathcal{B}(\mathbb{R})$.

1. Donner des exemples de loi symétrique.
2. Montrer que si X est de loi symétrique ssi φ est à valeurs réelles.
3. Soit Y une variable aléatoire de même loi que X mais indépendante de X . Donner, en fonction de φ , la fonction caractéristique de $X - Y$. Que dire si X est symétrique ?
4. Soit U une variable indépendante de X et de loi $\mathbb{P}(U = 1) = \mathbb{P}(U = -1) = 1/2$. Donner, en fonction de φ , la fonction caractéristique de UX . Que dire si X est symétrique ?
5. Soient Y v.a. de fonction caractéristique φ , Z v.a. de fonction caractéristique ψ et $p \in [0, 1]$. Montrer que $p\varphi + (1 - p)\psi$ est encore une fonction caractéristique (i.e., construire une v.a. de fonction caractéristique $p\varphi + (1 - p)\psi$).

3.7 Notion d'indépendance

On se place sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

Définition 3.7.1 (Indépendance d'événements). Deux événements $A, B \in \mathcal{F}$ sont dits **indépendants** si

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

Une famille d'événements $(A_i)_{i \in I}$ est dite **(mutuellement) indépendante** si $\forall p \geq 1, \forall i_1, \dots, i_p \in I$

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_p}) = \prod_{k=1}^p \mathbb{P}(A_{i_k}).$$

On remarque que si A et B sont deux événements, avec $\mathbb{P}(B) > 0$, alors A et B sont indépendants si et seulement si $\mathbb{P}(A|B) = \mathbb{P}(A)$.



Indépendance et disjoint sont deux notions bien distinctes.

Si A et B disjoints alors $A \cap B = \emptyset$ et donc $\mathbb{P}(A \cap B) = 0$. A et B ne peuvent alors être indépendants sauf si l'un des deux est négligeable.

Définition 3.7.2 (Indépendance de tribus). Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Deux familles $\mathcal{M}_1$ et $\mathcal{M}_2$ d'ensemble mesurables sont dites **indépendantes** si $\forall A \in \mathcal{M}_1$ et $\forall B \in \mathcal{M}_2$

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

Une famille quelconque d'ensembles mesurables $(\mathcal{M}_i)_{i \in I}$ est dite **indépendante** si pour tout $A_i \in \mathcal{M}_i, i \in I$, la famille $(A_i)_{i \in I}$ est mutuellement indépendante.

Propriété 3.7.3. Soit $(\mathcal{A}_i)_{i \in I}$ une famille de Π -systèmes (ensembles stables par intersection finie), alors $(\mathcal{A}_i)_{i \in I}$ sont indépendants si et seulement si $(\sigma(\mathcal{A}_i))_{i \in I}$ sont indépendantes.

Preuve.

La réciproque est évidente : si les tribus engendrées sont indépendantes, comme $\mathcal{A}_i \subset \sigma(\mathcal{A}_i)$, on a $(\mathcal{A}_i)_{i \in I}$ sont indépendants.

Il suffit donc de montrer l'implication directe.

Cas de deux ensembles

On suppose $\mathcal{A}_1$ et $\mathcal{A}_2$ stables par intersection finie et indépendants.

Soit $A_1 \in \mathcal{A}_1$. On note

$$\mathcal{M} = \{A_2 \in \sigma(\mathcal{A}_2) : \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)\}.$$

On a $\mathcal{A}_2 \subset \mathcal{M}$. Par ailleurs, $\mathcal{M}$ contient Ω , est stable par différence (car $(B \setminus C) \cap A = (B \cap A) \setminus (C \cap A)$) et union croissante, donc c'est une classe monotone. $\mathcal{M}$ contient la classe monotone engendrée par $\mathcal{A}_2$ qui est égale à $\sigma(\mathcal{A}_2)$ car $\mathcal{A}_2$ est stable par intersection finie (voir théorème des classes monotones, Théorème 3.2.10). D'où $\mathcal{M} = \sigma(\mathcal{A}_2)$.

De même, pour $A_2 \in \sigma(\mathcal{A}_2)$, on définit

$$\mathcal{M}' = \{A_1 \in \sigma(\mathcal{A}_1) : \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)\}.$$

Par la même raison, $\mathcal{M}'$ est une classe monotone contenant $\mathcal{A}_1$ et donc $\sigma(\mathcal{A}_1)$, d'où $\mathcal{M}' = \sigma(\mathcal{A}_1)$. On en déduit que $\sigma(\mathcal{A}_1)$ et $\sigma(\mathcal{A}_2)$ sont indépendantes (car A_2 était choisi de façon quelconque).

Cas général Soit $p \geq 2$, $I_1, \dots, i_p \in I$.

On fixe $A_{i_k} \in \mathcal{A}_{i_k}$ pour $2 \leq k \leq p$ et on considère

$$\mathcal{M}_1 = \left\{ A_{i_1} \in \sigma(\mathcal{A}_{i_1}) : \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_p}) = \prod_{k=1}^p \mathbb{P}(A_{i_k}) \right\}$$

Par classe monotone, $\mathcal{M}_1 = \sigma(\mathcal{A}_{i_1})$.

Soit $A_{i_1} \in \sigma(\mathcal{A}_{i_1})$ et on fixe $A_{i_k} \in \mathcal{A}_{i_k}$ pour $3 \leq k \leq p$. On considère alors la famille

$$\mathcal{M}_2 = \left\{ A_{i_2} \in \sigma(\mathcal{A}_{i_1}) : \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_p}) = \prod_{k=1}^p \mathbb{P}(A_{i_k}) \right\}$$

Par classe monotone, $\mathcal{M}_2 = \sigma(\mathcal{A}_{i_2})$. On continue par récurrence et on montre alors que

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_p}) = \prod_{k=1}^p \mathbb{P}(A_{i_k})$$

pour tout $p \geq 1$ et $A_{i_k} \in \sigma(\mathcal{A}_{i_k})$ pour $1 \leq k \leq p$. □

Théorème 3.7.4. Soit $(\mathcal{F}_i)_{i \in I}$ une famille de sous-tribus de $\mathcal{F}$ indépendantes. Soit $I = \bigcup_{a \in A} I_a$ une partition de I ($I_a \cap I_b = \emptyset$ pour $a \neq b$) et on note $\mathcal{G}_a = \sigma(\bigcup_{i \in I_a} \mathcal{F}_i)$, alors $(\mathcal{G}_a)_{a \in A}$ sont des tribus indépendantes.

Preuve. La tribu $\mathcal{F}_i$ contient les informations portées par les événements constituant la tribu.

On note

$$\mathcal{C}_a = \{B \in \mathcal{F} : \exists J \text{ fini } \subset I_a : \exists A_j \in \mathcal{F}_j, j \in J, B = \bigcap_{j \in J} A_j\}$$

Pour tout $i \in I_a$, on a $\mathcal{F}_i \subset \mathcal{C}_a$, par conséquent $\bigcup_{i \in I_a} \mathcal{F}_i \subset \mathcal{C}_a \subset \mathcal{G}_a$. Par ailleurs, $\mathcal{C}_a$ est stable par intersection finie. En effet, si $B, B' \in \mathcal{C}_a$, alors on peut écrire

$$B \cap B' = \bigcap_{j \in J} A_j \cap \bigcap_{j \in J'} A'_j = \bigcap_{j \in J \cup J'} A''_j$$

avec $A''_j = A_j$ ou A'_j selon que $j \in J$ ou $j \in J'$. Par conséquent, $\sigma(\mathcal{C}_a) = \mathcal{G}_a$ et par hypothèse les $\mathcal{C}_a$ sont indépendants. On conclut à l'aide de la Propriété 3.7.3. □

Définition 3.7.5. On appelle **tribu engendré par une variable aléatoire** X la plus petite tribu rendant l'application $X : \Omega \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable, i.e.

$$\sigma(X) = \{X^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\}$$

La tribu $\sigma(X)$ est la tribu qui contient toutes les informations liées à la v.a. X .

Définition 3.7.6 (Indépendance de variables aléatoires). Deux variables aléatoires X et Y sont dites **indépendants** si $\sigma(X)$ et $\sigma(Y)$ sont des tribus indépendantes, i.e. si $\forall A, B \in \mathcal{B}(\mathbb{R})$,

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

Une famille de variables aléatoires $(X_i)_{i \in I}$ est **(mutuellement) indépendantes** si leur tribus associées sont indépendantes, i.e. $\forall p \geq 1, \forall i_1, \dots, i_p \in I, A_{i_1}, \dots, A_{i_p} \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}(X_{i_1} \in A_{i_1}, \dots, X_{i_p} \in A_{i_p}) = \prod_{k=1}^p \mathbb{P}(X_{i_k} \in A_{i_k}).$$

3.7.a) Critères d'indépendance

Proposition 3.7.7. Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire. Les propositions suivantes sont équivalentes

1. les variables $X_1, \dots, X_d$ sont mutuellement indépendantes,
2. la loi du vecteur $X = (X_1, \dots, X_d)$ est la loi produit des lois marginales :

$$\mathbb{P}_X = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_d}.$$

3. $\forall x = (x_1, \dots, x_d) \in \mathbb{R}^d$, $F_X(x) := \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d) = \prod_{k=1}^d F_{X_k}(x_k)$,
4. pour toute fonction $(h_k)_{1 \leq k \leq d}$ mesurables bornées (ou positives)

$$\mathbb{E}[h_1(X_1) \dots h_d(X_d)] = \prod_{k=1}^d \mathbb{E}[h_k(X_k)],$$

5. $\varphi_X(t) = \prod_{k=1}^d \varphi_{X_k}(t_k) \quad \forall t \in \mathbb{R}^d$ où $\varphi_X(t) = \mathbb{E}[e^{i\langle t, X \rangle}]$ avec $t = (t_1, \dots, t_d)$,
6. (cas discret) $\mathbb{P}(X_1 = x_1, \dots, X_d = x_d) = \prod_{k=1}^d \mathbb{P}(X_k = x_k) \quad \forall x \in \mathbb{R}^d$,
7. (cas à densité) si la densité de $X = (X_1, \dots, X_d)$ est à variables séparables, on a

$$f_X(x_1, \dots, x_d) = f_{X_1}(x_1) \dots f_{X_d}(x_d).$$

Preuve. 1. $\Leftrightarrow$ 2. : Comme l'ensemble des pavés engendrent $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R})^{\otimes d}$, et est stable par intersection finie, les marginales sont indépendantes si et seulement si pour tout pavé $A = A_1 \times \dots \times A_d$

$$\mathbb{P}(X_1 \in A_1, \dots, X_d \in A_d) = \prod_{k=1}^d \mathbb{P}(X_k \in A_k)$$

Or $\mathbb{P}(X_1 \in A_1, \dots, X_d \in A_d) = \mathbb{P}_X(A)$ et $\prod_{k=1}^d \mathbb{P}(X_k \in A_k) = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_d}(A)$.

Par conséquent, les marginales sont indépendantes si et seulement si $\mathbb{P}_X$ et $\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_d}$ coïncident sur les pavés (i.e sur $\mathcal{B}(\mathbb{R}^d)$). On conclut à l'aide du théorème de la classe monotone (Théorème 3.2.10)

2. $\Leftrightarrow$ 3. : L'ensemble $\{B =]-\infty, x_1] \times \dots \times]-\infty, x_d], x_i \in \mathbb{R}\}$ est stable par intersection finie et engendre $\mathcal{B}(\mathbb{R}^d)$, donc on conclut par le théorème de classe monotone.

2. $\Leftrightarrow$ 4. Si on suppose les variables indépendantes, par le théorème de transfert, puis indépendance + théorème de Fubini,

$$\mathbb{E}[h_1(X_1) \dots h_d(X_d)] = \int_{\mathbb{R}^d} h_1(x_1) \dots h_d(x_d) \mathbb{P}_X(x_1, \dots, x_d) = \prod_{k=1}^d \int h_k(x_k) \mathbb{P}_{X_k}(x_k) = \prod_{k=1}^d \mathbb{E}[h_k(X_k)].$$

Réciproquement, il suffit de prendre $h_k = \mathbb{1}_{A_k}$ avec $A_k \in \mathcal{B}(\mathbb{R})$.

1. $\Leftrightarrow$ 5., 6. : les fonctions caractéristiques s'expriment comme des espérances de fonctionnelles de X et caractérisent la loi.

1. $\Leftrightarrow$ 7. : (en dimension 2) $\mathbb{P}(X \in A, Y \in B) = \sum_{x \in A, y \in B} \mathbb{P}(X = x, Y = y)$ et par Fubini-Tonelli $\mathbb{P}(X \in A) \mathbb{P}(Y \in B) = \sum_{x \in A, y \in B} \mathbb{P}(X = x) \mathbb{P}(Y = y)$.

1. $\Leftrightarrow$ 8. : $\mathbb{P}(X \in A, Y \in B) = \int_{A \times B} f_{(X,Y)}(x, y) dx dy$ et $\mathbb{P}(X \in A) \mathbb{P}(Y \in B) = \int_A f_X(x) dx \int_B f_Y(y) dy$. $\square$

Remarque. Si X et Y sont indépendantes, connaître les lois marginales suffit à connaître la loi du couple (X, Y) , ce qui est faux en général.

Exemple. 1. Soit (X, Y) un couple de densité $f_{(X,Y)}(x, y) = \frac{1}{3} \mathbb{1}_{[0,1] \times [-1,2]}(x, y)$. La fonction est à variables séparables, il y a donc indépendance.

2. Soit (X, Y) un couple de densité $f_{(X,Y)}(x, y) = \frac{1}{3\pi} e^{-\frac{x^2+2xy+5y^2}{6}}$. Il n'y a pas indépendance, car $f_{(X,Y)}(x, y) \neq f_X(x)f_Y(y)$ où $f_X(x) = e^{-4x^2/30}/\sqrt{15\pi/2}$ et $f_Y(y) = e^{-4y^2/6}/\sqrt{3\pi/2}$.

3. Soit (X, Y) un couple de densité $f_{(X,Y)}(x, y) = \sqrt{x/y}$ si $0 < y \leq x < 1$ et 0 sinon. Les variables ne sont pas indépendantes.

Propriété 3.7.8. Soit $(X_i)_{i \in I}$ une famille de v.a.. Les variables sont mutuellement indépendantes si et seulement si pour toute sous-partie finie J de I les variables $(X_j)_{j \in J}$ sont mutuellement indépendantes.

3.7.b) Événements asymptotiques

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité.

Liminf et Limsup d'ensembles

On considère $(A_n)_{n \in \mathbb{N}}$ une suite d'événements. Dans ce qui suit, il faut voir $n \in \mathbb{N}$ comme le temps. On introduit deux événements.

Définition 3.7.9. On définit

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n \geq 1} \bigcap_{k \geq n} A_k \quad \text{et} \quad \limsup_{n \rightarrow \infty} A_n = \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k.$$

On peut mettre en parallèle les notions équivalentes pour les suites : Soit $(u_n)_{n \in \mathbb{N}}$ une suite réelle, alors $\liminf_{n \rightarrow \infty} u_n = \sup_{n \geq 0} \inf_{k \geq n} u_k$ et $\limsup_{n \rightarrow \infty} u_n = \inf_{n \geq 0} \sup_{k \geq n} u_k$ qui correspondent respectivement aux plus petite et plus grande valeur d'adhérence (auss appelé, point d'accumulation) de la suite.

On remarque que

$$\liminf_{n \rightarrow \infty} A_n \subset \limsup_{n \rightarrow \infty} A_n$$

$$\left(\limsup_{n \rightarrow \infty} A_n \right)^c = \liminf_{n \rightarrow \infty} A_n^c.$$

Preuve. soit n fixé de façon arbitraire. On a $\bigcap_{k \geq n} A_k \subset A_p$ pour tout $p > n$, donc $\bigcap_{k \geq n} A_k \subset \bigcup_{p > m} A_p$ pour tout $m \in \mathbb{N}$ et donc $\bigcap_{k \geq n} A_k \subset \bigcap_{m \geq 1} \bigcup_{p > m} A_p = \limsup_{n \rightarrow \infty} A_n$ et on conclut facilement. $\square$

Remarque. Soit $(A_n)_{n \in \mathbb{N}}$ une suite infinie d'événements.

1. Si $\omega \in \liminf_{n \rightarrow \infty} A_n$, alors $\exists n \geq 0, \forall k \geq n, \omega \in A_k$. Par conséquent, $\omega \in \liminf_{n \rightarrow \infty} A_n$ signifie à partir d'un certain rang, tous les A_n sont réalisés (i.e., tous les A_n sont réalisés sauf un nombre fini).
2. Si $\omega \in \limsup_{n \rightarrow \infty} A_n$, alors $\forall n \geq 0, \exists k \geq n, \omega \in A_k$. Par conséquent, $\omega \in \limsup_{n \rightarrow \infty} A_n$ signifie une infinité de A_n sont réalisés.

Propriété 3.7.10. Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements. On a

$$\liminf_{n \rightarrow \infty} \mathbb{1}_{A_n} = \mathbb{1}_{\liminf_{n \rightarrow \infty} A_n} \quad \text{et} \quad \limsup_{n \rightarrow \infty} \mathbb{1}_{A_n} = \mathbb{1}_{\limsup_{n \rightarrow \infty} A_n}$$

Preuve. Il suffit de le faire pour $\limsup$. On a $\omega \in \limsup_{n \rightarrow \infty} A_n$ si $\forall n, \exists k > n$, tel que $\omega \in A_k$, ie $\forall n, \exists k > n \mathbb{1}_{A_k}(\omega) = 1$, soit $\inf_{n \geq 0} \sup_{k \geq n} \mathbb{1}_{A_k}(\omega) = 1$, soit $\limsup_{n \rightarrow \infty} \mathbb{1}_{A_n}(\omega) = 1$. L'équivalence est claire. $\square$

Tribus du futur et tribu asymptotique

Définition 3.7.11. Soit $(\mathcal{A}_n)_{n \geq 0}$ une suite de sous-tribus de $\mathcal{F}$ ($\forall n \geq 0, \mathcal{A}_n \subset \mathcal{F}$ et $\mathcal{A}_n$ tribu). On appelle **tribus du futur**, pour $n \in \mathbb{N}$,

$$\mathcal{A}^n = \sigma\left(\bigcup_{k \geq n} \mathcal{A}_k\right).$$

On appelle **tribu asymptotique** la tribu

$$\mathcal{A}^\infty = \bigcap_{n \geq 0} \mathcal{A}^n.$$

Exemple. Si on lance une infinité de fois une pièce et X_n est le résultat du $n^{\text{ième}}$ lancer. Si $\mathcal{A}_n = \sigma(X_n)$ représente le présent, alors $\mathcal{A}^n = \sigma(X_k : k \geq n)$ contient toutes les informations de ce qui se passe après l'instant n : le futur, et $\mathcal{A}^\infty$ contient les événements asymptotiques.

Exemple. Soit $(X_n)_{n \geq 10}$ une suite de variables aléatoires. On pose $\mathcal{A}_n = \sigma(X_n)$.

- On considère l'événement $A = \{X_n \text{ passe une infinité de fois par } 0\}$. On a

$$A = \{\forall n \geq 1, \exists k \geq n : X_k = 0\} = \bigcap_{n \geq 0} \bigcup_{k \geq n} \{X_k = 0\} = \limsup \{X_n = 0\}.$$

On va montrer que $A \in \mathcal{A}^p$ pour tout $p \geq 0$.

Soit $p \geq 0$ fixé. Pour tout $n \geq p$, on a $\bigcup_{k \geq n} \{X_k = 0\} \in \mathcal{A}^p$. Comme la suite $(\bigcup_{k \geq n} \{X_k = 0\})_{n \geq 0}$ est décroissante, on a $\bigcap_{n \geq 1} \bigcup_{k \geq n} \{X_k = 0\} = \bigcap_{n \geq p} \bigcup_{k \geq n} \{X_k = 0\}$. Par conséquent, pour tout $p \geq 0$,

$$\limsup \{X_n = 0\} \in \mathcal{A}^p,$$

i.e. $\limsup \{X_n = 0\} \in \mathcal{A}^\infty$. L'événement A est un événement asymptotique.

- On considère maintenant l'événement $B = \{\text{les } X_n \text{ sont bornés par } M\}$. On a

$$B = \{\forall n \geq 0, |X_n| \leq M\} = \bigcap_{n \geq 0} \{|X_n| \leq M\} = \{|X_0| \leq M\} \cap \bigcap_{n \geq 1} \{|X_n| \leq M\}.$$

Cet événement n'appartient pas à la tribu asymptotique en général car $\{|X_0| \leq M\} \notin \mathcal{A}^n$ pour $n \geq 1$, car $\mathcal{A}^n = \sigma(X_n, X_{n+1}, \dots)$.

- On considère l'événement $C = \{\text{la suite } (X_n)_{n \geq 1} \text{ converge}\}$. On a

$$\begin{aligned} C &= \{\text{la suite } (X_n)_{n \geq 0} \text{ est de Cauchy}\} \\ &= \{\forall \varepsilon \in \mathbb{Q}^+, \exists N \geq 0, \forall n \geq N, q \geq 0 : |X_n - X_{n+q}| < \varepsilon\} \\ &= \bigcap_{\varepsilon \in \mathbb{Q}^+} \bigcup_{N \geq 0} \bigcap_{\substack{n \geq N \\ q \geq 0}} \{|X_n - X_{n+q}| < \varepsilon\}. \end{aligned}$$

Soit $p \geq 0$ fixé. On a $\{\text{la suite } (X_n)_{n \geq 0} \text{ converge}\} = \{\text{la suite } (X_{n+p})_{n \geq 0} \text{ converge}\}$, par conséquent

$$C = \bigcap_{\varepsilon \in \mathbb{Q}^+} \bigcup_{N \geq p} \bigcap_{\substack{n \geq N \\ q \geq 0}} \{|X_n - X_{n+q}| < \varepsilon\}.$$

Cependant, $\bigcap_{n \geq N, q \geq 0} \{|X_n - X_{n+q}| < \varepsilon\} \in \mathcal{A}^p$ pour tout $N \geq p$, donc

$$\bigcup_{N \geq p} \bigcap_{\substack{n \geq N \\ q \geq 0}} \{|X_n - X_{n+q}| < \varepsilon\} \in \mathcal{A}^p.$$

Par stabilité par intersection dénombrable, $C \in \mathcal{A}^p$ pour tout $p \geq 0$, et donc $C \in \mathcal{A}^\infty$.

Théorème 3.7.12 (Loi du 0-1 de Kolmogorov (Tout ou Rien)). *Soit $(\mathcal{A}_n)_{n \geq 1}$ une suite de sous-tribus indépendantes. Alors la tribu asymptotique est triviale, i.e.*

$$\forall A \in \mathcal{A}^\infty \quad \mathbb{P}(A) = 0 \text{ ou } \mathbb{P}(A) = 1.$$

Conséquence. Soit $(X_n)_{n \geq 1}$ une suite de variables (mutuellement) indépendantes. On pose $\mathcal{A}_n = \sigma(X_n)$. Les tribus sont indépendantes. Alors comme les événements précédents A et C sont asymptotiques, on a $\mathbb{P}(A) = 0$ ou 1 et $\mathbb{P}(C) = 0$ ou 1 . Par conséquent, soit p.s. la suite passe une infinité de fois par 0 soit p.s. la suite passe un nombre fini de fois par 0. De même, soit p.s. la suite $(X_n)_{n \geq 1}$ converge, soit p.s. la suite ne converge pas.

Preuve. On va montrer que $\mathcal{A}^\infty$ est indépendante d'elle-même.

En effet, si $\mathcal{A}^\infty$ est indépendante de $\mathcal{A}^\infty$, alors tout événement $A \in \mathcal{A}^\infty$ est indépendant de lui-même : $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$ et donc $\mathbb{P}(A) = 0$ ou 1 .

Soit $n \geq 0$.

On rappelle que $\mathcal{A}^n = \sigma(\bigcup_{k \geq n} \mathcal{A}_k)$, donc $\mathcal{A}^n$ est indépendant de $\sigma(\bigcup_{p < n} \mathcal{A}_p)$.

Comme $\mathcal{A}^\infty = \bigcap_{m \geq 0} \mathcal{A}^m$, on a $\mathcal{A}^\infty \subset \mathcal{A}^n$. $\mathcal{A}^\infty$ est donc indépendant de $\sigma(\bigcup_{p < n} \mathcal{A}_p)$ pour tout $n \geq 0$.

Par conséquent, $\mathcal{A}^\infty$ est donc indépendant de $\bigcup_{n \geq 1} \sigma(\bigcup_{p < n} \mathcal{A}_p)$.

L'ensemble $\mathcal{C} = \bigcup_{n \geq 0} \sigma(\bigcup_{p < n} \mathcal{A}_p)$ est stable par intersection finie. En effet si $A, B \in \mathcal{C}$, il existe n, m tels que $A \in \sigma(\bigcup_{p < n} \mathcal{A}_p)$ et $B \in \sigma(\bigcup_{p < m} \mathcal{A}_p)$, par conséquent $A \cap B \in \sigma(\bigcup_{p < \max(n, m)} \mathcal{A}_p) \subset \mathcal{C}$.

$\mathcal{A}^\infty$ étant stable par intersection finie, on en déduit que $\mathcal{A}^\infty$ et $\sigma(\mathcal{C})$ sont indépendantes.

Par ailleurs, pour tout $p \geq 0$, $\mathcal{A}_p \subset \mathcal{C}$ et donc $\bigcup_{p \geq 0} \mathcal{A}_p \subset \mathcal{C}$.

Comme $\mathcal{A}^\infty \subset \mathcal{A}^0 = \sigma(\bigcup_{p \geq 0} \mathcal{A}_p) \subset \sigma(\mathcal{C})$, on en déduit que la tribu est indépendante d'elle-même. $\square$

3.7.c) Lemmes de Borel-Cantelli

Ce sont des lemmes très importants en probabilité. L'un est trivial et l'autre est plus subtil.

Théorème 3.7.13 (Lemme de Borel-Cantelli 1). *Soit $(A_n)_{n \geq 0}$ une suite d'événements telle que*

$$\sum_{n=0}^{\infty} \mathbb{P}(A_n) < \infty,$$

alors $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0$. Par conséquent, seul un nombre fini d'événements A_n sont réalisés.

Preuve. On rappelle que $\limsup A_n = \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k$. La suite $(\bigcup_{k \geq n} A_k)_{n \geq 1}$ est décroissante, donc

$$\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{k \geq n} A_k\right).$$

Par ailleurs, $\mathbb{P}(\bigcup_{k \geq n} A_k) \leq \sum_{k \geq n} \mathbb{P}(A_k)$. Comme la série $\sum_{n=0}^{\infty} \mathbb{P}(A_n)$ converge, son reste converge vers 0 et donc $\mathbb{P}(\limsup A_n) = 0$. $\square$

Théorème 3.7.14 (Lemme de Borel-Cantelli 2). *Soit $(A_n)_{n \geq 1}$ une suite d'événements **indépendants** telle que*

$$\sum_{n=0}^{\infty} \mathbb{P}(A_n) = +\infty,$$

alors $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 1$. Par conséquent, ps. une infinité d'événements A_n sont réalisés.

Preuve. On a $\mathbb{P}(\limsup A_n) = 1 - \mathbb{P}(\liminf A_n^c)$. Comme la suite $(\bigcap_{k \geq n} A_k^c)_{n \geq 1}$ est croissante, on a

$$\begin{aligned} \mathbb{P}(\liminf A_n^c) &= \mathbb{P}\left(\bigcup_{n \geq 0} \bigcap_{k \geq n} A_k^c\right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcap_{k \geq n} A_k^c\right) \\ &= \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \mathbb{P}\left(\bigcap_{k=n}^m A_k^c\right) \text{ car } \bigcap_{k \geq n} A_k^c = \bigcap_{m \geq n} \bigcap_{k=n}^m A_k^c \text{ (intersection décroissante)} \\ &= \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \prod_{k=n}^m \mathbb{P}(A_k^c) \text{ par indépendance} \\ &= \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \prod_{k=n}^m (1 - \mathbb{P}(A_k)). \end{aligned}$$

Or comme $0 \leq 1 - x \leq e^{-x}$ pour $x \in [0, 1]$, on a $\prod_{k=n}^m (1 - \mathbb{P}(A_k)) \leq \exp(-\sum_{k=n}^m \mathbb{P}(A_k))$. Comme $\sum_{n \geq 1} \mathbb{P}(A_n) = +\infty$, on a pour tout $n \geq 1$, $\sum_{k=n}^{\infty} \mathbb{P}(A_k) = +\infty$, et donc on en déduit que

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

$\square$

3.7.d) Exercices sur l'indépendance

Exercice 3.7.1. Montrer que la fonction $t \mapsto \cos^n t$ ($t \in \mathbb{R}, n \geq 1$ entier) est une fonction caractéristique.

Exercice 3.7.2. 1. On considère une variable X de loi exponentielle de paramètre $\alpha > 0$.

Vérifier que la loi exponentielle satisfait la propriété d'absence de mémoire, ie

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

2. Soit X une variable positive à densité satisfaisant la propriété d'absence de mémoire. Montrer que X est de loi exponentielle dont on précisera le paramètre.

Exercice 3.7.3. [Couple aléatoire] Soit $f(x, y) = Ce^{-x(1+y^2)}$ définie pour $(x, y) \in \mathbb{R}^+ \times \mathbb{R}$.

1. Calculer C pour que f soit la densité de probabilité d'un couple aléatoire (X, Y) .

2. Exprimer les lois marginales.
3. Les variables X et Y sont-elles indépendantes ?
4. Les variables aléatoires X, Y ont-elles une espérance ? Si oui, les calculer.
5. Reprendre avec $f(x, y) = \frac{\lambda}{x^2 y} \mathbb{1}_{x \geq 1} \mathbb{1}_{\frac{1}{x} \leq y \leq x}$

Exercice 3.7.4. [Min et max] Soient $X_1, \dots, X_n$ des variables aléatoires réelles indépendantes de même fonction de répartition F . On note $U = \min_{1 \leq i \leq n} X_i$ et $V = \max_{1 \leq i \leq n} X_i$.

1. Calculer les fonctions de répartitions de U et V à l'aide de F .
2. Si F admet une densité f , montrer que U et V admettent des densités et les calculer.

Exercice 3.7.5. [Exponentielles] Un appareil électronique est composé de n éléments (montés en série) dont les durées de fonctionnement sont indépendantes et de lois exponentielles.

1. Déterminer la loi de la durée de fonctionnement de l'appareil.
2. Quelle est la probabilité que le composant n^o soit responsable de la défaillance ?

Exercice 3.7.6. [Limites inférieure et supérieure] Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. On considère une suite d'ensembles mesurables $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$ et on note

$$\liminf_n A_n = \bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} A_m, \quad \limsup_n A_n = \bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} A_m.$$

1. Montrer que $\mathbb{P}(\liminf_n A_n) \leq \liminf_{n \rightarrow +\infty} \mathbb{P}(A_n) \leq \limsup_{n \rightarrow +\infty} \mathbb{P}(A_n) \leq \mathbb{P}(\limsup_n A_n)$.
2. On dit que la suite $(A_n)_{n \in \mathbb{N}}$ est convergente si $\liminf_n A_n = \limsup_n A_n$. Montrer que si la suite est croissante (respectivement décroissante) alors elle est convergente et

$$\lim_n A_n = \bigcup_{n \in \mathbb{N}} A_n \quad (\text{respectivement } \lim_n A_n = \bigcap_{n \in \mathbb{N}} A_n).$$

3. Montrer que si la suite $(A_n)_{n \in \mathbb{N}}$ est convergente, on a la propriété de continuité de la mesure : $\mathbb{P}(\lim_n A_n) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n)$.

Exercice 3.7.7. [Comparaison de limite supérieure et inférieure d'ensembles et de v.a.] Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. réelles définies sur un espace mesurable $(\Omega, \mathcal{F})$.

1. Comparer les ensembles $\{\limsup_n X_n > 1\}$, $\limsup_n \{X_n > 1\}$, $\{\limsup_n X_n \geq 1\}$ et $\limsup_n \{X_n \geq 1\}$.
2. Comparer les ensembles $\{\liminf_n X_n > 1\}$, $\liminf_n \{X_n > 1\}$, $\{\liminf_n X_n \geq 1\}$ et $\liminf_n \{X_n \geq 1\}$.

Exercice 3.7.8. [Tribu asymptotique] On considère $(X_n)_{n \geq 1}$ une suite de variables aléatoires définies sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

1. On définit alors la suite $(S_n)_{n \geq 0}$ par $S_0 = 0$ et $S_n = X_1 + \dots + X_n$.
Vérifier que $A_0 = \{X_n \rightarrow 0\}$, $A_1 = \{\limsup_n (X_n) < +\infty\}$ et $A_2 = \{(\frac{S_n}{n}) \text{ converge dans } \mathbb{R}\}$ sont des événements de la tribu asymptotique $\mathcal{F}^\infty$ associée à $(X_n)_{n \geq 1}$.
2. On suppose maintenant que les variables X_n sont indépendantes.
 - (a) Que peut-on dire de $\mathbb{P}((X_n) \text{ converge})$?
 - (b) On suppose que X_n converge $\mathbb{P}$ -presque sûrement vers une variable X . Que peut-on dire de X ?

Exercice 3.7.9. [Lemmes de Borel-Cantelli] Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires de même loi. Soit $p \geq 1$ et $c > 0$.

1. Montrer que si $\mathbb{E}[|X_0|^p] < \infty$, alors $\mathbb{P}(\limsup_n \{|X_n| > cn^{1/p}\}) = 0$.
2. Montrer la réciproque dans le cas indépendant.

Exercice 3.7.10. [Poisson composé] Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes de même loi et dans L^2 . On considère N une variable aléatoire indépendante de loi $\mathcal{P}(\alpha)$ et on pose $Y = \sum_{i=1}^N X_i$ (on convient que la somme est nulle si elle est vide).

1. Calculer l'espérance et la variance de Y .
2. Déterminer la fonction caractéristique de Y .
3. Reconnaître la loi de Y lorsque $X_1 \sim \mathcal{B}(p)$.

Exercice 3.7.11. [Quotient d'exponentielles] Soient X et Y deux variables aléatoires indépendantes de même loi exponentielle $\mathcal{E}(1)$. On note $Z = \frac{X}{Y}$ et $U = \ln(X)$, $V = -\ln(Y)$.

1. Exprimer $\mathbb{P}(Z \leq t)$ à l'aide des variables aléatoires U et V .
2. Montrer que la densité de U est donnée par $f_U(x) = \exp(x - e^x)$.
3. En déduire la densité de V .
4. Montrer que $U + V$ admet pour densité $f_{U+V}(x) = \frac{e^{-x}}{(1+e^{-x})^2}$.
5. Calculer $\mathbb{P}(Z \leq t)$.

Exercice 3.7.12. [Sommes de variables aléatoires] Soient les variables aléatoires indépendantes X et Y . Calculer la loi de la somme $X + Y$ lorsque

1. $X \sim \mathcal{B}(n, p)$ et $Y \sim \mathcal{B}(m, p)$.
2. $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$.
3. $X \sim \mathcal{N}(m, \sigma^2)$ et $Y \sim \mathcal{N}(n, \tau^2)$.
4. $X \sim \text{Gamma}(p, \lambda)$ et $Y \sim \text{Gamma}(q, \lambda)$.

Indice : pensez à la fonction caractéristique.

Exercice 3.7.13. [Box-Muller] On considère un point (aléatoire) M du plan cartésien. On note (X, Y) ses coordonnées cartésiennes et (R, Θ) ses coordonnées polaires. Montrer que X, Y sont indépendantes de même loi $\mathcal{N}(0, 1)$ ssi R^2 et Θ indépendantes avec $R^2 \sim \mathcal{E}(1/2)$ et $\Theta \sim \mathcal{U}([0, 2\pi])$.

Exercice 3.7.14. [Loi beta] Soient $a, b > 0$. Désignons par Z_a, Z_b deux variables aléatoires indépendantes de lois respectivement $\text{Gamma}(a, 1)$, $\text{Gamma}(b, 1)$.

1. Trouver la densité du vecteur aléatoire $(U, V) = \left(Z_a + Z_b, \frac{Z_a}{Z_a + Z_b}\right)$.
2. Les variables aléatoires U et V sont-elles indépendantes ?
(On remarquera que V suit une loi beta $B(a, b)$).

Exercice 3.7.15. [Bernstein] Soient X et Y deux v.a. indépendantes de même loi.

1. Montrer que si X et Y sont deux variables normales centrées réduites alors $X + Y$ et $X - Y$ sont indépendantes.
2. *Théorème de Bernstein.* Réciproquement, on suppose que X et Y sont de carrés intégrables et que $X + Y$ et $X - Y$ sont indépendantes. On veut montrer que X et Y sont deux variables normales. Pour cela :

- (a) Montrer qu'on peut supposer que X et Y sont centrées et de variance 1.
- (b) Montrer que φ , la fonction caractéristique commune de X et de Y satisfait l'égalité : $\varphi(2t) = \varphi(t)^3 \varphi(-t)$.
- (c) En utilisant la continuité de φ en 0, en déduire que φ ne s'annule nulle part.
- (d) On pose $\psi(t) = \varphi(t)/\varphi(-t)$. Montrer que $\psi(2t) = \psi(t)^2$.
- (e) En étudiant le comportement de φ au voisinage de 0, en déduire que $\psi(t) = 1 \ \forall t \in \mathbb{R}$.
- (f) En déduire que $\varphi(t) = e^{-t^2/2}$.
Indication : On pourra montrer que $v(t) = \ln \varphi(t)$ de classe C^2 et calculer $v(0)$, $v'(0)$, $v''(0)$.
 Puis montrer que $t \in \mathbb{R}$ et $p \in \mathbb{N}$, $v''(t) = v''(t/2^p)$.

Exercice 3.7.16. [Retours en zéro, Borel-Cantelli] Soit $p \in]0, 1[$ avec $p \neq 1/2$.

Soit $(X_n)_{n \geq 1}$ une suite de v.a. indépendantes de même loi que X avec $\mathbb{P}(X = 1) = p$ et $\mathbb{P}(X = -1) = 1 - p$.

On considère la marche aléatoire $Y_n = \sum_{i=1}^n X_i$ et $Y_0 = 0$.

On considère $A_n = \{Y_n = 0\}$ un retour en zéro.

1. Que représente l'événement $\limsup A_n$?
2. Montrer que $\mathbb{P}(\limsup A_n) = 0$.

Chapitre 4

Différentes notions de convergence et théorèmes limites

On se place sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On considère une suite de v.a. $(X_n)_{n \geq 0}$ et X une variable aléatoire à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. On note $|\cdot|$ la norme euclidienne dans $\mathbb{R}^d$.

4.1 Rappels sur les ensembles négligeables

Définition 4.1.1. On dit qu'une partie N de Ω est un **ensemble négligeable** pour $\mathbb{P}$ ($\mathbb{P}$ -négligeable) s'il existe $A \in \mathcal{F}$ tel que $N \subset A$ et $\mathbb{P}(A) = 0$.

Propriété 4.1.2. Une union au plus dénombrable d'ensembles négligeables est négligeable.

Preuve. Soit $(N_n)_{n \geq 0}$ des ensembles négligeables. Alors il existe pour tout $n \geq 0$, $A_n \in \mathcal{A}$ tel que $N_n \subset A_n$. Comme $\bigcup_{n \geq 0} N_n \subset \bigcup_{n \geq 0} A_n$ et $\bigcup_{n \geq 0} A_n \in \mathcal{A}$ avec

$$\mathbb{P}\left(\bigcup_{n \geq 0} A_n\right) \leq \sum_{n \geq 0} \mathbb{P}(A_n)$$

et $\mathbb{P}(A_n) = 0$ pour tout $n \geq 0$, donc $\bigcup_{n \geq 0} N_n$ est un ensemble négligeable. $\square$

On dit qu'une propriété **P** est **vraie $\mathbb{P}$ -presque sûrement** ($\mathbb{P}$ -p.s. ou encore juste p.s.) si l'ensemble des points de Ω où elle est fautive est $\mathbb{P}$ -négligeable, i.e. $\{\omega \in \Omega : \mathbf{P} \text{ n'est pas satisfaite}\}$ est $\mathbb{P}$ -négligeable.

Remarque.

- Soit X une v.a. réelle intégrable, i.e. $\mathbb{E}[|X|] < +\infty$. Alors X est finie p.s.
- Une variable X est négligeable ($X = 0$ p.s.) si et seulement si $\mathbb{E}[|X|] = 0$.
- Si X v.a. et $A \in \mathcal{F}$ avec $\mathbb{P}(A) = 0$, alors $\mathbb{E}[X \mathbf{1}_A] = 0$.

4.2 Convergences trajectorielles

4.2.a) Convergence presque-sûre

Cette notion est proche de la convergence simple pour les fonctions, cependant exiger d'avoir $\forall \omega \in \Omega$, $X_n(\omega) \rightarrow X(\omega)$ n'est pas nécessaire en probabilité.

Définition 4.2.1 (Convergence p.s.). Une suite de v.a. $(X_n)_{n \geq 0}$ converge vers X p.s. s'il existe $N \in \mathcal{F}$ tel que $\mathbb{P}(N) = 0$ et

$$\forall \omega \notin N, \quad X_n(\omega) \xrightarrow{n \rightarrow +\infty} X(\omega).$$

On note $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$.

Remarque. 1. Si $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$, alors

$$\mathbb{P}(\lim_{n \rightarrow +\infty} X_n = X) = 1,$$

car, en reprenant les notations de la définition, $N^c \subset \{\omega \in \Omega : \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)\}$ et $\mathbb{P}(N^c) = 1$.

2. $X_n = (X_n^1, \dots, X_n^d) \xrightarrow[n \rightarrow \infty]{p.s.} X = (X^1, \dots, X^d)$ si et seulement si $\forall k \in \{1, \dots, d\}$, $X_n^k \xrightarrow[n \rightarrow \infty]{p.s.} X^k$.

Exemple. On se place sur $(\Omega, \mathcal{F}, \mathbb{P}) = ([-1, 1], \mathcal{B}([-1, 1]), \frac{1}{2} \mathbb{1}_{[-1, 1]} d\lambda)$ et on considère les variables $X_n(\omega) = \omega^n$.

On remarque que pour $\omega \in]-1, 1[$, $(X_n(\omega))_{n \geq 0}$ converge vers 0, pour $\omega = 1$, la suite converge vers 1 et la suite ne converge pas pour $\omega = -1$.

Comme $\lambda(\{-1, 1\}) = 0$, la suite converge ps. et sa limite est ps égale à $X = 0$.

Proposition 4.2.2. $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$ si et seulement si $\forall \varepsilon > 0$, $\mathbb{P}(\limsup_n \{|X_n - X| > \varepsilon\}) = 0$.

Preuve. Supposons que $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$. Il existe $N \in \mathcal{F}$ tel que $\mathbb{P}(N) = 0$ et $\forall \omega \notin N$, $X_n(\omega) \rightarrow X(\omega)$. Soit $\varepsilon > 0$.

Si $\omega \in \limsup_n \{|X_n - X| > \varepsilon\} = \bigcap_{n \geq 0} \bigcup_{k \geq n} \{|X_k - X| > \varepsilon\}$, alors il existe une infinité de n tels que $|X_n(\omega) - X(\omega)| > \varepsilon$, ce qui implique que la suite $X_n(\omega)$ ne converge pas vers $X(\omega)$. Par conséquent $\limsup_n \{|X_n - X| > \varepsilon\} \subset N$, et donc $\mathbb{P}(\limsup_n \{|X_n - X| > \varepsilon\}) = 0$.

Supposons que $\forall \varepsilon > 0$, $\mathbb{P}(\limsup_n \{|X_n - X| > \varepsilon\}) = 0$.

On pose $N = \bigcup_{p \geq 0} \limsup_n \{|X_n - X| > 2^{-p}\}$. N est un ensemble négligeable comme union dénombrable d'ensembles négligeables.

Soit $\omega \notin N$. Alors $\forall p \geq 0$, $\omega \in \liminf_n \{|X_n - X| \leq 2^{-p}\}$. Il existe n_ω tel que $\forall k \geq n_\omega$, $|X_k(\omega) - X(\omega)| \leq 2^{-p}$. Par conséquent, pour $\omega \notin N$, $\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)$. $\square$

Corollaire 4.2.3. Si pour tout $\varepsilon > 0$, $\sum_{n \geq 0} \mathbb{P}(|X_n - X| > \varepsilon) < +\infty$, alors la suite $(X_n)_{n \geq 0}$ converge vers X p.s.

Preuve. Il suffit d'appliquer le premier lemme de Borel-Cantelli avec $A_n = \{|X_n - X| > \varepsilon\}$ pour avoir $\forall \varepsilon > 0$,

$$\mathbb{P}\left(\limsup_n \{|X_n - X| > \varepsilon\}\right) = 0.$$

$\square$

Proposition 4.2.4. On suppose les variables $(X_n)_{n \geq 0}$ mutuellement indépendantes.

La suite $(X_n)_{n \geq 0}$ converge presque sûrement vers 0 si et seulement si $\forall \varepsilon > 0$, $\sum_{n \geq 0} \mathbb{P}(|X_n| > \varepsilon) < +\infty$.

Preuve. Il suffit de montrer l'implication dans un sens (car l'autre découle du résultat précédent). Supposons qu'il existe $\varepsilon > 0$ tel que $\sum_{n \geq 0} \mathbb{P}(|X_n| > \varepsilon) = +\infty$. Les événements $A_n = \{|X_n| > \varepsilon\}$ étant indépendants, par le second lemme de Borel-Cantelli on a $\mathbb{P}(\limsup_n \{|X_n| > \varepsilon\}) = 1$. Donc $(X_n)_{n \geq 0}$ ne converge pas vers 0. Contradiction.

Remarque. La preuve de la proposition précédente ne fonctionne que si la valeur de la limite est nulle (ou une constante) pour pouvoir utiliser l'indépendance des A_n .

□

Proposition 4.2.5. Soit $(\varepsilon_n)_{n \geq 0}$ une suite de réels positifs telle que $\sum_{n \geq 0} \varepsilon_n < +\infty$. Si

$$\sum_{n \geq 0} \mathbb{P}(|X_{n+1} - X_n| > \varepsilon_n) < +\infty,$$

alors $(X_n)_{n \geq 0}$ converge p.s.

Preuve. D'après le premier lemme de Borel-Cantelli, on a $\mathbb{P}(\limsup \{|X_{n+1} - X_n| > \varepsilon_n\}) = 0$.

On remarque que $X_n = X_0 + \sum_{k=0}^{n-1} (X_{k+1} - X_k)$.

Soit $\omega \in \liminf \{|X_{n+1} - X_n| \leq \varepsilon_n\}$, alors il existe $n_\omega \geq 0$ tel que pour tout $k \geq n_\omega$, $|X_{k+1}(\omega) - X_k(\omega)| \leq \varepsilon_k$.

Comme la série de $(\varepsilon_n)_{n \geq 0}$ converge, la série $\sum_{k \geq 0} (X_{k+1}(\omega) - X_k(\omega))$ est absolument convergente. Donc $(X_n(\omega))$ converge pour tout $\omega \in \liminf \{|X_{n+1} - X_n| \leq \varepsilon_n\}$, qui est de probabilité 1. La suite converge p.s. □

On énonce maintenant quelques propriétés immédiates :

Propriété 4.2.6. Soit $(X_n)_{n \geq 0}$ une suite de v.a..

1. Si $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$ et $X_n \xrightarrow[n \rightarrow \infty]{p.s.} Y$, alors $X = Y$ p.s. (unicité p.s. de la limite).
2. Si $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$ et $f : \mathbb{R}^d \rightarrow \mathbb{R}^q$ continue, alors $f(X_n) \xrightarrow[n \rightarrow \infty]{p.s.} f(X)$.
3. Si $X_n = Y_n$ p.s. pour tout $n \geq 0$, alors $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$ si et seulement si $Y_n \xrightarrow[n \rightarrow \infty]{p.s.} X$.

Preuve. En exercice. □

4.2.b) Convergence dans L^p

Soit $p \in [1, \infty[$.

On rappelle que l'espace L^p est l'espace vectoriel des v.a. X telles que $|X|^p$ intégrable, muni de la norme

$$\|X\|_p = (\mathbb{E}[|X|^p])^{1/p}.$$

Définition 4.2.7. Soit $(X_n)_{n \geq 0}$ une suite de v.a. de L^p . On dit que la suite converge dans L^p vers X , noté $X_n \xrightarrow[n \rightarrow \infty]{L^p} X$, si $\lim_{n \rightarrow \infty} \|X_n - X\|_p = 0$, c'est à dire

$$\lim_{n \rightarrow +\infty} \mathbb{E}[|X_n - X|^p] = 0.$$

Remarque. 1. Si $X_n \xrightarrow[n \rightarrow \infty]{L^p} X$ alors $\lim_{n \rightarrow +\infty} \mathbb{E}[|X_n|^p] = \mathbb{E}[|X|^p]$.

2. $X_n = (X_n^1, \dots, X_n^d) \xrightarrow[n \rightarrow \infty]{L^p} X = (X^1, \dots, X^d)$ ssi $X_n^k \xrightarrow[n \rightarrow \infty]{L^p} X^k, \forall k \in \{1, \dots, d\}$.
3. Quand $p = 1$, on parle de convergence en moyenne et quand $p = 2$ on parle de convergence quadratique.
4. Une conséquence de l'inégalité triangulaire est l'unicité de la limite, au sens p.s..

Proposition 4.2.8. Soit $(X_n)_{n \geq 0}$ une suite de v.a. de L^q . Si $X_n \xrightarrow[n \rightarrow \infty]{L^q} X$, alors $X_n \xrightarrow[n \rightarrow \infty]{L^p} X$ pour tout $p \leq q$.

Preuve. Si $p \leq q$, on utilise l'inégalité de Hölder avec $p' = \frac{p}{q}$ et q' tel que $\frac{1}{p'} + \frac{1}{q'} = 1$. On a

$$\mathbb{E}[|X_n - X|^p] = \mathbb{E}[|X_n - X|^p \times 1] \leq \mathbb{E}[|X_n - X|^q]^{p/q} \times 1.$$

□

Liens entre convergence p.s et convergence dans L^p

Théorème 4.2.9 (Théorème de Beppo-Levi, aussi appelé convergence monotone).

Soit $(X_n)_{n \in \mathbb{N}}$ une suite **croissante** de v.a. **positives**. Alors

$$\lim_{n \rightarrow +\infty} \mathbb{E}[X_n] = \mathbb{E}\left[\lim_{n \rightarrow +\infty} X_n\right].$$

Le Théorème de Beppo-Levi est bien adapté aux suites croissantes de v.a., cependant cette hypothèse est très restrictive. Si on enlève l'hypothèse de croissance alors l'existence de limite n'est plus assurée. On peut alors travailler avec la limite inférieure, qui elle est toujours définie.

Théorème 4.2.10 (Lemme de Fatou).

Si X_n suite de v.a. **positives** ($X_n \geq 0 \forall n \geq 0$) alors

$$\mathbb{E}[\liminf_{n \rightarrow \infty} X_n] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_n].$$

Preuve. Posons $X = \liminf_{n \rightarrow +\infty} X_n = \sup_{n \geq 0} \inf_{k \geq n} X_k$.

La suite de v.a. $Y_n = \inf_{k \geq n} X_k$ est croissante. Par le théorème de Beppo-Levi, on a

$$\mathbb{E}\left[\lim_{n \rightarrow +\infty} Y_n\right] = \lim_{n \rightarrow +\infty} \mathbb{E}[Y_n].$$

Cependant, on remarque que $\lim_{n \rightarrow +\infty} Y_n = \sup_{n \geq 0} Y_n = X$ et donc

$$\mathbb{E}[X] = \lim_{n \rightarrow +\infty} \mathbb{E}[Y_n].$$

Par ailleurs, pour tout $n \geq 0$, on a $Y_n \leq X_n$, donc $\mathbb{E}[Y_n] \leq \mathbb{E}[X_n]$. La suite $(\mathbb{E}[X_n])_{n \geq 0}$ n'a pas forcément de limite, mais sa limite inférieure existe toujours. Par passage à la limite inférieure on obtient,

$$\liminf_{n \rightarrow +\infty} \mathbb{E}[Y_n] \leq \liminf_{n \rightarrow +\infty} \mathbb{E}[X_n].$$

Le terme de gauche admettant une limite, on a $\liminf_{n \rightarrow +\infty} \mathbb{E}[Y_n] = \lim_{n \rightarrow +\infty} \mathbb{E}[Y_n]$. Par conséquent, le théorème est prouvé. □

Théorème 4.2.11 (Théorème de convergence dominée).

Soit $p \geq 1$. Soient $(X_n)_{n \geq 0}$ et X des v.a. avec $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$. S'il existe Y v.a. positive telle que

$$\forall n \geq 0, |X_n| \leq Y \quad \mathbb{P} - p.s. \text{ et } Y \in L^p,$$

alors $X_n \in L^p$ et $(X_n)_{n \geq 0}$ converge vers X dans L^p .

Remarque. Ce théorème existe en théorie de la mesure générale. La preuve est identique et il se formule de la façon suivante.

Soit $(E, \mathcal{A}, \mu)$ un espace mesuré, avec μ une mesure positive.

Soient $(f_n)_{n \geq 0}$ et f des fonctions mesurables à valeurs réelles telles que

$$f_n(x) \xrightarrow[n \rightarrow +\infty]{} f(x) \quad \text{pour } \mu\text{-presque tout } x \in E.$$

On suppose qu'il existe une fonction positive g intégrable telle que

$$\forall n \geq 0, |f_n(x)| \leq g(x) \quad \mu\text{-presque tout } x \in E.$$

Alors pour tout $n \geq 0$, f_n est intégrable et

$$\lim_{n \rightarrow +\infty} \int |f_n - f| d\mu = 0.$$

Par conséquent, on a

$$\lim_{n \rightarrow +\infty} \int f_n d\mu = \lim_{n \rightarrow +\infty} \int f d\mu.$$

Preuve. Considérons le cas $p = 1$. Le cas général $p \geq 1$ est une conséquence, en utilisant Y^p à la place de Y .

Supposons tout d'abord que la convergence de $(X_n)_{n \in \mathbb{N}}$ vers X a lieu partout et que l'inégalité $|X_n(\omega)| \leq Y(\omega)$ est vraie pour tout $\omega \in E$.

On pose $Y_n = 2Y - |X_n - X|$. Alors $(Y_n)_{n \in \mathbb{N}}$ est une suite de v.a. positives. D'après le lemme de Fatou, on a

$$2\mathbb{E}[Y] \leq \liminf_{n \rightarrow +\infty} \mathbb{E}[Y_n] = 2\mathbb{E}[Y] - \limsup_{n \rightarrow +\infty} \mathbb{E}[|X_n - X|]$$

Comme $\mathbb{E}[Y] < \infty$, on déduit donc que $\limsup_{n \rightarrow +\infty} \mathbb{E}[|X_n - X|] \leq 0$ et donc

$$\lim_{n \rightarrow +\infty} \mathbb{E}[|X_n - X|] = 0.$$

Comme $|\mathbb{E}[X_n - X]| \leq \mathbb{E}[|X_n - X|]$, on a donc

$$\lim_{n \rightarrow +\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

Dans le cas général, on introduit N ensemble négligeable tel que $\forall \omega \notin N \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)$ et pour $n \geq 0$ N_n ensemble négligeable tel que $\forall \omega \notin N_n |X_n(\omega)| \leq Y(\omega)$. On définit $M = N \cup (\cup_n N_n)$. Comme M est une union dénombrable d'ensembles négligeables, M est négligeable. Soit $A \in \mathcal{F}$ tel que $M \subset A$ et $\mathbb{P}(A) = 0$. On définit $Z_n = X_n \mathbb{1}_{A^c}$ et $Z = X \mathbb{1}_{A^c}$. Pour tout $\omega \in \Omega$ on a

$$\lim Z_n(\omega) = Z(\omega) \quad \text{et} \quad |Z_n(\omega)| \leq Y(\omega).$$

D'après ce qu'on vient de démontrer, on a donc

$$\lim_{n \rightarrow +\infty} \mathbb{E}[|Z_n - Z|] = 0.$$

On conclut en remarquant que $Z_n = X_n$ p.s. et $Z = X$ p.s.. Comme $\mathbb{E}[X_n - X] = \mathbb{E}[Z_n - Z]$, le théorème est prouvé. $\square$

Exemple. Soit $(X_n)_{n \geq 0}$ une suite de v.a. indépendantes de loi

$$\mathbb{P}(X_n = x_n) = p_n, \quad \mathbb{P}(X_n = 0) = 1 - p_n,$$

avec $p_n \in]0, 1[$ et $x_n \geq 1$.

La variables étant indépendantes, la suite $(X_n)_{n \geq 0}$ converge p.s. vers 0 si et seulement si $\forall \varepsilon > 0$, $\sum_{n \geq 0} \mathbb{P}(|X_n| > \varepsilon) < \infty$, soit $\sum_{n \geq 0} p_n < \infty$.

La suite $(X_n)_{n \geq 0}$ converge dans L^p vers 0 si et seulement si $\mathbb{E}[|X_n|^p] = p_n x_n^p \xrightarrow{n \rightarrow \infty} 0$.

Par conséquent,

- si $p_n = \frac{1}{n^2}$ et $x_n = 1$, alors $X_n \xrightarrow[n \rightarrow \infty]{p.s.} 0$ et $X_n \xrightarrow[n \rightarrow \infty]{L^p} 0$ pour tout $p \geq 1$,
- si $p_n = 2^{-n}$ et $x_n = 2^n$, on a $X_n \xrightarrow[n \rightarrow \infty]{p.s.} 0$ et $(X_n)_{n \geq 0}$ ne converge pas dans L^1 ,
- si $p_n = 1/n$ et $x_n = 1$, on a $X_n \xrightarrow[n \rightarrow \infty]{L^1} 0$ et $(X_n)_{n \geq 0}$ ne converge pas ps vers 0,
- si $p_n = 1/n^2$ et $x_n = n$, on a $X_n \xrightarrow[n \rightarrow \infty]{p.s.} 0$, $X_n \xrightarrow[n \rightarrow \infty]{L^1} 0$, mais $(X_n)_{n \geq 0}$ ne converge pas dans L^2 .

4.2.c) Exercices

Exercice 4.2.1. Démontrer la Propriété 4.2.6.

Exercice 4.2.2. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles indépendantes et identiquement distribuées suivant la loi gaussienne $\mathcal{N}(0, 1)$. On rappelle que $\mathbb{E}[e^{X_1}] = e^{1/2}$. On note pour $n \geq 1$

$$S_n = X_1 + \dots + X_n, \quad M_n = \exp\left(S_n - \frac{n}{2}\right).$$

1. (a) Étudier la convergence presque sûre de $(\frac{S_n}{n})_{n \geq 1}$ et préciser la limite.
(b) Que dire de $(M_n)_{n \geq 1}$?
2. La convergence de $(M_n)_{n \geq 1}$ a-t-elle lieu dans L^1 ?

Exercice 4.2.3. On considère la suite de variables aléatoires $(X_n)_{n \geq 1}$ telle que

$$\mathbb{P}(X_n = 0) = 1 - \frac{1}{n^2} \quad \text{et} \quad \mathbb{P}(X_n = c_n) = \frac{1}{n^2},$$

où $c_n > 0$ pour chaque $n = 1, 2, \dots$

Montrer que $\mathbb{P}\left(\sum_{n \geq 1} X_n < \infty\right) = 1$.

(Indice : Utiliser le lemme de Borel-Cantelli.)

Exercice 4.2.4. [Théorème de Weierstrass]

Le théorème de Weierstrass est le suivant

Soit f une fonction continue sur $[a, b]$. Il existe une suite (P_n) de polynômes convergeant uniformément vers f sur $[a, b]$.

Soit f une fonction continue sur $[a, b]$.

1. Soit $x \in \mathbb{R}$ et X_n une suite de v.a. à valeurs dans $[a, b]$, d'espérance x et de variance $\sigma_n^2(x)$. Montrer que si $\sigma_n^2(x)$ converge uniformément vers 0 sur l'intervalle $[a, b]$, alors $\mathbb{E}[f(X_n)] \rightarrow f(x)$ quand $n \rightarrow \infty$ uniformément sur $[a, b]$.
2. On considère f continue sur $[0, 1]$. Dédurre le théorème de Weierstrass de la question précédente. Indic : considérer $X_n = \frac{1}{n}(Y_1 + \dots + Y_n)$ avec Y_i indépendantes de loi $\mathcal{B}(x)$.
3. Comment étendre le résultat précédent à n'importe quel intervalle $[a, b]$?

Exercice 4.2.5. [Loi forte des grands nombres dans le cas L^4]

1. Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles indépendantes et de même loi telles que $\mathbb{E}[X_1] = 0$ et $\mathbb{E}[X_1^4] < \infty$.
Montrer que $\frac{1}{n}(X_1 + \dots + X_n)$ converge presque sûrement vers 0.
2. Montrer que dire si les v.a. X_n ne sont plus de même loi mais $\sup_{n \geq 1} \mathbb{E}[X_n^4] < \infty$?

4.2.d) Convergence en probabilité

Définition 4.2.12. Une suite de v.a. $(X_n)_{n \geq 0}$ converge en probabilité vers X , noté $X_n \xrightarrow[n \rightarrow \infty]{\text{proba}} X$ ou $X_n \xrightarrow{\mathbb{P}} X$, si pour tout $\varepsilon > 0$,

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

Proposition 4.2.13. $X_n = (X_n^1, \dots, X_n^d) \xrightarrow[n \rightarrow \infty]{\text{proba}} X = (X^1, \dots, X^d)$ ssi $X_n^k \xrightarrow[n \rightarrow \infty]{\text{proba}} X^k, \forall k \in \{1, \dots, d\}$.

Preuve. Si $X_n \xrightarrow[n \rightarrow \infty]{\text{proba}} X$, alors pour tout $\varepsilon > 0$,

$$\mathbb{P}(|X_n^k - X^k| > \varepsilon) \leq \mathbb{P}(|X_n - X| > \varepsilon)$$

$$\text{car } |X_n^k - X^k| \leq |X_n - X| = \left(\sum_{k=1}^d |X_n^k - X^k|^2 \right)^{1/2}.$$

Supposons maintenant que $X_n^k \xrightarrow[n \rightarrow \infty]{\text{proba}} X^k, \forall k \in \{1, \dots, d\}$. On a

$$\{|X_n - X| > \varepsilon\} \subset \bigcup_{1 \leq k \leq d} \{|X_n^k - X^k| > \varepsilon/\sqrt{d}\}$$

En effet, si pour tout $k \in \{1, \dots, d\}$, $|X_n^k - X^k| \leq \varepsilon/\sqrt{d}$ alors $|X_n - X| \leq \varepsilon$.
Par conséquent,

$$\begin{aligned} \mathbb{P}(\{|X_n - X| > \varepsilon\}) &\leq \mathbb{P}\left(\bigcup_{1 \leq k \leq d} \{|X_n^k - X^k| > \varepsilon/\sqrt{d}\}\right) \\ &\leq \sum_{k=1}^d \mathbb{P}(|X_n^k - X^k| > \varepsilon/\sqrt{d}). \end{aligned}$$

Par hypothèse, on en déduit donc que $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$. □

Remarque. Si $X_n \xrightarrow[n \rightarrow \infty]{\text{proba}} X$ et $X_n \xrightarrow[n \rightarrow \infty]{\text{proba}} Y$, alors $X = Y$ ps.

Preuve. Soit $\varepsilon > 0$, $|X - Y| \leq |X - X_n| + |X_n - Y|$ et donc on a

$$\mathbb{P}(|X - Y| > \varepsilon) \leq \lim_{n \rightarrow \infty} (\mathbb{P}(|X_n - X| > \varepsilon/2) + \mathbb{P}(|X_n - Y| > \varepsilon/2)) = 0.$$

Comme $\mathbb{P}(X \neq Y) = \mathbb{P}(\bigcup_{k \geq 1} |X - Y| > 1/k) = 0$, on en déduit $X = Y$ ps. $\square$

Proposition 4.2.14. Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}^q$ continue. Si $X_n \xrightarrow[n \rightarrow \infty]{\text{proba}} X$, alors $f(X_n) \xrightarrow[n \rightarrow \infty]{\text{proba}} f(X)$.

Par conséquent, si $X_n \xrightarrow[n \rightarrow \infty]{\text{proba}} X$ et $Y_n \xrightarrow[n \rightarrow \infty]{\text{proba}} Y$, alors $X_n + Y_n \xrightarrow[n \rightarrow \infty]{\text{proba}} X + Y$ et $X_n Y_n \xrightarrow[n \rightarrow \infty]{\text{proba}} XY$, si $(\alpha_n)_{n \geq 0}$ suite de nombres réels tels que $\alpha_n \rightarrow \alpha$, alors $X_n + \alpha_n Y_n \xrightarrow[n \rightarrow \infty]{\text{proba}} X + \alpha Y$, etc....

Preuve. On utilise l'uniforme continuité de f sur les compacts.

Soit $\varepsilon > 0$ et $a > 0$ fixés. On restreint la fonction f au sous-ensemble compact de $\mathbb{R}^d : \{x \in \mathbb{R}^d : |x| \leq a\}$. Il existe $\eta = \eta(\varepsilon, a) > 0$ tel que $\forall x, y$ avec $|x| \leq a$, $|x - y| \leq \eta$, on a $|f(x) - f(y)| \leq \varepsilon$. Par conséquent

$$\{|X| \leq a\} \cap \{|X_n - X| \leq \eta\} \subset \{|f(X_n) - f(X)| \leq \varepsilon\}.$$

Par passage au complémentaire, on a

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

Par conséquent

$$\limsup_{n \rightarrow \infty} \mathbb{P}(|f(X_n) - f(X)| > \varepsilon) \leq \mathbb{P}(|X| > a) + \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \eta) = \mathbb{P}(|X| > a).$$

Comme X est une va à valeurs dans $\mathbb{R}^d$ (finie), par convergence monotone $\lim_{a \rightarrow \infty} \mathbb{P}(|X| > a) = 0$ et donc $\mathbb{P}(|f(X_n) - f(X)| > \varepsilon) \rightarrow 0$ quand $n \rightarrow \infty$. $\square$

Proposition 4.2.15 (Convergence ps ou L^1 vers convergence en probabilité).

Si $(X_n)_{n \geq 0}$ converge vers X p.s ou dans L^1 alors $(X_n)_{n \geq 0}$ converge en probabilité vers X .

Preuve. Soit $\varepsilon > 0$.

Supposons que $X_n \xrightarrow[n \rightarrow \infty]{p.s} X$. On a

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{E}[\mathbf{1}_{|X_n - X| > \varepsilon}].$$

La variable $\mathbf{1}_{|X_n - X| > \varepsilon}$ est dominée par 1 et converge ps. vers 0, on conclut par convergence dominée.

Supposons que $X_n \xrightarrow[n \rightarrow \infty]{L^1} X$. Alors par l'inégalité de Markov,

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \frac{\mathbb{E}[|X_n - X|]}{\varepsilon},$$

qui tend vers 0 quand $n \rightarrow +\infty$. $\square$

Exemple. Soit $(X_n)_{n \geq 0}$ une suite de v.a. indépendantes de loi

$$\mathbb{P}(X_n = x_n) = p_n, \quad \mathbb{P}(X_n = 0) = 1 - p_n,$$

avec $p_n \in]0, 1[$ et $x_n \geq 1$. On a $\mathbb{P}(|X_n| > \varepsilon) = p_n \forall \varepsilon \in]0, 1[$.

Pour $p_n = 1/n$ et $x_n = n$, on a $X_n \xrightarrow[n \rightarrow \infty]{\text{proba}} 0$ mais X_n ne converge ni ps ni dans L^1 .

Proposition 4.2.16 (Convergence en probabilité vers convergence ps).

Si $(X_n)_{n \geq 0}$ converge en probabilité vers X alors il existe une sous-suite $(X_{n_k})_{k \geq 0}$ qui converge ps vers X .

Preuve. Pour tout $k \geq 0$, on a $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > 2^{-(k+1)}) = 0$. Il existe donc $n_k \in \mathbb{N}$ tel que pour tout $n \geq n_k$, $\mathbb{P}(|X_n - X| > 2^{-(k+1)}) \leq 2^{-k+1}$.

On peut supposer que la suite est strictement croissante, car si n_k convient, alors $n_k + 1$ convient aussi. Par conséquent, pour tout $k \in \mathbb{N}$,

$$\mathbb{P}(|X_{n_{k+1}} - X_{n_k}| > 2^{-k}) \leq \mathbb{P}(|X_{n_{k+1}} - X| > 2^{-(k+1)}) + \mathbb{P}(|X_{n_k} - X| > 2^{-(k+1)}) \leq 2^{-k}.$$

La série $\sum \mathbb{P}(|X_{n_{k+1}} - X_{n_k}| > 2^{-k})$ converge, donc d'après la Proposition 4.2.5, la suite $(X_{n_k})_{k \geq 0}$ converge ps.

On note Y sa limite. Comme $X_{n_k} \xrightarrow[n \rightarrow \infty]{p.s} Y$ on a $X_{n_k} \xrightarrow[n \rightarrow \infty]{\text{proba}} Y$. Or $X_{n_k} \xrightarrow[n \rightarrow \infty]{\text{proba}} X$ car sous-suite de $(X_n)_{n \geq 0}$. Par conséquent, $Y = X$ ps. $\square$

Remarque. $X_n \xrightarrow[n \rightarrow \infty]{\text{proba}} X$ si et seulement si de toute sous-suite on peut extraire une sous-suite qui converge ps vers X .

Preuve. Si $X_n \xrightarrow[n \rightarrow \infty]{\text{proba}} X$, il suffit d'utiliser la proposition précédente. Réciproquement, on raisonne par l'absurde. Si X_n ne converge pas en proba vers X , alors il existe $\varepsilon > 0$ et $\alpha > 0$ tel que $\forall n \geq 0 \exists n_k > n$ tel que $\mathbb{P}(|X_{n_k} - X| > \varepsilon) \geq \alpha$. On construit ainsi une suite strictement croissante d'entiers telle que $\mathbb{P}(|X_{n_k} - X| > \varepsilon) \geq \alpha$. Par hypothèse, on peut extraire une sous-suite $X_{n'_k}$ qui converge ps vers X et donc en probabilité. Mais c'est incompatible avec $\forall k \in \mathbb{N} \mathbb{P}(|X_{n_k} - X| > \varepsilon) \geq \alpha$. $\square$

4.2.e) Uniforme intégrabilité

Définition 4.2.17 (Uniforme intégrabilité). Une suite de v.a. $(X_n)_{n \geq 1}$ intégrables est dite uniformément intégrable (ou équi-intégrable) si

$$\lim_{a \rightarrow \infty} \sup_{n \geq 1} \mathbb{E}[|X_n| \mathbf{1}_{|X_n| > a}] = 0.$$

Remarque. • Une suite $(X_n)_{n \geq 1}$ bornée par une v.a positive Z intégrable est uniformément intégrable.

- Un n -uplet de v.a. $(X_k)_{k \in \{1, \dots, n\}}$ intégrables est uniformément intégrable.
- Considérons deux suites $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ avec $|X_n| \leq |Y_n|$ pour tout $n \in \mathbb{N}$. Si $(Y_n)_{n \in \mathbb{N}}$ est uniformément intégrable, alors $(X_n)_{n \in \mathbb{N}}$ est uniformément intégrable.

Preuve.

- Supposons que $|X_n| \leq Z$, alors $|X_n| \mathbf{1}_{|X_n| > a} \leq Z \mathbf{1}_{Z > a}$ et donc

$$\sup_{n \geq 0} \mathbb{E}[|X_n| \mathbf{1}_{|X_n| > a}] \leq \mathbb{E}[Z \mathbf{1}_{Z > a}].$$

On conclut en utilisant le théorème de convergence dominée.

- Le n -uplet est dominée par la variable $Z = \sum_{k=1}^n |X_k|$ qui est intégrable.
- On utilise le fait que $|X_n| \mathbb{1}_{|X_n| > a} \leq |Y_n| \mathbb{1}_{|Y_n| > a}$ pour tout $n \in \mathbb{N}$.

□

Définition 4.2.18. On dit qu'une suite $(X_n)_{n \geq 0}$ est bornée dans L^p s'il existe $M > 0$ tel que

$$\sup_{n \geq 0} \mathbb{E}[|X_n|^p] \leq M.$$

Remarque. Si la suite $(X_n)_{n \geq 1}$ est uniformément intégrable, alors elle est bornée dans L^1 .

Par contre, avoir une suite bornée dans L^1 ne suffit pas pour qu'elle soit uniformément intégrable. En effet,

$$\text{si } X_n = \begin{cases} n & \text{avec probabilité } 1/n \\ 0 & \text{avec probabilité } 1 - 1/n \end{cases} \text{ on a } \mathbb{E}[|X_n|] = 1 \text{ et } \sup_n \mathbb{E}[|X_n| \mathbb{1}_{|X_n| > a}] = 1 \text{ pour tout } a \geq 0.$$

Preuve. En effet, soit $\varepsilon > 0$, il existe $a > 0$ tel que $\sup_{n \geq 0} \mathbb{E}[|X_n| \mathbb{1}_{|X_n| > a}] \leq \varepsilon$. Par conséquent,

$$\sup_{n \geq 0} \mathbb{E}[|X_n|] = \sup_{n \geq 0} (\mathbb{E}[|X_n| \mathbb{1}_{|X_n| \leq a}] + \mathbb{E}[|X_n| \mathbb{1}_{|X_n| > a}]) \leq a + \varepsilon.$$

□

Proposition 4.2.19. Si $(X_n)_{n \geq 1}$ est une suite de variable de L^1 telle qu'elle soit bornée dans $L^{1+\varepsilon}$ pour un $\varepsilon > 0$, alors $(X_n)_{n \geq 1}$ est uniformément intégrable.

Preuve. On remarque que $|X_n| \mathbb{1}_{|X_n| > a} \leq |X_n| \frac{|X_n|^\varepsilon}{a^\varepsilon} \mathbb{1}_{|X_n| > a} \leq \frac{1}{a^\varepsilon} |X_n|^{1+\varepsilon}$. Par conséquent,

$$\sup_{n \in \mathbb{N}} \mathbb{E}[|X_n| \mathbb{1}_{|X_n| > a}] \leq \frac{1}{a^\varepsilon} \sup_{n \in \mathbb{N}} \mathbb{E}[|X_n|^{1+\varepsilon}],$$

qui converge bien vers 0 quand $a \rightarrow \infty$.

□

Remarque. Si $X \in L^1$, alors $\forall \varepsilon > 0, \exists \eta > 0$ tel que si $A \in \mathcal{F}$ avec $\mathbb{P}(A) < \eta$ alors $\mathbb{E}[|X| \mathbb{1}_A] < \varepsilon$.

Preuve. En effet, par convergence dominée, à $\varepsilon > 0$ fixé, on a pour a assez grand $\mathbb{E}[|X| \mathbb{1}_{|X| > a}] < \varepsilon/2$ et donc

$$\begin{aligned} \mathbb{E}[|X| \mathbb{1}_A] &= \mathbb{E}[|X| \mathbb{1}_{A \cap \{|X| \leq a\}}] + \mathbb{E}[|X| \mathbb{1}_{A \cap \{|X| > a\}}] \\ &\leq a \mathbb{P}(A) + \varepsilon/2. \end{aligned}$$

Par conséquent, pour tout $A \in \mathcal{F}$ avec $\mathbb{P}(A) < \varepsilon/(2a)$, on a $\mathbb{E}[|X| \mathbb{1}_A] < \varepsilon$.

□

Ce résultat se généralise pour les suites uniformément intégrables.

Proposition 4.2.20 (Critère d'uniforme intégrabilité). Une suite de v.a. $(X_n)_{n \geq 1}$ est uniformément intégrable ssi

- $\forall \varepsilon > 0, \exists \eta > 0$ tel que pour $A \in \mathcal{F}$ avec $\mathbb{P}(A) < \eta$ on a $\mathbb{E}[|X_n| \mathbb{1}_A] < \varepsilon$ pour tout $n \geq 1$, et
- $\sup_{n \geq 1} \mathbb{E}[|X_n|] < \infty$.

Preuve. $\Rightarrow$] Supposons $(X_n)_{n \geq 1}$ est uniformément intégrable. On sait déjà que la suite est bornée dans L^1 .

Soit $\varepsilon > 0$, il existe $a > 0$ tel que $\sup_{n \geq 1} \mathbb{E}[|X_n| \mathbf{1}_{|X_n| > a}] < \varepsilon/2$.

Soit $A \in \mathcal{F}$, on a pour tout $n \geq 1$

$$\begin{aligned} \mathbb{E}[|X_n| \mathbf{1}_A] &= \mathbb{E}[|X_n| \mathbf{1}_{A \cap \{|X_n| \leq a\}}] + \mathbb{E}[|X_n| \mathbf{1}_{A \cap \{|X_n| > a\}}] \\ &\leq a\mathbb{P}(A) + \varepsilon/2. \end{aligned}$$

Il suffit de choisir $\eta = \varepsilon/(2a)$.

[$\Leftarrow$ Soit $\varepsilon > 0$ et η tel que pour tout $A \in \mathcal{F}$ avec $\mathbb{P}(A) < \eta$ on a $\mathbb{E}[|X_n| \mathbf{1}_A] < \varepsilon \forall n \geq 1$.

On pose $M = \sup_{n \geq 1} \mathbb{E}[|X_n|]$. D'après l'inégalité de Markov, on a

$$\mathbb{P}(|X_n| > a) \leq \frac{M}{a}.$$

Par conséquent, pour a assez grand on a $\mathbb{P}(|X_n| > a) < \eta$ et donc pour tout $n \geq 1$

$$\mathbb{E}[|X_n| \mathbf{1}_{|X_n| > a}] < \varepsilon.$$

La suite est bien uniformément intégrable. □

Proposition 4.2.21 (Critère de Vitali). *Soit $(X_n)_{n \geq 1}$ une suite de v.a. intégrables. Il y a équivalence entre*

1. $(X_n)_{n \geq 1}$ converge dans L^1 ,
2. $(X_n)_{n \geq 1}$ est uniformément intégrable et $(X_n)_{n \geq 1}$ converge en probabilité.

Preuve. (partielle) 1. $\Rightarrow$ 2. On sait déjà que la convergence dans L^1 implique la convergence en proba. Notons X_0 sa limite qui est dans L^1 . On va utiliser la proposition précédente pour montrer que la suite est uniformément intégrable. Soit $\varepsilon > 0$, il existe $n_0 \geq 1$ tel que pour tous $n \geq n_0$, $\mathbb{E}[|X_n - X_0|] < \varepsilon/2$.

Soit $A \in \mathcal{F}$, alors pour $n \geq n_0$, on a

$$\begin{aligned} \mathbb{E}[|X_n| \mathbf{1}_A] &\leq \mathbb{E}[|X_n - X_0| \mathbf{1}_A] + \mathbb{E}[|X_0| \mathbf{1}_A] \\ &\leq \frac{\varepsilon}{2} + \mathbb{E}[|X_0| \mathbf{1}_A] \end{aligned}$$

Par conséquent, $\sup_{n \geq n_0} \mathbb{E}[|X_n| \mathbf{1}_A] \leq \frac{\varepsilon}{2} + \mathbb{E}[|X_0| \mathbf{1}_A]$ et donc

$$\sup_{n \geq 1} \mathbb{E}[|X_n| \mathbf{1}_A] \leq \frac{\varepsilon}{2} + \sup_{0 \leq k \leq n_0} \mathbb{E}[|X_k| \mathbf{1}_A]$$

En prenant $A = \Omega$, on obtient $\sup_{n \geq 1} \mathbb{E}[|X_n|] < \infty$ et comme $(|X_k|)_{0 \leq k \leq n_0}$ est uniformément intégrable, il existe $\eta > 0$ tel que si $\mathbb{P}(A) < \eta$ alors $\sup_{0 \leq k \leq n_0} \mathbb{E}[|X_k| \mathbf{1}_A] < \varepsilon/2$. On en déduit que (X_n) est uniformément intégrable.

2. $\Rightarrow$ 1. Comme (X_n) converge en probabilité vers X , on peut extraire une sous-suite (n_k) qui converge p.s.. D'après le lemme de Fatou, on a

$$\mathbb{E}[|X|] = \mathbb{E}[\liminf |X_{n_k}|] \leq \liminf \mathbb{E}[|X_{n_k}|] \leq \sup_n \mathbb{E}[|X_n|] < \infty.$$

Donc $X \in L^1$.

Soit $\varepsilon > 0$. On a

$$\begin{aligned}\mathbb{E}[|X_n - X|] &\leq \varepsilon + \mathbb{E}[|X_n - X| \mathbf{1}_{|X_n - X| > \varepsilon}] \\ &\leq \varepsilon + \mathbb{E}[|X_n| \mathbf{1}_{|X_n - X| > \varepsilon}] + \mathbb{E}[|X| \mathbf{1}_{|X_n - X| > \varepsilon}]\end{aligned}$$

Comme $(X, X_n)_{n \geq 1}$ est uniformément intégrable, il existe $\eta > 0$ tel que si $\mathbb{P}(A) < \eta$, on a pour tout $n \geq 1$,

$$\mathbb{E}[|X_n| \mathbf{1}_A] < \varepsilon \quad \text{et} \quad \mathbb{E}[|X| \mathbf{1}_A] < \varepsilon.$$

Comme la suite converge en probabilité, il existe $n_0 \geq 1$ tel que pour tout $n \geq n_0$, $\mathbb{P}(|X_n - X| > \varepsilon) < \eta$. Par conséquent pour tout $n \geq n_0$, $\mathbb{E}[|X_n - X|] < 3\varepsilon$. La suite converge bien dans L^1 . $\square$

4.2.f) Critères de convergence

Les critères de type Cauchy sont parfois bien utiles pour établir la convergence d'une suite sans avoir à expliciter la valeur de la limite.

Proposition 4.2.22. Soit $(X_n)_{n \geq 1}$ une suite de v.a. à valeurs dans $\mathbb{R}^d$. Alors

1. $(X_n)_{n \geq 1}$ converge presque sûrement ssi $\forall \varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\sup_{k \geq 1} |X_{n+k} - X_n| > \varepsilon\right) = 0.$$

2. $(X_n)_{n \geq 1}$ converge en probabilité ssi $\forall \varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \sup_{k \geq 1} \mathbb{P}(|X_{n+k} - X_n| > \varepsilon) = 0.$$

3. $(X_n)_{n \geq 1}$ une famille de v.a. de L^p converge dans L^p ssi

$$\lim_{n \rightarrow \infty} \sup_{k \geq 1} \mathbb{E}(|X_{n+k} - X_n|^p) = 0.$$

Preuve. 3. Le dernier point est juste le théorème de Riesz-Fisher : l'espace L^p est complet.

1. Regardons la convergence p.s.. Si $(X_n)_{n \geq 1}$ converge p.s., la suite est ps. de Cauchy. Par conséquent, $\sup_{k \geq 1} |X_{n+k} - X_n|$ converge vers 0 p.s. quand $n \rightarrow +\infty$ et donc en probabilité, d'où le résultat. Réciproquement, on pose $Y_n = \sup_{p, q \geq n} |X_p - X_q|$ qui est décroissante en n et minorée par 0, donc converge ps. vers une v.a. notée Y . Par ailleurs d'après l'inégalité triangulaire $Y_n \leq 2 \sup_{k \geq 1} |X_{n+k} - X_n|$ ce qui implique que $(Y_n)_{n \geq 1}$ converge en probabilité vers 0 et donc $Y = 0$ ps. La suite $(X_n)_{n \geq 1}$ est par conséquent p.s. de Cauchy et donc converge p.s. par complétude de $\mathbb{R}^d$.

2. Regardons maintenant la convergence en probabilité. Si $(X_n)_{n \geq 1}$ converge en probabilité, comme $|X_{n+k} - X_n| \leq |X_{n+k} - X| + |X_n - X|$, on a

$$\mathbb{P}(|X_{n+k} - X_n| > \varepsilon) \leq \mathbb{P}(|X_{n+k} - X| > \varepsilon/2) + \mathbb{P}(|X_n - X| > \varepsilon/2) \leq 2 \sup_{k \geq n} \mathbb{P}(|X_k - X| > \varepsilon/2)$$

Le majorant convergeant vers 0, on a la première implication. Supposons la réciproque, on construit une suite strictement croissante (n_q) d'entiers telle que

$$\sup_{k \geq 1} \mathbb{P}(|X_{n_q+k} - X_{n_q}| > 2^{-q}) \leq 2^{-q}.$$

En particulier, $\mathbb{P}(|X_{n_q+1} - X_{n_q}| > 2^{-q}) \leq 2^{-q}$ et par conséquent la sous-suite (X_{n_q}) converge p.s. et donc en probabilité vers X . On a alors,

$$\begin{aligned} \mathbb{P}(|X_q - X| > \varepsilon) &\leq \mathbb{P}(|X_q - X_{n_q}| > \varepsilon/2) + \mathbb{P}(|X_{n_q} - X| > \varepsilon/2) \\ &\leq \sup_{k \geq 1} \mathbb{P}(|X_{q+k} - X_q| > \varepsilon/2) + \mathbb{P}(|X_{n_q} - X| > \varepsilon/2). \end{aligned}$$

Les deux termes de droite convergent vers 0 et donc la suite converge en probabilité. $\square$

4.3 Convergence en loi

Soit $(X_n)_{n \geq 0}$ une suite de v.a et X une v.a. à valeurs dans $\mathbb{R}^d$

On note $\mathcal{C}_b$ l'ensemble des fonctions continues bornées $\mathbb{R}^d \rightarrow \mathbb{R}$ et $\mathcal{C}_c$ l'ensemble des fonctions continues à support compact $\mathbb{R}^d \rightarrow \mathbb{R}$.

Définition 4.3.1. La suite $(X_n)_{n \geq 0}$ converge en loi vers X si pour tout fonction $f \in \mathcal{C}_b$ continue bornée

$$\lim_{n \rightarrow +\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)].$$

On note alors $X_n \xrightarrow[n \rightarrow \infty]{\text{Loi}} X$.

On dit aussi que la suite de mesure (μ_n) converge étroitement vers μ , où $\mu_n = \mathbb{P}_{X_n}$ et $\mu = \mathbb{P}_X$:

$$\forall f \in \mathcal{C}_b, \quad \int f d\mu_n \xrightarrow[n \rightarrow +\infty]{} \int f d\mu.$$

Remarque. On peut définir la convergence en loi de $(X_n)_{n \geq 0}$ vers X sans que les variables soient définies sur le même espace de probabilité.

Proposition 4.3.2 (Admis). Soit $\mathcal{H} \subset \mathcal{C}_b$ tel que $\mathcal{C}_c \subset \overline{\mathcal{H}}$ (adhérence pour la norme $\|\cdot\|_\infty$).

On a $X_n \xrightarrow[n \rightarrow \infty]{\text{Loi}} X$ si et seulement si

$$\forall f \in \mathcal{H}, \quad \lim_{n \rightarrow +\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)].$$

Remarque. La continuité de f est essentielle, dans la convergence en loi. En général, même si (X_n) converge en loi vers X , $\mathbb{P}(X_n \in A) = \mathbb{E}[\mathbf{1}_A(X_n)]$ ne converge pas vers $\mathbb{P}(X \in A)$, pour $A \in \mathcal{F}$.

Théorème 4.3.3. On a équivalence entre

1. $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X ,
2. Pour tout fermé F , $\limsup \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$,
3. Pour tout ouvert O , $\liminf \mathbb{P}(X_n \in O) \geq \mathbb{P}(X \in O)$,
4. Pour tout borélien $B \in \mathcal{B}(\mathbb{R}^d)$ tel que $\mathbb{P}(X \in \overline{B} \setminus \overset{\circ}{B}) = 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n \in B) = \mathbb{P}(X \in B).$$

ssi $\forall t \in \mathbb{R}$ **point de continuité** de la fonction de répartition F_X , $\lim_{n \rightarrow \infty} F_{X_n}(t) = F_X(t)$.

Preuve.(Partielle)

1 $\Rightarrow$ 2 Soit F un fermé de $\mathbb{R}^d$. Pour $k \geq 1$, on considère les fonctions

$$f_k(x) = \frac{1}{(1 + d(x, F))^k}$$

Ces fonctions sont continues, bornées par 1. La suite décroît vers $\mathbb{1}_F(x)$. Par conséquent, pour $k \geq 1$,

$$\limsup_{n \rightarrow \infty} \mathbb{P}(X_n \in F) = \limsup_{n \rightarrow \infty} \mathbb{E}[\mathbb{1}_F(X_n)] \leq \limsup_{n \rightarrow \infty} \mathbb{E}[f_k(X_n)] = \mathbb{E}[f_k(X)].$$

Par ailleurs, par convergence dominée, $\mathbb{E}[f_k(X)]$ converge vers $\mathbb{E}[\mathbb{1}_F(X)] = \mathbb{P}(X \in F)$ quand $k \rightarrow \infty$. On en déduit que $\limsup_{n \rightarrow \infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$.

2 $\Leftrightarrow$ 3 Par passage au complémentaire.

2 + 3 $\Rightarrow$ 4 On a $\mathring{B} \subset B \subset \overline{B}$, alors

$$\begin{aligned} \mathbb{P}(X \in \mathring{B}) &\leq \liminf \mathbb{P}(X_n \in \mathring{B}) \leq \liminf \mathbb{P}(X_n \in B) \\ &\leq \limsup \mathbb{P}(X_n \in B) \leq \limsup \mathbb{P}(X_n \in \overline{B}) \leq \mathbb{P}(X \in \overline{B}) \end{aligned}$$

Dans le cas où $\mathbb{P}(X \in \overline{B} \setminus \mathring{B}) = 0$, on en déduit que

$$\mathbb{P}(X \in B) = \liminf \mathbb{P}(X_n \in B) = \limsup \mathbb{P}(X_n \in B).$$

4 $\Leftrightarrow$ 5 On considère $B =]-\infty, t]$. Si F_X est continue en t , alors $\mathbb{P}(X \in \overline{B} \setminus \mathring{B}) = \mathbb{P}(X = t) = 0$. La convergence en loi implique donc la convergence des fonctions de répartition au point t . Réciproquement, on considère $f \in \mathcal{C}_c^\infty$, alors pour tout $x \in \mathbb{R}$,

$$f(x) = \int_{-\infty}^x f'(t) dt = \int_{\mathbb{R}} f'(t) \mathbb{1}_{]-\infty, x[}(t) dt.$$

Par Fubini (f est bornée car à support compact), on a

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f'(t) \mathbb{E}[\mathbb{1}_{]-\infty, X[}(t)] X dt = \int_{\mathbb{R}} f'(t) (1 - F_X(t)) dt$$

Comme l'ensemble des points de discontinuité de F_X est au plus dénombrable, $F_{X_n} \rightarrow F_X$ λ -pp sur $\mathbb{R}$.

En effet, par le théorème de convergence dominée (car $|f'(t)(1 - F_{X_n}(t))| \leq |f'(t)| \in L^1(\mathbb{R}, dt)$ car $f' \in \mathcal{C}_c^\infty$),

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \int_{\mathbb{R}} f'(t) \lim_{n \rightarrow \infty} (1 - F_{X_n}(t)) X dt = \mathbb{E}[f(X)].$$

On conclut en utilisant le fait que $\mathcal{C}_c^\infty$ est dense dans $\mathcal{C}_c$.

□

Proposition 4.3.4. Si $X_n \xrightarrow[n \rightarrow \infty]{Loi} X$ et $g : \mathbb{R}^d \rightarrow \mathbb{R}^q$ une fonction continue, alors $g(X_n) \xrightarrow[n \rightarrow \infty]{Loi} g(X)$.

Preuve. Soit $f : \mathbb{R}^q \rightarrow \mathbb{R}$ une fonction continue bornée, alors $f \circ g : \mathbb{R}^d \rightarrow \mathbb{R}$ est continue bornée. Par ailleurs, $\mathbb{E}[f(g(X_n))] = \mathbb{E}[f \circ g(X_n)]$. On conclut en utilisant le fait que $X_n \xrightarrow[n \rightarrow \infty]{Loi} X$. $\square$

Corollaire 4.3.5. Si $X_n = (X_n^1, \dots, X_n^d) \xrightarrow[n \rightarrow \infty]{Loi} X = (X^1, \dots, X^d)$, alors pour tout $k \in \{1, \dots, d\}$, $X_n^k \xrightarrow[n \rightarrow \infty]{Loi} X^k$.

Attention : La réciproque est fausse !

Exemple. Soit X de loi $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$. On pose $X_n = X$ si n est pair et $X_n = -X$ si n impair et $Y_n = X$ pour tout n . Comme $X \stackrel{Loi}{=} -X$, on a $X_n \stackrel{Loi}{=} X$ et $Y_n \stackrel{Loi}{=} X$ pour tout $n \geq 1$, et donc $X_n \xrightarrow[n \rightarrow \infty]{Loi} X$ et $Y_n \xrightarrow[n \rightarrow \infty]{Loi} X$. Par contre (X_n, Y_n) ne converge pas en loi vers (X, X) , car sinon $X_n + Y_n$ convergerait en loi vers $2X$. Or $X_n + Y_n = 2X$ si n est pair et 0 si n impair (de même $X_n Y_n = 1$ ou -1 selon la parité de n).

Théorème 4.3.6. La convergence loi $X_n \xrightarrow[n \rightarrow \infty]{Loi} X$ est équivalente à la convergence simple des fonctions caractéristiques.

Preuve.(en dimension 1.)

Si $X_n \xrightarrow[n \rightarrow \infty]{Loi} X$. Comme $f_t = e^{itx}$ est $\mathcal{C}_b$, on a de façon évidente la convergence simple des fonctions caractéristiques.

Si les fonctions caractéristiques convergent simplement. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction $\mathcal{C}_c^\infty$. Alors sa transformée de Fourier $\hat{f}(t) = \int_{\mathbb{R}} e^{itx} f(x) dx \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ car $\widehat{f''}(t) = \int e^{itx} f''(x) dx = -t^2 \hat{f}(t)$ (par intégration par parties, la fonction étant à support compact) et $\|\widehat{f''}\|_\infty \leq \|f''\|_1$, ce qui implique $\hat{f}(t) = O(1/t^2) \in L^1(\mathbb{R}, dt)$.

Par le théorème d'inversion de Fourier,

$$f(x) = \frac{1}{2\pi} \int e^{-itx} \hat{f}(t) dt.$$

Par conséquent, par Fubini,

$$\mathbb{E}[f(X_n)] = \frac{1}{2\pi} \int \mathbb{E}[e^{-itX_n}] \hat{f}(t) dt$$

et on conclut par convergence dominée. $\square$

Théorème 4.3.7 (Paul Lévy). On considère une suite de fonctions caractéristiques $(\varphi_n)_{n \geq 0}$ de v.a. X_n réelles. On suppose que pour tout $t \in \mathbb{R}$, $\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t)$ où φ est une fonction continue en $t = 0$. Alors il existe une v.a. X telle que φ soit sa fonction caractéristique et $X_n \xrightarrow[n \rightarrow \infty]{Loi} X$.

Liens avec les autres convergences

Proposition 4.3.8. Si $X_n \xrightarrow[n \rightarrow \infty]{proba} X$ alors $X_n \xrightarrow[n \rightarrow \infty]{Loi} X$.

Attention, la réciproque de la proposition est fausse. Reprenons l'exemple précédent. Soit X de loi $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$ et considérons la suite $X_n = X$ si n est pair et $X_n = -X$ si n impair. Alors $\mathbb{P}(|X_{n+1} - X_n| > \varepsilon) = \mathbb{P}(2|X| > \varepsilon) = 1$ pour tout $\varepsilon < 1$.

Preuve. Soit $t \in \mathbb{R}^d$ fixé. On a

$$|\varphi_{X_n}(t) - \varphi_X(t)| \leq \mathbb{E}[|e^{itX_n} - e^{itX}|] = \mathbb{E}[|e^{it(X_n - X)} - 1|] \leq \mathbb{E}[\min(2, |t||X_n - X|)].$$

En effet, $|e^{ix} - 1| \leq 2$ et $|e^{ix} - 1| = |i \int_0^x e^{is} ds| \leq |x|$.

Soit $\varepsilon > 0$, comme $1 = \mathbf{1}_{|X_n - X| \leq \varepsilon} + \mathbf{1}_{|X_n - X| > \varepsilon}$, on a

$$\begin{aligned} |\varphi_{X_n}(t) - \varphi_X(t)| &\leq \varepsilon |t| \mathbb{P}(|X_n - X| \leq \varepsilon) + 2\mathbb{P}(|X_n - X| > \varepsilon) \\ &\leq \varepsilon |t| + 2\mathbb{P}(|X_n - X| > \varepsilon) \end{aligned}$$

Par conséquent, $\limsup_{n \rightarrow \infty} |\varphi_{X_n}(t) - \varphi_X(t)| \leq \varepsilon |t|$ pour tout $\varepsilon > 0$. Il suffit de faire tendre ε vers 0 pour conclure. $\square$

Proposition 4.3.9. Soit $c \in \mathbb{R}^d$. Supposons que les variables X_n sont construites sur le même espace de probabilité et que $X_n \xrightarrow[n \rightarrow \infty]{Loi} c$. Alors $X_n \xrightarrow[n \rightarrow \infty]{proba} c$.

Preuve. (en dimension 1). Soit $\varepsilon > 0$.

$$\begin{aligned} \mathbb{P}(|X_n - c| > \varepsilon) &= \mathbb{P}(X_n < c - \varepsilon) + \mathbb{P}(X_n > c + \varepsilon) \\ &\leq F_{X_n}(c - \varepsilon) + 1 - F_{X_n}(c + \varepsilon). \end{aligned}$$

Comme $\varepsilon > 0$, la fonction de répartition F_c de la constante c est continue en $c + \varepsilon$ et $c - \varepsilon$. Comme $X_n \xrightarrow[n \rightarrow \infty]{Loi} c$, alors $F_{X_n}(c - \varepsilon) \rightarrow F_c(c - \varepsilon) = 0$ et $F_{X_n}(c + \varepsilon) \rightarrow F_c(c + \varepsilon) = 1$ quand $n \rightarrow \infty$. Donc $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - c| > \varepsilon) = 0$ pour tout $\varepsilon > 0$. $\square$

Lemme 4.3.10 (Lemme de Slutsky). On considère deux suites $(X_n)_{n \geq 0}$ et $(Y_n)_{n \geq 0}$ de variables aléatoires avec

$$X_n \xrightarrow[n \rightarrow \infty]{Loi} X \quad \text{et} \quad Y_n \xrightarrow[n \rightarrow \infty]{proba} c.$$

Alors on a la convergence en loi du couple $(X_n, Y_n) \xrightarrow[n \rightarrow \infty]{Loi} (X, c)$.

Preuve. La fonction caractéristique de (X, c) est $\varphi_{(X,c)}(t, s) = \varphi_X(t)e^{isc}$. Soit $\varepsilon > 0$. On a

$$\begin{aligned} |\varphi_{(X_n, Y_n)}(t, s) - \varphi_{(X,c)}(t, s)| &= |\mathbb{E}[e^{itX_n}(e^{isY_n} - e^{isc})] + e^{isc}\mathbb{E}[e^{itX_n} - e^{itX}]| \\ &\leq \mathbb{E}[|e^{isY_n} - e^{isc}|] + |\mathbb{E}[e^{itX_n} - e^{itX}]| \\ &= \varepsilon |s| + \mathbb{P}(|Y_n - c| > \varepsilon) + |\varphi_{X_n}(t) - \varphi_X(t)|. \end{aligned}$$

Comme $Y_n \xrightarrow[n \rightarrow \infty]{proba} c$ et $X_n \xrightarrow[n \rightarrow \infty]{Loi} X$, on a $\limsup |\varphi_{(X_n, Y_n)}(t, s) - \varphi_{(X,c)}(t, s)| \leq \varepsilon |s|$ pour tout $\varepsilon > 0$. On conclut en faisant $\varepsilon \rightarrow 0$. $\square$

Conséquence. Si $X_n \xrightarrow[n \rightarrow \infty]{Loi} X$ et $X_n - Y_n \xrightarrow[n \rightarrow \infty]{proba} 0$, alors $Y_n \xrightarrow[n \rightarrow \infty]{Loi} X$.

En effet, d'après le lemme précédent $(X_n, X_n - Y_n) \xrightarrow[n \rightarrow \infty]{Loi} (X, 0)$ et $Y_n = f(X_n, X_n - Y_n)$ avec $f(u, v) = u - v$ fonction continue.

Exemple. Soit X_n des v.a. de loi géométrique de paramètre respectif $\mathcal{G}(\lambda/n)$ avec $\lambda > 0$. La suite X_n/n converge en loi vers la loi exponentielle $\mathcal{E}(\lambda)$.

En effet, la fonction caractéristique de X_n est

$$\varphi_{X_n}(t) = \frac{\frac{\lambda}{n}e^{it}}{1 - (1 - \frac{\lambda}{n})e^{it}}.$$

Par conséquent,

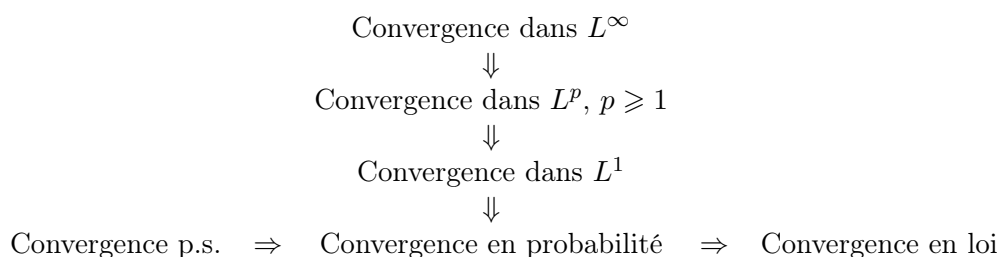
$$\begin{aligned}\varphi_{X_n/n}(t) &= \varphi_{X_n}(t/n) = \frac{\frac{\lambda}{n}e^{it/n}}{1 - (1 - \frac{\lambda}{n})e^{it/n}} \\ &= \frac{\lambda}{n(e^{-it/n} - 1) + \lambda}.\end{aligned}$$

Or

$$n(e^{-it/n} - 1) = n(\cos(t/n) - 1) - in \sin(t/n) \rightarrow -it \quad \text{quand } n \rightarrow \infty.$$

Par conséquent, $\varphi_{X_n/n}(t) \rightarrow \lambda/(\lambda - it)$. X_n/n converge vers la loi exponentielle quand $n \rightarrow \infty$.

4.4 Résumé des liens entre les différentes convergences



4.5 Exercices sur les différentes convergence

Exercice 4.5.1. [Convergence en probabilité]

Soit $(X_n)_{n \geq 1}$, X et Y des v.a. réelles avec $(X_n)_{n \geq 1}$ converge en probabilité vers X .

- On suppose dans cette question que pour tout $n \geq 1$, $|X_n| \leq Y$ p.s.
 - Montrer que $|X| \leq Y$ p.s.
 - Montrer que si Y est bornée, alors $(X_n)_{n \geq 1}$ converge vers X dans L^p , $\forall p \geq 1$.
- Montrer que $\mathbb{E}[|X|] \leq \liminf \mathbb{E}[|X_n|]$. (Indic. Si $x \geq 0$ et $k \geq 0$, $x \geq \min(x, k)$).

Exercice 4.5.2. [Convergence de variables aléatoires exponentielles]

Soit $(\theta_n)_{n \geq 1}$ une suite de réels strictement positifs telle que $\lim_{n \rightarrow +\infty} \theta_n = +\infty$. On considère $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes de loi exponentielle $\mathcal{E}(\theta_n)$.

- Montrer que la suite $(X_n)_{n \geq 1}$ converge en probabilité.
- La suite $(X_n)_{n \geq 1}$ converge-t-elle dans L^1 ?
- Étudier la convergence presque sûre dans les deux cas suivants : $\theta_n = n$; $\theta_n = \ln n$.

Exercice 4.5.3. [Lois binomiales]

Soit $\lambda > 0$ et $(X_n)_{n \geq 0}$ une suite de v.a.r. de loi binomiale $X_n \sim \mathcal{B}(n, \lambda/n)$. Montrer que X_n converge en loi. Préciser la limite.

Exercice 4.5.4. [Lois uniformes]

Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. de loi uniforme $X_n \sim \mathcal{U}(1, \dots, n)$. Montrer que $\frac{X_n}{n}$ converge en loi. Préciser la limite.

Exercice 4.5.5. [Convergence du min de v.a.i. uniformes]

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes ayant toutes une loi uniforme sur $[0, 1]$. On pose

$$M_n = \min(X_1, \dots, X_n), \quad U_n = nM_n.$$

1. Calculer les fonctions de répartition de M_n et U_n .
2. Montrer que la suite $(U_n)_{n \geq 1}$ converge en loi. Préciser la limite.
3. Étudier la convergence de M_n : en loi, puis en proba, puis p.s., puis dans L^p , et enfin dans L^∞ .

Exercice 4.5.6. [Renormalisation et maximum de variables aléatoires iid, cas borné]

Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes de même loi avec un support borné. On note F leur fonction de répartition commune et $M_n = \max(X_1, \dots, X_n)$.

1. Montrer qu'il existe $x_0 \in \mathbb{R}$ tel que $\forall x < x_0, F(x) < 1$, et $F(x_0) = 1$.
2. Montrer que $\lim_{n \rightarrow +\infty} \mathbb{P}(M_n \in]a, b]) = 0$ si $b < x_0$ ou si $a \geq x_0$ et que la limite est égale à 1 si $a < x_0 \leq b$.
3. En déduire que M_n converge en loi vers x_0 .

On suppose désormais $x_0 = 1$ et $F(x) = 1 - (1 - x)^\alpha$ si $x \in [0, 1]$ (où $\alpha > 0$).

4. Montrer que si $\alpha = 1$ les v.a. X_i sont uniformément distribuées sur $[0, 1]$.
5. Soit $Z_n = n^{1/\alpha}(M_n - 1)$. Montrer que

$$\lim_{n \rightarrow +\infty} \mathbb{P}(Z_n < x) = \begin{cases} \exp(-(-x)^\alpha) & \text{si } x \leq 0 \\ 1 & \text{si } x > 0. \end{cases}$$

Exercice 4.5.7. [Renormalisation et maximum de variables aléatoires iid, cas non borné]

Comme dans l'exercice 4.5.6, $X_1, \dots, X_n$ sont des v.a. iid de fonction de répartition F . On suppose que $\forall x \in \mathbb{R}, F(x) < 1$. On pose $M_n = \max(X_1, \dots, X_n)$.

1. Montrer que, pour tout $x \in \mathbb{R}, \lim_{n \rightarrow +\infty} \mathbb{P}(M_n > x) = 1$.
2. On suppose que les variables aléatoires X_i sont de loi exponentielle $\mathcal{E}(\theta)$ et on pose $Z_n = \theta M_n - \ln n$. Déterminer, pour tout $x \in \mathbb{R}, \lim_{n \rightarrow +\infty} \mathbb{P}(Z_n \leq x)$.
3. On suppose maintenant que les variables aléatoires X_i sont de loi de Cauchy $\mathcal{C}(\theta)$ et on pose $Z_n = \frac{\pi M_n}{n\theta}$. Montrer que

$$\lim_{n \rightarrow +\infty} \mathbb{P}(Z_n \leq x) = \begin{cases} \exp(-x^{-1}) & \text{si } x > 0 \\ 0 & \text{sinon.} \end{cases}$$

Indication. On utilisera l'identité trigonométrique

$$\forall x > 0, \quad \arctan(x) + \arctan(1/x) = \pi/2.$$

4.6 Théorèmes limites classiques

4.6.a) Loi des grands nombres : LGN

On se place sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On considère une suite de v.a. réelles $(X_n)_{n \geq 0}$ indépendantes et de même loi : i.i.d. (indépendantes et identiquement distribuées)

Théorème 4.6.1 (Kolmogorov). Soit $(X_n)_{n \geq 0}$ une suite de v.a. i.i.d. dans L^1 de moyenne $m = \mathbb{E}[X_1]$. Alors

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{} m \quad \text{p.s. et dans } L^1.$$

Remarque. Si $(X_n)_{n \geq 0}$ une suite de v.a. i.i.d. telle que il existe $c \in \mathbb{R}$ avec

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{p.s.} c,$$

alors $\mathbb{E}[|X_1|] < \infty$ et $c = \mathbb{E}[X_1]$.

Il y a donc équivalence entre la convergence presque sûre de la moyenne empirique et l'existence d'un moment d'ordre 1 pour les suites de v.a. i.i.d.

Preuve. On pose $S_n = \sum_{i=1}^n X_i$.

Si on a convergence, alors $X_n/n = S_n/n - S_{n-1}/(n-1) \times \frac{n-1}{n} \xrightarrow[n \rightarrow \infty]{p.s.} c - c = 0$. D'après Borel Cantelli, la suite étant i.i.d., on a $\forall \varepsilon > 0, \sum \mathbb{P}\left(\frac{|X_n|}{n} > \varepsilon\right) < \infty$. Par conséquent,

$$\sum \mathbb{P}\left(\frac{|X_n|}{n} > \varepsilon\right) = \sum \mathbb{P}(|X_n| > n\varepsilon) = \sum \mathbb{P}(|X_1| > n\varepsilon)$$

On conclut en remarquant que $\sum_{n \geq 1} \mathbb{P}(|X_1| > n) \leq \mathbb{E}[|X_1|] \leq \sum_{n \geq 0} \mathbb{P}(|X_1| > n)$ (voir Exercice du Chapitre 3). $\square$

Preuve. (Loi forte des Grands Nombres dans le cas L^2)

On suppose que $X_1 \in L^2$. On note $x^+ = \sup(x, 0)$ et $x^- = \sup(-x, 0)$.

On peut écrire $X_n = X_n^+ - X_n^-$ avec X_n^+ et X_n^- satisfaisants les mêmes hypothèses que X_n . Par ailleurs, on peut écrire $M_n = \frac{1}{n} \sum_{i=1}^n X_i = M_n^+ - M_n^-$ où $M_n^+ = \frac{1}{n} \sum_{i=1}^n X_i^+$, $M_n^- = \frac{1}{n} \sum_{i=1}^n X_i^-$ et $m = \mathbb{E}[X_1^+] - \mathbb{E}[X_1^-]$.

On peut donc supposer sans restriction que $X_n \geq 0$.

On pose $Z_n = M_{n^2}$. On a $\mathbb{E}[Z_n] = m$ et $\text{Var}(Z_n) = \sigma^2/n^2$

On souhaite montrer que $Z_n \xrightarrow[n \rightarrow \infty]{p.s.} m$. Soit $\varepsilon > 0$, par l'inégalité de Bienaymé-Chebychev,

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

Donc $\sum \mathbb{P}(|Z_n - m| > \varepsilon) < +\infty$ et $Z_n \xrightarrow[n \rightarrow \infty]{p.s.} m$.

Montrons maintenant que $M_n \xrightarrow[n \rightarrow \infty]{p.s.} m$. On pose $q_n = \lfloor \sqrt{n} \rfloor$. On a $q_n \leq \sqrt{n} \leq q_n + 1$, par conséquent, $q_n^2 \leq n \leq (q_n + 1)^2$.

Comme les variables sont positives, en posant $S_n = \sum_{i=1}^n X_i$, on a

$$\frac{S_{q_n^2}}{n} \leq \frac{S_n}{n} \leq \frac{S_{(q_n+1)^2}}{n}$$

$$\frac{q_n^2}{n} Z_{q_n} \leq M_n \leq \frac{(q_n + 1)^2}{n} Z_{q_n+1}$$

Comme $\sqrt{n} - 1 \leq q_n \leq \sqrt{n}$, $q_n^2/n \rightarrow 1$ quand $n \rightarrow \infty$. Comme $Z_n \xrightarrow[n \rightarrow \infty]{p.s.} m$, par encadrement $M_n \xrightarrow[n \rightarrow \infty]{p.s.} m$.

□

Simulations Considérons l'exemple du jeu "Pile ou Face" avec une pièce bien équilibrée. La probabilité de tomber sur "Pile" est alors $p = 1/2$. Le joueur mise sur "Pile". On répète plusieurs fois l'expérience et on observe dans la Figure 4.1 l'évolution de la fréquence du nombre de succès en fonction du nombre de lancers :

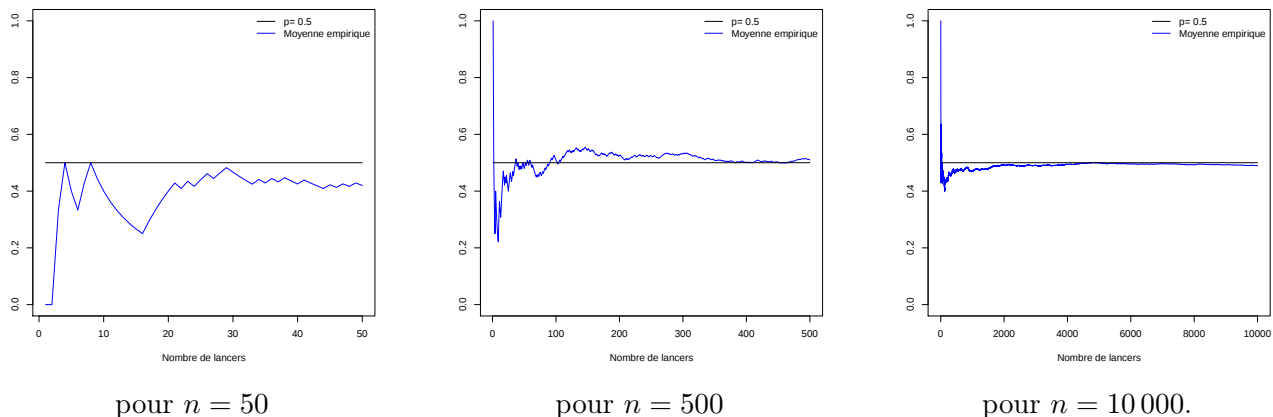


FIGURE 4.1 – Fréquence de pile dans un jeu de pile ou face.

Application de la loi des grands nombres : MÉTHODE DE MONTE-CARLO

Soit $(X_n)_{n \geq 0}$ une suite de v.a. i.i.d à valeurs dans $\mathbb{R}^d$. et soit $g : \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction mesurable telle que $\mathbb{E}[|g(X_1)|] < \infty$.

D'après la loi des grands nombres $\frac{1}{n} \sum_{i=1}^n g(X_i) \xrightarrow[n \rightarrow \infty]{p.s.} \mathbb{E}[g(X_1)]$.

Si X_1 est une v.a. à densité f , alors $\mathbb{E}[g(X_1)] = \int g(x)f(x)dx$. Par conséquent, $\frac{1}{n} \sum_{i=1}^n g(X_i)$ est une approximation de l'intégrale $\int g(x)f(x)dx$.

Par exemple, si $X_i \sim \mathcal{U}([a, b])$, alors $\frac{1}{n} \sum_{i=1}^n g(X_i)$ est une approximation de l'intégrale $\frac{1}{b-a} \int_a^b g(x)dx$.

La méthode de Monte-Carlo permet d'approcher numériquement une intégrale que l'on ne sait pas calculer. Le point fort de cette méthode par rapport aux méthodes numériques (rectangles, trapèze, ...) est que l'on n'a pas besoin de régularité sur la fonction g et elle est très facile à mettre en oeuvre quelque soit la dimension $d \geq 1$.

On peut alors se poser la question de la vitesse de convergence.

Simulations Observons la vitesse de convergence à l'aide de simulations.

On reprend notre exemple et on trace dans la Figure 4.2 plusieurs réalisations de la trajectoire *aléatoire* $n \mapsto M_n$ (chaque réalisation ayant une couleur différente sur le graphique). On obtient le résultat suivant. On observe que la convergence est plutôt lente. Essayons différentes fonctionnelles pour évaluer la vitesse de convergence dans la Figure 4.3 (courbe tracée en noir).

Le théorème central limite va nous fournir la vitesse de convergence de la loi des grands nombres.

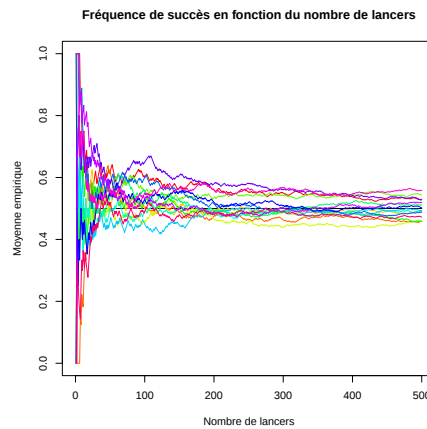
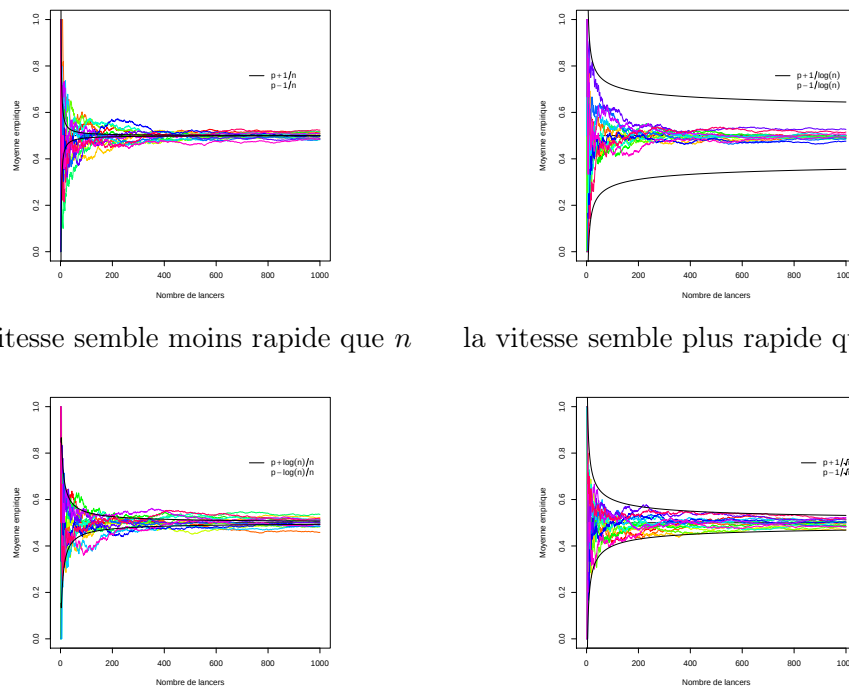


FIGURE 4.2 – Plusieurs trajectoires de $M_n = \frac{X_1 + \dots + X_n}{n}$ lorsque les (X_i) sont des variables i.i.d $\mathcal{B}(1/2)$.



la vitesse semble moins rapide que n la vitesse semble plus rapide que $\ln(n)$

la vitesse semble moins rapide que $\frac{n}{\ln(n)}$ la vitesse semble de l'ordre de $\sqrt{n}$.

FIGURE 4.3 – Trouver la vitesse de convergence de façon heuristique.

4.6.b) Théorème Central Limite (TCL)

Théorème 4.6.2. Soit $(X_n)_{n \geq 0}$ une suite de v.a. réelles i.i.d. dans L^2 d'espérance $m = \mathbb{E}[X_1]$ et de variance $\sigma^2 = \text{Var}(X_1) > 0$. On pose $S_n = \sum_{i=1}^n X_i$. Alors

$$\frac{\sqrt{n}}{\sigma} \left(\frac{1}{n} \sum_{i=1}^n X_i - m \right) \xrightarrow[n \rightarrow \infty]{\text{Loi}} \mathcal{N}(0, 1).$$

ce qu'on peut réécrire sous la forme,

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{Loi} \mathcal{N}(0, 1)$$

Conséquence. Pour tout $a, b \in \mathbb{R}$, sous les hypothèses du théorème, on a

$$\mathbb{P}\left(\frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}\sum_{i=1}^n X_i - m\right) \in [a, b]\right) \xrightarrow[n \rightarrow \infty]{} \mathbb{P}(Z \in [a, b]), \quad \text{avec } Z \sim \mathcal{N}(0, 1).$$

Preuve. [Preuve du TCL. dans le cas où $X \in L^3$.]

En remplaçant X_i par $\frac{X_i - m}{\sigma}$ on peut se ramener à $m = 0$ et $\sigma = 1$. On veut alors montrer que

$$\frac{S_n}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{Loi} \mathcal{N}(0, 1).$$

On va montrer la convergence des fonctions caractéristique.

Tout d'abord, on remarque que

$$\lim_{n \rightarrow +\infty} (1 - t^2/2n)^n = \lim_{n \rightarrow +\infty} e^{n \ln(1 - \frac{t^2}{2n})} = e^{-t^2/2},$$

qui est la fonction caractéristique de la loi $\mathcal{N}(0, 1)$.

Lemme 4.6.3. Pour tout $x \in \mathbb{R}$,

$$\left| e^{ix} - \sum_{k=0}^p \frac{(ix)^k}{k!} \right| \leq \min\left(\frac{|x|^{p+1}}{(p+1)!}, \frac{2|x|^p}{p!}\right).$$

Preuve. On utilise la formule de Taylor avec reste intégral ;

$$e^{ix} = \sum_{k=0}^p \frac{(ix)^k}{k!} + \frac{i^{p+1}}{p!} \int_0^x (x-s)^p e^{is} ds.$$

Par conséquent,

$$\left| e^{ix} - \sum_{k=0}^p \frac{(ix)^k}{k!} \right| \leq \frac{1}{p!} \left| \int_0^x (x-s)^p ds \right| = \frac{|x|^{p+1}}{(p+1)!}.$$

Par ailleurs, par intégration par parties (en posant $u = (x-s)^p$ et $dv = e^{is} ds$),

$$\left| e^{ix} - \sum_{k=0}^p \frac{(ix)^k}{k!} \right| = \left| \frac{i^p}{(p-1)!} \int_0^x (x-s)^{p-1} e^{is} ds - \frac{(ix)^p}{p!} \right|$$

□

Comme les variables sont i.i.d, $\varphi_{s_n/\sqrt{n}}(t) = \varphi(t/\sqrt{n})^n$. On remarque que $|a^n - b^n| \leq n|a - b|$ si $|a| \leq 1$ and $|b| \leq 1$, et donc en posant $a_i = \varphi(t/\sqrt{n})$ et $b_i = (1 - t^2/2n)$, on a

$$|\varphi(t/\sqrt{n})^n - (1 - t^2/2n)^n| \leq n|\varphi(t/\sqrt{n}) - (1 - t^2/2n)|$$

Par ailleurs, d'après le lemme 4.6.3, et en remarquant que $\mathbb{E}[X] = 0$,

$$n|\varphi(t/\sqrt{n}) - (1 - t^2/2n)| \leq n\mathbb{E}\left[\min\left(\frac{t^3|X|^3}{6n^{3/2}}, \frac{t^2|X|^2}{n}\right)\right] = \mathbb{E}\left[\min\left(\frac{t^3|X|^3}{6n^{1/2}}, t^2|X|^2\right)\right].$$

Par le théorème de convergence dominée, on en déduit que $\lim_{n \rightarrow \infty} |\varphi(t/\sqrt{n})^n - (1 - t^2/2n)^n| = 0$.

Comme $\lim_{n \rightarrow +\infty} (1 - t^2/2n)^n = e^{-t^2/2}$ qui est la fonction caractéristique de la loi $\mathcal{N}(0, 1)$. On conclut en utilisant le théorème de Paul Lévy.

□

Application

On considère des variables X_i i.i.d. dans L^2 . On peut toujours approcher la loi de $S_n = \sum_{i=1}^n X_i$ par la loi normale quelque soit la loi initiale des X_i .

Par conséquent, si $X_n \sim \mathcal{B}(p)$, on sait que $S_n \sim \mathcal{B}(n, p)$. Comme $\frac{S_n - np}{\sqrt{np(1-p)}}$ est proche de la loi $\mathcal{N}(0, 1)$, on approche souvent la loi $\mathcal{B}(n, p)$ par la loi $\mathcal{N}(np, np(1-p))$ pour n grand.

Par conséquent, si $X_n \sim \mathcal{P}(\lambda)$, on sait que $S_n \sim \mathcal{P}(n\lambda)$. Comme $\frac{S_n - n\lambda}{\sqrt{n\lambda}}$ est proche de la loi $\mathcal{N}(0, 1)$, on approche la loi $\mathcal{P}(n\lambda)$ par la loi $\mathcal{N}(n\lambda, n\lambda)$ pour n grand.

Simulations Observons dans la Figure 4.4 l'évolution de la loi Binomiale $\mathcal{B}(n, p)$ en fonction de n vers la "cloche gaussienne", pour $p = 1/2$:

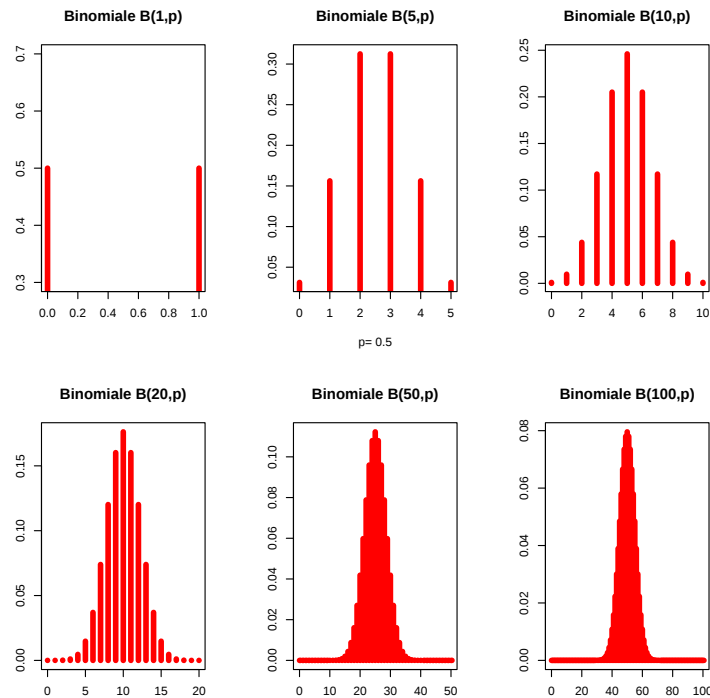


FIGURE 4.4 – Illustration de la convergence de la loi Binomiale avec la loi gaussienne

4.6.c) Preuve de la loi forte des grands nombres

Cette section est pour votre curiosité, mais la preuve de la loi des grands nombre dans le cas L^1 est vraiment complexe.

Théorème 4.6.4 (Théorème des séries centrées). *Soit $(X_n)_{n \geq 1}$ une suite de variables réelles indépendantes telle que $\forall n \geq 1$ $X_n \in L^2$ et $\mathbb{E}[X_n] = 0$.*

Si $\sum_{n \geq 1} \mathbb{E}[X_n^2] < \infty$, alors $(S_n)_{n \geq 1}$ converge p.s. et dans L^2 vers une v.a. réelle.

Preuve. On montre tout d'abord le lemme suivant :

Lemme 4.6.5. *Soit $(X_n)_{n \geq 1}$ une suite de variables réelles indépendantes telle que $\forall n \geq 1$ $X_n \in L^2$ et $\mathbb{E}[X_n] = 0$. Alors pour tout $\varepsilon > 0$, on a*

$$\mathbb{P}\left(\sup_{1 \leq k \leq n} |S_k| \geq \varepsilon\right) \leq \frac{\text{Var}(S_n)}{\varepsilon^2}.$$

Preuve. On introduit les événements disjoints

$$A_k = \{|S_1| < \varepsilon, \dots, |S_{k-1}| < \varepsilon, |S_k| \geq \varepsilon\}.$$

On a

$$\begin{aligned} \mathbb{E}[S_n^2] &\geq \sum_{k=1}^n \mathbb{E}[S_n^2 \mathbf{1}_{A_k}] = \sum_{k=1}^n \mathbb{E}[(S_n - S_k + S_k)^2 \mathbf{1}_{A_k}] \\ &\geq \sum_{k=1}^n \mathbb{E}[(2(S_n - S_k)S_k + S_k^2) \mathbf{1}_{A_k}]. \end{aligned}$$

Comme $A_k \in \sigma(X_1, \dots, X_k)$, S_k est $\sigma(X_1, \dots, X_k)$ -mesurable et $S_n - S_k = \sum_{i=k+1}^n X_i$ est indépendant de $\sigma(X_1, \dots, X_k)$, on a $\mathbb{E}[2(S_n - S_k)S_k \mathbf{1}_{A_k}] = 2\mathbb{E}[S_n - S_k]\mathbb{E}[S_k \mathbf{1}_{A_k}] = 0$ car les X_i sont centrées. Par conséquent,

$$\mathbb{E}[S_n^2] \geq \sum_{k=1}^n \mathbb{E}[S_k^2 \mathbf{1}_{A_k}] \geq \varepsilon^2 \sum_{k=1}^n \mathbb{P}(A_k) = \varepsilon^2 \mathbb{P}\left(\sup_{1 \leq k \leq n} |S_k| \geq \varepsilon\right).$$

Démontrons maintenant le théorème des séries centrées. □

D'après les critères de convergence, on a $(X_n)_{n \geq 1}$ converge presque sûrement ssi $\forall \varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\sup_{k \geq 1} |S_{n+k} - S_n| > \varepsilon\right) = 0.$$

Soit $\varepsilon > 0$, par monotonie on a

$$\mathbb{P}\left(\sup_{k \geq 1} |S_{n+k} - S_n| > \varepsilon\right) = \lim_{p \rightarrow \infty} \mathbb{P}\left(\sup_{1 \leq k \leq p} |S_{n+k} - S_n| > \varepsilon\right).$$

Par ailleurs,

$$S_{n+k} - S_n = \sum_{i=n+1}^{n+k} X_i = \sum_{i=1}^k X_{n+i}$$

Par conséquent, en utilisant le lemme et l'indépendance des variables centrées X_i ,

$$\mathbb{P}\left(\sup_{1 \leq k \leq p} |S_{n+k} - S_n| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} \text{Var}\left(\sum_{i=1}^p X_{n+i}\right) = \frac{1}{\varepsilon^2} \sum_{i=n+1}^{n+p} \mathbb{E}[X_i^2].$$

On obtient ainsi,

$$\mathbb{P}\left(\sup_{k \geq 1} |S_{n+k} - S_n| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} \sum_{i=n+1}^{\infty} \mathbb{E}[X_i^2].$$

Comme la série $\sum \mathbb{E}[X_n^2]$ converge, on obtient que la suite $(S_n)_{n \geq 1}$ converge ps. □

Dans le cas des sommes de v.a. indépendantes, on a un résultat remarquable sur les différents types de convergence.

Théorème 4.6.6. *Soit $(X_n)_{n \geq 1}$ une suite de variables réelles indépendantes. Pour $n \geq 1$, on pose $S_n = \sum_{k=1}^n X_k$. On a équivalence entre*

1. $(S_n)_{n \geq 1}$ converge p.s. vers une v.a. réelle,
2. $(S_n)_{n \geq 1}$ converge en proba vers une v.a. réelle,
3. $(S_n)_{n \geq 1}$ converge en loi vers une v.a. réelle.

Preuve. Voir Probability and Measure de P. Billingsley. □

Donnons maintenant une preuve de la loi forte des grands nombres. On rappelle tout d'abord quelques lemmes bien utiles en analyse.

Lemme 4.6.7. *Soit $(b_n)_{n \geq 1}$ une suite croissante de réels strictement positifs telle que $\lim_{n \rightarrow +\infty} b_n = +\infty$.*

On pose $b_0 = 0$.

Soit $(x_n)_{n \geq 1}$ une suite de réels.

Cesaro : *Si la suite $(x_n)_{n \geq 1}$ converge vers x alors $\lim_{n \rightarrow +\infty} \frac{1}{b_n} \sum_{k=1}^n (b_k - b_{k-1})x_k = x$.*

Kronecker : *Si la série $\sum_{n \geq 1} \frac{1}{b_n} x_n$ converge dans $\mathbb{R}$ alors $\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n x_k = 0$.*

Lemme 4.6.8. *Pour $\alpha > 1$, $k \geq 1$, $\sum_{n \geq k+1} n^{-\alpha} \leq \frac{1}{(\alpha - 1)k^{\alpha-1}}$.*

Preuve. Comme la fonction $x \mapsto x^{-\alpha}$ est décroissante sur $]0, \infty[$, on a

$$\sum_{n \geq k+1} n^{-\alpha} \leq \sum_{n \geq k+1} \int_{n-1}^n x^{-\alpha} dx = \int_k^{\infty} x^{-\alpha} dx = \frac{k^{1-\alpha}}{(\alpha - 1)}.$$

□

Preuve. (Loi des grands nombres)

On peut supposer sans perte de généralité que $m = \mathbb{E}[X_1] = 0$ et on souhaite montrer que $M_n = \frac{1}{n} \sum_{k=1}^n X_k$ converge p.s. vers 0.

On pose

$$\begin{aligned}\widehat{X}_n &= X_n \mathbb{1}_{|X_n| < n}, & \widehat{M}_n &= \frac{1}{n} \sum_{k=1}^n \widehat{X}_k; \\ \widetilde{X}_n &= \widehat{X}_n - \mathbb{E}[\widehat{X}_n], & \widetilde{M}_n &= \frac{1}{n} \sum_{k=1}^n \widetilde{X}_k.\end{aligned}$$

Les suites $(\widehat{X}_n)_{n \geq 1}$ et $(\widetilde{X}_n)_{n \geq 1}$ sont toutes les deux des suites de v.a. indépendantes. On va procéder en deux étapes :

(a) On a les équivalences suivantes

1. $M_n \xrightarrow[n \rightarrow \infty]{p.s.} 0$ si et seulement si $\widehat{M}_n \xrightarrow[n \rightarrow \infty]{p.s.} 0$.
2. $\widehat{M}_n \xrightarrow[n \rightarrow \infty]{p.s.} 0$ si et seulement si $\widetilde{M}_n \xrightarrow[n \rightarrow \infty]{p.s.} 0$.

(b) On montre que $\widetilde{M}_n \xrightarrow[n \rightarrow \infty]{p.s.} 0$.

Montrons le premier point.

(a) 1. On montre que $M_n - \widehat{M}_n \xrightarrow[n \rightarrow \infty]{p.s.} 0$.

On a $\forall n \geq 1$, $M_n - \widehat{M}_n = \frac{1}{n} \sum_{k=1}^n X_k \mathbb{1}_{|X_k| \geq k}$. Comme les variables sont de même loi, on a

$$\sum_{n \geq 1} \mathbb{P}(|X_n| \geq n) = \sum_{n \geq 1} \mathbb{P}(|X_1| \geq n) \leq 1 + \mathbb{E}[|X_1|] < \infty.$$

Donc d'après le premier lemme de Borel-Cantelli,

$$\mathbb{P}(\limsup \{|X_n| \geq n\}) = 0.$$

Soit $\omega \in \liminf \{|X_n| < n\}$, il existe alors $n_\omega \geq 1$ tel que $\forall n \geq n_\omega$, $|X_n(\omega)| < n$, i.e. $\widehat{X}_n(\omega) = X_n(\omega)$. Par conséquent, $M_n(\omega) - \widehat{M}_n(\omega) = \frac{1}{n} \sum_{k=1}^{n_\omega} X_k(\omega) \mathbb{1}_{|X_k(\omega)| \geq k}$. On en déduit que pour tout $\omega \in \liminf \{|X_n| < n\}$, événement de probabilité 1, $M_n(\omega) - \widehat{M}_n(\omega) \rightarrow 0$ quand $n \rightarrow +\infty$.

2. On va maintenant montrer que $\widehat{M}_n - \widetilde{M}_n \xrightarrow[n \rightarrow \infty]{p.s.} 0$. Du fait que les variables soient de même loi, on remarque que

$$\widehat{M}_n - \widetilde{M}_n = \frac{1}{n} \sum_{k=1}^n \mathbb{E}[X_k \mathbb{1}_{|X_k| \geq k}] = \frac{1}{n} \sum_{k=1}^n \mathbb{E}[X_1 \mathbb{1}_{|X_1| \geq k}].$$

Comme $X_1 \in L^1$, par convergence dominée, on a $\lim_{n \rightarrow \infty} \mathbb{E}[X_1 \mathbb{1}_{|X_1| \geq n}] = 0$. En utilisant le lemme de Cesaro avec $b_n = n$ et $x_n = \mathbb{E}[X_1 \mathbb{1}_{|X_1| \geq n}]$, on en déduit que $\widehat{M}_n - \widetilde{M}_n$ converge vers 0.

(b) On étudie la convergence p.s. de $(\widetilde{M}_n)_{n \geq 1}$.

Comme $\widetilde{M}_n = \frac{1}{n} \sum_{k=1}^n \widetilde{X}_k$, d'après le Lemme de Kronecker, il suffit de montrer que $\sum_{k=1}^n \widetilde{X}_k/k$ converge dans $\mathbb{R}$ p.s.

D'après le théorème des séries centrées, il suffit de montrer que $\sum_{n \geq 1} \mathbb{E}[\widetilde{X}_n^2]/n^2 < \infty$. On a

$$\sum_{n \geq 1} \frac{1}{n^2} \mathbb{E}[\widetilde{X}_n^2] = \sum_{n \geq 1} \frac{1}{n^2} \mathbb{E}\left[\left(\widehat{X}_n - \mathbb{E}[\widehat{X}_n]\right)^2\right] = \sum_{n \geq 1} \frac{1}{n^2} \text{Var}(\widehat{X}_n) \leq \sum_{n \geq 1} \frac{1}{n^2} \mathbb{E}[\widehat{X}_n^2]$$

Comme les variables sont de même loi, on a pour tout $n \geq 1$,

$$\mathbb{E}[\hat{X}_n^2] = \mathbb{E}[X_n^2 \mathbf{1}_{|X_n| < n}] = \mathbb{E}[X_1^2 \mathbf{1}_{|X_1| < n}].$$

Par conséquent, par Fubini-Tonelli et en utilisant le lemme 4.6.8,

$$\begin{aligned} \sum_{n \geq 1} \frac{1}{n^2} \mathbb{E}[\hat{X}_n^2] &\leq \mathbb{E}\left[X_1^2 \sum_{n \geq 1} \frac{1}{n^2} \mathbf{1}_{|X_1| < n}\right] \leq \mathbb{E}\left[X_1^2 \sum_{n \geq [|X_1|] + 1} \frac{1}{n^2}\right] \\ &\leq \mathbb{E}\left[\frac{X_1^2}{[|X_1|] + 1}\right] \leq \mathbb{E}[|X_1|] < \infty. \end{aligned}$$

Par conséquent, $\widetilde{M}_n \xrightarrow[n \rightarrow \infty]{p.s.} 0$ et donc $M_n \xrightarrow[n \rightarrow \infty]{p.s.} 0$.

Montrons maintenant la convergence de $(M_n)_{n \geq 1}$ vers 0 dans L^1 . Soit $p \geq 1$, on a

$$|M_n| = \min(|M_n|, p) + (|M_n| - p)^+,$$

où $(x - p)^+ = \sup(x - p, 0)$. Comme la fonction $x \mapsto (x - p)^+$ est convexe croissante, on a

$$(|M_n| - p)^+ \leq \frac{1}{n} \sum_{k=1}^n (|X_k| - p)^+.$$

Par convergence dominée, à p fixé, on a $\mathbb{E}[\min(|M_n|, p)] \rightarrow 0$ quand $n \rightarrow \infty$ et comme les variables sont de même loi

$$\mathbb{E}[(|M_n| - p)^+] \leq \frac{1}{n} \sum_{k=1}^n \mathbb{E}[(|X_k| - p)^+] = \mathbb{E}[(|X_1| - p)^+].$$

Par convergence dominée par $|X_1|$, on a $\mathbb{E}[(|X_1| - p)^+] \rightarrow 0$ quand $p \rightarrow \infty$. On en déduit la convergence vers 0 dans L^1 de $(M_n)_{n \geq 1}$. $\square$

4.6.d) Extensions

On peut se demander s'il existe des extensions à la loi des grands nombres ou au TCL pour des suites de v.a. non i.i.d. Le théorème des séries centrées permet de travailler avec des suites de v.a. indépendantes mais pas nécessairement de même loi. Si les variables sont faiblement dépendantes, dans le cas des chaînes de Markov, on a alors un théorème similaire qui s'appelle le théorème ergodique.

Dans le cas du théorème central limite, on peut affaiblir les conditions. Le théorème reste notamment vrai sous les conditions de Lindeberg : supposons que les v.a. (X_n) sont indépendantes dans L^2 , mais pas nécessairement de même loi. En notant $m_n = \mathbb{E}[X_n]$, $\sigma_n^2 = \text{Var}(X_n)$ et $s_n^2 = \sum_{k=1}^n \sigma_k^2$, si pour tout $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{k=1}^n \mathbb{E}[(X_k - m_k)^2 \mathbf{1}_{|X_k - m_k| > \varepsilon s_n}] = 0,$$

alors

$$\frac{1}{s_n} \sum_{k=1}^n (X_k - m_k) \xrightarrow[n \rightarrow \infty]{Loi} \mathcal{N}(0, 1).$$

Il existe par ailleurs un théorème de Loi des grands nombres et un théorème Central Limite pour certains types de processus $(X_n)_{n \geq 1}$, appelés martingales. Pour ces processus, la dépendance entre X_{n+1} et $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ est contrôlée par la condition suivante : connaissant le passé $\mathcal{F}_n$, la loi de X_{n+1} est centrée en X_n .

Chapitre 5

Esperance Conditionnelle et Martingales

5.1 L'espérance conditionnelle

5.1.a) Heuristique

Lorsqu'on réalise une expérience, dans l'espace $(\Omega, \mathcal{F}, \mathbb{P})$:

- Ω représente l'ensemble de tous les résultats possibles,
- $\omega \in \Omega$ représente une réalisation de l'expérience,
- $\mathcal{F}$ représente l'ensemble des résultats observables,
- $\mathbb{P}$ permet d'évaluer la probabilité d'apparition de chaque événement de $\mathcal{F}$.

Une expérience est réalisée et on a accès qu'à une information partielle. On souhaiterait, au vue de cette information, estimer au mieux et de façon synthétique le résultat réel.

L'**espérance conditionnelle** est un outil adapté à ce genre de question.

Exemples d'utilisation

1. Séries chronologiques :

On considère une série de données dans le temps (le cours du CAC40, le nombre journalier de recherche du mot "bikini" dans google, l'évolution du niveau de la mer, ...). Connaissant les valeurs jusqu'à aujourd'hui, on aimerait prédire les valeurs futurs.

Si on note $(\mathcal{F}_n)_{n \geq 0}$ la filtration naturelle du processus temporel (X_n) , définie par

$$\sigma(\mathcal{F}_n) = \sigma(X_0, \dots, X_n),$$

la tribu engendrée par les variables jusqu'à l'instant n . On aimerait alors prédire X_{n+1} (le futur) sachant $\mathcal{F}_n$ (le passé).

2. Information bruitée :

On réalise une expérience, mais l'outil de mesure ne nous permet pas de bien quantifier le résultat. Cela peut être du à un défaut de fabrication ou de précision de l'outil de mesure ou de données dénaturées par l'environnement.

Le but est par conséquent de reconstruire le signal initial au fur et à mesure que les données arrivent. Il existe des techniques de filtrage permettant cette extrapolation, comme par exemple le filtre de Kalman Bucy qui consiste à effectuer successivement des étapes de prédiction et de correction. Si X est le signal initial (inconnu) et Y le signal bruité (connu), on suit le schéma suivant : prédire X_n sachant $Y_0, \dots, Y_{n-1}$, corriger ajoutant la donnée de Y_n , puis prédire X_{n+1} en connaissant $Y_0, \dots, Y_n$, etc.

3. Information cachée :

On réalise une expérience dont on sait à l'avance qu'elle ne nous donnera pas directement des informations sur l'objet d'étude. On espère à l'aide de cette expérience obtenir tout de même des informations sur cet objet.

Par exemple, on peut regarder le cours d'un portefeuille d'actions et en fonction de la volatilité en déduire si les placements choisis sont risqués ou non. Une question intéressante en génétique est d'identifier les zones codantes et les zones non codantes d'un brin d'ADN. Le seul objet que l'on peut observer est le brin lui-même et donc on aimerait au vu de la suite de nucléotides $\{a, c, g, t\}$ en déduire la localisation des zones codantes et des zones non-codantes.

5.1.b) Définitions

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. On rappelle que l'espace $L^2(\Omega, \mathcal{F}, \mathbb{P})$ munit du produit scalaire $\langle X, Y \rangle = \mathbb{E}[XY]$ est un espace de Hilbert.

Soit $\mathcal{A}$ une sous-tribu de $\mathcal{F}$. On note $L^2(\Omega, \mathcal{A}, \mathbb{P})$ l'espace des variables X telles que

1. $\mathbb{E}[|X|^2] < +\infty$,
2. $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est mesurable (on dit que X est $\mathcal{A}$ -mesurable) : $\forall B \in \mathcal{B}(\mathbb{R}), X^{-1}(B) \in \mathcal{A}$.

Comme $\mathcal{A} \subset \mathcal{F}$, on a facilement que si $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, alors $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Proposition 5.1.1. $L^2(\Omega, \mathcal{A}, \mathbb{P})$ est un sous-espace vectoriel fermé de $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Preuve. Il est facile de voir que $L^2(\Omega, \mathcal{A}, \mathbb{P})$ est un espace vectoriel, par propriétés des fonctions mesurables. Montrons que c'est un sous-ensemble fermé. Soit $(X_n)_{n \geq 0}$ une suite convergente de $L^2(\Omega, \mathcal{A}, \mathbb{P})$. On note X sa limite. On sait déjà que $\mathbb{E}[X^2] < +\infty$. La convergence dans L^2 implique la convergence en probabilité, et donc il existe une sous-suite de $(X_n)_{n \geq 0}$ qui converge p.s. vers X . Comme tous les X_n sont $\mathcal{A}$ -mesurables, on déduit que X est $\mathcal{A}$ -mesurable. $\square$

D'après les propriétés sur les espaces de Hilbert, on peut alors définir la projection orthogonale sur $L^2(\Omega, \mathcal{A}, \mathbb{P})$.

Définition 5.1.2. Soit $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ une v.a. et $\mathcal{A}$ une sous-tribu de $\mathcal{F}$. On appelle *espérance conditionnelle de X sachant $\mathcal{A}$* , notée $\mathbb{E}[X|\mathcal{A}]$, la projection orthogonale de X sur $L^2(\Omega, \mathcal{A}, \mathbb{P})$.

L'espérance conditionnelle est une variable aléatoire $\mathcal{A}$ -mesurable.

$\mathbb{E}[X|\mathcal{A}]$ est la meilleure approximation au sens de la distance L^2 de X par une v.a. $\mathcal{A}$ -mesurable ou encore la meilleure prévision de X si on ne dispose que de l'information apportée par la tribu $\mathcal{A}$. On a donc

$$\mathbb{E}[X|\mathcal{A}] = \operatorname{argmin} \left\{ \mathbb{E}[(X - Z)^2]^{1/2} : Z \in L^2(\Omega, \mathcal{A}, \mathbb{P}) \right\}.$$

Propriété 5.1.3. 1. L'espérance conditionnelle $\mathbb{E}[X|\mathcal{A}]$ est l'unique variable Y de $L^2(\Omega, \mathcal{A}, \mathbb{P})$ telle que pour toute variable $Z \in L^2(\Omega, \mathcal{A}, \mathbb{P})$,

$$\mathbb{E}[XZ] = \mathbb{E}[YZ].$$

2. L'espérance conditionnelle est une application linéaire.

3. Positivité : si $X \geq 0$, alors $\mathbb{E}[X|\mathcal{A}] \geq 0$ p.s.

4. Croissance : si $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ avec $X \leq Y$ p.s. alors $\mathbb{E}[X|\mathcal{A}] \leq \mathbb{E}[Y|\mathcal{A}]$ p.s.

Preuve.

1. En décomposant la v.a. X en sa composante $\mathbb{E}[X|\mathcal{A}]$ sur $\mathcal{A}$ et son orthogonal W , $X = \mathbb{E}[X|\mathcal{A}] + W$, on obtient facilement que pour toute v.a. $Z \in L^2(\Omega, \mathcal{A}, \mathbb{P})$,

$$\mathbb{E}[XZ] = \mathbb{E}[\mathbb{E}[X|\mathcal{A}]Z] + 0.$$

Réciproquement, si on a $\mathbb{E}[XZ] = \mathbb{E}[YZ]$, alors forcément $(X - Y)$ est orthogonal à $L^2(\Omega, \mathcal{A}, \mathbb{P})$ et par unicité de la projection orthogonale on a $Y = \mathbb{E}[X|\mathcal{A}]$ p.s.

2. L'application est linéaire par linéarité de la projection.
3. Si $X \geq 0$, on choisit $Z = \mathbb{1}_{\mathbb{E}[X|\mathcal{A}] < 0}$ qui est une v.a. $\mathcal{A}$ -mesurable. Comme X et Z sont positives, $\mathbb{E}[XZ] \geq 0$.
Or $\mathbb{E}[XZ] = \mathbb{E}[\mathbb{E}[X|\mathcal{A}]Z] = \mathbb{E}[\mathbb{E}[X|\mathcal{A}]\mathbb{1}_{\mathbb{E}[X|\mathcal{A}] < 0}]$ et $\mathbb{E}[\mathbb{E}[X|\mathcal{A}]\mathbb{1}_{\mathbb{E}[X|\mathcal{A}] < 0}] < 0$ si $\mathbb{P}(\mathbb{E}[X|\mathcal{A}] < 0) > 0$.
Par conséquent $\mathbb{P}(\mathbb{E}[X|\mathcal{A}] < 0) = 0$.
4. La croissance se déduit de la positivité et de la linéarité.

□

On étend maintenant la notion d'espérance conditionnelle aux v.a. dans $L^1(\Omega, \mathcal{F}, \mathbb{P})$:

Théorème 5.1.4. Soit $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ une v.a. et $\mathcal{A}$ une sous-tribu de $\mathcal{F}$.

Il existe une unique variable aléatoire $Y \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ telle que $\forall A \in \mathcal{A}$,

$$\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A].$$

Cette v.a. coïncide avec l'espérance conditionnelle dans le cas $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Définition 5.1.5. Par extension, l'unique variable aléatoire de $L^1(\Omega, \mathcal{A}, \mathbb{P})$ du théorème précédent est appelée espérance conditionnelle de X sachant $\mathcal{A}$ et est notée $\mathbb{E}[X|\mathcal{A}]$.

Notation. 1. Soit $B \in \mathcal{F}$, on définit $\mathbb{P}(B|\mathcal{A}) := \mathbb{E}[\mathbb{1}_B|\mathcal{A}]$. On prouve facilement que $\mathbb{P}(\cdot|\mathcal{A})$ est une mesure de probabilité sur Ω .

2. Si Y est une v.a. et $\mathcal{A} = \sigma(Y)$, alors on note $\mathbb{E}[X|Y] := \mathbb{E}[X|\sigma(Y)]$.

Preuve.

Existence : On se restreint à $X \geq 0$ en utilisant $X = X^+ - X^-$ et en remarquant que $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[X^+\mathbb{1}_A] - \mathbb{E}[X^-\mathbb{1}_A]$, car $X \in L^1$.

Montrons qu'il existe une v.a. Y $\mathcal{A}$ -mesurable positive telle que $\forall A \in \mathcal{A}$,

$$\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A].$$

Si $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, on a $\mathbb{1}_A \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ pour tout $A \in \mathcal{A}$, il suffit de prendre $Y = \mathbb{E}[X|\mathcal{A}]$.

Si $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, on définit $X_n = X \wedge n \in L^2$. On pose $Y_n = \mathbb{E}[X_n|\mathcal{A}]$. On remarque que les suite (X_n) et (Y_n) sont positives croissantes (par croissance de l'espérance conditionnelle dans L^2). Par convergence monotone, pour tout $A \in \mathcal{A}$,

$$\mathbb{E}[X\mathbb{1}_A] = \lim_n \mathbb{E}[X_n\mathbb{1}_A] = \lim_n \mathbb{E}[\mathbb{E}[X_n|\mathcal{A}]\mathbb{1}_A] = \mathbb{E}\left[\lim_n \mathbb{E}[X_n|\mathcal{A}]\mathbb{1}_A\right]$$

On pose alors $Y = \lim_n \mathbb{E}[X_n|\mathcal{A}]$ (au sens p.s., bien défini car suite croissante).

On remarque, en prenant $A = \Omega$, que $\mathbb{E}[X] = \mathbb{E}[Y]$. Comme X, Y sont des v.a. positives et X intégrable,

on déduit que Y est intégrable.

Unicité : Si Y, Y' deux variables $\mathcal{A}$ -mesurables intégrables telles que $\forall A \in \mathcal{A}$,

$$\mathbb{E}[Y\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[Y'\mathbf{1}_A].$$

On pose $\Lambda = \{Y < Y'\} \in \mathcal{A}$. On a alors $\mathbb{E}[Y\mathbf{1}_\Lambda] = \mathbb{E}[Y'\mathbf{1}_\Lambda]$, d'où $\mathbb{E}[(Y' - Y)\mathbf{1}_\Lambda] = 0$. Or $(Y' - Y)\mathbf{1}_\Lambda$ est positive p.s. et donc $(Y' - Y)\mathbf{1}_\Lambda = 0$ p.s.. De même, on montre que $(Y' - Y)\mathbf{1}_{\Lambda^c} = 0$ ps. Par conséquent, $Y = Y'$ p.s.. $\square$

Remarque. A-t-on $\mathbb{E}[XZ] = \mathbb{E}[\mathbb{E}[X|\mathcal{A}]Z]$ pour toute v.a. $Z \in L^1(\Omega, \mathcal{A}, \mathbb{P})$?

La réponse est oui si $XZ \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, mais sinon $\mathbb{E}[XZ]$ n'est pas bien défini.

On montre ce résultat, d'abord pour les v.a. indicatrices, puis étagées, puis mesurables positives et enfin intégrables.

5.1.c) Propriétés fondamentales de l'espérance conditionnelle

Propriété 5.1.6. Soient X, Y deux v.a. intégrables et $\mathcal{A}$ une sous tribu de $\mathcal{F}$.

1. On a $\mathbb{E}[\mathbb{E}[X|\mathcal{A}]] = \mathbb{E}[X]$.
2. Si X est $\mathcal{A}$ -mesurable, alors $X = \mathbb{E}[X|\mathcal{A}]$ p.s.
3. $\forall a, b \in \mathbb{R}$ $\mathbb{E}[aX + bY|\mathcal{A}] = a\mathbb{E}[X|\mathcal{A}] + b\mathbb{E}[Y|\mathcal{A}]$ p.s.
4. Si $X \leq Y$ p.s., alors $\mathbb{E}[X|\mathcal{A}] \leq \mathbb{E}[Y|\mathcal{A}]$ p.s.
5. Si X est indépendante de $\mathcal{A}$, alors $\mathbb{E}[X|\mathcal{A}] = \mathbb{E}[X]$ p.s.
6. Si $\mathbb{E}[|XY|] < \infty$ et Y est $\mathcal{A}$ -mesurable, $\mathbb{E}[XY|\mathcal{A}] = Y\mathbb{E}[X|\mathcal{A}]$ p.s.

Preuve. Les propriétés 1., 2., 3., 4. découlent de l'unicité et du fait que l'espérance conditionnelle dans L^1 est un prolongement de celle de L^2 .

5. Si on a indépendance, $\mathbb{E}[XZ] = \mathbb{E}[X]\mathbb{E}[Z] = \mathbb{E}[\mathbb{E}[X|\mathcal{A}]Z]$ pour toute v.a. Z $\mathcal{A}$ -mesurable telle que $XZ \in L^1$. De plus, $\mathbb{E}[X]$ une v.a. $\mathcal{A}$ -mesurable. On conclut par unicité.
6. $\mathbb{E}[XY\mathbf{1}_A] = \mathbb{E}[X(Y\mathbf{1}_A)] = \mathbb{E}[\mathbb{E}[X|\mathcal{A}](Y\mathbf{1}_A)] = \mathbb{E}[(\mathbb{E}[X|\mathcal{A}]Y)\mathbf{1}_A]$ pour tout $A \in \mathcal{A}$, avec $\mathbb{E}[X|\mathcal{A}]Y$ une v.a. $\mathcal{A}$ -mesurable. On conclut par unicité. $\square$

Propriété 5.1.7. Soit X une v.a. intégrable et $\mathcal{A}, \mathcal{B}$ deux sous-tribus de $\mathcal{F}$.

1. Si $\mathcal{A} \subset \mathcal{B}$, alors $\mathbb{E}[\mathbb{E}[X|\mathcal{B}]|\mathcal{A}] = \mathbb{E}[\mathbb{E}[X|\mathcal{A}]|\mathcal{B}] = \mathbb{E}[X|\mathcal{A}]$ p.s..
2. π -SYSTÈMES : Soit $\mathcal{M}$ un π -système qui engendre $\mathcal{A}$. S'il existe une v.a. Z , $\mathcal{A}$ -mesurable, telle que $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[Z\mathbf{1}_A]$ pour tout $A \in \mathcal{M}$, alors $Z = \mathbb{E}[X|\mathcal{A}]$ p.s.
3. Si X et $\mathcal{A}$ sont indépendantes de $\mathcal{B}$, alors $\mathbb{E}[X|\mathcal{A} \vee \mathcal{B}] = \mathbb{E}[X|\mathcal{A}]$ p.s., où $\mathcal{A} \vee \mathcal{B} = \sigma(\mathcal{A} \cup \mathcal{B})$.

Preuve. 1. Théorème des trois perpendiculaires dans le cadre de la géométrie si on considère $X \in L^2$. On étend ensuite le résultat aux variables L^1 .

2. Preuve par classe monotone.

3. On a $\mathcal{A} \vee \mathcal{B} = \sigma(\mathcal{A} \cup \mathcal{B}) = \sigma\{A \cup B : A \in \mathcal{A}, B \in \mathcal{B}\} = \sigma\{A \cap B : A \in \mathcal{A}, B \in \mathcal{B}\}$.

Comme $\mathcal{A} \vee \mathcal{B}$ est engendrée par $A \cap B$ avec $A \in \mathcal{A}$ et $B \in \mathcal{B}$ (qui forme un π -système) et $\mathbf{1}_{A \cap B} = \mathbf{1}_A \mathbf{1}_B$, on a par indépendance $\mathbb{E}[X\mathbf{1}_A \mathbf{1}_B] = \mathbb{E}[\mathbb{E}[X\mathbf{1}_A|\mathcal{B}]\mathbf{1}_B] = \mathbb{E}[X\mathbf{1}_A]\mathbb{E}[\mathbf{1}_B] = \mathbb{E}[\mathbb{E}[X|\mathcal{A}]\mathbf{1}_A]\mathbb{E}[\mathbf{1}_B] = \mathbb{E}[\mathbb{E}[X|\mathcal{A}]\mathbf{1}_A \mathbf{1}_B]$. $\square$

Propriété 5.1.8. Inégalité de Jensen

Soient $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{A}$ une sous-tribu de $\mathcal{F}$ et $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction convexe.

Si $\mathbb{E}[|\varphi(X)|] < \infty$, alors $\varphi(\mathbb{E}[X|\mathcal{A}]) \leq \mathbb{E}[\varphi(X)|\mathcal{A}]$ p.s.

Preuve. Même preuve que l'inégalité de Jensen pour l'espérance (basée sur la linéarité de l'espérance). $\square$

On montre aussi les résultats suivants de la même manière que pour l'espérance.

Propriété 5.1.9. On considère $(X_n)_{n \geq 0}$ une suite de v.a. de $L^1(\Omega, \mathcal{F}, \mathbb{P})$, X une v.a. et $\mathcal{A}$ une sous-tribu de $\mathcal{F}$.

1. Convergence monotone

Si $(X_n)_{n \geq 0}$ suite croissante de v.a. positives qui converge vers X p.s. et $\mathbb{E}[X] < \infty$, alors $(\mathbb{E}[X_n|\mathcal{A}])_{n \geq 0}$ suite croissante de v.a. positives qui converge vers $\mathbb{E}[X|\mathcal{A}]$ p.s.

2. Convergence dominée

Si $(X_n)_{n \geq 0}$ converge vers X p.s. et $\mathbb{E}[\sup_{n \geq 0} |X_n|] < \infty$, alors

$$\mathbb{E}[|X_n - X||\mathcal{A}] \xrightarrow[n \rightarrow \infty]{} 0 \text{ p.s.}$$

3. Lemme de Fatou

Si $(X_n)_{n \geq 0}$ suite de v.a. positives intégrables avec $\mathbb{E}[\liminf_n X_n] < \infty$, alors

$$\mathbb{E}\left[\liminf_n X_n|\mathcal{A}\right] \leq \liminf_n \mathbb{E}[X_n|\mathcal{A}] \text{ p.s.}$$

5.2 La notion de loi conditionnelle

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ et X une v.a. à valeurs dans $\mathbb{R}^d$ intégrable.

Soit $\mathcal{A}$ une sous-tribu de $\mathcal{F}$. Alors il existe une application, notée $\mathbb{P}_X(\cdot|\mathcal{A})$, sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, appelée loi de probabilité conditionnelle de X sachant $\mathcal{A}$ telle que

(i) pour tout $B \in \mathcal{B}(\mathbb{R}^d)$, $\mathbb{P}_X(B|\mathcal{A}) = \mathbb{P}(X \in B|\mathcal{A})$ avec $\mathbb{P}(X \in B|\mathcal{A}) := \mathbb{E}[\mathbf{1}_{X \in B}|\mathcal{A}]$

(ii) pour tout $\omega \in \Omega$, $\mathbb{P}_X(\cdot|\mathcal{A})(\omega)$ est une probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$

$\mathbb{E}[X|\mathcal{A}]$ est alors l'espérance de la loi conditionnelle $\mathbb{P}_X(\cdot|\mathcal{A})$.

Remarque. Une loi conditionnelle est une loi aléatoire.

Comme $\mathbb{E}[X|Y]$ est une variable $\sigma(Y)$ -mesurable, il existe une fonction mesurable φ telle que $\mathbb{E}[X|Y] = \varphi(Y)$. La quantité $\mathbb{E}[X|Y = y]$ n'a pas de sens en général. La notation $\mathbb{E}[X|Y = y]$ désigne juste $\varphi(y)$.

On retrouve l'espérance de la loi conditionnelle basée sur la formule de Bayes dans le cas discret.

Proposition 5.2.1. Si Y est une variable discrète à valeurs dans un ensemble discret $\mathbb{D}_Y$, on a pour tout $y \in \mathbb{D}_Y$ (i.e. $\mathbb{P}(Y = y) \neq 0$),

$$\mathbb{E}[X|Y = y] = \frac{1}{\mathbb{P}(Y = y)} \mathbb{E}[X \mathbf{1}_{Y=y}].$$

Preuve. Soit $h : \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction mesurable bornée. On note $\varphi(Y) = \mathbb{E}[X|Y]$. Par définition de l'espérance conditionnelle, on a $\mathbb{E}[Xh(Y)] = \mathbb{E}[\varphi(Y)h(y)]$. De plus, en utilisant Fubini,

$$\begin{aligned} \mathbb{E}[Xh(Y)] &= \sum_{y \in \mathbb{D}_Y} \mathbb{E}[X \mathbf{1}_{Y=y}] h(y) \\ \mathbb{E}[\varphi(Y)h(y)] &= \sum_{y \in \mathbb{D}_Y} \varphi(y) \mathbb{P}(Y = y) h(y). \end{aligned}$$

Comme ceci est vrai pour toute fonction h , on déduit que

$$\mathbb{E}[X|Y = y] = \varphi(y) = \frac{\mathbb{E}[X\mathbb{1}_{Y=y}]}{\mathbb{P}(Y = y)}.$$

□

Proposition 5.2.2 (Vecteurs à densité). *Soit (X, Y) un vecteur aléatoire de $\mathbb{R}^p \times \mathbb{R}^q$ à densité $f_{(X,Y)}$ par rapport à la mesure de Lebesgue et $\psi : \mathbb{R}^p \rightarrow \mathbb{R}$ telle que $\mathbb{E}[|\psi(X)|] < \infty$ alors $\mathbb{P}_Y$ – p.s.*

$$\mathbb{E}[\psi(X)|Y = y] = \int_{\mathbb{R}^p} \psi(x) \frac{f_{(X,Y)}(x, y)}{f_Y(y)} dx.$$

Preuve. En effet, pour toute fonction h mesurable bornée, par Fubini, on a

$$\mathbb{E}[\psi(X)h(Y)] = \iint \psi(x)h(y)f_{(X,Y)}(x, y)dxdy = \int \left[\int \psi(x) \frac{f_{(X,Y)}(x, y)}{f_Y(y)} dx \right] h(y)f_Y(y)dy.$$

Par ailleurs, si $\varphi(Y) = \mathbb{E}[\psi(X)|Y]$, on a

$$\mathbb{E}[\psi(X)h(Y)] = \mathbb{E}[\varphi(Y)h(Y)] = \int \varphi(y)h(y)f_Y(y)dy.$$

De l'égalité pour toutes fonctions h , on en déduit que $\varphi(y) = \int_{\mathbb{R}^p} \psi(x) \frac{f_{(X,Y)}(x, y)}{f_Y(y)} dx$.

La densité conditionnelle de X sachant $\{Y = y\}$ est alors définie comme

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)}.$$

□

Proposition 5.2.3 (v.a. indépendantes). *Soit X, Y deux v.a. de $\mathbb{R}^p$ et $\mathbb{R}^q$ respectivement et $\psi : \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}$ telle que $\mathbb{E}[|\psi(X, Y)|] < \infty$. Si X et Y indépendantes, alors p.s.*

$$\mathbb{E}[\psi(X, Y)|Y] = \int_{\mathbb{R}^p} \psi(x, Y)\mathbb{P}_X(dx).$$

Preuve. Par Fubini, pour toute fonction h bornée $\mathbb{E}[\psi(X, Y)h(Y)] = \iint \psi(x, y)h(y)\mathbb{P}_X(dx) \otimes \mathbb{P}_Y(dy) = \int \left[\int \psi(x, y)\mathbb{P}_X(dx) \right] h(y)\mathbb{P}_Y(dy) = \mathbb{E}[\int \psi(x, Y)\mathbb{P}_X(dx)h(Y)]$. □

5.2.a) Cas gaussien

Définition 5.2.4. *Soit $Z = (Z_1, \dots, Z_d)$ un vecteur aléatoire de $\mathbb{R}^d$. On dit que Z est un vecteur gaussien si pour tout vecteur $a \in \mathbb{R}^d$, $a^t Z$ est une variable aléatoire gaussienne :*

$\forall a = (a_1, \dots, a_d) \in \mathbb{R}^d$, la variable $a_1 Z_1 + \dots + a_d Z_d$ est une variable gaussienne.

5.3 Exercices sur l'espérance conditionnelle

Exercice 5.3.1. Loi Binomiale, Loi de Poisson

- Soient X_1, X_2 deux variables indépendantes de loi binomiale de paramètres respectifs (n_1, p) et (n_2, p) .
 - Déterminer la loi de X_1 sachant $X_1 + X_2 = n$.
 - Déterminer $\mathbb{E}[X_1 | X_1 + X_2]$.
- On suppose que $X_1, \dots, X_p$ sont p v.a. de loi de Poisson indépendantes de paramètre respectifs $\lambda_1, \dots, \lambda_p$.
 - Déterminer la loi de $X_1, \dots, X_{p-1}$ conditionnellement à $X_1 + \dots + X_p = n$.
 - Déterminer $\mathbb{E}[X_1 | X_1 + X_2]$.

Exercice 5.3.2. Soit $X_1, \dots, X_n$ des variables aléatoires i.i.d. et intégrables. On définit $S_n = X_1 + \dots + X_n$.

- Montrer que $\mathbb{E}[X_1 | S_n] = \mathbb{E}[X_i | S_n]$ pour tout $i \in \{1, \dots, n\}$.
- En déduire l'expression de $\mathbb{E}[X_1 | S_n]$.

Exercice 5.3.3. Variance conditionnelle

On définit la variance conditionnelle

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

Vérifier que

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X|Y)] + \text{Var}(\mathbb{E}[X|Y]).$$

Exercice 5.3.4. Somme aléatoires de v.a.r.

On souhaite étudier le nombre personne ayant subi un accident au travail au cours d'une année. On note N le nombre d'accidents sur l'année et on introduit $(X_i)_{i \geq 1}$ une suite de variables i.i.d où chaque X_i représente le nombre de travailleurs touchés par l'accident numéro i . On suppose les variables N et $(X_i)_{i \geq 1}$ indépendantes. Donner une expression du nombre moyen de personnes accidentées sur son lieu de travail au cours d'une année et de sa variance.

Exercice 5.3.5. Cas discret

On considère $(X_1, \dots, X_n)$ des variables i.i.d de loi uniforme sur $\{-1, 0, 1\}$ et on pose

$$T_0 = \sum_{i=1}^n \mathbb{1}_{X_i=0} \quad T_1 = \sum_{i=1}^n \mathbb{1}_{X_i=1}.$$

- Calculer $\mathbb{E}[f(X_i) | T_0, T_1]$ où f est une fonction mesurable.
- On pose $U = T_0 + T_1$. Calculer $\mathbb{E}[f(X_i) | U]$.
- Déterminer la loi conditionnelle de T_1 sachant $U = k$ pour $k \in \{0, \dots, n\}$.

Exercice 5.3.6. Loi conditionnelle du minimum sachant le maximum

Soient $X_1, \dots, X_n$ des v.a.r. indépendantes de densité $f(x)$. On pose $X = \max(X_1, \dots, X_n)$ et $Y = \min(X_1, \dots, X_n)$.

- Déterminer la loi conditionnelle de Y sachant $X = x$.
- Exprimer $\mathbb{E}[Y|X]$ lorsque f est la densité uniforme sur $[0, 1]$.

Exercice 5.3.7. Indépendance et conditionnement

1. Soient X et Y deux v.a.r.. Montrer que X et Y sont indépendantes si et seulement si pour toute fonction g borélienne bornée, on a

$$\mathbb{E}[g(Y)|X] = \mathbb{E}[g(Y)] \text{ p.s.}$$

2. Soit (X, Y) un couple de densité $f(x, y) = e^{-y} \mathbb{1}_{0 < x < y}$. Calculer la loi conditionnelle de Y sachant $X = x$. En déduire que X et $Y - X$ sont indépendantes.

Exercice 5.3.8. Soient $X, Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ telles que

$$\mathbb{E}[X|Y] = Y \text{ p.s. et } \mathbb{E}[Y|X] = X \text{ p.s.}$$

Montrer que $\mathbb{P}(X = Y) = 1$.

Exercice 5.3.9. Soit X et Y deux variables dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$, et $\mathcal{A}$ une sous-tribu de $\mathcal{F}$. Établir l'inégalité de Cauchy-Schwartz conditionnelle :

$$\mathbb{E}[XY|\mathcal{A}]^2 \leq \mathbb{E}[X^2|\mathcal{A}]\mathbb{E}[Y^2|\mathcal{A}] \text{ p.s.s}$$

On pourra considérer la quantité : $\mathbb{E}[(X + aY)^2|\mathcal{A}]$.

Exercice 5.3.10. Soit (X, Y, Z) un vecteur gaussien centré de matrice de covariance

$$\Gamma = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

1. Montrer que $\mathbb{E}[X|Y, Z] = \frac{Y+Z}{3}$.
2. Montrer que $\mathbb{E}[X|Z] = \frac{Z}{2}$.
3. Montrer que le vecteur $(Y + Z, Z)$ est gaussien et donner sa matrice de covariance.
4. En déduire $\mathbb{E}[Y + Z|Z]$.
5. Conclure que l'on a bien $\mathbb{E}[\mathbb{E}[X|Y, Z]|Z] = \mathbb{E}[X|Z]$.

Exercice 5.3.11. Calculer $\mathbb{E}[X|Y]$ lorsque la loi du couple (X, Y) admet la densité suivante par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$:

1. $f_{(X,Y)}(x, y) = 4y(x - y)e^{-(x+y)} \mathbb{1}_{0 \leq y \leq x}(x, y)$,
2. $f_{(X,Y)}(x, y) = \frac{1}{y} e^{-\frac{x}{y} - y} \mathbb{1}_{(0, +\infty)^2}(x, y)$.

5.4 Les martingales

5.4.a) Définition et Premières propriétés

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité.

- Définition 5.4.1.**
1. Une **filtration** est une suite croissante $(\mathcal{F}_n)_{n \geq 0}$ de sous-tribus de $\mathcal{F}$: $\mathcal{F}_n \subset \mathcal{F}_{n+1}$.
 2. Un processus $(X_n)_{n \geq 0}$ est dit **adapté** à une filtration $(\mathcal{F}_n)_{n \geq 0}$ si pour tout $n \geq 0$, X_n est $\mathcal{F}_n$ -mesurable.
 3. On appelle **filtration naturelle** d'un processus $(X_n)_{n \geq 0}$ la filtration définie par $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$.

En modélisation, n représente souvent le temps et $\mathcal{F}_n$ l'historique jusqu'à l'instant n d'une expérience aléatoire.

Définition 5.4.2. Soit $(\mathcal{F}_n)_{n \geq 0}$ une filtration. Un processus $(M_n)_{n \geq 0}$ à valeurs réelles est une $(\mathcal{F}_n)_{n \geq 0}$ -*(sur ; sous)* martingale si

1. le processus $(M_n)_{n \geq 0}$ est $(\mathcal{F}_n)_{n \geq 0}$ -adapté,
2. $\forall n \geq 0$, M_n est intégrable,
3. $\forall n \geq 0$, $M_n = \mathbb{E}[M_{n+1} | \mathcal{F}_n]$ p.s. ($\geq$; $\leq$)

En cas de sur-martingale : M_n sur-estime l'espérance conditionnelle, et

en cas de sous-martingale : M_n sous-estime l'espérance conditionnelle.

Propriété 5.4.3. 1. Si $(M_n)_{n \geq 0}$ est une $(\mathcal{F}_n)_{n \geq 0}$ -*(sur ; sous)* martingale, alors $\forall p \geq 0$, $M_n = \mathbb{E}[M_{n+p} | \mathcal{F}_n]$ ($\geq$; $\leq$).

En particulier l'espérance est constant, décroissante ou encore croissante, selon qu'on a une martingale, sur-martingale ou sous-martingale.

2. Soit Z une v.a. intégrable et $(\mathcal{F}_n)_{n \geq 0}$ une filtration. Alors le processus $(\mathbb{E}[Z | \mathcal{F}_n])_{n \geq 0}$ est une $(\mathcal{F}_n)_{n \geq 0}$ -martingale.
3. Soit φ une fonction convexe et $(M_n)_{n \geq 0}$ une $(\mathcal{F}_n)_{n \geq 0}$ -martingale, alors $(\varphi(M_n))_{n \geq 0}$ est une $(\mathcal{F}_n)_{n \geq 0}$ -sous martingale.

Remarque. Si $(M_n)_{n \geq 0}$ est une $(\mathcal{F}_n)_{n \geq 0}$ -martingale, alors $(M_n^+)_{n \geq 0}$, $(|M_n|)_{n \geq 0}$ et $(M_n \vee C)_{n \geq 0}$ sont des $(\mathcal{F}_n)_{n \geq 0}$ -sousmartingales, pour $C \in \mathbb{R}$.

5.4.b) Quelques exemples de martingales

Marche aléatoire Soit $(X_n)_{n \geq 0}$ des variables i.i.d. dans L^1 de moyenne m et $(\mathcal{F}_n)_{n \geq 0}$ la filtration naturelle associée.

On considère la marche aléatoire $(S_n)_{n \geq 0}$ avec pour tout $n \geq 0$

$$S_n = X_0 + \dots + X_n$$

Le processus $(S_n)_{n \geq 0}$ est une sur-martingale, sous-martingale ou martingale en fonction du signe de m . Si $m = 0$, on dit que la marche aléatoire est symétrique.

Urne de Polya Une urne contient N boules de couleurs $1, 2, \dots, m$ en quantités $N^1, N^2, \dots, N^m$ avec $N^i \geq 1$ et $N^1 + \dots + N^m = N$. À chaque instant $n \geq 1$, on tire une boule au hasard puis on la remet dans l'urne en ajoutant une boule de même couleur.

On note N_n^i le nombre de boules de couleur i et X_n^i la proportion de boules i à l'instant n .

On note $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ où $X_n = (X_n^1, \dots, X_n^m)$. Alors $(X_n^i)_{n \geq 0}$ est une martingale bornée. $\mathcal{F}_n$ est la filtration engendrée par toutes les boules dans l'urne jusqu'à l'instant n .

En effet, on a $X_n^i = N_n^i / (n + N)$. Par construction, sachant $\mathcal{F}_n$, $N_{n+1}^i - N_n^i = 1$ avec proba X_n^i et $N_{n+1}^i - N_n^i = 0$ avec proba $1 - X_n^i$, d'où

$$\mathbb{E}[N_{n+1}^i - N_n^i | \mathcal{F}_n] = X_n^i,$$

et donc $\mathbb{E}[X_{n+1}^i - X_n^i | \mathcal{F}_n] = 0$.

Processus de Galton-Watson En 1873, Galton s'est inquiété de la disparition des noms de familles nobles en Angleterre. Avec l'aide du mathématicien Watson, il a étudié ce problème. Ce modèle est très utilisé en dynamique des populations et en génétique.

On suppose que

- à l'instant initial, $t = 0$ on a un seul ancêtre
- les générations ne se chevauchent pas
- chaque individu à la génération n donne naissance à des enfants de façon indépendante et selon la même loi $p = (p_0, p_1, \dots)$ pour tout le monde, loi que l'on connaît, avec $p_0 > 0$.

On note μ le nombre moyen d'enfant par individu et σ^2 la variance de la loi p .

Le nombre d'individu Z_{n+1} à la génération $n + 1$ est égal à la somme des enfants de chacun des individus de la génération n , soit

$$Z_{n+1} = \sum_{k=1}^{Z_n} X_k^n,$$

où X_k^n est le nombre d'enfant du $k^{\text{ème}}$ individu de la génération n . Les variables $(X_k^n)_{k \geq 0}$ sont indépendantes de Z_n . Le but est d'étudier l'évolution de la taille de la population $(Z_n)_{n \geq 0}$.

Ce type de processus est aussi appelé **processus de branchement** ou **de ramification**.

La suite $(Z_n)_{n \geq 0}$ est une chaîne de Markov à espace d'état $S = \mathbb{N}$ et l'état 0 est absorbant.

Si $p_0 > 0$, alors tous les états non nuls sont transitoires, car on a une probabilité non nulle de passer de l'état i à l'état 0 égale à $(p_0)^i$ (les i individus n'ont aucun enfant) et de cet état on ne peut pas retourner en i (l'état 0 est absorbant).

La taille moyenne de la population et sa variance vérifient

$$\begin{aligned} \mathbb{E}[Z_n] &= \mathbb{E}[X] \mathbb{E}[Z_{n-1}] = \mu \mathbb{E}[Z_{n-1}] \\ &= \mu^n. \end{aligned}$$

$$\text{Var}(Z_n) = \mu^2 \text{Var}(Z_{n-1}) + \sigma^2 \mathbb{E}[Z_{n-1}] = \mu^2 \text{Var}(Z_{n-1}) + \sigma^2 \mu^{n-1}$$

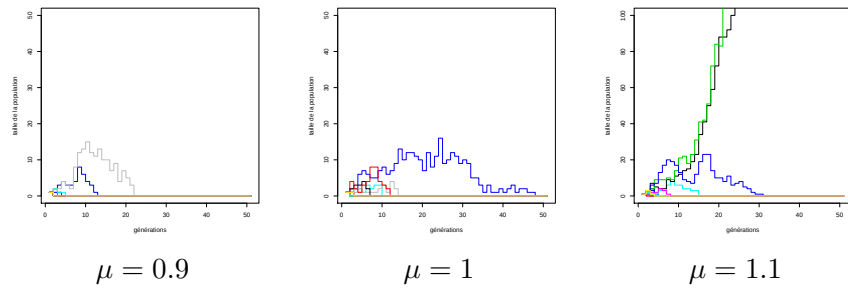
Par conséquent, comme $X_0 = 1$

$$\begin{aligned} \text{Var}(Z_1) &= \sigma^2 \\ \text{Var}(Z_2) &= \mu^2 \sigma^2 + \sigma^2 \mu = \sigma^2 \mu (\mu + 1) \\ \text{Var}(Z_3) &= \mu^2 \sigma^3 (\mu + 1) + \sigma^2 \mu^2 = \sigma^2 \mu^2 (\mu^2 + \mu + 1) \\ &\vdots \\ \text{Var}(Z_n) &= \sigma^2 \mu^{n-1} (\mu^{n-1} + \mu^{n-2} + \dots + 1) = \sigma^2 \mu^{n-1} \sum_{k=0}^{n-1} \mu^k \end{aligned}$$

On a donc

$$\text{Var}(Z_n) = \begin{cases} \sigma^2 \mu^{n-1} \frac{1-\mu^n}{1-\mu} & \text{si } \mu \neq 1 \\ n \sigma^2 & \text{si } \mu = 1 \end{cases}$$

Dans les graphes ci-dessous on a tracé plusieurs trajectoires de X_n en fonction de n pour différentes valeurs de μ :



Si $\mu < 1$, le nombre moyen d'individu dans la population et sa variance tendent vers 0 quand $n \rightarrow \infty$.

Si $\mu = 1$, le nombre moyen d'individu dans la population reste constant, mais sa variance part à l'infini quand $n \rightarrow \infty$.

Si $\mu > 1$, les deux quantités tendent vers $+\infty$ quand $n \rightarrow \infty$.

On remarque que $M_n = \frac{Z_n}{\mu^n}$ est une martingale positive pour la filtration naturelle de $(Z_n)_{n \geq 0}$.

On va introduire dans la suite des théorèmes limites permettant de montrer que

- si $\mu \leq 1$ la population s'éteint p.s.,
- si $\mu > 1$ la probabilité d'extinction de la population est strictement inférieure à 1.

5.5 Martingales et temps d'arrêt

Soit $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbb{P})$ un espace de probabilité filtré.

Définition 5.5.1. Une variable aléatoire T à valeurs dans $\bar{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$ est appelée $(\mathcal{F}_n)_{n \geq 0}$ -temps d'arrêt si

$$\forall n \geq 0 \quad \{T = n\} \in \mathcal{F}_n.$$

De manière équivalente, T est un $(\mathcal{F}_n)_{n \geq 0}$ -temps d'arrêt si $\forall n \geq 0 \quad \{T \leq n\} \in \mathcal{F}_n$, ou encore si $\{T > n\} \in \mathcal{F}_n$.

On définit la tribu $\mathcal{F}_T$ la tribu des événements antérieurs à T

$$\begin{aligned} \mathcal{F}_T &= \{A \in \mathcal{F} : A \cap \{T = n\} \in \mathcal{F}_n \quad \forall n \geq 0\} \\ &= \{A \in \mathcal{F} : A \cap \{T \leq n\} \in \mathcal{F}_n \quad \forall n \geq 0\}. \end{aligned}$$

Il s'agit bien d'une tribu, car

- $\Omega \in \mathcal{F}_T$ car T est un temps d'arrêt ;
- si $A \in \mathcal{F}_T$, on a $A^c \cap \{T \leq n\} = (A \cap \{T \leq n\})^c \cap \{T \leq n\} \in \mathcal{F}_n$;
- si $(A_i)_{i \geq 0} \in \mathcal{F}_T$, alors $(\cup_i A_i) \cap \{T \leq n\} = \cup_i (A_i \cap \{T \leq n\}) \in \mathcal{F}_n$.

Exemple. Temps d'atteinte d'un borélien.

Soit $(X_n)_{n \geq 0}$ un processus $(\mathcal{F}_n)_{n \geq 0}$ -adapté et B un borélien de $\mathbb{R}$ alors $T = \inf \{n \geq 0 : X_n \in B\}$ est un $(\mathcal{F}_n)_{n \geq 0}$ -temps d'arrêt.

Par contre $M = \sup \{n \geq 0 : X_n \in B\}$ n'est pas un temps d'arrêt en général.

On note $x \wedge y = \min(x, y)$ et $x \vee y = \max(x, y)$.

Propriété 5.5.2. Soient S, T deux $(\mathcal{F}_n)_{n \geq 0}$ -temps d'arrêts. On a

1. $S \wedge T$, $S \vee T$, et $S + T$ sont des $(\mathcal{F}_n)_{n \geq 0}$ -temps d'arrêts.

2. Si $Y \in L^1$, alors

$$\mathbb{E}[Y|\mathcal{F}_T] = \sum_{n \in \bar{\mathbb{N}}} \mathbb{E}[Y|\mathcal{F}_n] \mathbb{1}_{\{T=n\}}.$$

3. Si $(X_n)_{n \geq 0}$ $(\mathcal{F}_n)_{n \geq 0}$ -adapté, alors X_T est une v.a. $\mathcal{F}_T$ -mesurable, où $\forall \omega \in \Omega$, $X_T(\omega) := X_{T(\omega)}(\omega)$.

4. $\mathcal{F}_{S \wedge T} = \mathcal{F}_S \cap \mathcal{F}_T$. Par conséquent, si $S \leq T$ p.s., alors $\mathcal{F}_S \subset \mathcal{F}_T$.

Preuve.

1. En exercice.

2. En effet, soit $A \in \mathcal{F}_T$, par Fubini,

$$\mathbb{E}[Y \mathbb{1}_A] = \sum_{n \in \bar{\mathbb{N}}} \mathbb{E}[Y \mathbb{1}_{A \cap \{T=n\}}] = \mathbb{E} \left[\sum_{n \in \bar{\mathbb{N}}} \mathbb{E}[Y|\mathcal{F}_n] \mathbb{1}_{\{T=n\}} \mathbb{1}_A \right]$$

3. Soit A un borélien, on veut montrer que $\{X_T \in A\} \in \mathcal{F}_T$, i.e. $\{X_T \in A\} \cap \{T = n\} \in \mathcal{F}_n$ ce qui est clair car le processus est $(\mathcal{F}_n)_{n \geq 0}$ -adapté.

4. Soit $A \in \mathcal{F}_{S \wedge T}$, alors $A \cap \{S \leq n\} = A \cap \{S \wedge T \leq n\} \cap \{S \leq n\}$. Comme $\{S \leq n\} \in \mathcal{F}_n$ et $A \cap \{S \wedge T \leq n\} \in \mathcal{F}_n$, on a $A \cap \{S \leq n\} \in \mathcal{F}_n$ et alors $A \in \mathcal{F}_S$. De même $A \in \mathcal{F}_T$.

Réciproquement, si $A \in \mathcal{F}_S \cap \mathcal{F}_T$, alors $A \cap \{S \leq n\} \in \mathcal{F}_n$ et $A \cap \{T \leq n\} \in \mathcal{F}_n$, donc $A \cap \{S \wedge T \leq n\} = A \cap \{S \leq n\} \cap \{T \leq n\} \in \mathcal{F}_n$.

□

Proposition 5.5.3. Si $(M_n)_{n \geq 0}$ une $(\mathcal{F}_n)_{n \geq 0}$ -martingale (*sur* ou *sous*) et T un $(\mathcal{F}_n)_{n \geq 0}$ -temps d'arrêt, alors $(M_n^T)_{n \geq 0}$ une $(\mathcal{F}_n)_{n \geq 0}$ -martingale (*sur* ou *sous*), où $M_n^T := M_{T \wedge n}$.

Preuve. $M_n^T = M_n \mathbb{1}_{T > n} + M_T \mathbb{1}_{T \leq n}$ est bien $\mathcal{F}_n$ -mesurable et dans L^1 et

$$\mathbb{E}[M_{n+1}^T | \mathcal{F}_n] = \mathbb{E}[M_{n+1} | \mathcal{F}_n] \mathbb{1}_{T > n} + M_T \mathbb{1}_{T \leq n} = M_n^T$$

□

Théorème 5.5.4 (Théorème d'arrêt). On considère $(M_n)_{n \geq 0}$ est une $(\mathcal{F}_n)_{n \geq 0}$ -martingale (*sur* ou *sous*) et deux $(\mathcal{F}_n)_{n \geq 0}$ -temps d'arrêt, T et S , avec T borné.

Alors sur $\{S \leq T\}$, $M_S = \mathbb{E}[M_T | \mathcal{F}_S]$ (resp. $\geq$ et $\leq$).

Preuve.

1. Supposons que $S \leq T$ p.s.

Si $S = p$ déterministe et supposons T borné par K . $(M_n^T)_{n \geq 0}$ est une martingale, donc pour $n \geq p$ $\mathbb{E}[M_n^T | \mathcal{F}_p] = M_p^T = M_{T \wedge p}$. Prenons $n = K$, comme $T \geq p$ et $M_K^T = M_T$ on a $\mathbb{E}[M_T | \mathcal{F}_p] = M_{T \wedge p} = M_p$.

Si S est aléatoire,

$$\mathbb{E}[M_T | \mathcal{F}_S] = \sum \mathbb{E}[M_T | \mathcal{F}_n] \mathbb{1}_{S=n} = \sum M_n \mathbb{1}_{S=n} = M_S.$$

2. Cas général. On pose $\tilde{S} = S \wedge T$. On a $\mathbb{E}[M_T | \mathcal{F}_{\tilde{S}}] = M_{\tilde{S}}$. Comme sur $\{S \leq T\}$ on a $S = \tilde{S}$, on obtient le résultat.

Conséquence : Si $S \leq T$ p.s., on a $\mathbb{E}[M_S] = \mathbb{E}[M_T]$. □

Remarque. Si (M_n) est une martingale bornée, il suffit d'avoir T fini p.s. pour utiliser le théorème d'arrêt (introduire $\tau_n = T \wedge n$ et utiliser le théorème de convergence dominée).

Exemple. On considère $(M_n)_{n \geq 0}$ la marche aléatoire symétrique issue de 0 sur $\mathbb{Z}$. On définit $T = \inf \{n \geq 0 : M_n = 1\}$. Alors $T < \infty$ p.s.

Par contre, $\mathbb{E}[M_T] = 1 \neq \mathbb{E}[M_0]$ donc l'hypothèse de "bornitude" est essentielle.

5.6 Inégalités sur les martingales

On remarque $M_n^* = \sup_{k \leq n} M_k$ est une sous-martingale car suite croissante de v.a. intégrables et adaptée. Par l'inégalité de Markov, on a

$$\mathbb{P}(M_n^* \geq C) \leq \frac{\mathbb{E}[M_n^*]}{C}$$

On va voir qu'avec les martingales $\mathbb{E}[M_n^*]$ peut-être remplacé par $\mathbb{E}[|M_n|]$.

Proposition 5.6.1 (Lemme Maximal). *Soit $C > 0$ fixé.*

Si $(M_n)_{n \geq 0}$ est une $(\mathcal{F}_n)_{n \geq 0}$ -martingale, alors

$$P(\sup_{k \leq n} |M_k| \geq C) \leq \frac{1}{C} \mathbb{E}[|M_n|].$$

Preuve. On considère le temps d'arrêt $T = \inf \{n \geq 0 : |M_n| \geq C\}$. On a le lien suivant $\{T \leq n\} = \{M_n^* \geq C\}$. T est un temps d'arrêt, mais a priori pas borné, on introduit donc $\tau_n = T \wedge n$. On remarque que $|M_n|$ est une sous martingale. On a par l'inégalité de Markov

$$\mathbb{P}(M_n^* \geq C) = \mathbb{P}(T \leq n, |M_T| \geq C) \leq \mathbb{E}[|M_T| \mathbf{1}_{T \leq n}] / C$$

Comme $M_T = M_{\tau_n}$ sur $\{T \leq n\}$,

$$\mathbb{P}(M_n^* \geq C) \leq \mathbb{E}[|M_{\tau_n}| \mathbf{1}_{T \leq n}] / C \leq \mathbb{E}[|M_{\tau_n}|] / C \leq \mathbb{E}[|M_n|] / C$$

par le théorème d'arrêt (avec $S = \tau_n$ et $T = n$).

□

Conséquence. Si $(M_n)_{n \geq 0}$ est une $(\mathcal{F}_n)_{n \geq 0}$ -martingale, alors

$$P(\sup_{k \geq 0} |M_k| \geq C) \leq \frac{1}{C} \sup_{k \geq 0} \mathbb{E}[|M_k|].$$

Proposition 5.6.2 (Inégalité du nombre de montées).

1. Soit $(M_n)_{n \geq 0}$ est une surmartingale et $\overline{m}_{a,b}^M(n)$ le nombre de montées de a à b effectuées par M avant l'instant n . Alors

$$\mathbb{E}[\overline{m}_{a,b}^M(n)] \leq \frac{1}{b-a} \mathbb{E}[(M_n - a)^-].$$

2. Soit $(M_n)_{n \geq 0}$ est une sousmartingale et $\underline{m}_{a,b}^M(n)$ le nombre de descentes de b à a effectuées par M avant l'instant n . Alors

$$\mathbb{E}[\underline{m}_{a,b}^M(n)] \leq \frac{1}{b-a} \mathbb{E}[(M_n - b)^+].$$

Preuve. On remarque que 2. implique 1. car $-M_n$ est une sousmartingale et $\overline{m}_{a,b}^M(n) = \underline{m}_{a,-b}^{-M}(n)$.

Prouvons 2. On décrit les périodes de descente avec la suite croissante de temps d'arrêt suivants : $T_0 = 0$, $T_1 = \inf\{n \leq 0 : M_n > b\}$, $T_2 = \inf\{n \leq T_1 : M_n < a\}$, $\dots$, $T_{2p-1} = \inf\{n \leq T_{2p-2} : M_n > b\}$, $T_{2p} = \inf\{n \leq T_{2p-1} : M_n < a\}$, $\dots$.

Pour tout p , $M_{n \wedge T_{2p-1}} - b$ est une sousmartingale positive sur $\{T_{2p-1} \leq n\}$. On a donc

$$\begin{aligned} 0 &\leq \mathbb{E}[(M_{n \wedge T_{2p-1}} - b)\mathbb{1}_{T_{2p-1} \leq n}] \leq \mathbb{E}[(M_{n \wedge T_{2p}} - b)\mathbb{1}_{T_{2p-1} \leq n}] \quad (\text{théorème d'arrêt}) \\ 0 &\leq \mathbb{E}[(M_{T_{2p}} - b)\mathbb{1}_{T_{2p} \leq n}] + \mathbb{E}[(M_n - b)\mathbb{1}_{T_{2p-1} \leq n < T_{2p}}] \\ 0 &\leq (a - b)\mathbb{P}(T_{2p} \leq n) + \mathbb{E}[(M_n - b)^+ \mathbb{1}_{T_{2p-1} \leq n < T_{2p}}] \end{aligned}$$

Par conséquent,

$$(b - a) \sum_p \mathbb{P}(\underline{m}_{a,b}^M(n) \leq p) \leq \sum_p \mathbb{E}[(M_n - b)^+ \mathbb{1}_{T_{2p-1} \leq n < T_{2p}}]$$

c'est à dire

$$(b - a)\mathbb{E}[\underline{m}_{a,b}^M(n)] \leq \mathbb{E}[(M_n - b)^+].$$

□

Proposition 5.6.3 (Inégalité de Doob).

Soit $(M_n)_{n \geq 0}$ est une $(\mathcal{F}_n)_{n \geq 0}$ -martingale (ou sousmartingale positive), $\forall p, q > 1$ tels que $1 = \frac{1}{p} + \frac{1}{q}$. Alors

$$\| \sup_{0 \leq k \leq n} |M_k| \|_p \leq q \sup_{0 \leq k \leq n} \|M_k\|_p,$$

où $\|\cdot\|_p$ est la norme L^p .

Démonstration. $(|M_n|)_{n \geq 0}$ est une sous martingale positive par Jensen conditionnel. On sait que, si X v.a. positive

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

On pose $X_n = M_n \mathbb{1}_{|M_n| > a/2}$ et on définit la martingale,

$$Z_k = \mathbb{E}[X_n | \mathcal{F}_k], \quad 0 \leq k \leq n$$

Par ailleurs, $M_n^* \leq Z_n^* + a/2$. En effet,

$$\begin{aligned} |M_k| &= |\mathbb{E}[M_n | \mathcal{F}_k]| = |\mathbb{E}[X_n | \mathcal{F}_k] + \mathbb{E}[M_n \mathbb{1}_{|M_n| \leq a/2} | \mathcal{F}_k]| \\ &\leq |\mathbb{E}[X_n | \mathcal{F}_k]| + a/2 = |Z_k| + a/2 \end{aligned}$$

Par le lemme maximal,

$$\begin{aligned} \mathbb{P}(M_n^* > a) &\leq \mathbb{P}(Z_n^* > a/2) \leq \frac{2}{a} \mathbb{E}[|Z_n|] \\ &\leq \frac{2}{a} \mathbb{E}[|X_n|] = \frac{2}{a} \mathbb{E}[|M_n| \mathbb{1}_{|M_n| > a/2}] \end{aligned}$$

Par conséquent, par ce qui précède et Fubini

$$\begin{aligned} \mathbb{E}[M_n^{*p}] &= p \int_0^\infty z^{p-1} \mathbb{P}(M_n^* > z) dz \leq p \int_0^\infty z^{p-1} \frac{2}{z} \mathbb{E}[|M_n| \mathbb{1}_{|M_n| > z/2}] dz \\ &\leq 2p \mathbb{E} \left[\int_0^{2|M_n|} z^{p-2} dz |M_n| \right] = \frac{2^p p}{p-1} \mathbb{E}[|M_n|^p] \end{aligned}$$

Il faut travailler encore un peu plus pour avoir la bonne constante.

□

5.7 Décomposition de Doob

Définition 5.7.1. Un processus $(X_n)_{n \geq 0}$ est dit $(\mathcal{F}_n)_{n \geq 0}$ -**prévisible** si pour tout $n \geq 0$, X_n est $\mathcal{F}_{n-1}$ -mesurable (avec la convention $\mathcal{F}_{-1} = \{\emptyset, \Omega\}$).

Théorème 5.7.2 (Décomposition de Doob).

Soit $(X_n)_{n \geq 0}$ un processus $(\mathcal{F}_n)_{n \geq 0}$ -adapté avec $X_n \in L^1, \forall n \geq 0$.

Il existe une unique décomposition

$$X_n = X_0 + M_n + A_n$$

avec

- $(M_n)_{n \geq 0}$ une $(\mathcal{F}_n)_{n \geq 0}$ -martingale nulle en zéro,
- $(A_n)_{n \geq 0}$ un processus prévisible nul en zéro.

Par ailleurs, $(X_n)_{n \geq 0}$ est une sous martingale ssi $(A_n)_{n \geq 0}$ est un processus croissant.

Preuve. Si X a une décomposition de Doob, comme M martingale et A prévisible

$$\mathbb{E}[X_n - X_{n-1} | \mathcal{F}_{n-1}] = 0 + (A_n - A_{n-1})$$

Par conséquent,

$$A_n = \sum_{k=1}^n \mathbb{E}[X_k - X_{k-1} | \mathcal{F}_{k-1}].$$

ce qui permet de définir A .

Unicité : s'il existe deux décomposition (M, A) et (N, B) alors $M_n - N_n = B_n - A_n$ alors $M_n - N_n$ est une martingale $\mathcal{F}_{n-1}$ mesurable donc constante. $\square$

Remarque. Considérons un jeu de hasard (loto, casino, investissement boursier,...).

À chaque instant n , le joueur joue une certaine somme d'argent ϕ_n . Le résultat du jeu à l'instant n est une certaine variable aléatoire X_n qui dépend d'événements plus ou moins complexes. La mise du joueur peut-être le fruit d'une stratégie compliquée mais dans la mesure où le joueur ne connaît pas le résultat du jeu avant de l'avoir joué, le processus (ϕ_n) est forcément prévisible.

5.8 Théorèmes limites et Martingales

5.8.a) Convergence p.s.

Théorème 5.8.1. Soit $(M_n)_{n \geq 0}$ une sous-martingale telle que $\sup_n \mathbb{E}[M_n^+] < \infty$. Alors il existe une variable intégrable M_∞ telle que $M_n \rightarrow M_\infty$ p.s.

On remarque que l'hypothèse est équivalente à (M_n) est bornée dans L^1 , i.e. $\sup_n \mathbb{E}[|M_n|] < \infty$, alors $\sup_n \mathbb{E}[M_n^+] < \infty$.

Corollaire 5.8.2. Une sur-martingale positive converge p.s. vers une variable intégrable.

Preuve. $(-M_n)_{n \geq 0}$ est une sous martingale négative et $M_n^+ = 0$ donc converge p.s. vers une v.a. intégrable. Par ailleurs, en utilisant le lemme de Fatou, on remarque que $(M_n)_{n \in \mathbb{N}}$ est une surmartingale :

$$\mathbb{E}[M_\infty | \mathcal{F}_n] \leq \liminf_{p \rightarrow \infty} \mathbb{E}[M_{n+p} | \mathcal{F}_n] \leq M_n.$$

Attention! La convergence p.s. n'implique pas la convergence dans L^1 . $\square$

Exemple. Soit $(X_i)_{i \geq 1}$ une suite de v.a. iid telles que $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 2) = 1/2$.

Alors $M_n = \prod_{i=1}^n X_i$ est une martingale positive, donc qui converge p.s.. De plus, elle est bornée dans L^1 , mais par contre, elle ne converge pas dans L^1 .

Preuve. (Théorème 5.8.1)

D'après la Proposition 5.6.2, le nombre de descentes vérifie $\mathbb{E}[\underline{m}_{a,b}^M(\infty)] < \infty$, donc $\underline{m}_{a,b}^M(\infty) < \infty$ p.s.

Soit N l'ensemble de mesure nulle sur lequel $\underline{m}_{a,b}^M(\infty) = \infty$ pour tout $a, b \in \mathbb{Q}$. Soit $\omega \notin N$. Si $M_n(\omega)$ ne converge pas, alors il existe $a, b \in \mathbb{Q}$:

$$\liminf M_n(\omega) < a < b < \limsup M_n(\omega).$$

Contradiction, donc la martingale converge p.s.. Notons M_∞ sa limite.

Par ailleurs, comme $|M_n| = 2M_n^+ - M_n$ et $-M$ est une surmartingale, on a

$$\mathbb{E}[|M_n|] \leq 2\mathbb{E}[M_n^+] - \mathbb{E}[M_0]$$

Donc $\sup_n \mathbb{E}[|M_n|] < \infty$. D'après le lemme de Fatou,

$$\mathbb{E}[|M_\infty|] \leq \liminf \mathbb{E}[|M_n|] \leq \sup_n \mathbb{E}[|M_n|] < \infty.$$

Donc $M_\infty \in L^1$. □

5.8.b) Convergence dans L^1

Théorème 5.8.3. Soit $(M_n)_{n \geq 0}$ une martingale. Il y a équivalence entre les propriétés suivantes :

1. $\exists Z \in L^1$ tel que $\forall n \in \mathbb{N} \ M_n = \mathbb{E}[Z | \mathcal{F}_n]$,
2. $(M_n)_{n \geq 0}$ est U.I.,
3. $(M_n)_{n \geq 0}$ converge dans L^1 .

On a alors $Z = M_\infty$, la martingale est dite fermée, i.e. $(M_n)_{n \in \mathbb{N}}$ est une martingale.

Preuve. $\underline{1} \Rightarrow \underline{2}$: On a

$$\mathbb{E}[|M_n| \mathbb{I}_{|M_n| > a}] \leq \mathbb{E}[|Z| \mathbb{I}_{|M_n| > a}].$$

Par ailleurs, en utilisant l'inégalité de Markov

$$\mathbb{P}(|M_n| > a) \leq \frac{\mathbb{E}[|Z|]}{a}.$$

Comme $Z \in L^1$, $\lim_{a \rightarrow +\infty} \sup_{n \leq 0} \mathbb{E}[|M_n| \mathbb{I}_{|M_n| > a}] = 0$.

$\underline{2} \Rightarrow \underline{3}$: d'après le théorème précédent, on a M_n converge p.s. vers M_∞ , donc converge dans L^1 car U.I.

$\underline{3} \Rightarrow \underline{1}$: On a $M_n = \mathbb{E}[M_{n+p} | \mathcal{F}_n]$ pour tout $p \geq 0$. Comme l'espérance conditionnelle est continue sur L^1 , en faisant tendre $p \rightarrow +\infty$, d'où le résultat.

Conséquence $M_n = \mathbb{E}[M_\infty | \mathcal{F}_n]$ □

5.8.c) Martingales dans L^2

Théorème 5.8.4. Soit (M_n) une martingale bornée dans L^2 , alors elle converge p.s. et dans L^2 vers une v.a. M_∞ .

Preuve. (M_n) une martingale bornée dans L^2 signifie

$$\sup_n \mathbb{E}[M_n^2] = \mathbb{E}[M_0^2] + \sum_{k \geq 1} \mathbb{E}[(M_k - M_{k-1})^2] < \infty$$

En effet $M_n = M_0 + \sum_{k \geq 1} (M_k - M_{k-1})$ qui est une somme de termes orthogonaux, on conclut par Pythagore. Comme bornée dans L^2 , elle est bornée dans L^1 . Par notre premier thm de CV, elle converge p.s. vers M_∞ . Par ailleurs, par Pythagore

$$\mathbb{E}[(M_{n+p} - M_n)^2] = \sum_{k=n+1}^{n+p} \mathbb{E}[(M_k - M_{k-1})^2]$$

On prend $p \rightarrow \infty$ et par le Lemme de Fatou ($\mathbb{E}[\liminf] \leq \liminf \mathbb{E}[\]$), on a

$$\mathbb{E}[(M_\infty - M_n)^2] \leq \sum_{k=n+1}^{\infty} \mathbb{E}[(M_k - M_{k-1})^2]$$

Comme la série CV, $\lim_n \mathbb{E}[(M_\infty - M_n)^2] = 0$ cqfd. $\square$

Lemme 5.8.5 (Lemme de Kronecker). Soit $(y_n)_{n \geq 1}$ une suite de nombres réels. Si la série $\sum_{k \geq 1} y_k/k$ converge alors

$$\lim_{n \rightarrow +\infty} \frac{\sum_{k=1}^n y_k}{n} = 0.$$

Théorème 5.8.6 (Loi des grands nombres). Soit (M_n) une martingale dans L^2 et $Y_n = M_n - M_{n-1}$ telle que

$$\sum_{n \geq 1} \frac{1}{n^2} \mathbb{E}[Y_n^2] < \infty.$$

Alors $\frac{M_n}{n} \xrightarrow[n \rightarrow \infty]{} 0$ p.s. et dans L^2 .

Preuve. On remarque que $M_n = \sum_{k \leq n} Y_k - M_0$. Par Kronecker, il suffit de montrer que $\sum_{k \geq 1} Y_k/k$ CV p.s. Soit $W_n = \sum_{k=0}^n \frac{Y_k}{k}$. C'est une martingale. Par ailleurs, $\mathbb{E}[W_n^2] = \sum_{k=0}^n \mathbb{E}[\frac{Y_k^2}{k^2}]$ qui est bornée. Donc W_n est une martingale bornée dans L^2 , par conséquent elle converge p.s. et dans L^2 . $\square$

5.9 Exercices sur les martingales

Exercice 5.9.1. Soit $(\mathcal{F}_n)_{n \geq 0}$ une filtration. Soit $(X_n)_{n \geq 0}$ une sous-martingale telle que les variables X_n ont toutes la même loi.

1. Montrer que la suite $(X_n)_{n \geq 0}$ est en fait une martingale, ainsi que $(X_n \vee a)_{n \geq 0}$ et $(X_n \wedge a)_{n \geq 0}$, où a est un réel fixé.
2. Soit $n > m$ deux entiers et $a \in \mathbb{R}$. Montrer que sur l'ensemble $\{X_m > a\}$, on a $X_n \geq a$ p.s. En déduire que la suite $(X_n)_{n \geq 0}$ est constante p.s.

Exercice 5.9.2. Soient S, T deux $(\mathcal{F}_n)_{n \geq 0}$ -temps d'arrêts. Montrer que $S \wedge T$, $S \vee T$, et $S + T$ sont des $(\mathcal{F}_n)_{n \geq 0}$ -temps d'arrêts.

Exercice 5.9.3. [Processus de Galton-Watson]

Vérifier que le processus $M_n = Z_n/\mu^n$ est une martingale, où (Z_n) est le processus de Galton Watson introduit dans la section 5.4.b). Étudier son comportement quand $n \rightarrow \infty$.

Exercice 5.9.4. [Modèle de Wright-Fisher]

Soit k, N deux entiers tels que $0 < k < N$.

On souhaite connaître l'évolution de la fréquence d'un allèle dans une population de taille N , pour un gène se présentant sous deux allèles seulement. Chaque individu possède un seul exemplaire du gène, sous la forme d'un des deux allèles A ou B . On fait les hypothèses suivantes :

- la population reste de taille constante N au cours du temps,
- les générations ne se chevauchent pas : à chaque instant n , la n ème génération meurt et donne naissance aux N individus de la génération $n + 1$,
- chaque enfant choisit son parent uniformément parmi tous les individus de la génération précédente et indépendamment des autres ;
- la reproduction à l'instant n ne dépend pas des reproductions précédentes.

On introduit la suite de variables $(X_n)_{n \geq 0}$, où X_n est le nombre d'allèles A dans la population à l'instant n . On suppose qu'initialement il y a k individus avec l'allèle A : $X_0 = k$.

1. Vérifier la suite $(X_n)_{n \geq 0}$ est définie par induction de la façon suivante : $X_0 = k$, et pour tout $n \geq 0$, $i \in \{0, \dots, N\}$, la loi conditionnelle de X_{n+1} sachant $X_n = i$ est la loi binomiale $\mathcal{B}(N, \frac{i}{N})$.
2. Montrer que $(X_n)_{n \geq 0}$ est une martingale.
3. Montrer que lorsque n tend vers l'infini, $(X_n)_{n \geq 0}$ converge p.s. et dans L^p pour tout $p \geq 1$.
4. Décrire la loi de la variable limite X_∞ .

Exercice 5.9.5. *Théorème de Wald*

Considérons (M_n) une $(\mathcal{F}_t)_{t \geq 0}$ -martingale dans L^2 et T un $(\mathcal{F}_t)_{t \geq 0}$ -temps d'arrêt tel que $\mathbb{E}[M_T^2] < \infty$. Vérifier que $\mathbb{E}[M_T] = \mathbb{E}[M_0]$.

Exercice 5.9.6. *Ruine du joueur*

Un joueur a une richesse égale initiale égale à a et dont le but est d'atteindre la richesse $b > a$ avant d'être ruiné, $a, b \in \mathbb{N}^*$. Il répète de manière indépendante un jeu où il gagne 1 euro avec probabilité $p \in]0, 1[$ et perd 1 euro avec probabilité $q = 1 - p$. Le gain du joueur à l'issue de la n ème épreuve est $S_n = \sum_{i=1}^n X_i$ où X_i i.i.d. de loi $\mathcal{B}(p)$.

1. Trouver λ et β tels que $M_n = \lambda^{S_n}$ et $N_n = S_n - n\beta$ soit des martingales pour la filtration naturelle.
2. Montrer que la durée du jeu T est fini p.s., sachant que le joueur s'arrête dès qu'il est ruiné ou dès qu'il atteint la richesse b .
3. Calculer la probabilité que le joueur soit ruiné à l'issue du jeu et en déduire la durée moyenne du jeu.

Bibliographie

- [BL07] Philippe Barbe and Michel Ledoux. *Probabilité*. EDP Sciences, 2007.
- [BP06] Marc Briane and Gilles Pages. *Théorie de l'intégration*. Vuibert, 2006.
- [Dur19] Rick Durrett. *Probability—theory and examples*, volume 49 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, fifth edition, 2019.
- [JP03] Jean Jacod and Philip Protter. *Probability essentials*. Universitext. Springer-Verlag, Berlin, second edition, 2003.
- [Wil91] David Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.