Université Paris Dauphine
Master 1 de Mathématiques et Applications
2024-2025

Pré-rentrée analyse (et application aux probabilités)

Pierre Cardaliaguet

Table des matières

0. Quelques rappels d'algèbre linéaire 2
0.1. Théorème du rang 2
0.2. Diagonalisation 2
0.3. Formes quadratiques 3
1. Topologie et analyse fonctionnelle 3
1.1. Espaces métriques 3
1.2. Espaces vectoriels normés, espaces de Banach 5
1.3. Espaces de Hilbert 7
1.4. Quelques espaces fonctionnels classiques 8
2. Calcul différentiel 9
2.1. Différentiabilité 9
2.2. Inversion locale et fonctions implicites 11
2.3. Equations différentielles 12
3. Intégration 13
3.1. Espaces mesurés, mesures 13
3.2. Contruction de l'intégrale de Lebesgue 14
3.3. Propriétés centrales de l'intégrale de Lebesgue 16
3.4. Les espaces $\mathcal{L}^p$ et L^p 18
3.5. Fonctions définies par des intégrales 19
3.6. Formule de changement de variables 20
3.7. Transformée de Fourier d'une fonction 21
3.8. Produit de convolution 21
4. Application au calcul des probabilités 22
4.1. Vocabulaire 22
4.2. Variance 23
4.3. Indépendance 24
4.4. Convergence de variables aléatoires 24
4.5. Vecteurs gaussiens 26
Annexe A. Exercices supplémentaires 27
A.1. Exercices de topologie et d'analyse fonctionnelle 27
A.2. Exercices de calcul différentiel 28
A.3. Exercices d'intégration 28
A.4. Exercices de probabilité 29
Index 32

Il est bien évidemment impossible de résumer en 12h et quelques pages les 3 années du programme d'analyse de Licence. Les notes qui suivent ne sont que des rappels sommaires de ces cours. Comme ce sont des rappels, les notions ne sont pas nécessairement abordées dans un ordre naturel (en particulier, les exemples et les exercices font souvent appel à des notions évoquées plus loin).

Dans ce qui suit : Les “notions de base”, non détaillées, sont supposées très bien connues. Si ce n’est pas le cas, il est vivement conseillé au lecteur de consulter les cours cités en référence pour les revoir au plus vite. Les “théorèmes centraux” figurent en encadré. Enfin, les exercices proposés sont très classiques et du niveau d’un étudiant de fin de L3.

Notation : (Fonction indicatrice) Si A est un sous-ensemble d’un ensemble X, on note $\mathbf{1}_A : X \to \{0,1\}$ la fonction indicatrice de A, i.e., la fonction définie par $\mathbf{1}_A(x) := 1$ si $x \in A$ et $\mathbf{1}_A(x) := 0$ sinon, pour tout $x \in X$.

0. Quelques rappels d’algèbre linéaire

Dans cette partie, on rappelle quelques résultats de base en algèbre linéaire. *Les espaces vectoriels sont ici toujours de dimension finie*. Cette partie est très largement empruntée au cours de Guillaume Legendre “Algèbre linéaire 3”.

0.1. Théorème du rang.

Notions de base : théorie de la dimension (espace vectoriel, sous-espace vectoriel, famille libre, famille génératrice, base, dimension) ; applications linéaires, somme et composition d’applications linéaires, endomorphismes.

Théorème 0.1. *Soient E et F deux espaces vectoriels* **de dimension finie** *(sur $\mathbb{R}$ ou sur $\mathbb{C}$) et $f : E \to F$ une application linéaire. Alors*

— *f est injective, si et seulement si, $Ker(f) := f^{-1}(\{0\}) = \{0\}$ (vrai aussi en dimension infinie).*

— *(Théorème du rang)* $\dim(Im(f)) + \dim(Ker(f)) = \dim(E)$.

— *(Conséquence du théorème du rang) Si* $\dim(E) = \dim(F)$, *alors f est injective, si et seulement si, f est sujective (auquel cas f est bien sûr bijective).*

0.2. Diagonalisation.

Notions de base : matrice représentant une application linéaire, lien entre produit matriciel et composition. Matrice inversible et lien avec les applications linéaires bijectives, déterminant. Matrices semblables. Valeurs propres et vecteurs propres d’un endomorphisme, lien avec le polynôme caractéristique.

On note $M_{n,p}(\mathbb{R})$ (resp. $M_{n,p}(\mathbb{C})$) l’ensemble des matrices réelles (resp. complexes) à n lignes et p colonnes. Si $p = n$, on note $M_n(\mathbb{R}) := M_{n,n}(\mathbb{R})$ et $M_n(\mathbb{C}) := M_{n,n}(\mathbb{C})$. Si $A \in M_{n,p}(K)$ (où $K = \mathbb{R}$ ou $K = \mathbb{C}$), A^T est la matrice transposée de A (i.e., $A^T \in M_{p,n}(K)$ avec, si $A = (a_{i,j})_{1 \le i \le n,\ 1 \le j \le p}$, $A^T = (a_{j,i})_{1 \le i \le p,\ 1 \le j \le n}$).

Le **polynôme caractéristique** d’une matrice $A \in M_n(K)$ (où $K = \mathbb{R}$ ou $K = \mathbb{C}$) est le polynôme $\chi_A(X) = \det(XI_n - A)$ (où I_n est la matrice identité de format $n \times n$). C’est un polynôme unitaire de degré n. On a

$$\chi_{A^T} = \chi_A, \qquad \chi_{AB} = \chi_{BA} \qquad \text{si } A, B \in M_n(K).$$

En particulier, deux matrices semblables ont même polynôme caractéristique, et on peut ainsi définir le polynôme caractéristique χ_f d’un endomorphisme $f : E \to E$ comme le polynôme caractéristique de n’importe quelle matrice représentant f. Rappelons que λ est une valeur propre de A, si et seulement si, λ est une racine de χ_A. On appelle alors **multiplicité géométrique de** λ la dimension de $\mathrm{Ker}(\lambda I_d - A)$. La **multiplicité algébrique de** λ est l’ordre de multiplicité de λ comme racine de χ_A (i.e., l’entier $r \in \{1, \ldots, n\}$ tel que $(X - \lambda)^r$ divise χ_A et $(X - \lambda)^{r+1}$ ne divise par χ_A).

Théorème 0.2 (Diagonalisation)**.** *Soit E un espace de dimension finie n sur $K = \mathbb{R}$ ou $K = \mathbb{C}$, et $f : E \to E$ un endomorphisme. On dit que f est diagonalisable si f vérifie une des (et donc toutes les) assertions équivalentes suivantes :*

(1) Il existe une base de E dans laquelle la matrice de f est diagonale (cette base est nécessairement constituée de vecteurs propres de f).

(2) Il existe une base de E constituée de vecteurs propres de f.

(3) Si $A \in M_n(K)$ est la matrice représentant f dans une base $\mathcal{B}$ de E, A est diagonalisable, i.e., il existe une matrice $P \in M_n(K)$ inversible telle que la matrice $P^{-1}AP$ est diagonale (alors les colonnes de la matrice P sont les coefficients dans la base $\mathcal{B}$ des vecteurs d'une base de vecteurs propres de f).

(4) Si $A \in M_n(K)$ est la matrice représentant f dans une base $\mathcal{B}$ de E, χ_A est scindé[1] et toute valeur propre de A a une multiplicité géométrique égale à sa multiplicité algébrique.

Trigonalisation. On dit qu'un endomorphisme $f : E \to E$ est **trigonalisable** s'il existe une base dans laquelle la matrice de f est triangulaire supérieure. On montre que f est trigonalisable, si et seulement si, son polynôme caractéristique est scindé (en particulier, dans $\mathbb{C}$, tout endomorphisme est trigonalisable).

Polynôme d'endomorphisme. Si $f : E \to E$ est un endomorphisme et $P(X) = \sum_{k=0}^{m} a_k X^k$, on note $P(f)$ l'endomorphisme $P(f) = \sum_{k=0}^{m} a_k f^{(k)}$ (où $f^{(k)} = \underbrace{f \circ \cdots \circ f}_{(k \text{ fois})}$ et, par convention $f^{(0)} = id_E$). Le théorème de **Cayley-Hamilton** affirme que $\chi_f(f) = 0$. Le **polynôme minimal** de f est le plus petit (au sens de la division euclidienne) polynôme unitaire P tel que $P(f) = 0$. Le théorème de Cayley-Hamilton affirme donc que le polynôme minimal divise le polynôme caractéristique. De plus, on peut montrer que f est diagonalisable, si et seulement si, le polynôme minimal de f est scindé et n'a que des racines simples.

0.3. **Formes quadratiques.**

Notions de base : Application bilinéaire, application bilinéaire symétrique, représentation matricielle et changement de base. Forme quadratique, forme polaire associée. Signature d'une forme quadratique.

Théorème 0.3 (spectral). *Toute matrice $A \in M_n(\mathbb{R})$ symétrique admet une base orthonormale de vecteurs propres réels : il existe $O \in M_n(\mathbb{R})$ telle que $OO^T = O^T\, O = I_n$ (dont les colonnes sont des vecteurs propres de A) et $D := O^T AO$ est diagonale (les coefficients de D étant les valeurs propres de A).*

Positivité. On rappelle qu'une forme quadratique $q : E \to \mathbb{R}$ est positive si $q(x) \geq 0$ pour tout $x \in E$. Elle est définie positive si $q(x) > 0$ pour tout $x \in E^*$. Une conséquence du théorème spectral est que q est positive (resp. définie positive), si et seulement si, sa matrice (dans n'importe quelle base) a toutes ses valeurs propres positives (resp. strictement positives).

1. Topologie et analyse fonctionnelle

Cette partie est largement empruntée au cours de Paul Pegon "Mise à niveau en analyse".

1.1. **Espaces métriques.**

1.1.1. *Notions de base.* Distance, ouverts, fermés, intersections d'ouverts ou de fermés, adhérence, convergence de suites, fonctions continues, fonctions lipschitziennes, ...

Exemples d'espaces métriques : $\mathbb{R}$ muni de la valeur absolue, $\mathbb{R}^n$ ou $\mathbb{C}^n$ (avec $n \in \mathbb{N}^*$) muni de la norme euclidienne, les espaces vectoriels normés (EVN, dont les espaces fonctionnels classiques, voir plus bas), les sous-ensembles d'espaces vectoriels normés munis de la distance induite par la norme de l'EVN, etc...

1. On dit qu'un polynôme est scindé s'il est le produit de polynômes de degré 1. Dans $\mathbb{C}$, tout polynôme est scindé.

1.1.2. *Densité, séparabilité.*

Définition 1.1. *Soit (X, d) un espace métrique. On dit qu'un sous-ensemble A de X est dense dans X si tout ouvert de X contient un point de A.*

De façon équivalente, A est dense dans X si, pour tout $x \in X$, il existe une suite d'éléments de A qui tend vers x.

Les exemples classiques sont : $\mathbb{Q}$ est dense dans $\mathbb{R}$ (et donc $\mathbb{Q}^N$ est dense dans $\mathbb{R}^N$) ; la densité des polynômes dans l'ensemble des fonctions continues sur un intervalle $[a, b]$ de $\mathbb{R}$ pour la norme uniforme, et la densité des polynômes trigonométriques dans l'espace des fonctions continues $2\pi-$périodiques (voir le théorème 1.20, version du théorème de Stone-Weierstrass) ; les fonctions étagées sont denses dans les L^p (voir le théorème 3.10) ; de même, l'ensemble $C_c^\infty(\mathbb{R})$ des fonctions C^∞ à support compact est dense dans $L^p(\mathbb{R}, dx)$ (muni de la mesure de Lebesgue, voir le théorème 3.11).

Définition 1.2. *Soit (X, d) un espace métrique. On dit que X est séparable s'il existe un ensemble dénombrable dense dans X (autrement dit s'il existe une suite $(x_n)_{n\in\mathbb{N}}$ d'éléments de X qui est dense dans X).*

Les exemples classiques sont : $\mathbb{R}^N$ est séparable (pour $N \in \mathbb{N}^*$) puisque $\mathbb{Q}^N$, qui est dénombrable, est dense dans $\mathbb{R}^N$; l'ensemble des fonctions continues sur un intervalle $[a, b]$ de $\mathbb{R}$ muni de la norme uniforme ; les espaces L^p pour $p \in [1, \infty[$ (par contre L^∞ n'est pas séparable).

Exercice 1. Montrer que si (X, d) un espace métrique séparable et si $Y \subset X$ est non vide, alors (Y, d) est également séparable.

Exercice 2. En utilisant la densité des polynômes dans $L^1([0, 1])$, montrer que, si $f \in L^1([0, 1])$ est telle que $\int_0^1 x^k f(x)dx = 0$ pour tout $k \in \mathbb{N}$, alors $f = 0$ p.p.

Des applications classiques de la densité sont :
- le fait que les coefficients de Fourier $c_n(f)$ d'une fonction de $f \in L^1([a, b])$ tend vers 0 lorsque $|n| \to \infty$
- l'extension à $L^2(\mathbb{R})$ de la transformée de Fourier (par densité de $L^1 \cap L^2$ dans L^2).
- en M1, dans le cours de "Processus discrets", on verra aussi l'extension de l'espérance conditionnelle de L^2 à L^1 (par densité de L^2 dans L^1).

Exercice 3. En admettant que l'espace $C_c^1(\mathbb{R})$ des fonctions de classe C^1 nulles en dehors d'un compact, est dense dans $L^1(\mathbb{R})$ (pour la norme L^1) montrer le lemme de Riemann-Lebesgue : si $f \in L^1(\mathbb{R})$,

$$\lim_{n\to+\infty} \int_{\mathbb{R}} \cos(nx)f(x)dx = 0.$$

1.1.3. *Complétude.*

Définition 1.3. *Une suite (x_n) d'un espace métrique (X, d) est dite de Cauchy si, pour tout $\varepsilon > 0$, il existe $N \in \mathbb{N}$ tel que $d(x_n, x_m) \leq \varepsilon$ pour tout $n, m \geq N$.*

Exercice 4. Montrer que toute suite convergente est de Cauchy.

Définition 1.4. *Un espace métrique (X, d) est dit complet si toute suite de Cauchy y possède une limite.*

Exercice 5. Montrer que tout sous-ensemble fermé d'un espace métrique complet est lui même complet (pour la distance induite).

Théorème 1.5 (de prolongement)**.** *Soit A une partie dense d'un espace métrique (X, d_X) et (Y, d_Y) un espace métrique complet. Si une fonction $f : A \to Y$ est uniformément continue sur A (pour la métrique d_X), alors il existe une unique fonction continue $\tilde{f} : X \to Y$ telle que $\tilde{f}(x) = f(x)$ pour tout $x \in A$. De plus $\tilde{f}$ est uniformément continue sur X.*

Théorème 1.6 (du point fixe)**.** *Soit (X, d) un espace métrique complet. On suppose que $f : X \to X$ est une application contractante, i.e., il existe un nombre réel k avec $0 \leq k < 1$ tel que*

$$d(f(x), f(y)) \leq kd(x, y) \qquad \forall x, y \in X.$$

Alors f possède un unique point fixe. De plus, si $x_0 \in X$ et si on définit la suite (x_n) par récurrence en posant $x_{n+1} = f(x_n)$ pour tout $n \in \mathbb{N}$, alors (x_n) converge vers l'unique point fixe de f.

Applications : le théorème 2.2 d'inversion locale ; le théorème 2.4 de Cauchy-Lipschitz, sur l'existence et l'unicité de la solution d'une équation différentielle (utilisé dans les cours "Optimisation" et "Analyse numérique, problèmes dépendant du temps").

1.1.4. *Compacité.* Soit (x_n) une suite dans un espace métrique (X, d). On rappelle qu'une suite extraite de (x_n) est n'importe quelle suite de la forme $(x_{\phi(n)})$ où $\phi : \mathbb{N} \to \mathbb{N}$ est strictement croissante. On note souvent une telle suite extraite (x_{n_k}). On voit facilement que, si (x_n) converge, alors toute suite extraite de (x_n) converge vers la limite de (x_n).

Définition 1.7. *On dit qu'un espace métrique (X, d) est compact si toute suite d'éléments de X possède une suite extraite qui converge dans X.*

Exercice 6. Montrer les assertions suivantes :

(1) tout sous-ensemble fermé d'un compact est compact.
(2) tout sous-ensemble compact d'un ensemble est fermé est borné (réciproque fausse en général).
(3) toute fonction continue sur un compact est bornée et atteint ses bornes (autrement dit, elle possède un maximum et un minimum).

Exercice 7 (un critère de convergence)**.** Soient (X, d) un espace métrique compact et (x_n) une suite d'éléments de X. On suppose qu'il existe $x \in X$ tel que toute suite extraite **convergente** de (x_n) converge vers x. Montrer que la suite (x_n) converge vers x.

1.2. **Espaces vectoriels normés, espaces de Banach.**

1.2.1. *Notions de base.* Norme, normes équivalentes, convergence en terme de norme, continuité en terme de norme. Noter qu'un espace vectoriel normé (EVN) est un espace métrique.

1.2.2. *Le cas de la dimension finie.*

Théorème 1.8. *Dans un espace vectoriel de dimension fini, toutes les normes sont équivalentes.*

Ce n'est pas le cas pour les espaces de dimension infinie.

Exercice 8. Par exemple, montrer que sur $X = C^0([0, 1])$ l'espace des fonctions continues sur $[0, 1]$, les normes $\|\cdot\|_\infty$ et $\|\cdot\|_1$ définies par

$$\|f\|_\infty = \sup_{x \in [0,1]} |f(x)| \qquad \text{et} \qquad \|f\|_1 := \int_0^1 |f(x)| dx \qquad \forall f \in X,$$

ne sont pas équivalentes (on pourra construire une suite de fonctions continues $f_n : [0, 1] \to [0, 1]$ telles que $f_n(0) = 1$ pour tout $n \in \mathbb{N}$ mais $\|f_n\|_1 \to 0$).

Une autre particularité des espaces de dimension finie est que les sous-ensembles compacts y sont assez “nombreux”.

Théorème 1.9. *Dans un espace vectoriel normé de dimension fini, les sous-ensembles compacts sont les sous-ensembles fermés et bornés.*

En dimension infinie, cette caractérisation est fausse (en fait, le théorème de Riesz affirme même qu'un EVN dont la boule unité fermée est compacte est de dimension finie). Dans le cours d' “Analyse fonctionnelle” du M1 sera introduite une topologie plus faible, comprenant plus de compacts. La notion de convergence en loi (en probabilité) est un moyen similaire de contourner cette difficulté.

Exercice 9. Montrer que, dans $\ell^1 := \{x = (x_k)_{k\in\mathbb{N}},\ \sum_{k=0}^{\infty} |x_k| < +\infty\}$ muni de la norme $\|x\|_1 = \sum_{k=0}^{\infty} |x_k|$, la suite (e^n) d'éléments de ℓ^1 définie par $e^n_k = \delta_{kn}$ (où $\delta_{kn} = 0$ si $k \neq n$ et $= 1$ si $k = n$) ne possède aucune sous-suite qui converge.

1.2.3. *Applications linéaires dans un EVN.*

Soient $(E, \|\cdot\|_E)$ et $(F, \|\cdot\|_F)$ deux EVN et $f : E \to F$ une application linéaire.

Théorème 1.10. *Les assertions suivantes sont équivalentes :*

(1) il existe $x_0 \in E$ tel que f est continue en x_0,

(2) f est continue sur E,

(3) $\sup_{\|x\|_E\le 1} \|f(x)\|_F$ est fini.

Si l'une de ces assertions est vraie, alors f est uniformément continue sur E. De plus $f \to \|f\|_{L(E,F)} := \sup_{\|x\|_E\le 1} \|f(x)\|_F$ est une norme sur l'espace vectoriel $L(E, F)$ des applications linéaires continues de E dans F.

Exercice 10. Soient $(E, \|\cdot\|_E)$ et $(F, \|\cdot\|_F)$ deux EVN, avec F complet. Montrer que l'ensemble $L(E, F)$ des applications linéaires continues de E dans F, muni de la norme $\|f\|_{L(E,F)}$, est également complet.

Un cas particulier est l'espace des formes linéaires continues de E (i.e., les applications linéaires continues de E dans $F := \mathbb{R}$), noté souvent E^*. Cet espace joue un rôle central dans le cours d'“Analyse fonctionnelle” de M1.

1.2.4. *Espaces de Banach.*

Définition 1.11. *Un espace de Banach est un EVN complet.*

La plupart des espaces fonctionnels classiques sont des espaces de Banach (voir la partie “Exemples d'espaces fonctionnels” plus bas).

Une propriété centrale des espaces de Banach est la convergence normale des séries :

Théorème 1.12 (Convergence normale)**.** *Soit $(X, \|\cdot\|)$ un espace de Banach. Si une suite (x_n) de X est normalement convergente, i.e., si la série de terme général $(\|x_n\|)$ converge, alors la série de terme général (x_n) converge (dans X) : autrement dit,*

$$\sum_{i=0}^{\infty} \|x_n\| < +\infty \qquad \Longrightarrow \qquad \lim_{m\to+\infty} \sum_{n=0}^{m} x_n \text{ existe (dans } X).$$

Exercice 11. Soit E un espace de Banach. On rappelle (cf. Exercice 10) que $L(E) := L(E, E)$ est également un espace de Banach. Pour $f \in L(E)$, on note $f^{(0)} = Id_E$ (l'application identité de E), $f^{(1)} = f$, $f^{(2)} = f \circ f$ et $f^{(n+1)} = f \circ f^{(n)}$ pour tout $n \in \mathbb{N}$. On suppose que $\|f\|_{L(E)} < 1$. Montrer que $Id_E - f$ est bijective et d'inverse continue avec $(Id_E - f)^{-1} = \lim_{N\to+\infty} \sum_{n=0}^{N} f^{(n)}$.

1.3. **Espaces de Hilbert.**

1.3.1. *Notions de base.* Produit scalaire, norme associée au produit scalaire, inégalité de Cauchy-Schwarz.

Définition 1.13. *Un espace de Hilbert est un espace vectoriel muni d'un produit scalaire, qui est complet pour la norme associée à ce produit scalaire. Lorsque cet espace est de dimension finie, on parle d'espace euclidien.*

Les espaces de Hilbert les plus souvent rencontrés en pratique sont les espaces euclidiens $\mathbb{R}^N$ muni de la norme euclidienne, les espaces L^2 en intégration et l'espace ℓ^2 des suites de carré sommable.

Exercice 12 (Identité du parallèlogramme)**.** Soit $(E, \|\cdot\|)$ un espace de Banach. Montrer que E est un espace de Hilbert (au sens où il existe un produit scalaire $\langle \cdot, \cdot \rangle$ tel que $\|x\|^2 = \langle x, x\rangle$ pour tout $x \in H$), si et seulement si, E vérifie l'identité du parallélogramme

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2) \qquad \forall x, y \in E.$$

1.3.2. *Le théorème de projection.*

Théorème 1.14 (de projection (cas d'un convexe fermé))**.** *Soit C un sous-ensemble convexe fermé d'un espace de Hilbert $(H, \langle \cdot, \cdot \rangle)$. Alors tout $x \in H$ possède une unique projection $\Pi_C(x)$ sur C, c'est-à-dire qu'il existe un unique élément $\Pi_C(X) \in C$ tel que*

$$\|x - y\| \geq \|x - \Pi_C(x)\| \qquad \forall y \in C.$$

De plus, $\Pi_C(x)$ est l'unique élément de C vérifiant l'inégalité variationnelle

$$\langle x - \Pi_C(x), y - \Pi_C(x)\rangle \leq 0 \qquad \forall y \in C.$$

L'aspect remarquable de ce résultat est que C n'est pas supposé compact. De plus il fournit une caractérisation de la projection. Ce résultat sera utilisé fréquemment en M1 Math, notamment en optimisation. La version qui suit, où C est un sous-espace vectoriel fermé, sera utilisé en M1 Math à peu près dans tous les cours (pour la définition de l'espérance conditionnelle en "Processus discrets", en statistique (moindres carrés), en "Série temporelle", etc...).

Exercice 13. On suppose que C est un sous-ensemble convexe fermé d'un espace de Hilbert $(H, \langle \cdot, \cdot \rangle)$. Montrer que Π_C est une application linéaire continue de H dans H, si et seulement si, C est un sous-espace vectoriel de H.

Théorème 1.15 (de projection (cas d'un sous-espace vectoriel fermé))**.** *Si C un sous-espace vectoriel fermé d'un espace de Hilbert H, alors l'application $x \to \Pi_C(x)$ (définie dans le théorème précédent) est une application linéaire continue (de norme 1). De plus, $\Pi_C(x)$ est l'unique élément de C vérifiant l'égalité*

$$\langle x - \Pi_C(x), y\rangle = 0 \qquad \forall y \in C.$$

(autrement dit $x - \Pi_C(x)$ appartient à l'orthogonal $C^\perp$ de C).

Exercice 14. Montrer le théorème 1.15 à partir du théorème 1.14.

Théorème 1.16 (de représentation de Riesz)**.** *Soit H un espace de Hilbert et $f : H \to \mathbb{R}$ une application linéaire continue. Alors il existe un unique élément $x_0 \in H$ tel que $f(x) = \langle x_0, x\rangle$ pour tout $x \in H$.*

1.3.3. *Orthogonal et bases hilbertiennes.*

Définition 1.17 (Orthogonal)**.** *Soit H un espace de Hilbert et A un sous-ensemble de A. L'orthogonal de A, noté $A^\perp$, est l'ensemble défini par*

$$A^\perp = \{x \in H \text{ tel que } \langle x, y\rangle = 0 \, \forall y \in A\}.$$

Exercice 15. Soit A un sous-ensemble non vide d'un espace de Hilbert H. Montrer que

(1) l'ensemble $A^{\perp}$ est un sous-espace vectoriel fermé de H,
(2) $A^{\perp} = (Vect(A))^{\perp}$, où $Vect(A)$ est le sous-espace vectoriel engendré par A,
(3) $H^{\perp} = \{0\}$

La dernière égalité dans l'exercice caractérise en fait H (c'est ici qu'on utilise le fait que H est complet) :

Théorème 1.18. *Soit A un* sous-espace vectoriel *de H. Alors*
$$(A^{\perp})^{\perp} = \overline{A} \qquad (\text{où } \overline{A} \text{ est la fermeture de } A).$$
En particulier, si $A^{\perp} = \{0\}$, alors A est dense dans H. Si A est fermé et $A^{\perp} = \{0\}$, alors $A = H$.

Définition 1.19 (Base hilbertienne)**.** *Une suite $(x_n)_{n\in\mathbb{N}}$ est une base hilbertienne d'un espace de Hilbert H si elle est orthonormée*
$$\langle x_n, x_m\rangle = \begin{cases} 0 & si\, n \neq m \\ 1 & si\, n = m \end{cases}$$
et totale :
$$(\{x_n,\ n \in \mathbb{N}\})^{\perp} = \{0\}.$$

Si H est séparable, alors H possède une base hilbertienne (cela se montre par le procédé d'orthonormalisation de Schmidt).

L'exemple emblématique de base hilbertienne est la famille constituée des fonctions $t \to \sqrt{2}\cos(2\pi nt)$ $(n \in \mathbb{N})$ et $t \to \sqrt{2}\sin(2\pi nt)$ $(n \in \mathbb{N}^*)$, qui est une base Hilbertienne de $L^2([0,1])$.

Exercice 16 (Inégalité et égalité de Bessel)**.** Vérifier que, si $(e_n)_{n\in\mathbb{N}}$ est une famille orthonormée d'un espace de Hilbert H, alors pour tout $x \in H$ on a $\displaystyle\sum_{n=0}^{\infty}\langle x, e_n\rangle^2 \leq \|x\|^2$. Montrer que si $(e_n)_{n\in\mathbb{N}}$ est une base hilbertienne de H, alors $\displaystyle\sum_{n=0}^{\infty}\langle x, e_n\rangle^2 = \|x\|^2$.

1.4. **Quelques espaces fonctionnels classiques.**

1.4.1. *Espaces des fonctions continues.* Si (X, d) est un espace métrique compact, alors l'ensemble $C^0(X)$ des applications continues de X dans $\mathbb{R}$, muni de la norme uniforme :
$$\|f\|_{\infty} = \sup_{x\in X}|f(x)| \qquad \forall f \in C^0(X),$$
est un espace de Banach [2].

Théorème 1.20 (Densité)**.**

— **(densité des polynômes)** *Si X est un sous-ensemble compact de $\mathbb{R}^n$ (où $n \in \mathbb{N}^*$), l'ensemble des (restrictions à X des) polynômes* [a] *forme une partie dense de $C^0(X)$ pour la norme $\|\cdot\|_{\infty}$. En particulier, $(C^0(X), \|\cdot\|_{\infty})$ est séparable.*

— **(densité des polynômes trigonométriques)** *L'ensemble des polynômes trigonométriques* [b] *est dense dans l'espace des fonctions continues et $2\pi-$périodiques de $\mathbb{R}$ dans $\mathbb{C}$.*

a. Un polynôme $P : \mathbb{R}^n \to \mathbb{R}$ est une expression de la forme $P(x) = \sum_{i\in\mathbb{N}^n} a_i x_1^{i_1}\dots x_n^{i_n}$ où seuls un nombre fini de coefficients $a_i \in \mathbb{R}$ sont non nuls et où $i = (i_1, \dots, i_n)$.

b. Un polynôme trigonométriques $P : \mathbb{R} \to \mathbb{C}$ est une expression de la forme $P(x) = \sum_{i=0}^{n} a_i e^{2i\pi x}$ où $n \in \mathbb{N}$, $a_i \in \mathbb{C}$.

2. Dans cette assertion, on peut remplacer $\mathbb{R}$ par n'importe quel espace de Banach $(E, \|\cdot\|_E)$. Dans ce cas, $\|f\|_{\infty} := \sup_{x\in X}\|f(x)\|_E$.

Ces deux résultats sont une conséquence du théorème de Stone-Weierstrass ci-dessous (et sont souvent connus sous ce nom). Rappelons qu'une sous-algèbre de $C^0(X)$ est un sous-espace vectoriel de $C^0(X)$ qui est stable par produit. On dit que cette sous-algèbre sépare les points si, pour tout $x_1, x_2 \in X$, il existe un élément f de cette sous-algèbre tel que $f(x_1) \neq f(x_2)$.

Théorème 1.21 (de Stone-Weierstrass)**.** *Soient (X, d) un espace compact et $C^0(X)$ l'ensemble fonctions continues de X dans $\mathbb{R}$, muni de la norme uniforme. Une sous-algèbre de $C^0(X)$ est dense dans $C^0(X)$, si et seulement si, elle sépare les points et contient, pour tout point x de X, une fonction qui ne s'annule pas en x.*

1.4.2. *Les espaces ℓ^p.* Ces espaces sont utilisés fréquemment, par exemple dans le cours de "Séries temporelles".

Définition 1.22. *Pour $p \in [1, \infty[$, $\ell^p(\mathbb{R})$ (resp. $\ell^p(\mathbb{C})$) est l'ensemble des suites réelles (resp. complexes) $x = (x_n)_{n\in\mathbb{N}}$ telles que $\|x\|_p := \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p} < \infty$. Pour $p = \infty$, $\ell^\infty(\mathbb{R})$ (resp. $\ell^\infty(\mathbb{C})$) est l'ensemble des suites réelles (resp. complexes) bornées.*

Pour $p \in [1, \infty[$, $(\ell^p(\mathbb{R}), \|\cdot\|_p)$ et $(\ell^p(\mathbb{C}), \|\cdot\|_p)$ sont des espaces vectoriels normés. Il en est de même pour $\ell^\infty(\mathbb{R})$ et $\ell^\infty(\mathbb{C})$ munis de la norme $\|x\|_\infty := \sup_n |x_n|$.

Théorème 1.23. *Les ℓ^p sont des espaces de Banach pour $p \in [1, \infty]$ et, si $1 \leq p \leq q \leq \infty$, alors*

$$\ell^p \subset \ell^q.$$

Enfin, $\ell^2(\mathbb{R})$ est un espace de Hilbert pour le produit scalaire

$$\langle x, y\rangle = \sum_{n=0}^{\infty} x_n y_n \qquad \forall x = (x_n), y = (y_n) \in \ell^2(\mathbb{R}).$$

1.4.3. *Les espaces L^p.* Ces espaces seront discutés plus loin, dans la section 3.4.

2. Calcul différentiel

Cette partie est très largement empruntée au cours de Jacques Fejoz "Calcul différentiel et optimisation".

2.1. **Différentiabilité.**

2.1.1. *Définition et propriétés de base.* Soient $(E, \|\cdot\|_E)$ et $(F, \|\cdot\|_F)$ deux espaces vectoriels normés, $\mathcal{O}$ un ouvert de E et $x_0 \in \mathcal{O}$. On dit que $f : \mathcal{O} \to F$ est différentiable en x_0 s'il existe une application linéaire continue $df(x_0) : E \to F$ telle que

$$\lim_{x\to x_0} \frac{\|f(x) - f(x_0) - df(x_0)(x - x_0)\|_F}{\|x - x_0\|_E} = 0 \tag{1}$$

La différentielle $df(x_0)$ est alors unique. A savoir :

(1) L'égalité (1) s'écrit aussi
$$f(x) = f(x_0) + df(x_0)(x - x_0) + \|x - x_0\|_E \varepsilon(x - x_0),$$
où $\varepsilon : E \to F$ est une application qui tend vers 0 lorsque x tend vers x_0.

(2) La différentielle, notée ici $df(x_0)$, peut aussi être appelée simplement la dérivée et être notée $f'(x_0)$ ou $Df(x_0)$, etc...

(3) Si f est différentiable en x_0, alors f est continue en x_0.

(4) Si $\mathcal{O} = E$ et f est une application linéaire continue de E dans F, alors f est différentiable en tout point $x_0 \in E$ et $df(x_0) = f$ (en particulier, $df(x_0)$ ne dépend pas de x_0).

(5) **Stabilité de la différentiabilité par addition et par multiplication par un scalaire :** si f_1, f_2 sont différentiables en $x_0 \in E$ et $\alpha \in \mathbb{R}$, alors $\alpha f_1 + f_2$ est différentiable en x_0 et

$$d(\alpha f_1 + f_2)(x_0) = \alpha df_1(x_0) + df_2(x_0).$$

(6) **Notion de gradient :** Lorsque $E = \mathbb{R}^n$ et $F = \mathbb{R}$, $df(x_0)$ est une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}$ et peut donc être identifiée à un vecteur de $\mathbb{R}^n$, noté $\nabla f(x_0)$ (appelé *gradient* de f en x_0). On a donc

$$df(x_0)(v) = \langle \nabla f(x_0), v \rangle \qquad \forall v \in \mathbb{R}^n.$$

(cette notation se généralise au cas où E est un espace de Hilbert, grâce au théorème de représentation de Riesz (Théorème 1.15)).

On dit que f est C^1 sur $\mathcal{O}$ si f est différentiable en tout point de $\mathcal{O}$ et si l'application $x \to df(x)$ est continue de $\mathcal{O}$ dans l'espace $L(E, F)$ des applications linéaires continues de E dans F.

A savoir :

(1) (Intégrale de la différentielle) Si f est de classe C^1 sur $\mathcal{O}$ à valeurs dans $\mathbb{R}^m$ ($m \in \mathbb{N}^*$) et $[a, b] \subset \mathcal{O}$, alors

$$f(b) - f(a) = \int_0^1 df((1-t)a + tb)(b-a)dt$$

(2) (Inégalité des accroissements finis) Sous les même hypothèses,

$$\|f(b) - f(a)\|_F \leq \sup_{z \in [a,b]} \|df(z)\|_{L(E,F)} \|b - a\|_E$$

(où le segment $[a, b]$ est le sous-ensemble convexe fermé de E défini par $[a, b] = \{(1-t)a + tb,\ t \in [0, 1]\}$).

2.1.2. *Le cas de la dimension finie.* On suppose ici que $E = \mathbb{R}^n$ et $G = \mathbb{R}^p$ (avec $n, p \in \mathbb{N}^*$), $\mathcal{O}$ est un ouvert de $\mathbb{R}^n$ et $f : \mathcal{O} \to \mathbb{R}^p$ avec $f = (f_1, \ldots, f_p)$. On suppose que f admet des dérivées partielles continues sur $\mathcal{O}$, i.e., que pour tout $x \in \mathcal{O}$, $i \in \{1, \ldots, p\}$ et $j \in \{1, \ldots, n\}$, la limite suivante existe

$$\frac{\partial f_i(x)}{\partial x_j} = \lim_{h \to 0} \frac{f_i(x_1, \ldots, x_{j-1}, x_j + h, x_{j+1}, \ldots, x_n) - f_i(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n)}{h},$$

et que les applications $\frac{\partial f_i(x)}{\partial x_j} : \mathcal{O} \to \mathbb{R}$ sont continues sur $\mathcal{O}$. Alors f est de classe C^1 sur $\mathcal{O}$ et on a, pour tout $v \in \mathbb{R}^n$ (en notation matricielle pour le terme droit de l'égalité)

$$df(x)(v) = J_f(x)v \qquad \text{où } J_f(x) = \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_p(x)}{\partial x_1} & \cdots & \frac{\partial f_p(x)}{\partial x_n} \end{pmatrix} \quad v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}.$$

La matrice $J_f(x)$ est appelée **la matrice jacobienne de f en x.**

2.1.3. *Différentiabilité et composition.*

Théorème 2.1. *Soient $(E, \|\cdot\|_E)$, $(F, \|\cdot\|_F)$ et $(G, \|\cdot\|_G)$ trois espaces vectoriels normés, $\mathcal{O}$ un ouvert de E, $\mathcal{O}'$ un ouvert de F, et $x_0 \in \mathcal{O}$. On suppose que $f : \mathcal{O} \to \mathcal{O}'$ et $g : \mathcal{O}' \to G$ sont différentiables en x_0 et en $f(x_0)$ respectivement. Alors $g \circ f$ est différentiable en x_0 et*

$$d(g \circ f)(x_0) = dg(f(x_0)) \circ df(x_0).$$

En dimension finie, on obtient l'égalité matricielle :

$$J_{g \circ f}(x_0) = J_g(f(x_0)) J_f(x_0).$$

En particulier, si f est bijective de $\mathcal{O}$ dans $\mathcal{O}'$ et si $g = f^{-1}$ est différentiable en $y_0 := f(x_0)$, alors $df(x_0)$ est inversible et

$$d(f^{-1})(y_0) = (df(x_0))^{-1}$$

(comparer avec le théorème d'inversion locale ci-dessous).

2.1.4. *Différentielles d'ordre supérieur.* Soient $(E, \|\cdot\|_E)$ et $(F, \|\cdot\|_F)$ deux espaces vectoriels normés, $\mathcal{O}$ un ouvert de E et $f : \mathcal{O} \to F$ une application. On dit que f est de classe C^2 si f est de classe C^1 et si $df : E \to L(E, F)$ est également de classe C^1. La dérivée de df en un point $x_0 \in E$, notée $d^2f(x_0)$, est un élément de $L(E, L(E, F))$ et s'identifie donc à une application bilinéaire continue symétrique (la symétrie venant du **lemme de Schwarz**). Si $[a, b] \subset \mathcal{O}$, on a l'égalité

$$f(b) = f(a) + df(a)(b-a) + \frac{1}{2}\int_0^1 d^2f(a + t(b-a))(b-a, b-a)(1-t)dt.$$

En particulier,

$$\|f(b) - f(a) - df(a)(b-a)\|_F \leq \frac{1}{2} \sup_{z\in[a,b]} |||d^2f(z)|||\ \|b-a\|^2$$

(où $|||h||| = \sup_{\|x\|_E\leq 1, \|y\|_E\leq 1} \|h(x,y)\|_F$ si $h : E \times E \to F$ est bilinéaire).

Si $E = \mathbb{R}^n$, $F = \mathbb{R}$ et $x_0 \in \mathbb{R}^n$, $d^2f(x_0)$ est une forme bilinéaire symétrique sur $\mathbb{R}^n$ et peut être représenté par une matrice symétrique, la matrice hessienne de f en x_0 :

$$\mathrm{Hess}_f(x_0) = \left(\frac{\partial^2 f(x_0)}{\partial x_i \partial x_j}\right)_{1\leq i,j\leq n}.$$

Par récurrence, on peut également définir la notion de fonction de classe C^k (où $k \in \mathbb{N}^*$), où $d^kf(x_0)$ est une application $k-$linéaire symétrique. La formule de Taylor d'ordre k s'écrit alors, pour tout $a, b \in \mathcal{O}$ tels que $[a, b] \subset \mathcal{O}$,

$$\begin{aligned} f(b) =& f(a) + df(a)(b-a) + \frac{1}{2}d^2f(a)(b-a)^2 + \cdots + \frac{1}{(k-1)!}d^{(k-1)}f(a)(b-a)^{k-1} \\ &+ \int_0^1 d^kf(a + t(b-a))(b-a)^k \frac{(1-t)^{k-1}}{(k-1)!}dt \end{aligned}$$

2.1.5. *Application à la convexité.* Soient E un espace vectoriel réel et $\mathcal{O} \subset E$ un sous-ensemble convexe de E, i.e.,

$$\forall x, y \in \mathcal{O},\ \forall\lambda \in [0, 1], \qquad (1-\lambda)x + \lambda y \in \mathcal{O}.$$

On dit que $f : \mathcal{O} \to \mathbb{R} \cup \{+\infty\}$ est convexe dans $\mathcal{O}$ si

$$\forall x, y \in \mathcal{O},\ \forall\lambda \in [0, 1], \qquad f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y).$$

Supposons que E soit normé, que $\mathcal{O}$ soit un ouvert convexe de E et que $f : \mathcal{O} \to \mathbb{R}$ soit de classe C^1. On a alors l'équivalence

$$f \text{ est convexe dans } \mathcal{O} \qquad \Longleftrightarrow \qquad [\forall x, y \in \mathcal{O}, \qquad (df(x) - df(y))(y-x) \geq 0].$$

Si f est de classe C^2, alors

$$f \text{ est convexe dans } \mathcal{O} \qquad \Longleftrightarrow \qquad \left[\forall x \in \mathcal{O},\ \forall\xi \in E, \qquad d^2f(x)(\xi,\xi) \geq 0\right].$$

2.2. **Inversion locale et fonctions implicites.** On suppose que $\mathcal{O}$ est un ouvert de $\mathbb{R}^n$ (où $n \in \mathbb{N}^*$) et $f : \mathcal{O} \to \mathbb{R}^n$ est de classe C^1.

Théorème 2.2 (d'inversion locale)**.** *Si $x_0 \in \mathcal{O}$ et $df(x_0) : \mathbb{R}^n \to \mathbb{R}^n$ est bijective, alors il existe un ouvert $\mathcal{O}' \subset \mathcal{O}$ de $\mathbb{R}^n$ contenant x_0 et un ouvert $\mathcal{O}''$ de $\mathbb{R}^n$ contenant $f(x_0)$ tels que f est un difféomorphisme de $\mathcal{O}'$ dans $\mathcal{O}''$, c'est-à-dire que f est bijective de $\mathcal{O}'$ dans $\mathcal{O}''$ avec f et f^{-1} de classe C^1.*

La preuve repose sur le théorème du point fixe. Le résultat se prolonge donc sans difficulté au cas où f est définie sur des espaces de Banach.

Supposons maintenant que $f : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^p$ (où $n, p \in \mathbb{N}$, $n, p \geq 1$) est de classe C^1.

Théorème 2.3 (des fonctions implicites)**.** *Soit* $(x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^p$*. On suppose que* $f(x_0, y_0) = 0$ *et que la dérivée partielle* $d_y f(x_0, y_0)$ *est inversible. Alors il existe un ouvert* $\mathcal{O}$ *de* $\mathbb{R}^{n+p}$ *contenant* (x_0, y_0)*, un ouvert* $\mathcal{O}'$ *de* $\mathbb{R}^n$ *contenant* x_0 *et une application* $g : \mathcal{O}' \to \mathbb{R}^p$ *de classe* C^1 *tels que, pour tout* $(x, y) \in \mathcal{O}$*,*

$$\Big[\, f(x,y) = 0 \,\Big] \qquad \Leftrightarrow \qquad \Big[\, y = g(x) \,\Big]$$

En particulier, $g(x_0) = y_0$ *et* $dg(x_0) = -(d_y f(x_0, y_0))^{-1} \circ d_x f(x_0, y_0)$*.*

La dernière relation s'obtient en dérivant l'égalité $f(x, g(x)) = 0$: par dérivation des fonctions composées, on a $d_x f(x, g(x)) + d_y f(x, g(x)) \circ dg(x) = 0$, ce qui donne le résultat (et aussi un moyen mnémotechnique pour se souvenir des hypothèses du théorème).

2.3. **Equations différentielles.** Cette partie sera notamment utile dans le cours d'optimisation de M1, pour la partie concernant le contrôle optimal.

2.3.1. *Le théorème de Cauchy-Lipschitz.* Soit $\mathcal{O}$ un ouvert de $\mathbb{R} \times \mathbb{R}^n$, $f : \mathcal{O} \to \mathbb{R}^n$ *un champ de vecteurs* et (t_0, x_0) *une condition initiale*. On cherche un couple (I, x) où I est un intervalle ouvert contenant t_0 et $x : I \to \mathbb{R}^n$ est de classe C^1 et vérifie

$$x'(t) = f(t, x(t)) \ \forall t \in I, \qquad x(t_0) = x_0.$$

On dit que la solution (I, x) est maximale si toute autre solution $(\tilde{I}, \tilde{x})$ telle que $\tilde{x} = x$ sur $I \cap \tilde{I}$ vérifie $\tilde{I} \subset I$. Autrement dit, on ne peut pas prolonger la solution x à un intervalle plus grand.

Théorème 2.4 (de Cauchy-Lipschitz)**.** *On suppose que* f *est continue sur* $\mathcal{O}$ *et que* f *est* $L-$*Lipschitzienne en* x*, i.e., qu'il existe une constante* $L > 0$ *telle que*

$$\|f(t,x) - f(t,y)\| \ \leq L\|x - y\| \qquad \forall (t,x), (t,y) \in \mathcal{O}.$$

Alors il existe une unique solution maximale à l'équation différentielle. Si de plus $\mathcal{O} =]t_1, t_2[\times \mathbb{R}^n$ *avec* $t_1 < t_0 < t_2$ *et si* f *est à croissance au plus linéaire, i.e., s'il existe une constante* $C > 0$ *telle que*

$$\|f(t,x)\| \leq C(\|x\| + 1) \qquad \forall (t,x) \in]t_1, t_2[\times \mathbb{R}^n,$$

alors la solution maximale (I, x) *est définie sur l'intervalle* $I =]t_1, t_2[$ *tout entier.*

2.3.2. *Le lemme de Gronwall.* La seconde partie du théorème 2.4 repose sur le lemme de Gronwall, très utile dans de nombreux contextes. Soit $a, b : [0, +\infty[\to \mathbb{R}$ deux fonctions continues avec $a(t) \geq 0$ pour tout $t \in [0, +\infty[$.

Lemme 2.5 (de Gronwall)**.** *Si une fonction continue* $x : [0, +\infty[\to \mathbb{R}$ *vérifie*

$$x(t) \leq b(t) + \int_0^t a(s)x(s)ds \qquad \forall t \geq 0,$$

alors

$$x(t) \leq b(t) + \int_0^t b(s)a(s)\exp\{\int_s^t a(\tau)d\tau\}ds \qquad \forall t \geq 0.$$

2.3.3. *Equations différentielles linéaires.* Dans cette partie, on confond allègrement vecteurs et matrices pour simplifier les notations. Soit $A : \mathbb{R} \to M_n(\mathbb{R})$ une application continue (avec $n \in \mathbb{N}^*$) et $b : \mathbb{R} \to \mathbb{R}^n$ continue. On s'intéresse aux solutions de l'équation différentielle linéaire

$$x'(t) = A(t)x(t) + b(t) \qquad t \in \mathbb{R}. \tag{2}$$

Une conséquence du théorème de Cauchy-Lipschitz est le résultat suivant :

Théorème 2.6. *L'ensemble des solutions de l'équation différentielle* (2) *est un espace affine de dimension n.*

Si A ne dépend pas de t et $b \equiv 0$, alors les solutions de (2) *sont de la forme $t \to e^{At}v$ où $v \in \mathbb{R}^n$ et où l'exponentielle d'une matrice carrée B est définie par $e^B = \sum_{m=0}^{\infty} \frac{B^m}{m!}$.*

3. Intégration

Cette partie est très largement empruntée au polycopié de Jacques Fejoz "Intégrale de Lebesgue et Probabilités"

3.1. **Espaces mesurés, mesures.**

3.1.1. *Notions de base :* Tribu, tribu engendrée par une partie, tribu borélienne, espace mesurable (= ensemble muni d'une tribu), fonction mesurable, image réciproque d'une tribu par une application mesurable, fonction indicatrice (d'un ensemble A, notée $\mathbf{1}_A$).

3.1.2. *Mesures.*

Définition 3.1. *Si $(E, \mathcal{A})$ est un espace mesurable, une application $\mu : \mathcal{A} \to [0, +\infty]$ est une mesure si μ est $\sigma-$additive, c'est-à-dire*

$$\mu\Big(\bigcup_{n=0}^{\infty} A_n\Big) = \sum_{n=0}^{\infty} \mu(A_n)$$

pour toute famille **dénombrable** (A_n) **disjointe** *d'éléments de $\mathcal{A}$. En particulier, $\mu(\emptyset) = 0$.*

Rappelons que, si μ est une mesure, alors

— μ est croissante pour l'inclusion : si $A_1 \subset A_2$ avec $A_1, A_2 \in \mathcal{A}$, alors $\mu(A_1) \leq \mu(A_2)$.

— si (A_n) est une famille croissante d'éléments de $\mathcal{A}$ (i.e., $A_n \subset A_{n+1}$ pour tout n), alors $\mu(\bigcup_n A_n) = \lim_n \mu(A_n)$.

— si (A_n) est une famille décroissante d'éléments de $\mathcal{A}$ *et si un des $\mu(A_n)$ est fini*, alors $\mu(\bigcap_n A_n) = \lim_n \mu(A_n)$.

L'espace $(E, \mathcal{A}, \mu)$ s'appelle alors un **espace mesuré**. Lorsque $\mu(E) = 1$, on dit que μ est une **probabilité**.

Ensemble de mesure nulle : On dit qu'un sous-ensemble B de E est de mesure nulle (pour μ) s'il existe $A \in \mathcal{A}$ tel que $B \subset A$ et $\mu(A) = 0$.
Propriété vraie $\mu-$presque partout. On dit qu'une propriété P_x (indexée par $x \in E$) est vraie $\mu-$presque partout s'il existe un ensemble de mesure nulle $\mathcal{N} \subset E$ telle que P_x est vraie pour tout $x \in E \setminus \mathcal{N}$. Par exemple, une application $f : E \to \mathbb{R}$ (pas forcément mesurable) est nulle presque partout si l'ensemble $\{x \in E,\ f(x) \neq 0\}$ est un ensemble de mesure nulle (ici P_x est la propriété "$f(x) = 0$").

Tribu complétée : Etant donné un espace mesuré $(E, \mathcal{A})$ et une mesure μ sur cette espace, il est souvent utile de considérer la tribu $\overline{\mathcal{A}}$ (tribu complétée de $\mathcal{A}$) dont les éléments sont donnés par l'union d'un élément de $\mathcal{A}$ et d'un ensemble de mesure nulle pour μ.

Les exemples standards de mesure sont :

(1) Si $(E,\mathcal{A})$ est un espace mesuré et $a \in E$, la mesure de Dirac δ_a est définie par
$$\delta_a(A) = \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{sinon} \end{cases}$$
C'est une probabilité.

(2) Sur l'espace mesuré $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$ (où $\mathcal{B}(\mathbb{R}^N)$ est la tribu borélienne sur $\mathbb{R}^N$), la mesure de Lebesgue λ est l'unique mesure sur $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$ vérifiant $\lambda(]a_1, b_1[\times\cdots\times]a_N, b_N[) = (b_1 - a_1)\cdots(b_N - a_N)$ pour tous les réels $a_1 < b_1, \ldots, a_N < b_N$. L'existence et l'unicité de cette mesure sont des résultats non triviaux.

(3) Si $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace mesuré et $X : \Omega \to \mathbb{R}$ une variable aléatoire, on définit la mesure image $\mathbb{P}_X$ de $\mathbb{P}$ par X sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ en posant
$$\mathbb{P}_X(B) = \mathbb{P}\left[\{\omega \in \Omega,\ X(\omega) \in B\}\right] \qquad \forall B \in \mathcal{B}(\mathbb{R}) \qquad (\text{noté le plus souvent } \mathbb{P}\left[X \in B\right].)$$

(4) Bien sûr toutes les lois classiques en probabilité :
  (a) les lois discrètes :
  - Bernouilli $Ber(p)$, où $p \in [0,1]$ (qui est $(1-p)\delta_0 + p\delta_1$),
  - Binômiale $Bin(n,p)$, où $n \in \mathbb{N}$, $p \in [0,1]$ (= somme de n Bernoulli indépendantes),
  - Poisson $\mathcal{P}(\lambda)$, où $\lambda \in]0,+\infty[$ ($= e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}\delta_n$),
  - Géométrique $\mathcal{G}(p)$ avec $p \in]0,1[$ ($= \sum_{n=1}^{\infty} p(1-p)^{n-1}\delta_n$),
  (b) les lois continues :
  - Loi uniforme $U(]a,b[)$, de densité $(b-a)^{-1}\mathbf{1}_{[a,b]}$, où $a < b$.
  - Loi normale $\mathcal{N}(m,\sigma^2)$, de densité $(2\pi\sigma)^{-1/2}\exp\{-(x-m)^2/(2\sigma^2)\}$, où $m \in \mathbb{R}$, $\sigma \in]0,+\infty[$,
  - Loi exponentielle $\mathcal{E}(\lambda)$, de densité $\lambda e^{-\lambda x}\mathbf{1}_{]0,+\infty[}$, ...

Exercice 17 (Borel-Cantelli)**.** Soit $(E,\mathcal{A},\mu)$ un espace mesuré. Soit (A_n) une suite d'éléments de $\mathcal{A}$. On définit $\limsup_n A_n$ et $\liminf_n A_n$ par
$$\limsup_n A_n = \bigcap_{n\in\mathbb{N}}\bigcup_{k\geq n} A_k, \qquad \liminf_n A_n = \bigcup_{n\in\mathbb{N}}\bigcap_{k\geq n} A_k.$$
Montrer que $\limsup_n A_n$ et $\liminf_n A_n$ sont dans $\mathcal{A}$, que $\liminf_n A_n \subset \limsup_n A_n$ et que, si $\sum_n \mu(A_n) < +\infty$, alors $\mu(\limsup_n A_n) = 0$.

3.2. **Contruction de l'intégrale de Lebesgue.** La construction qui suit doit être bien connue : elle joue un rôle important dans la construction de l'espérance conditionnelle (du cours "Processus discret"). Dans toute la suite, $(E,\mathcal{A},\mu)$ est un espace mesuré et $\mathbb{R}$ est muni de la tribu borélienne.

3.2.1. *Intégration de fonctions étagées.* On dit qu'une application mesurable $f : E \to [0,+\infty]$ est **étagée** si elle ne prend qu'un nombre fini de valeurs. Il existe alors un entier $n \in \mathbb{N}^*$, un nombre fini d'ensembles $A_i \in \mathcal{A}$ et de réels $a_i \in [0,+\infty]$ (pour $i = 1,\ldots,n$) tels que
$$f(x) = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}(x)$$
où $\mathbf{1}_A$ est la fonction indicatrice d'un ensemble A.

Lorsque f est étagée et positive, on définit l'intégrale de f par
$$\int_E f(x)\mu(dx) = \sum_{i=1}^{n} a_i\mu(A_i), \tag{3}$$
avec la convention que $0\times\infty = 0$. Le terme de droite est bien défini puisque qu'il ne fait intervenir que des sommes et produits de quantités positives (éventuellement égales à $+\infty$). On utilise aussi fréquemment la notation $\int_E f d\mu$. On peut montrer (et ce n'est pas immédiat) que l'expression de droite ne dépend pas de la représentation de f choisie.

3.2.2. *Intégration de fonctions positives.* Soit $f : E \to \mathbb{R}$ une application mesurable **positive**. On définit alors

$$\int_E f(x)\mu(dx) = \sup_{g \text{ étagée, } 0 \le g \le f} \int_E g(x)\mu(dx).$$

C'est un élément de $[0, +\infty]$.

On peut montrer que cette définition redonne la formule (3) si f est étagée. En particulier, si $A \in \mathcal{A}$, alors $\int_E \mathbf{1}_A(x)\mu(dx) = \mu(A)$.

3.2.3. *Intégration de fonctions intégrables.* Soit $f : E \to \mathbb{R}$ une application mesurable. Posons $f^+(x) = \max\{0, f(x)\}$ et $f^-(x) = \max\{0, -f(x)\}$. Les fonctions f^+ et f^- sont des fonctions mesurables positives avec $f = f^+ - f^-$.

On dit que f **est intégrable** si l'application $|f|$ (qui est également mesurable) vérifie

$$\int_E |f(x)|\mu(dx) < \infty.$$

Bien noter qu'alors $\displaystyle\int_E f^+(x)\mu(dx) < \infty \qquad \text{et} \qquad \int_E f^-(x)\mu(dx) < \infty.$

Si f est intégrable, on pose

$$\int_E f(x)\mu(dx) = \int_E f^+(x)\mu(dx) - \int_E f^-(x)\mu(dx).$$

(on utilise aussi la notation $\int_E f d\mu$).

Bien retenir que l'on peut définir $\int_E f d\mu$

(1) lorsque $f : E \to \mathbb{R} \cup \{+\infty\}$ est **mesurable positive** et, dans ce cas, $\int_E f d\mu \in [0, +\infty]$;

(2) ou bien lorsque $f : E \to \mathbb{R} \cup \{+\infty\}$ est **intégrable** et, dans ce cas, $\int_E f d\mu \in \mathbb{R}$.

Propriétés de base. L'espace $\mathcal{L}^1(E, \mathcal{A}, \mu)$ des fonctions intégrables est un espace vectoriel et l'intégrale est une application linéaire sur cet espace vectoriel. De plus l'intégrale est croissante au sens où, si f_1 et f_2 sont intégrables et telles que $f_1 \le f_2$, alors $\int_E f_1 d\mu \le \int_E f_2 d\mu$.

Attention, $\mathcal{L}^1(E, \mathcal{A}, \mu)$ n'est pas un espace vectoriel normé en général, car on peut avoir $\int_E |f(x)|\mu(dx) = 0$ sans que f soit nulle (on discutera de cette question plus loin, à la sous-section 3.4).

Voici quelques inégalités usuelles en intégration : pour $p \ge 1$, notons $\mathcal{L}^p(E)$ l'ensemble des fonction mesurables $f : E \to \mathbb{R}$ telles que $\int_E |f|^p d\mu < \infty$. L'ensemble $\mathcal{L}^\infty(E)$ est l'ensemble des fonctions mesurables $f : E \to \mathbb{R}$ pour lesquelles il existe $M > 0$ tel que $|f(x)| \le M$ pour μ−presque tout $x \in E$.

Théorème 3.2.

(1) (Inégalité triangulaire) Soit $(E, \mathcal{A}, \mu)$ un espace mesuré. Si f_1 et f_2 sont intégrables, alors

$$\left|\int_E (f_1+f_2)(x)\mu(dx)\right| \leq \int_E |f_1(x)|\mu(dx) + \int_E |f_2(x)|\mu(dx).$$

(2) (Inégalité de Hölder) Soit $(E, \mathcal{A}, \mu)$ un espace mesuré, $p, q \in [1, +\infty]$ avec $1/p+1/q = 1$ (avec la convention $1/\infty = 0$). Si $f \in \mathcal{L}^p(E)$ et $g \in \mathcal{L}^q(E)$, alors $fg \in \mathcal{L}^1(E)$ et

$$\left|\int_E f(x)g(x)\mu(dx)\right| \leq \left(\int_E |f(x)|^p \mu(dx)\right)^{1/p} \left(\int_E |g(x)|^q \mu(dx)\right)^{1/q}$$

(3) (Inégalité de Minkowski) Soit $(E, \mathcal{A}, \mu)$ un espace mesuré et $p \geq 1$. alors pour tout $f, g \in \mathcal{L}^p(E)$, on a $f + g \in \mathcal{L}^p(E)$ et

$$\left(\int_E |f(x)+g(x)|^p \mu(dx)\right)^{1/p} \leq \left(\int_E |f(x)|^p d\mu(dx)\right)^{1/p} + \left(\int_E |g(x)|^p d\mu(dx)\right)^{1/p}.$$

(4) (Inégalité de Jensen) Si $(E, \mathcal{A}, \mu)$ est un espace probabilisé (= espace mesuré avec μ probabilité), alors pour toute fonction convexe $\phi : \mathbb{R} \to \mathbb{R}$ et pour toute fonction intégrable f sur E on a

$$\phi\left(\int_E f(x)\mu(dx)\right) \leq \int_E \phi(f(x))\mu(dx).$$

Exercice 18. (1) Démontrer l'inégalité de Hölder : on pourra supposer dans un premier temps que $\int_E |f|^p d\mu = \int_E |g|^q d\mu = 1$ et utiliser l'inégalité de Young $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$ pour $a, b \geq 0$.

(2) Démontrer l'inégalité de Hölder à partir de l'inégalité de Jensen (on pourra utiliser la mesure de probabilité $\mu_f := |f|^p/(\int_E |f|^p d\mu)$).

(3) Démontrer l'inégalité de Minkowski à partir de l'inégalité de Hölder.

Cas des fonctions à valeurs complexes : Lorsque $f : E \to \mathbb{C}$ est à valeur complexe, on dit que f est intégrable si f est mesurable et si $|f|$ (le module de f) est intégrable. Dans ce cas les parties réelle et complexe $\mathcal{R}e(f)$ et $\mathcal{I}m(f)$ sont intégrables et on pose

$$\int_E f(x)\mu(dx) = \int_E \mathcal{R}e(f)(x)\mu(dx) + i\int_E \mathcal{I}m(f)(x)\mu(dx).$$

3.3. **Propriétés centrales de l'intégrale de Lebesgue.** Il s'agit du théorème de convergence monotone, du lemme de Fatou, du théorème de convergence dominée et du théorème de Fubini.

Théorème 3.3 (de convergence monotone)**.** *Si (f_n) est une suite croissante de fonctions mesurables positives sur E (au sens où $0 \leq f_n \leq f_{n+1}$ p.p. pour tout n), alors*

$$\lim_{n\to+\infty} \int_E f_n d\mu = \int_E \lim_n f_n d\mu.$$

L'intérêt de ce résultat est de ne demander aucune intégrabilité sur les f_n. Le prix à payer est la condition que les f_n sont positives. Attention, ce résultat est très dissymétrique : il n'est pas vrai si on suppose juste que (f_n) est une suite décroissante de fonctions positives par exemple.

Lemme 3.4 (de Fatou [a])**.** *Si (f_n) est une suite de fonctions mesurables positives sur E, alors*
$$\int_E \liminf_n f_n d\mu \leq \liminf_n \int_E f_n d\mu.$$

a. On rappelle que, si (a_n) est une suite réelle, alors $\liminf_n a_n = \lim_{k\to+\infty} \inf_{n\geq k} a_n$ et $\limsup_n a_n = \lim_{k\to+\infty} \sup_{n\geq k} a_n$. Bien sûr, si (a_n) a une limite (finie ou infinie), alors $\liminf_n a_n = \limsup_n a_n = \lim_n a_n$.

Comme pour le théorème de convergence monotone, la condition de signe est essentielle et conduit à un énoncé dissymétrique (avec une inégalité cette fois, mais sans supposer que la monotonie de la suite).

Exercice 19. (1) Montrer que le théorème de convergence monotone et le lemme de Fatou restent vrais sans condition de signe sur les (f_n) s'il existe g intégrable telle que $f_n \geq g$ pour tout n.

(2) Plus généralement, montrer que, s'il existe des fonctions intégrables g_1 et g_2 telles que $g_1 \leq f_n \leq g_2$ pour tout n, alors
$$\int_E \liminf_n f_n d\mu \leq \liminf_n \int_E f_n d\mu \leq \limsup_n \int_E f_n d\mu \leq \int_E \limsup_n f_n d\mu.$$

Théorème 3.5 (de convergence dominée ou de Lebesgue)**.** *Si f_n est une suite de fonctions mesurables de E dans $\mathbb{R}$ telle que*

(1) il existe une fonction intégrable $g : E \to \mathbb{R}$ vérifiant :
$$|f_n(x)| \leq g(x) \qquad \textit{pour } \mu\textit{-presque tout } x \in E,$$

(2) il existe une fonction (mesurable) $f : E \to \mathbb{R}$ telle que
$$\lim_{n\to+\infty} f_n(x) = f(x) \qquad \textit{pour } \mu\textit{-presque tout } x \in E,$$

alors f est intégrable et
$$\lim_{n\to+\infty} \int_E f_n d\mu = \int_E f d\mu.$$

Au contraire des deux résultats précédant, la condition de domination est parfaitement symétrique. Cela fait de ce résultat le principal outil de manipulation de l'intégrale de Lebesgue.

Exercice 20 (Intervertion somme et intégrale)**.** Soit (f_n) une suite de fonctions mesurables de E dans $\mathbb{R}$ pour laquelle il existe une fonction intégrable g telle que
$$\sum_n |f_n(x)| \leq g(x) \qquad \text{pour } \mu\text{-presque tout } x \in E.$$

Montrer que la fonction $x \to \sum_n f_n(x)$ est définie pour $\mu-$presque tout $x \in E$ et intégrable et que
$$\int_E (\sum_n f_n(x))\mu(dx) = \sum_n \int_E f_n(x)\mu(dx).$$

Exercice 21 (Convergence p.p. et convergence en mesure)**.** Soit $(E, \mathcal{A}, \mu)$ un espace mesuré avec $\mu(E) < \infty$. On dit qu'une suite $f_n : E \to \mathbb{R}$ de fonctions mesurables converge $\mu-$p.p. vers une fonction mesurable $f : E \to \mathbb{R}$ si l'ensemble des $x \in E$ tel que $(f_n(x))$ ne converge pas vers $f(x)$ est de mesure nulle. On dit que (f_n) converge en mesure vers f si
$$\forall \varepsilon > 0,\ \lim_{n\to+\infty} \mu(\{x \in E,\ |f_n(x) - f(x)| \geq \varepsilon\}) = 0.$$

(1) Montrer que, si (f_n) converge vers f $\mu-$p.p., alors (f_n) converge en mesure vers f.

(2) Donner un contre-exemple à l'assertion précédente si on supprime l'hypothèse que $\mu(E) < \infty$.

(3) Prouver que, si (f_n) converge en mesure vers f, alors il existe une sous-suite (f_{n_k}) qui converge $\mu-$p.p. vers f. (on pourra montrer qu'il existe une suite extraite (n_k) telle que, pour tout $k \geq 1$, $\mu\{x \in E,\ |f_{n_k}(x) - f(x)| \geq 1/k\} \leq 1/k^2$ et utiliser le lemme de Borel-Cantelli (Exercice 17)).

Le théorème de Fubini traite d'intégration sur des espaces produit. Soit $(E_1, \mathcal{A}_1, \mu_1)$ et $(E_2, \mathcal{A}_2, \mu_2)$ deux espaces mesurables. La tribu $\mathcal{A}_1 \otimes \mathcal{A}_2$ est la tribu sur $E_1 \times E_2$ engendrée par l'ensemble des produits $A_1 \times A_2$ avec $A_1 \in \mathcal{A}_1$ et $A_2 \in \mathcal{A}_2$, tandis que la mesure $\mu_1 \otimes \mu_2$ est la seule mesure sur $(E_1 \times E_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ telle que $\mu_1 \otimes \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ pour tout $A_1 \in \mathcal{A}_1$ et $A_2 \in \mathcal{A}_2$.

Théorème 3.6 (de Fubini)**.** *Le théorème comporte deux parties, la première pour les fonctions positives, et la seconde pour les fonctions intégrables :*

(1) Si $f : E_1 \times E_2 \to \mathbb{R}$ est mesurable et positive, alors

$$\begin{aligned} \int_{E_1 \times E_2} f(x_1, x_2)(\mu_1 \otimes \mu_2)(dx_1, dx_2) &= \int_{E_1} \Big(\int_{E_2} f(x_1, x_2)\mu_2(dx_2) \Big) \mu_1(dx_1) \\ &= \int_{E_2} \Big(\int_{E_1} f(x_1, x_2)\mu_1(dx_1) \Big) \mu_2(dx_2). \end{aligned} \tag{4}$$

(2) Si $f : E_1 \times E_2 \to \mathbb{R}$ est intégrable pour $\mu_1 \otimes \mu_2$, alors

(a) pour μ_1-presque tout $x_1 \in E_1$, $x_2 \to f(x_1, x_2)$ est intégrable pour μ_2,

(b) pour μ_2-presque tout $x_2 \in E_2$, $x_1 \to f(x_1, x_2)$ est intégrable pour μ_1,

(c) et enfin les deux égalités dans (4) *restent vraies.*

On utilise souvent la première partie du théorème pour montrer la condition d'intégrabilité de la seconde partie.

Exercice 22. En utilisant le théorème de Fubini, vérifier que, si $(E, \mathcal{A}, \mu)$ est un espace mesuré et si $f \in \mathcal{L}^p(E)$ (pour $p \geq 1$), alors

$$\int_E |f(x)|^p \mu(dx) = \int_0^\infty p t^{p-1} \mu(\{x \in E,\ |f(x)| \geq t\}) dt$$

3.4. **Les espaces $\mathcal{L}^p$ et L^p.**

Définition 3.7. *Soit $(E, \mathcal{A}, \mu)$ un espace mesuré. Pour $p \in [1, \infty[$, on définit les ensemble $\mathcal{L}^p(E)$ comme l'ensemble des fonctions mesurables $f : E \to \mathbb{R}$ (ou dans $\mathbb{C}$) telles que $\int_E |f(x)|^p \mu(dx)$ est fini. Pour $p = 0$, $\mathcal{L}^0(E)$ est l'ensemble des fonctions mesurables de E dans $\mathbb{R}$. Noter que $\mathcal{L}^p(E) \subset \mathcal{L}^0(E)$ pour tout $p \in [1, \infty[$.*

Les espaces $\mathcal{L}^p(E)$ sont des espaces vectoriels. Cependant, pour $p \in [1, +\infty[$ ces espace ne sont pas munis d'une structure d'espace vectoriel normé “naturelle” car on peut très bien avoir $\int_E |f(x)|^p \mu(dx) = 0$ sans que f soit nulle. En fait, $\int_E |f(x)|^p \mu(dx) = 0$, si et seulement si $f(x) = 0$ pour $\mu-$presque tout $x \in E$.

Pour surmonter cette (petite) difficulté, on introduit la relation d'équivalence suivante. Soit $f, g \in \mathcal{L}^0(E)$. On dit que $f \sim g$ si la fonction $f - g = 0$ $\mu-$presque-partout. On voit facilement que cette relation est bien une relation d'équivalence, compatible avec la structure d'espace vectoriel de $\mathcal{L}^0(E)$. De plus, pour $p \in [1, +\infty[$, si $f \in \mathcal{L}^p(E)$, $g \in \mathcal{L}^0(E)$ et $f \sim g$, alors $g \in \mathcal{L}^p(E)$. Donc $\sim$ est également une relation d'équivalence sur $\mathcal{L}^p(E)$.

Définition 3.8. *L'espace $L^p(E)$ est l'ensemble des classes d'équivalences des éléments de $\mathcal{L}^p(E)$ pour la relation $\sim$. Pour tout $\dot{f} \in L^p(E)$, on pose*

$$\|\dot{f}\|_p := \left(\int_E |f(x)|^p \mu(dx) \right)^{1/p}$$

où f est n'importe quel élément de la classe d'équivalence $\dot{f}$ (noté $f \in \dot{f}$).

De plus, si $f \in \dot{f}$, $g \in \dot{g}$ et $\alpha \in \mathbb{R}$, alors $\alpha \dot{f} + \dot{g} = \overbrace{\alpha f + g}^{\cdot}$. Par conséquent l'espace $L^p(E)$ a une structure d'espace vectoriel et on montre que $\|\cdot\|_p$ est une norme sur cet espace. **Par abus de notation, on note simplement par f, g etc... les éléments de $L^p(E)$ au lieu de $\dot{f}$, $\dot{g}$, etc..., même si ce sont en fait des classes d'équivalence et non des fonctions.**

Le cas $p = \infty$. On définit $L^\infty(E)$ comme étant l'ensemble des classes d'équivalence d'éléments f de $\mathcal{L}^0(E)$ pour lesquels il existe $M > 0$ avec $|f(x)| \leq M$ pour $\mu-$presque tout $x \in E$. On pose

$$\|f\|_\infty = \inf\{M > 0,\ |f(x)| \leq M \text{ pour } \mu-\text{presque tout } x \in E.\}.$$

L'espace $(L^\infty(E), \|\cdot\|_\infty)$ est un espace vectoriel normé.

Théorème 3.9. *Les espaces $(L^p(E), \|\cdot\|_p)$ sont des espaces de Banach pour $p \in [1, \infty]$. Pour $p = 2$, l'espace $(L^2(E), \|\cdot\|_2)$ est un espace de Hilbert, de produit scalaire associé*

$$\langle f, g\rangle = \int_E f(x)g(x)\mu(dx).$$

Enfin, si $\mu(E) < \infty$, alors

$$L^p(E) \subset L^q(E) \qquad \text{si } 1 \leq p \leq q \leq \infty.$$

Voici quelques propriétés de densité.

Théorème 3.10. *L'ensemble des fonctions étagées est dense dans $L^p(E)$ pour $p \in [1, \infty]$.*

Théorème 3.11. *On suppose que $E =]a, b[$ (avec $-\infty \leq a < b \leq +\infty$) et que μ est la mesure de Lebesgue sur $]a, b[$. Alors, pour tout $p \in [1, +\infty[$, l'ensemble $C_c^\infty(]a, b[)$ des fonctions de classe C^∞ à support compact dans $]a, b[$ est dense dans $L^p(]a, b[)$. En particulier, $L^p(]a, b[)$ est séparable.*

Plus généralement, si $\mathcal{O}$ est un ouvert de $\mathbb{R}^n$ (où $n \in \mathbb{N}^*$), μ est la mesure de Lebesgue sur $\mathbb{R}^n$ et $p \in [1, +\infty[$, alors $C_c^\infty(\mathcal{O})$ des fonctions de classe C^∞ à support compact dans $\mathcal{O}$ est dense dans $L^p(\mathcal{O})$. En particulier, $L^p(\mathcal{O})$ est séparable.

Attention, ce dernier résultat est faux dans L^∞.

3.5. **Fonctions définies par des intégrales.** Une conséquence classique du théorème de convergence dominée est l'analyse de fonction dépendant d'un paramètre. Dans cette partie on fixe $(E, \mathcal{A}, \mu)$ un espace mesuré, A un ouvert de $\mathbb{R}^N$ (où $N \in \mathbb{N}$) et $f : E \times A \to \mathbb{R}$ une application. On s'intéresse à l'application (lorsqu'elle est définie)

$$g(a) := \int_E f(x, a)\mu(dx).$$

Théorème 3.12 (Continuité)**.** *Soit $a_0 \in A$. On suppose que*

(1) pour $\mu-$presque tout $x \in E$, $a \to f(x, a)$ est continue en a_0,

(2) pour tout $a \in A$, $x \to f(x, a)$ est mesurable,

(3) il existe une fonction intégrable $h : E \to [0, +\infty[$ telle que, pour tout $a \in A$, pour $\mu-$presque tout $x \in E$,

$$|f(x, a)| \leq h(x).$$

Alors l'application g est continue en a_0.

Théorème 3.13 (Différentiabilité). *On suppose que*

(1) pour $\mu-$presque tout $x \in E$, l'application $a \to f(x,a)$ est différentiable sur A (de différentielle notée $d_a f(x,a)$),

(2) pour tout $a \in A$, $x \to f(x,a)$ est intégrable,

(3) il existe une fonction intégrable $h : E \to [0,+\infty[$ telle que, pour tout $a \in A$, pour $\mu-$presque tout $x \in E$,

$$\|d_a f(x,a)\| \leq h(x).$$

Alors l'application g est différentiable sur A et sa différentielle en un point $a_0 \in A$ est donnée par

$$dg(a_0) = \int_E d_a f(x,a_0)\mu(dx).$$

Enfin, si l'application l'application $a \to d_a f(x,a)$ est continue sur A pour $\mu-$presque tout $x \in E$, alors g est de classe C^1 sur A.

La transposition au cas où f prend ses valeurs dans $\mathbb{C}$ est immédiate (changer valeur absolue en module).

Les exemples d'application les plus courants sont la **régularisation par convolution**, la **transformée de Fourier** d'une fonction ou d'une mesure (en langage des probabilité, on parle de fonction caractéristique—à la conjugaison près).

Exercice 23. Soit μ une mesure de masse totale finie sur $\mathbb{R}$. On pose

$$\hat{\mu}(t) := \int_{\mathbb{R}} e^{-itx}\mu(dx) \qquad \forall t \in \mathbb{R}.$$

Montrer que l'application $t \to \hat{\mu}(t)$ est continue sur $\mathbb{R}$. Vérifier qu'elle est de classe C^k (où $k \in \mathbb{N}$, $k \geq 1$) si μ possède un moment d'ordre k fini, i.e., si $\int_{\mathbb{R}} |x|^k \mu(dx) < +\infty$ et qu'alors

$$\hat{\mu}^{(k)}(t) = \int_{\mathbb{R}} (-ix)^k e^{-itx}\mu(dx) \qquad \forall t \in \mathbb{R}.$$

Exercice 24. Soit $f : [0,+\infty[\to \mathbb{R}$ la fonction définie par

$$f(x) = \left(\int_0^x e^{-t^2}dt\right)^2 + \int_0^1 \frac{e^{-x^2(1+t^2)}}{1+t^2}dt.$$

Montrer que $f(0) = \pi/4$, que $\lim_{x\to+\infty} f(x) = (\int_0^\infty e^{-t^2}dt)^2$ et que f est dérivable de dérivée nulle. En déduire que $\int_0^\infty e^{-t^2}dt = \sqrt{\pi}/2$.

3.6. **Formule de changement de variables.** On travaille ici avec la mesure de Lebesgue sur $\mathbb{R}^n$, notée dx. Soient $\mathcal{O}$ et $\mathcal{O}'$ deux ouverts de $\mathbb{R}^n$ et $\phi : \mathcal{O} \to \mathcal{O}'$ un difféomorphisme (i.e. ϕ est une bijection de $\mathcal{O}$ dans $\mathcal{O}'$ avec ϕ et ϕ^{-1} de classe C^1). Alors, pour toute fonction positive ou intégrable $f : \mathcal{O} \to \mathbb{R}$ (pour la mesure de Lebesgue), la fonction $y \to f(\phi^{-1}(y)) \left|\det(J_\phi(\phi^{-1})(y)\right|$ est positive ou intégrable sur $\mathcal{O}'$ et

$$\int_{\mathcal{O}} f(x)dx = \int_{\mathcal{O}'} f(\phi^{-1}(y)) \left|\det(J_\phi(\phi^{-1}(y))\right| dy.$$

On verra une application classique de cette formule dans l'étude de la convolution.

Par exemple, on a la formule de **"passage en coordonnées polaires" dans le plan** : si $f : \mathbb{R}^2 \to \mathbb{R}$ est intégrable, alors

$$\int_{\mathbb{R}^2} f(x_1,x_2)dx_1dx_2 = \int_0^\infty \int_0^{2\pi} rf(r\cos(\theta), r\sin(\theta))d\theta dr.$$

3.7. **Transformée de Fourier d'une fonction.** Soit $f \in L^1(\mathbb{R}^n)$. On rappelle que la transformée de Fourier de f est définie par

$$\hat{f}(\xi) = \int_{\mathbb{R}^n} f(x)e^{-i\langle x,\xi\rangle}dx \qquad \forall \xi \in \mathbb{R}^n.$$

Proposition 3.14. *La fonction* $\xi \to \hat{f}(\xi)$ *est une fonction continue, qui tend vers* 0 *lorsque* $\|\xi\| \to +\infty$ *et vérifie*

$$\sup_{\xi\in\mathbb{R}^n} |f(\xi)| \leq \int_{\mathbb{R}^n} |f(x)|dx.$$

Exercice 25. Cet exercice montre comment quantifier la convergence de $\hat{f}$ vers 0. On suppose que $f : \mathbb{R} \to \mathbb{C}$ est de classe C^1 et que f et f' sont dans $L^1(\mathbb{R})$. Montrer que $\widehat{f'}(\xi) = (i\xi)\hat{f}(\xi)$ pour tout $\xi \in \mathbb{R}$ et en déduire que $\lim_{|\xi|\to\infty} |\xi||\hat{f}(\xi)| = 0$.

Exercice 26 (Une application classique du théorème de Fubini)**.** Montrer que, si $f, g \in L^1(\mathbb{R}^n)$ alors $\int_{\mathbb{R}^n} \hat{f}(\xi)g(\xi)d\xi = \int_{\mathbb{R}^n} f(\xi)\hat{g}(\xi)d\xi$.

Si $f \in L^1(\mathbb{R}^n)$, on pose

$$\check{f}(x) = \int_{\mathbb{R}^n} e^{i\langle x,\xi\rangle} f(\xi)d\xi \qquad \forall x \in \mathbb{R}^n.$$

Théorème 3.15 (Formule d'inversion)**.** *Si* $f \in L^1(\mathbb{R}^n)$ *et* $\hat{f} \in L^1(\mathbb{R}^n)$, *alors*

$$f = \frac{1}{(2\pi)^n}\check{f} \qquad p.p.$$

Plus important encore pour ses applications en probabilités :

Théorème 3.16. *L'application qui à une mesure borélienne* μ *sur* $\mathbb{R}^d$ *associe sa transformée de Fourier :*

$$\hat{\mu}(\xi) := \int_{\mathbb{R}^n} e^{-i\langle x,\xi\rangle}\mu(dx) \qquad \forall \xi \in \mathbb{R}^n$$

est injective.

Autrement dit, pour démontrer l'égalité entre deux mesures boréliennes μ_1 et μ_2 sur $\mathbb{R}^n$, il suffit de montrer que $\hat{\mu}_1(\xi) = \hat{\mu}_2(\xi)$ pour tout $\xi \in \mathbb{R}^n$.

3.8. **Produit de convolution.**

Théorème 3.17. *Soient* f *et* g *deux fonctions intégrables sur* $\mathbb{R}^n$ *(pour la mesure de Lebesgue). Alors l'application* $x \to f * g(x) := \int_{\mathbb{R}^n} f(x-y)g(y)dy$ *est finie pour presque tout* $x \in \mathbb{R}^n$ *et* $f * g$ *est intégrable sur* $\mathbb{R}^n$.

Remarques :

(1) Le produit de convolution est bilinéaire et symétrique : si $\alpha \in \mathbb{R}$ et si $f_1, f_2, g : \mathbb{R}^n \to \mathbb{R}$ sont intégrables, alors $(\alpha f_1 + f_2) * g = \alpha(f_1 * g) + f_2 * g$ et $f * g = g * f$.

(2) Le produit de convolution joue un rôle central en probabilité : si X et Y sont deux variables aléatoires indépendantes sur $\mathbb{R}^n$ de densités respectives f_X et f_Y, alors la somme $X + Y$ a également une densité sur $\mathbb{R}^n$ et cette densité est $f_X * f_Y$: en particulier, pour toute fonction $\phi : \mathbb{R}^n \to \mathbb{R}$ continue bornée sur $\mathbb{R}^n$, on a

$$\mathbb{E}\left[\phi(X+Y)\right] = \int_{\mathbb{R}^n} \phi(z)(f_X * f_Y)(z)dz.$$

Preuve du théorème. L'application $(x, y) \to f(y - x)g(y)$ est intégrable sur $\mathbb{R}^{2n}$ car

$$\int_{\mathbb{R}^{2n}} |f(x-y)g(y)|dxdy = \int_{\mathbb{R}^{2n}} |f(z)g(y)|dzdy = \left(\int_{\mathbb{R}^n} |f(z)|dz\right)\left(\int_{\mathbb{R}^n} |g(y)|dy\right) < +\infty,$$

où la première égalité vient par changement de variable $(x, y) \to (y - x, y)$, dont le déterminant du jacobien est de valeur absolue égale à 1, et la seconde égalité est donnée par le théorème de Fubini (première partie). Par conséquent, toujours par le théorème de Fubini (mais deuxième partie), l'application $x \to f * g(x) := \int_{\mathbb{R}^n} f(x - y)g(y)dy$ est finie pour presque tout $x \in \mathbb{R}^n$ et $f * g$ est intégrable sur $\mathbb{R}^n$. □

Exercice 27. Montrer que si $f \in L^1(\mathbb{R}^n)$ et $f \in L^p(\mathbb{R}^n)$ (avec $p \geq 1$), alors $f * g \in L^p(\mathbb{R}^n)$ avec $\|f * g\|_p \leq \|f\|_1 \|g\|_p$.

Exercice 28. Montrer que si $f, g \in L^1(\mathbb{R}^n)$, alors

$$\widehat{f * g}(\xi) = \hat{f}(\xi)\hat{g}(\xi) \qquad \forall \xi \in \mathbb{R}^n.$$

Une application plus analytique de la convolution est l'approximation de fonctions intégrables par des fonctions régulières. Soit $\phi : \mathbb{R}^n \to \mathbb{R}$ une application de classe C^∞, à support compact, positive et d'intégrale 1 : $\int_{\mathbb{R}^n} \phi(x)dx = 1$ (une telle application existe, mais la preuve de son existence n'est pas simple). Pour $\varepsilon > 0$, on pose $\phi^\varepsilon(x) = \varepsilon^{-n}\phi(x/\varepsilon)$.

Théorème 3.18. *Si $f : \mathbb{R}^n \to \mathbb{R}$ est intégrable sur $\mathbb{R}^n$, alors, pour tout $\varepsilon > 0$, la fonction $f^\varepsilon := \phi^\varepsilon * f$ est intégrable, de classe C^∞ sur $\mathbb{R}^n$ et*

$$\lim_{\varepsilon \to 0} \int_{\mathbb{R}^n} |f(x) - f^\varepsilon(x)|dx = 0.$$

4. Application au calcul des probabilités

4.1. **Vocabulaire.**

- **Un espace probabilisé** $(\Omega, \mathcal{A}, \mathbb{P})$ est la donnée d'un ensemble Ω muni d'une tribu $\mathcal{A}$ et d'une mesure de probabilité $\mathbb{P}$ (i.e., $\mathbb{P}$ est une mesure sur $\mathcal{A}$ telle que $\mathbb{P}(\Omega) = 1$). Les éléments de $\mathcal{A}$ sont appelés des événements.
- **Un vecteur aléatoire** X est une application mesurable de Ω dans $\mathbb{R}^n$, où $n \in \mathbb{N}^*$ et $\mathbb{R}^n$ est muni de la tribu borélienne ; si $n = 1$, on parle de **variable aléatoire**. Si $f : \mathbb{R}^n \to \mathbb{R}$ (ou dans $\mathbb{C}$) est borélienne et bornée, on note

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

  Noter que, bien que X dépende de ω, on n'écrit pratiquement jamais $X(\omega)$.

  Si X est une variable aléatoire, **la fonction de répartition de** X, est la fonction $F_X : \mathbb{R} \to [0, 1]$ définie par

  $$F_X(t) = \mathbb{P}[X \leq t].$$

  Enfin, si X est une variable aléatoire, $\sigma(X)$ est la tribu $\{X^{-1}(A),\ A \in \mathcal{A}\}$.
- **Loi d'un vecteur aléatoire.** En probabilité, on appréhende très souvent un vecteur aléatoire X (sur $\mathbb{R}^n$) par sa **loi** $\mathbb{P}_X$, qui est une probabilité sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, où $\mathcal{B}(\mathbb{R}^n)$ est la tribu borélienne ; $\mathbb{P}_X$ est définie par

  $$\mathbb{P}_X[B] = \mathbb{P}[X \in B] \qquad (= \mathbb{P}[\{\omega \in \Omega,\ X(\omega) \in B\}]) \qquad \forall B \in \mathcal{B}(\mathbb{R}^n)$$

  ou, de façon équivalente, par l'égalité

  $$\mathbb{E}[\phi(X)] = \int_{\mathbb{R}^n} \phi(x)\mathbb{P}_X(dx) \qquad \text{pour tout } \phi : \mathbb{R}^n \to \mathbb{R} \text{ borélienne bornée.}$$

(cette dernière égalité est parfois connue sous le nom de **théorème du transfer**). Par exemple, si X est une variable aléatoire de fonction de répartition F_X, alors

$$F_X(t) = \mathbb{P}_X(]-\infty, t]) \qquad \forall t \in \mathbb{R}.$$

— On dit que le vecteur aléatoire X sur $\mathbb{R}^n$ **admet une densité** s'il existe une fonction intégrable $f_X : \mathbb{R}^n \to \mathbb{R}$ telle que

$$\mathbb{P}_X[B] = \int_B f_X(x)dx \qquad \forall B \in \mathcal{B}(\mathbb{R}^n)$$

ou, de façon équivalente,

$$\mathbb{E}[\ \phi(X)] = \int_{\mathbb{R}^n} \phi(x) f_X(x)dx \qquad \forall \phi : \mathbb{R}^n \to \mathbb{R} \text{ borélienne bornée.}$$

La fonction f_X (qui est définie de façon unique dans $L^1(\mathbb{R}^n, dx)$ par une des relations ci-dessus) est positive et d'intégrale égale à 1.

— **La fonction caractéristique** Φ_X d'un vecteur aléatoire X sur $\mathbb{R}^n$ est (la conjuguée de) la transformée de Fourier de $\mathbb{P}_X$:

$$\Phi_X(t) = \mathbb{E}\left[e^{i\langle X,t\rangle}\right] = \int_{\mathbb{R}^n} e^{i\langle x,t\rangle}\mathbb{P}_X(dx) \qquad \forall t \in \mathbb{R}^n,$$

où $\langle\cdot,\cdot\rangle$ est le produit scalaire de $\mathbb{R}^n$.

La fonction caractéristique caractérise la loi : si X et Y sont deux vecteurs aléatoires tels que $\Phi_X = \Phi_Y$, alors X et Y ont même loi, i.e., $\mathbb{P}_X = \mathbb{P}_Y$.

4.2. **Variance.** Si X, Y sont deux variables aléatoires de carré intégrable (i.e., $X, Y \in L^2(\Omega)$), la covariance de (X, Y) est

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

La variance de X est

$$\operatorname{var}(X) = \operatorname{cov}(X, X) = \mathbb{E}\left[|X - \mathbb{E}(X)|^2\right] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2.$$

On a

$$\operatorname{var}(X + Y) = \operatorname{var}(X) + \operatorname{var}(Y) + 2\operatorname{cov}(X, Y).$$

Proposition 4.1 (Deux inégalités fondamentales)**.**

— *(Inégalité de Markov) Si $X \in L^1(\Omega)$, alors, pour tout $a > 0$,*

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

— *(Inégalité de Bienaymé-Tchebicheff) Si $X \in L^2(\Omega)$, alors pour tout $a > 0$,*

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

Plus généralement si $X = (X_1, \ldots, X_n) \in (L^2(\Omega))^n$ est un vecteur aléatoire, la matrice de covariance de X est la matrice symétrique

$$K_X := (\operatorname{cov}(X_i, X_j))_{1\leq i,j\leq n}.$$

Cette égalité s'écrit aussi (en identifiant un vecteur avec la matrice colonne associée)

$$K_X = \mathbb{E}\left[(X - \mathbb{E}[X])\ (X - \mathbb{E}[X])^T\right],$$

où l'espérance d'une matrice est simplement la matrice de l'espérance de ses coefficients.

Exercice 29. Soit $X = (X_1, \dots, X_n) \in (L^2(\Omega))^n$ un vecteur aléatoire, M une matrice $p \times n$ et $b \in \mathbb{R}^p$. Montrer que

$$K_{MX+b} = MK_X\ M^T.$$

Montrer que K_X est une matrice semi-définie positive.

Attention à ne pas confondre le fait que des variables aléatoires sont décorrélées (i.e., dans le cas de deux v.a., telles que $\text{cov}(X_1, X_2) = 0$) du fait qu'elles sont indépendantes. On verra juste après que la notion d'indépendance est beaucoup plus forte : en gros, X_1 et X_2 sont indépendantes si "toute" fonction de X_1 est décorrélée de "toute" fonction de X_2. Ces deux notions, bien distinctes en général, coïncident lorsque le vecteur (X_1, X_2) est gaussien (voir plus bas).

4.3. **Indépendance.** On dit que des événements A_1 et A_2 sont indépendants si $\mathbb{P}[A_1 \cap A_2] = \mathbb{P}[A_1]\mathbb{P}[A_2]$. Plus généralement, les événements $\{A_i\}_{i \in I}$ (où I est un ensemble quelconque) sont indépendants si, pour tout sous-ensemble fini $J \subset I$, on a

$$\mathbb{P}\left[\bigcap_{i \in J} A_i\right] = \prod_{i \in J} \mathbb{P}\left[A_i\right].$$

On dit que des sous-tribus $\mathcal{A}_1$ et $\mathcal{A}_2$ sont indépendantes si, pour tout $(A_1, A_2) \in \mathcal{A}_1 \times \mathcal{A}_2$, les événements A_1 et A_2 sont indépendants. Cette notion se généralise comme au-dessus à une famille arbitraire de sous-tribus.

Définition 4.2. *On dit que deux variables aléatoires X_1 et X_2 sont indépendantes si une des assertions équivalentes a lieu :*

(1) les tribus $\sigma(X_1)$ et $\sigma(X_2)$ engendrées par X_1 et X_2 sont indépendantes.

(2) Pour tout couple de fonctions mesurables bornées $f_1, f_2 : \mathbb{R} \to \mathbb{R}$, on a

$$\mathbb{E}\left[f_1(X_1) f_2(X_2)\right] = \mathbb{E}\left[f_1(X_1)\right] \mathbb{E}\left[f_2(X_2)\right].$$

(3) (caractérisation en terme de loi) $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$; autrement dit, pour toute fonction $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ borélienne bornée, on a

$$\mathbb{E}\left[f(X,Y)\right] = \int_{\mathbb{R}} (\int_{\mathbb{R}} f(x,y) \mathbb{P}_Y(dy)) \mathbb{P}_X(dx) = \int_{\mathbb{R}} (\int_{\mathbb{R}} f(x,y) \mathbb{P}_X(dx)) \mathbb{P}_Y(dy).$$

(4) (caractérisation en terme de fonctions caractéristiques)

$$\Phi_{(X_1,X_2)}(\xi_1, \xi_2) = \Phi_{X_1}(\xi_1)\Phi_{X_2}(\xi_2) \qquad \forall (\xi_1, \xi_2) \in \mathbb{R}^2.$$

Les assertions (2) et (4) permettent de *montrer* que les v.a. sont indépendantes ; on *utilise* le plus souvent cette indépendance via l'assertion (3). Ces équivalences se généralisent à un nombre quelconque de variables aléatoires. On dit que des variables aléatoires $(X_n)_{n \in \mathbb{N}}$ sont **i.i.d.** si elles sont **indépendantes et identiquement distribuées**.

Exercice 30 (Algorithme de Box-Muller)**.** Soit U_1 et U_2 deux variables indépendantes de loi uniforme sur $[0,1]$. Montrer que les variables aléatoires

$$Z_1 = \sqrt{-2\ln(U_1)} \cos(2\pi U_2), \qquad Z_2 = \sqrt{-2\ln(U_1)} \sin(2\pi U_2)$$

sont indépendante et de loi normale $\mathcal{N}(0,1)$ (Indication : passer en coordonnées polaires).

4.4. **Convergence de variables aléatoires.**

Définition 4.3. *Soit (X_n) une suite de variables aléatoires définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et X une variable aléatoire. On dit que (X_n) converge vers X*

— **presque sûrement** *si*

$$\mathbb{P}\left[\lim_n X_n = X\right] = 1 \qquad (\text{noté } X_n \xrightarrow{p.s.} X)$$

— **uniformément (ou dans L^∞)** *si*

$$\lim_n \|X_n - X\|_\infty \longrightarrow 0.$$

(on sous-entend ici que les X_n et X sont bornées).

— **en moyenne d'ordre p (ou dans L^p)** *(où $p \in [1, +\infty[$) si*

$$\lim_n \mathbb{E}\left[|X_n - X|^p\right] = 0 \qquad (\text{noté } X_n \xrightarrow{L^p} X)$$

(on sous-entend ici que les X_n et X sont dans $L^p(\Omega)$). Si $p = 1$, on parle de **convergence en moyenne** *et, si $p = 2$, de* **convergence en moyenne quadratique***.*

— **en probabilité** *si, pour tout $\varepsilon > 0$,*

$$\lim_n \mathbb{P}\left[|X_n - X| \geq \varepsilon\right] = 0. \qquad (\text{noté } X_n \xrightarrow{\mathbb{P}} X)$$

— **en loi** *si, pour toute fonction continue bornée $f : \mathbb{R} \to \mathbb{R}$,*

$$\lim_n \mathbb{E}\left[f(X_n)\right] = \mathbb{E}\left[f(X)\right]. \qquad (\text{noté } X_n \xrightarrow{\mathcal{L}} X)$$

Noter que, pour cette dernière forme de convergence, il n'est pas utile que les X_n et X soient définies sur le même espace de probabilité.

Les notions ci-dessus se généralise aisément au cas de vecteurs aléatoires (si les vecteurs sont dans $\mathbb{R}^d$, remplacer la valeur absolue par n'importe quelle norme sur $\mathbb{R}^d$, la fonction f dans la convergence en loi devant aller de $\mathbb{R}^d$ dans $\mathbb{R}$).

Il existe de très nombreuses relations entre ces différentes formes de convergence, difficile de les retenir toutes. Il vaut mieux connaître par coeur les relations suivantes (pour gagner du temps) :

Proposition 4.4 (Lien entre les convergence)**.**

(1) La convergence en probabilité implique la convergence en loi.

(2) La convergence p.s. ainsi que la convergence en moyenne d'ordre p (où $p \in [1, +\infty[$) impliquent la convergence en probabilité et la convergence en loi.

(3) La convergence en moyenne implique l'existence d'une sous-suite qui converge p.s.

(4) Si $1 \leq p \leq q$, la convergence dans L^q implique la convergence dans L^p.

(5) La convergence uniforme implique toute les autres convergences.

La caractérisation suivante de la convergence en loi en terme de fonctions de répartition et de fonctions caractéristiques est très utile :

Théorème 4.5. *Sous les hypothèses de la définition 4.3, les assertions suivantes sont équivalentes :*

- *(X_n) converge en loi vers X,*
- *$F_{X_n}(t) \to F_X(t)$ en tout point t de continuité de F_X,*
- *$\Phi_{X_n}(t) \to \Phi_X(t)$ pour tout $t \in \mathbb{R}$,*

où F_{X_n} (resp. Φ_{X_n}) est la fonction de répartition (resp. fonction caractéristique) de X_n.

La dernière équivalence est connue sous le nom de théorème de Lévy et reste vraie pour les vecteurs aléatoires. Alors qu'elle intervient très fréquemment en probabilité et en statistique, la convergence en loi est délicate à manipuler car totalement incompatible avec les opérations usuelles. Retenir cependant :

Théorème 4.6 (de Slutsky)**.** *Si (X_n) converge en loi vers une variable aléatoire X et (Y_n) converge en probabilité vers une constante c, alors (X_n, Y_n) converge en loi vers (X, c). En particulier, $(X_n + Y_n)$ converge en loi vers $(X + c)$ et (X_nY_n) converge en loi vers cX.*

4.5. **Vecteurs gaussiens.** Rappelons que la loi gaussienne (ou normale) $\mathcal{N}(m, \sigma^2)$ (de moyenne $m \in \mathbb{R}$ et d'écart-type $\sigma \geq 0$) est

— si $\sigma > 0$, la probabilité sur $\mathbb{R}$ de densité $\frac{1}{\sigma\sqrt{2\pi}} \exp\{-(x-m)^2/(2\sigma^2)\}$.

— si $\sigma = 0$, la masse de Dirac δ_m.

On dit que la loi est réduite si $\sigma = 0$, centrée si $m = 0$ et dégénérée si $\sigma = 0$.

Définition 4.7. *Un vecteur aléatoire $(X_1, \ldots, X_n)$ sur $(\Omega, \mathcal{A}, \mathbb{P})$ est dit gaussien si toute combinaison linéaire des $(X_i)_{i=1,\ldots,n}$ suit une loi gaussienne (autrement dit, pour tout $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, la variable aléatoire $\alpha_1 X_1 + \cdots + \alpha_n X_n$ suit une loi gaussienne).*

En particulier, si $(X_1, \ldots, X_n)$ est un vecteur gaussien, toutes les variables aléatoires X_i ($i = 1, \ldots, n$) sont gaussiennes.

Théorème 4.8 (Caractérisation par les fonctions caractéristiques)**.** *Un vecteur aléatoire $X = (X_1, \ldots, X_n)$ est gaussien si et seulement si il existe un vecteur (déterministe) $m_X \in \mathbb{R}^n$ et une matrice symétrique (déterministe) $K_X \in \mathbb{R}^{n\times n}$ tels que*

$$\Phi_X(\xi) = \exp\left\{ i\langle m_X, \xi\rangle - \frac{1}{2}\langle K_X\xi, \xi\rangle \right\} \qquad \forall \xi \in \mathbb{R}^n.$$

De plus, m_X est l'espérance du vecteur X et K_X est la matrice de covariance.

Proposition 4.9. *On suppose que le vecteur aléatoire $X = (X_1, \ldots, X_n)$ est gaussien. Alors les variables aléatoires $\{X_1, \ldots, X_n\}$ sont indépendantes, si et seulement si, la matrice de covariance K_X de X est diagonale.*

Ce résultat est une conséquence directe de la caractérisation de l'indépendance en terme de fonctions caractéristiques et de la forme spécifique des fonctions caractéristiques pour les vecteurs gaussiens.

Théorème 4.10 (Vecteurs gaussiens à densité)**.** *On suppose que le vecteur aléatoire $X = (X_1, \ldots, X_n)$ est gaussien. Alors ce vecteur admet une densité, si et seulement si* $\det(K_X) \neq 0$ *(où K_X est la matrice de covariance de X). Dans ce cas, la densité de X est donnée par*

$$f_X(x) = \frac{1}{(2\pi)^{n/2}(\det(K_X))^{1/2}} \exp\{-\frac{1}{2}\langle K_X^{-1}(x - m_X), (x - m_X)\rangle\} \qquad \forall x = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Annexe A. Exercices supplémentaires

A.1. **Exercices de topologie et d'analyse fonctionnelle.**

Exercice 31. Si (X, d) est un espace métrique, démontrer l'inégalité triangulaire inverse $\Big|d(x, y_1) - d(x, y_2)\Big| \leq d(y_1, y_2)$ pour tous $x, y_1, y_2 \in X$.

Exercice 32 (Valeur d'adhérence)**.** Soit (x_n) une suite dans un espace métrique (X, d) et $\ell \in X$. Montrer l'équivalence entre les assertions suivantes :

(1) Pour tout $\varepsilon > 0$, l'ensemble $\{n \in \mathbb{N},\ d(x_n, \ell) < \varepsilon\}$ est infini,

(2) $\ell \in \bigcap_n \overline{\{x_p : p \geq n\}}$ (où $\bar{A}$ est l'adhérence d'un ensemble $A \subset X$),

(3) il existe une sous-suite $(x_{\phi(n)})$ telle que $\lim x_{\phi(n)} = \ell$.

On dit alors que ℓ est valeur d'adhérence de (x_n).

Exercice 33. On suppose que $(X_1, d_1), (X_2, d_2)$ sont deux espaces métriques compacts. Montrer que le produit $X_1 \times X_2$, muni de la distance $d((x_1, x_2), (y_1, y_2)) = \max(d_1(x_1, y_1), d_2(x_2, y_2))$, est compact.

Exercice 34. Montrer qu'une bijection continue $f : X \to Y$ où X est métrique compact et Y métrique, est de réciproque continue. (on pourra utiliser l'exercice 7).

Exercice 35. Soit (X, d) un espace métrique et $A \subset X$ un sous-ensemble non vide X. On suppose que $f : A \to \mathbb{R}$ est une fonction L-lipschitzienne sur A et on définit pour tout $x \in X$

$$\bar{f}(x) = \inf_{y \in A} f(y) + Ld(x, y).$$

Montrer que $\bar{f}$ est une fonction $L-$Lipschitzienne sur X qui est égale à f sur A.

Exercice 36. Soit (X, d) un espace métrique.

(1) On suppose que A est un sous-ensemble fermé de X. Montrer qu'il existe une fonction continue $f : X \to [0, +\infty[$ qui s'annule exactement sur A.

(2) On suppose que A et B sont deux sous-ensembles fermés de X avec $A \cap B = \emptyset$. Montrer qu'il existe une fonction continue $g : X \to [0, 1]$ telle que $A = \{x \in X,\ g(x) = 0\}$ et $B = \{x \in X,\ g(x) = 1\}$.

Exercice 37. Soit $(E, \|\cdot\|)$ un espace vectoriel normé et F un sous-espace vectoriel de E. On suppose que F a un intérieur non vide. Montrer que $F = E$.

Exercice 38 (Une version du théorème de Cauchy-Lipschitz)**.** Soit $f : \mathbb{R}^n \to \mathbb{R}^n$ une fonction $L-$Lipschitzienne (où $L \in \mathbb{R}^*$), $x_0 \in \mathbb{R}^n$ et, pour $T > 0$, $E_T := C^0([-T, T], \mathbb{R}^n)$, muni de la norme

$$\|x\|_{E_T} = \sup_{t \in [-T, T]} \|x(t)\|.$$

Finalement on définit la fonction $\Phi : E_T \to E_T$ par

$$\Phi(x)(t) = x_0 + \int_0^t f(x(s))ds \qquad \forall t \in [-T, T],\ \forall x \in E_T.$$

(1) Montrer que, si $T > 0$ est suffisamment petit, l'application Φ_T est une contraction sur E_T.

(2) En déduire que, si $T > 0$ est suffisamment petit, il existe une unique solution à l'équation différentielle

$$x't) = f(x(t)),\ \forall t \in [-T, T], \qquad x(0) = x_0.$$

A.2. **Exercices de calcul différentiel.**

Exercice 39. Soit $f : \mathbb{R}^2 \to \mathbb{R}$ définie par $f(x,y) = \frac{xy^2}{x^2+y^4}$ si $(x,y) \neq (0,0)$ et $f(0,0) = 0$. Montrer que la restriction de f à toute droite passant par l'origine est continue, mais que f n'est pas continue en $(0,0)$.

Exercice 40. Soit $\mathcal{O} := \{(x,y) \in \mathbb{R}^2,\ x \neq y\}$. Montrer que l'application $f : \mathcal{O} \to \mathbb{R}$ définie par $f(x,y) = \frac{\sin(x)-\sin(y)}{x-y}$ est prolongeable par continuité à $\mathbb{R}^2$. Montrer que ce prolongement est de classe C^1 sur $\mathbb{R}^2$.

Exercice 41. Soit $f : \mathbb{R}^2 \to \mathbb{R}$ définie par $f(x,y) = \frac{xy^2}{x^2+y^2}$ si $(x,y) \neq (0,0)$ et $f(0,0) = 0$. Montrer que f est continue en $(0,0)$, que $\frac{\partial f}{\partial x}(0,0)$ et $\frac{\partial f}{\partial y}(0,0)$ existent, mais que f n'est pas différentiable en $(0,0)$.

Exercice 42. Soit $f : \mathbb{R}^n \to \mathbb{R}$ (où $n \in \mathbb{N}^*$) définie par $f(x) = \sin(\|x\|^2 + 1)$ pour tout $x \in \mathbb{R}^n$ ($\|x\|$ étant la norme euclidienne de $x \in \mathbb{R}^n$). Montrer que f est de classe C^2 sur $\mathbb{R}^n$, calculer son gradient et sa matrice hessienne en tout point.

Exercice 43 (Relation d'Euler)**.** Soit $f : (\mathbb{R}^n)^* \to \mathbb{R}$ une fonction de classe C^1 et $k \in \mathbb{R}^*$. Montrer que f est $k-$homogène (i.e., $f(\lambda x) = \lambda^k f(x)$ pour tout $x \in (\mathbb{R}^n)^*$ et pour tout $\lambda > 0$), si et seulement si, $\langle f(x), x\rangle = kf(x)$ pour tout $x \in (\mathbb{R}^n)^*$.

Exercice 44. Soit $M_n(\mathbb{R})$ l'ensemble des matrices de format $n \times n$ et $GL_n(\mathbb{R})$ l'ensemble des matrices de $M_n(\mathbb{R})$ inversibles.

(1) Montrer que $GL_n(\mathbb{R})$ est un ouvert de $M_n(\mathbb{R})$.

(2) Montrer que les applications suivantes sont de classe C^1 sur leur domaine de définition et déterminer leur différentielle.
$$f(A) = A^2, \qquad \forall A \in M_n(\mathbb{R}); \qquad g(A) = A^{-1}, \qquad \forall A \in GL_n(\mathbb{R}).$$

A.3. **Exercices d'intégration.**

Exercice 45. Soit $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, où λ est la mesure de Lebesgue.

(1) Soit $\mathcal{O}$ un ouvert de $\mathbb{R}$. Montrer que $\lambda(\mathcal{O}) > 0$.

(2) Soit $\mathcal{O}$ un ouvert de $\mathbb{R}$. On suppose que $\lambda(\mathcal{O}) < \infty$. L'ensemble $\mathcal{O}$ est-il nécessairement borné ?

(3) Construire un ouvert dense de $\mathbb{R}$ de mesure de Lebesgue égal à 3.

Exercice 46. Soient μ et ν deux mesures de Borel finies sur $\mathbb{R}^d$. On pose
$$\sigma(A) := \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbf{1}_{x+y\in A} (\mu \otimes \nu)(dx, dy), \quad \text{pour tout borélien } A \subseteq \mathbb{R}^d.$$

(1) Montrer que σ est une mesure positive sur $\mathbb{R}^d$, on note $\mu * \nu$.

(2) Remarquer que $\mu * \nu$ est une mesure finie et que $\mu * \nu = \nu * \mu$.

(3) Montrer que si $\mu = f\lambda^d, \nu = g\lambda^d$, où λ^d est la mesure de Lebesgue sur $\mathbb{R}^d$ et $f, g \geq 0$ sont intégrables, alors $\mu \star \nu = (f \star g)\lambda^d$.

Exercice 47. Dans les cas suivants (où $f_n : \mathbb{R}^+ \to \mathbb{R}$) montrer que la suite $(\int_{\mathbb{R}^+} f_n(x)dx)_{n\in\mathbb{N}}$ converge et déterminer sa limite :

(1) $f_n(x) = \dfrac{ne^{-x}}{\sqrt{1+n^2x^2}}$,

(2) $f_n(x) = \dfrac{ne^{-nx}}{\sqrt{1+n^2x^2}}$,

(3) $f_n(x) = \sin(nx)\mathbf{1}_{[0,n]}(x)$,

(4) $f_n(x) = |\cos(x)|^{1/n} e^{-x}$,

(5) $f_n(x) = \dfrac{ne^{-x}}{nx+1}\mathbf{1}_{[0,1]}$,

(6) $f_n(x) = \dfrac{\sin(nx^n)}{nx^{n+\frac{1}{2}}}$.

Exercice 48. Calculer la limite des suites suivantes :

$$\int_{\mathbb{R}} e^{-|x|/n} dx, \qquad \int_{\mathbb{R}} \frac{e^{-x^2}}{2\cos(\frac{x}{n})-1}\ \mathbf{1}_{\{3|\cos(\frac{x}{n})|\geq 2\}}\ dx, \qquad \sum_{m\geq 1} \frac{n}{m}\sin(\frac{1}{nm}).$$

Exercice 49. (1) Montrer que l'application φ définie sur $\mathbb{R}^2$ par $\varphi(u,v) = (u^2+v^2, 2uv)$ est un C^1-difféomorphisme de $\Delta = \{(u,v) \in \mathbb{R}^2; u > v > 0\}$ sur $D = \{(x,y) \in \mathbb{R}^2; x > y > 0\}$.

(2) En déduire la valeur de $\int_{(\mathbb{R}_+)^2} |u^4 - v^4| e^{-(u+v)^2} dudv$.

Exercice 50. Calculer l'intégrale

$$I = \int_{y>x>0} e^{-y+x} \frac{\sqrt{y-x}}{y^2} dxdy.$$

[Indication : on pourra considérer le changement de variable $u = y - x, v = y/x$.]

A.4. **Exercices de probabilité.**

Exercice 51. Soient X, Y deux variables aléatoires indépendantes réelles. Calculer la loi de $X+Y$ dans les cas suivants :

(1) X et Y ont loi uniforme sur $[-1,1]$.

(2) X et Y ont respectivement une loi de densité $\gamma_{a,\lambda}$ et $\gamma_{b,\lambda}$ où

$$\gamma_{a,\lambda}(x) = \frac{\lambda^a}{\Gamma(a)} e^{-\lambda x} x^{a-1} \mathbf{1}_{\mathbb{R}_+}(x), \quad \Gamma(a) := \int_0^\infty e^{-x}\, x^{a-1}\, dx.$$

On pourra vérifier et utiliser le fait que $\int_{\mathbb{R}_+} \gamma_{a,\lambda} = 1$.

Exercice 52. Soit (X_1, X_2) un couple de variables aléatoires de densité

$$f(x_1, x_2) = e^{-(x_1+x_2)} \mathbf{1}_{\mathbb{R}_+}(x_1) \mathbf{1}_{\mathbb{R}_+}(x_2) \qquad \forall (x_1, x_2) \in \mathbb{R}^2.$$

(1) Quelle est la densité de $Y_1 = X_1 + X_2$?

(2) Quelle est la densité de $Y_2 = X_1/X_2$?

(3) Les variables aléatoire Y_1 et Y_2 sont-elles indépendantes ?

Exercice 53. Soient $(X_n)_{n\in\mathbb{N}}$ une suite de variable aléatoire i.i.d. telles que $X_0 \in L^2(\Omega)$. Montrer que

$$\bar{X}_N := \frac{1}{N}\sum_{n=1}^N X_n \text{ tend vers } \mathbb{E}[X_0] \text{ dans } L^2(\Omega).$$

Exercice 54. Soit X un vecteur aléatoire dans $\mathbb{R}^d$ (où $d \in \mathbb{N}^*$) et M une matrice réelle de format $k \times d$.

(1) Montrer que

$$\mathbb{E}[MX] = M\mathbb{E}[X] \qquad \text{et} \qquad K_{MX} = M\ K_X\ M^T.$$

(2) On suppose que X est un vecteur gaussien de loi $\mathcal{N}(0, I_d)$. Quelle est la loi de MX ?

(3) On suppose que X est un vecteur gaussien sur $\mathbb{R}^3$ de loi $\mathcal{N}(0, D)$ avec

$$D = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Montrer qu'il existe une matrice orthogonale $P \in \mathbb{R}^{3\times 3}$ telle que $P^{-1}DP$ soit diagonale. Déterminer la loi du vecteur $P^{-1}X$.

Exercice 55. Soit $X_1, \dots, X_n$ des variables aléatoires indépendantes de même loi $\mathcal{N}(0,1)$ et $a_1, \dots, a_n$ et $b_1, \dots, b_n$ des réels. On pose $Y = \sum_{i=1}^n a_i X_i$ et $Z = \sum_{i=1}^n b_i X_i$.

(1) Montrer que (Y, Z) est un vecteur gaussien.

(2) Montrer que Y et Z sont indépendantes, si et seulement si, $\sum_{i=1}^{n} a_i b_i = 0$.

(3) Montrer que les variables aléatoires $X_1 + X_2 + X_3$, $2X_1 - X_2 - X_3$ et $X_2 - X_3$ sont indépendantes.

Liens vers quelques cours en ligne

De très nombreux cours des L1-L2-L3 MIDO sont disponibles en ligne, et ces notes sont largement tributaires des polycopiés que le lecteur peut trouver sur les sites de :

— Guillaume Carlier, pour la partie "Calcul différentiel et optimisation" (et le très beau cours d'analyse fonctionnelle, plutôt de niveau M1)

— Jacques Fejoz, pour la partie "Calcul différentiel et optimisation" et "Intégrale de Lebesgue et probabilités"

— Guillaume Legendre, pour la partie "Algèbre linéaire"

— Stéphane Mischler, pour la partie "Intégration-Probabilités"

— Paul Pegon, pour le cours précédent de rappels d'Analyse de L3 (bien plus complet que ces notes), ainsi que pour les exercices correspondants (la partie "Exercice supplémentaires" lui est également très largement redevable).

— Eric Séré, pour la partie "Topologie et analyse fonctionnelle".

Le cours en ligne de Jean-François Legall sur l'intégration, les probabilités et les processus aléatoires est également une excellente source (la dernière partie étant particulièrement utile en M1).

Je me suis abstenu de proposer une bibliographie : le lecteur trouvera de très bonnes listes de références classiques dans les cours pré-cités.

Index

Applications linéaires continues, 6

Base hilbertienne, 8

Convergence
  en loi, 25
  en moyenne, 25
  en moyenne quadratique, 25
  en probabilité, 25
  normale, 6
  presque sûre, 25
Convolution, 21
Covariance, 23

Densité, 4
  dans L^p, 19
  dans l'espace des fonctions continues, 9
Densité d'un vecteur aléatoire, 23
Différentiabilité, 9

Ensemble
  de mesure nulle, 13
Equation différentielle, 12
  linéaire, 13
Espace
  L^p, 19
  ℓ^p, 9
  $\mathcal{L}^p$, 15, 18
  compact, 5
  compact en dimension finie, 6
  complet, 4
  de Banach, 6
  de Hilbert, 7
  de probabilité, 13
  des fonctions continues, 8
  mesuré, 13
  métrique, 3
  probabilisé, 22
  séparable, 4

Fonction
  étagée, 14
  intégrable, 15
Fonction caractéristique, 23
Fonction indicatrice, 2

Gradient, 10

Indépendance, 24

Inégalité
  de Bienaymé-Tchebicheff, 23
  de Hölder, 16
  de Jensen, 16
  de Markov, 23
  de Minkowski, 16
  triangulaire, 16
Intégrale
  de fonctions à valeurs complexes, 16
  de fonctions étagées, 14
  de fonctions intégrables, 15
  de fonctions positives, 15

Lemme
  de Borel-Cantelli, 14
  de Fatou, 17
  de Gronwall, 12
Loi d'un vecteur aléatoire, 23

Matrice jacobienne, 10

Orthogonal, 7

Suite de Cauchy, 4

Théorème
  d'inversion de la transformée de Fourier, 21
  d'inversion locale, 12
  de Cauchy-Lipschitz, 12
  de continuité de fonctions définie par une intégrale, 19
  de convergence dominée, 17
  de convergence monotone, 16
  de différentiabilité de fonctions définie par une intégrale, 20
  de Fubini, 18
  de projection sur un convexe fermé, 7
  de projection sur un sous-espace, 7
  de régularisation par convolution, 22
  de représentation de Riesz, 7
  de Slutsky, 26
  de Stone-Weierstrass, 9
  des fonctions implicites, 12
  du point fixe, 5

Variable aléatoire, 22
Variance, 23
Vecteur aléatoire, 22
Vecteur gaussien, 26