

Théorie de la mesure et probabilité

Basile de Loynes

25 septembre 2023

Ce document est mis à disposition selon les termes de la licence [Creative Commons](#) “Attribution – Partage dans les mêmes conditions 4.0 International”.



Table des matières

Introduction	vii
I Topologie et théorie de la mesure	1
1 Rappels et compléments d'analyse	3
1.1 Espaces vectoriels normés	3
1.2 Espaces métriques	4
1.2.1 Métrique, boule ouverte, boule fermée, parties bornées	4
1.2.2 Topologie des espaces métriques	5
1.2.3 Notion de limites	9
1.2.4 Continuité	12
1.2.5 Topologies et opérations ensemblistes	13
1.2.6 Compacité	15
1.2.7 Espaces métriques complets	20
1.3 Espaces polonais	23
2 Tribus, applications mesurables et mesures	25
2.1 Tribus et Applications mesurables	25
2.1.1 Tribu	25
2.1.2 Tribu borélienne	27
2.1.3 La droite achevée	28
2.1.4 Applications mesurables, applications boréliennes	29
2.1.5 Approximation des fonctions mesurables	31
2.2 Mesures positives	32
2.2.1 Définitions et propriétés élémentaires	32
2.2.2 Quelques exemples de mesures : mesures discrètes et mesure de Lebesgue	35
2.2.3 Théorème des classes monotones, caractérisation des mesures et théorème de prolongement de Carathéodory	37
2.2.4 Régularité des mesures, mesures de Borel et espaces polonais	46
3 Intégrale au sens de Lebesgue	51
3.1 Construction de l'intégrale de Lebesgue	51
3.1.1 Intégration des fonctions étagées positives	51
3.1.2 Intégration des fonctions mesurables positives	52
3.1.3 Intégration des fonctions mesurables	54
3.2 L'intégrale de Lebesgue en pratique	56
3.2.1 L'intégrale de Lebesgue contre des mesures discrètes	56
3.2.2 Mesures à densité	57
3.2.3 Mesure image et théorème de transfert	58
3.2.4 Intégrale de Riemann et intégrale de Lebesgue	59

4	Théorèmes limites	63
4.1	Lemme de Fatou	63
4.2	Ensembles et fonctions mesurables négligeables	63
4.3	Théorème de convergence dominée	65
4.4	Intégrale à paramètres	66
5	Mesure produit	69
5.1	Mesure produit	69
5.2	Théorèmes de Fubini-Tonelli et de Fubini-Lebesgue	71
5.3	La mesure produit en application	73
5.4	Mesure image et changement de variables	75
6	Espaces \mathcal{L}^p et L^p	79
6.1	Généralités	79
6.2	Inégalités de Hölder et de Minkowski	80
6.3	Théorème de Radon-Nikodym	82
6.3.1	Un peu d'espace de Hilbert	82
6.3.2	Lemme de Fréchet-Riesz	85
6.3.3	Théorème de Radon-Nikodym, cas des mesures positives	86
6.3.4	Théorème de Radon-Nikodym, cas des mesures signées	88
6.4	Approximation dans les espaces L^p , $p \in [1, \infty)$	90
6.4.1	Approximation par des fonctions étagées mesurables	90
6.4.2	Approximation par des fonctions continues à support compact	91
6.4.3	Convolution	92
II	Probabilités générales	97
7	Variables aléatoires réelles et vecteurs aléatoires	99
7.1	Variables aléatoires	99
7.2	Variables aléatoires réelles	101
7.2.1	Intégration des variables aléatoires réelles	101
7.2.2	Caractérisation de la loi d'une <i>v.a.r.</i>	104
7.2.3	Exemples de calcul de lois	108
7.2.4	Classification des lois de probabilités sur \mathbb{R}	110
7.2.5	Simulation de lois	110
7.3	Vecteurs aléatoires	112
7.3.1	Généralités	112
7.3.2	Loi d'un vecteur aléatoire, lois marginales	112
7.3.3	Moments	112
7.3.4	Lois à densité	113
7.3.5	Fonction de répartition	113
7.3.6	Transformation des vecteurs aléatoires à densité	114
8	Indépendance	115
8.1	Tribus indépendantes	115
8.2	Lemme de Borel-Cantelli	117
8.3	Variables aléatoires indépendantes	118
8.3.1	Définition et caractérisation élémentaire	118
8.3.2	Constructions de variables aléatoires indépendantes	118
8.3.3	Caractérisation de l'indépendance de <i>v.a.r.</i>	120
8.4	Une application du second lemme de Borel-Cantelli	122

9 Fonctions caractéristiques	123
9.1 Fonction caractéristique d'une <i>v.a.r.</i>	123
9.2 Fonctions caractéristiques et moments	125
9.3 Fonctions caractéristiques de vecteurs aléatoires	126
9.4 Fonctions caractéristiques et indépendance	126
10 Vecteurs gaussiens	127
10.1 Manipulation des vecteurs gaussiens	127
10.2 Loi du χ^2 , moyenne et variance empiriques	130
11 Convergences de suites de variables aléatoires	133
11.1 Convergences trajectorielles	133
11.1.1 Convergence presque sûre ou presque partout	133
11.1.2 Convergence dans \mathbf{L}^p	134
11.1.3 Convergence en probabilité	135
11.1.4 Convergence trajectorielle et critère de type Cauchy	137
11.2 Convergence étroite et convergence en loi	138
11.2.1 Convergence étroite	139
11.2.2 Convergence en loi	143
11.3 Loi du 0-1 de Kolmogorov et séries aléatoires	146
12 Loi des grands nombres et Théorème Central Limite	153
12.1 Loi des grands nombres	153
12.2 Théorème Central Limite	158
12.3 TCL multivarié	158
12.4 Applications de la loi des grands nombres	159
13 Espérance conditionnelle	163
13.1 Conditionnement par un événement	163
13.2 Espérance conditionnelle	165
13.3 Propriétés de l'espérance conditionnelle	166
13.4 Inégalité de Jensen et de Markov conditionnelles	168
13.5 Conditionnement des vecteurs gaussiens	168
13.6 Point de vue hilbertien des espérances conditionnelles	169
13.7 Lois conditionnelles régulières	169
13.7.1 Densité conditionnelle	169
13.7.2 Noyau de transition et loi conditionnelle régulière	171
Lois usuelles	173
13.8 Lois discrètes	173
13.9 Lois continues	173

Introduction

Ces notes constituent le support d'un cours dispensé en première année de l'Ensay. L'objectif principal de ce cours est d'introduire le formalisme moderne de la théorie des probabilités. Un cours de probabilité moderne ne saurait se dispenser des bases solides données par la théorie de la mesure. C'est ainsi que ce cours est découpé en deux parties : la première est dédiée la construction de l'intégrale de Lebesgue dans le formalisme de la théorie de la mesure ; la seconde quant à elle s'attachera à introduire les concepts fondamentaux de probabilité.

Dans l'axiomatique de la théorie de la mesure, on se donne un triplet $(\mathbb{X}, \mathcal{X}, \mu)$ où

- \mathbb{X} est un ensemble ;
- \mathcal{X} est une collection de parties de \mathbb{X} dites parties mesurables ;
- μ est une fonction d'ensembles de \mathcal{X} dans \mathbb{R}_+ .

On s'attachera dans la première partie du cours à définir proprement la notion de partie mesurable et mesure. Puis, on donnera un sens aux notations

$$\int_{\mathbb{X}} f \, d\mu, \quad \int_{\mathbb{R}} f(x) \, \mu(dx) \quad \text{ou encore} \quad \int_{\mathbb{R}} f(x) \, dx.$$

Au vu de la notation utilisée ci-dessus à droite, on peut s'interroger sur l'utilité de construire une nouvelle intégrale. D'autant plus que, comme nous le verrons, toute fonction numérique réelle, intégrable au sens de Riemann sur un intervalle $[a, b]$ de \mathbb{R} , est en particulier intégrable au sens de Lebesgue et les deux intégrales coïncident. Avant de décrire le contenu de ce cours, prenons le temps de discuter les raisons nous poussant à construire une nouvelle intégrale.

Pour cela, rappelons succinctement la construction de l'intégrale de Riemann. Soit $f : [a, b] \rightarrow \mathbb{R}$ une fonction bornée que l'on supposera positive pour simplifier. On considère une subdivision de l'intervalle $[a, b]$, notée σ , $a = t_0 < t_1 < \dots < t_{n-1} < t_n = b$. Les sommes de Darboux inférieure et supérieure relativement à la subdivision σ sont définies respectivement par

$$s(f, \sigma) = \sum_{i=0}^{n-1} (t_{i+1} - t_i) \inf_{x \in [t_i, t_{i+1}]} f(x) \quad \text{et} \quad S(f, \sigma) = \sum_{i=0}^{n-1} (t_{i+1} - t_i) \sup_{x \in [t_i, t_{i+1}]} f(x).$$

En notant \mathfrak{S} l'ensemble des subdivisions de l'intervalle $[a, b]$, une fonction $f : [a, b] \rightarrow \mathbb{R}$ est dite intégrable au sens de Riemann si

$$\inf\{S(f, \sigma) : \sigma \in \mathfrak{S}\} \leq \sup\{s(f, \sigma) : \sigma \in \mathfrak{S}\}.$$

La valeur commune de cet infimum et ce supremum est alors notée $\int_a^b f(x) \, dx$.

Cette construction a l'avantage de la simplicité, quelques lignes suffisent à définir l'intégrale de Riemann. Cette simplicité est aussi son principal défaut : l'ensemble des fonctions intégrables au sens de Riemann est trop restreint. La raison en est que l'intégrabilité au sens de Riemann impose une certaine régularité sur la fonction f , celle-ci ne doit pas trop osciller au risque que les sommes de Darboux inférieure et supérieure ne puissent coïncider à la limite. C'est le cas par exemple pour $f = \mathbf{1}_{[0,1] \cap \mathbb{Q}}$. Nous verrons que cette dernière fonction est intégrable au sens de Lebesgue et d'intégrale nulle. D'une manière plus générale, pratiquement toute fonction positive peut être intégrée (dans un sens large, c'est à dire l'intégrale peut être infinie) au sens de Lebesgue ; de telles fonctions sont dites mesurables positives ; en fait, il y a bien un exemple de fonction de non mesurable mais sa construction utilise l'axiome du choix.

Cependant, s'il n'était seulement question que d'intégrer plus de fonctions, ce serait un peu court. Le réel défaut de la notion d'intégrale au sens de Riemann est qu'elle n'est pas préservée par passage à la limite : on peut exhiber une suite $(f_n)_{n \geq 0}$ croissante de fonctions Riemann intégrables qui converge

simplement vers une fonction non intégrable au sens de Riemann¹. Dans le contexte de l'intégrale de Lebesgue, la positivité des fonctions f_n suffira à donner un sens à l'intégrale de la limite. Une condition tout aussi simple pour des fonctions non partout positives existe bien entendu. Dans le contexte Riemann, la bonne hypothèse est en général la continuité et la convergence uniforme; celle-ci est beaucoup trop technique et surcharge souvent inutilement les preuves.

Ses bonnes propriétés de convergence sont sans doute à mettre à l'actif de ce que l'intégrale de Lebesgue est définie comme une borne supérieure; en ce sens, on approche l'intégrale de Lebesgue par valeurs inférieures, contrairement à l'intégrale de Riemann qui est définie via la convergence de deux suites adjacentes. Notons en outre que lors de la démonstration du théorème de convergence monotone, nous introduisons la fonction $v = u\mathbf{1}_{u \leq f}$. Lorsque u est une fonction étagée (la fonction u admet un nombre fini de valeurs distinctes) alors v sera également une fonction étagée. Au contraire, si u est une fonction en escalier, c'est à dire une combinaison linéaire d'indicatrice d'ouverts, et que f est raisonnablement pathologique il est à peu près sûr que v n'hériterait pas de cette propriété.

L'intégrale de Lebesgue est aussi plus maniable pour traiter le cas de fonctions présentant des singularités ou lorsque l'intervalle $[a, b]$ considéré n'est plus borné. Considérons l'exemple de la fonction $f(x) = \frac{1}{\sqrt{x}}$ que l'on souhaite intégrer sur $[0, 1]$. Dans le contexte de l'intégrale de Riemann, on définit

$$\int_0^1 \frac{dx}{\sqrt{x}} = \lim_{\varepsilon \rightarrow 0} \int_\varepsilon^1 \frac{dx}{\sqrt{x}},$$

dès que la limite à droite existe. Dans ce cas précis, le théorème fondamental de l'analyse montre que cette limite existe effectivement. Dans le contexte de l'intégrale de Lebesgue, la positivité et la régularité de f^2 sur $]0, 1]$ suffit à donner un sens à $\int_0^1 dx/\sqrt{x} \in [0, \infty) \cup \{\infty\}$ ³. Le principe consiste à approcher f par en-dessous par des fonctions étagées positives et passer à la limite. En approchant f par en-dessous, la singularité en 0 n'est plus véritablement un problème.

La théorie de Lebesgue date du début du siècle dernier, l'axiomatique de Kolmogorov formalisant la théorie des probabilités dates des années trente (évidemment, nous avons pas attendu ce formalisme pour faire du calcul de probabilités). Cette formalisation a le très grand avantage de rendre transparente la distinction artificielle entre les probabilités discrètes et diffuses (ou continues). Ceci a coûté, celui de l'appréhension de ce morceau conceptuel que l'on appelle théorie de la mesure. Outre cette unification, ce formalisme est rendu nécessaire lorsque les modèles probabilistes deviennent plus complexes : comment faire du calcul de probabilités proprement sur des espaces tels que $\{0, 1\}^{\mathbb{N}}$ ou encore sur l'espace de fonctions $C^0([0, 1], \mathbb{R})$? La théorie de la mesure met ainsi à disposition des concepts clairs pour définir et manipuler proprement des objets aléatoires comme le mouvement brownien — Figure 1a — qui est à la base de beaucoup de modèle de dynamiques réelles perturbées et/ou bruitées — Figures 1b, 1c et 1d.

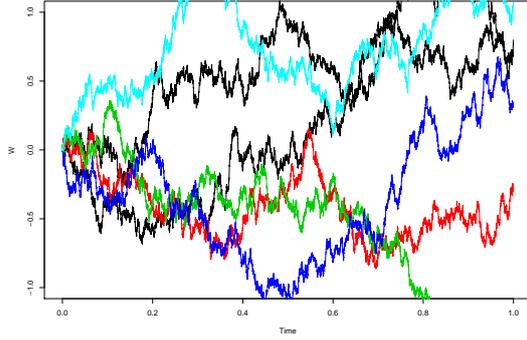
Si nous devons citer un inconvénient de l'intégrale de Lebesgue, ce serait l'absence en tant que tel d'un théorème fondamental de l'analyse. En réalité, c'est un faux problème puisque nous verrons que toute fonction Riemann intégrable est Lebesgue intégrable et les intégrales coïncident. Cela donne un moyen simple de calculer explicitement, dans certains cas, l'intégrale de Lebesgue d'une fonction réelle.

Dans les cas concrets, l'ensemble \mathbb{X} est naturellement muni d'une topologie permettant de définir les notions de parties ouvertes, fermées, compactes ou encore les notions de convergences. La plupart du temps, la topologie et la mesure sont définies de manière consistante si bien que des propriétés liées aux mesures et des propriétés de nature topologique se trouvent mêlées. Nous insisterons ici essentiellement sur le cas un peu plus restrictif mais souvent largement suffisant des espaces métriques. Néanmoins, afin de bien distinguer les notions intrinsèquement topologiques de celles propres aux espaces métriques, nous définirons la notion d'espace topologique et démontrerons autant que possible les résultats dans le contexte général des espaces topologiques. Ces notions sont introduites dans le chapitre 1 donnant quelques rappels et des compléments d'analyse. On rappelle en particulier la notion d'espace vectoriel normé. Les espaces vectoriels normés sont en particulier des espaces métriques. On termine ce chapitre par considérer la propriété de compacité ainsi que la notion d'espace complet.

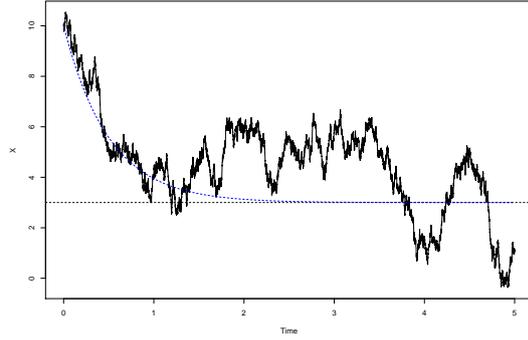
1. Soit en effet $(q_n)_{n \geq 0}$ une énumération de $[0, 1] \cap \mathbb{Q}$ et posons, pour tout $n \geq 0$, $f_n = \sum_{k=0}^n \mathbf{1}_{\{q_k\}}$. La suite $(f_n)_{n \geq 0}$ est bien monotone croissante et pour chaque $n \geq 0$ est intégrable au sens de Riemann d'intégrale nulle. Enfin, la suite $(f_n)_{n \geq 0}$ converge simplement vers $\mathbf{1}_{[0, 1] \cap \mathbb{Q}}$.

2. La fonction f est continue, mais le raisonnement est valide plus généralement pour des fonctions qui seront dites mesurables. La fonction $\mathbf{1}_{\mathbb{Q} \cap [0, 1]}$ est un exemple de telle fonction.

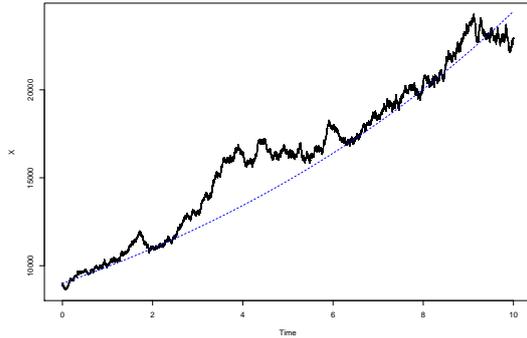
3. Dans la suite, l'intervalle semi-ouvert $[a, b[$ sera toujours noté à la mode anglo-saxonne $[a, b)$.



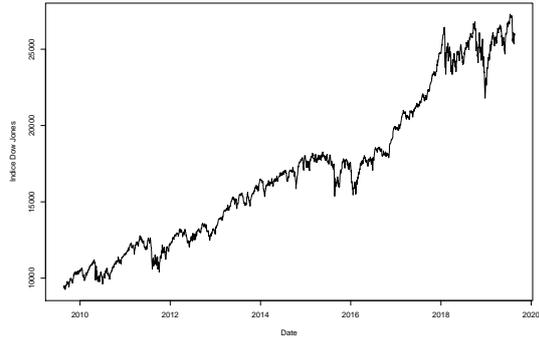
(a) Processus de Wiener.



(b) Processus d'Orstein-Uhlenbeck.



(c) Modèle de Black-Scholes.



(d) Indice du Dow Jones : août 2009 à août 2019.

FIGURE 1 – (a) Cinq réalisations du mouvement brownien $W = (W_t)_{t \geq 0}$ aussi appelé processus de Wiener. La variable aléatoire W est à valeurs dans l'espace des fonctions continue muni de la topologie de la convergence uniforme sur les compacts. Le processus de Wiener est la loi limite naturelle du théorème central limite fonctionnelle, il a un rôle similaire à la loi normale dans le cas réel. (b) Processus d'Orstein-Uhlenbeck : $(X_t)_{t \geq 0}$ satisfait l'Équation Différentielle Stochastique $dX_t = -\theta(X_t - \mu)dt + \sigma dW_t$ ($\theta = 2$, $\mu = 3$, $\sigma = 3$). Le paramètre θ définit l'intensité de la force de rappel vers μ . En bleu, la version sans bruit ($\sigma = 0$). (c) Modèle de Black-Scholes : l'EDS définissant cette dynamique est $dX_t = \tau X_t dt + \sigma X_t dW_t$ ($\tau = 0.1$, $\sigma = 0.07$). Le paramètre τ modélise le taux d'intérêt. En bleu, la version non bruitée, *i.e.* $\sigma = 0$. (d) Indice du Dow Jones sur la période août 2009 à août 2019.

Le chapitre 2 est consacré à la théorie de la mesure abstraite. Les axiomes ensemblistes de cette théorie sont pour l'essentielle la traduction de propriétés intuitives du calcul d'aire ou de volume. L'un des axiomes de la théorie de la mesure traduit cette assertion bien connue énonçant que le tout est la somme de ses parties. En substance, cela signifie que pour calculer l'aire de la réunion de deux parties disjointes il suffit de sommer les aires de chacune des deux parties. La théorie de la mesure tire toute sa puissance de ses théorèmes limites. Ces derniers sont en réalité conséquence directe la propriété de σ -additivité qui n'est autre que l'extension du dicton ci-dessus à des réunions *dénombrables* d'ensembles disjoints. On pourra se questionner sur cette restriction aux réunions dénombrables qui paraît de prime abord arbitraire. Il s'agit en fait d'un compromis : cette restriction est suffisante car elle permet de déduire les théorèmes limites usuels et elle est nécessaire pour des questions de consistance de la théorie. Pour ce dernier point, sans cet artifice, il arrive que le tout ne soit pas la somme de ses parties.

À ce stade, nous avons donc à disposition une collection de parties mesurables et une mesure qui permet de les mesurer. Ces deux ingrédients permettent de construire de façon abstraite une intégrale. Il est important de préciser que cette construction ne donne pas de moyen pratique de calculer l'intégrale d'une fonction arbitraire ; celle-ci est définie comme une limite qui peut être ardue à calculer explicitement. Par contre cette construction est complètement agnostique quant à la nature discrète ou diffuse de la mesure considérée ce qui permet une théorie unifiée de l'intégrale. On termine ce chapitre par préciser

le sens que l'on donne à cette intégrale lorsque la mesure sous-jacente est discrète ou diffuse. Dans le premier cas, il s'agit simplement d'une somme (une série) et dans le second, on peut considérer très grossièrement qu'il s'agit de la l'intégrale au sens usuel. En particulier, les résultats d'interversion de limites, d'intégration par parties *etc* sont valables à la fois dans le cas diffus et dans le cas discret. Il est du reste assez remarquable que la méthode d'intégration par partie soit très souvent bien connue dans le cas des fonctions numériques réelles mais que sa contre-partie pour les suites, appelée transformée d'Abel, soit si souvent ignorée. Le lien entre ces deux outils sera clairement établi en exercice.

Le chapitre 4 est certainement le plus important en pratique puisqu'il établit les théorèmes d'interversion limite/intégrale les plus importants : théorème de convergence monotone ou théorème de Beppo-Lévy 4.1.1, lemme de Fatou 4.1.2 et théorème de convergence dominée de Lebesgue 4.3.1. Le premier d'entre eux est une conséquence directe de l'hypothèse de σ -additivité dans la définition d'une mesure. Ce théorème constitue le socle de nombreux théorème d'intégration dont le lemme de Fatou et le théorème de convergence dominée.

Le chapitre 5 permet de munir d'une mesure un espace mesurable produit. Cette mesure est à la racine de la notion d'intégrale multiple. De cette construction, qui n'a rien de complètement triviale, on déduit le théorème de Fubini qui permet d'écrire l'intégrale contre la mesure produit comme une intégrale itérée contre des mesures unidimensionnelles. On termine ce chapitre par quelques calculs pratiques d'intégrales multiples et l'énoncé du théorème de changement de variables multidimensionnels. Cette notion de mesure produit est à la racine de la notion d'indépendance en probabilité et sera donc retrouvée au chapitre 8.

Le dernier chapitre de cette partie (chapitre 6) est une introduction à l'analyse fonctionnelle. Outre les inégalités usuelles, on s'intéressera au caractère complet des espaces \mathbf{L}^p . On énoncera également le théorème de Radon-Nikodym particulièrement important pour le formalisme des modèles paramétriques en statistique. Ce théorème permet également de prouver l'existence de l'espérance conditionnelle qui est la généralisation de la notion de probabilité conditionnelle aux variables aléatoires diffuses : elle donne un sens au conditionnement par rapport à un événement de probabilité nulle (voir le chapitre 13). On étudiera enfin les propriétés d'approximation dans les espaces \mathbf{L}^p . À cette occasion on introduira la notion de convolution et d'approximation de l'identité. Ce dernier chapitre utilisera intensivement les propriétés vues au chapitre 1.

Le premier objectif de cette deuxième partie du cours est de démontrer deux théorèmes essentiels de la théorie des Probabilités : la loi des grands nombres (Figure 2a) et le théorème central limite (Figure 2b). Ces deux théorèmes justifie *a posteriori* l'axiomatique introduite par Kolmogorov.

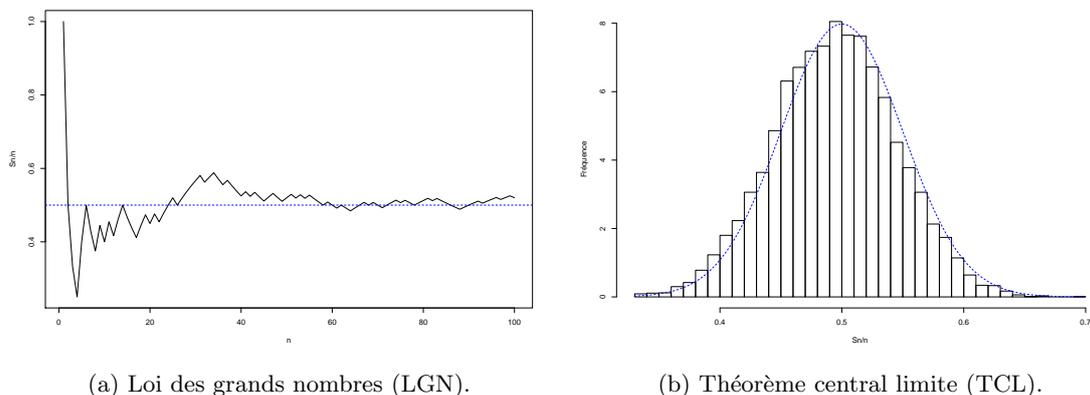


FIGURE 2 – (a) Loi des grands nombres : trajectoire de la moyenne S_n/n , $n = 1, \dots, 100$, où S_n est le nombre de face d'une pièce équilibrée. La moyenne théorique est matérialisée en bleu. (b) Théorème central limite : la distribution empirique de $(S_{100}^{(1)}/100, \dots, S_{100}^{(N)}/100)$, $N = 10^4$, se rapproche d'une distribution normale. En bleu est représentée la densité de la loi normale $\mathcal{N}(\mu, \sigma^2)$, $\mu = \frac{1}{2}$, $\sigma^2 = 4 \cdot 10^{-3}$, $x \rightarrow \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}}$.

Une notion centrale en théorie des probabilités est celle de variables aléatoires. Une variable aléatoire n'est rien d'autre qu'une application mesurable. Une problématique récurrente en probabilité consiste à caractériser la loi d'une variable aléatoire. Partant d'un espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$, la loi \mathbf{P}_X d'une

variable aléatoire $X : (\Omega, \mathcal{F}, \mathcal{P}) \rightarrow (E, \mathcal{E})$ est la mesure de probabilité image de \mathbf{P} par X . Le chapitre 7 est dédié à la caractérisation par différentes méthodes de la loi d'une variable aléatoire en particulier lorsque celle-ci est à valeur dans \mathbb{R} ou \mathbb{R}^d . Il sera fait usage d'un grand nombre de notions vues dans la première partie de ce cours.

Notons enfin que la notion de variable aléatoire est identique à celle d'observable en physique (classique). Lors d'une expérience aléatoire (ou de physique classique) l'expérimentateur n'a en général pas accès au triplet probabiliste mais plutôt à une observation. Le triplet $(\Omega, \mathcal{F}, \mathbf{P})$ est purement théorique et doit être considéré comme une "boîte noire" : ce que l'expérimentateur observe réellement est la valeur de la variable aléatoire X définie sur $(\Omega, \mathcal{F}, \mathbf{P})$. En ce sens, le triplet probabiliste n'est pas un objet canonique, plusieurs choix sont possibles : il est facile, par exemple, de construire deux triplets distincts et une variable aléatoire sur chacun des triplets décrivant la même expérience aléatoire du pile ou face.

Le chapitre 8 introduit la notion d'indépendance. Cette notion là-encore est purement théorique en ce sens qu'elle est difficile à exhiber dans la nature. Quoiqu'il en soit, c'est une hypothèse suffisante à la LGN et au TCL du chapitre 12. Néanmoins, d'autres hypothèses plus faibles peuvent être faites pour l'établissement de ces théorèmes comme par exemple dans [Kin73].

Le chapitre 9 introduit la notion de fonction caractéristique. Il s'agit ni plus ni moins de la notion de transformée de Fourier en analyse fonctionnelle appliquée à la théorie des probabilités. La démonstration du TCL au chapitre 12 utilise pleinement les fonctions caractéristiques. Elles permettent également de simplifier les calculs ainsi que l'établissement de résultats théoriques telles des convergences en loi. Au-delà de son usage en probabilités, l'analyse de Fourier ainsi que l'analyse en ondelettes, que l'on regroupe sous la terminologie analyse du signal, trouvent de nombreuses applications en ingénierie telles le débruitage d'un son, d'une image, la compression, l'analyse statistique de processus stochastiques.

Le chapitre 10 traite de vecteurs aléatoires, appelés vecteurs gaussiens, aux propriétés remarquables. Ceux-ci apparaissent naturellement en de nombreuses occasions et tout particulièrement dans le TCL multivarié énoncé au chapitre 12.

Le chapitre 11 introduit les notions de convergences trajectoires de variables aléatoires ainsi que celles de convergences des mesures et lois de probabilité. La première intervient dans la LGN alors que la seconde apparaît dans le TCL. De manière plus générique, la convergence trajectoire permet d'établir la convergence d'estimateurs statistiques et la convergence en loi permet d'établir les intervalles de confiance correspondants. En outre, le TCL est parfois interprété comme une vitesse de convergence dans la loi des grands nombres.

Enfin, le dernier chapitre introduit l'espérance conditionnelle. Ce chapitre est quelque peu à l'écart des autres. C'est également un outils primordiale en théorie des probabilités et théorie des processus. Cela permet notamment d'étudier des processus non *i.i.d.*.

Première partie

Topologie et théorie de la mesure

Chapitre 1

Rappels et compléments d'analyse

Sauf mention contraire, dans la suite, \mathbb{K} représente le corps des nombres réels \mathbb{R} ou le corps des nombres complexes \mathbb{C} . On notera indifféremment $|\cdot|$ la valeur absolue ou le module selon que $\mathbb{K} = \mathbb{R}$ ou $\mathbb{K} = \mathbb{C}$.

1.1 Espaces vectoriels normés

Définition 1.1.1 (Norme). Soit E un \mathbb{K} -espace vectoriel. Une norme sur E est une application $\|\cdot\| : E \rightarrow \mathbb{R}_+$ satisfaisant

1. $\|x\| = 0$ si et seulement si $x = 0$;
2. pour tout $\lambda \in \mathbb{K}$ et tout $x \in E$, $\|\lambda x\| = |\lambda| \|x\|$;
3. pour tout $x, y \in E$, $\|x + y\| \leq \|x\| + \|y\|$.

Un espace vectoriel normé est la donnée d'un couple $(E, \|\cdot\|)$ où E est un \mathbb{K} -espace vectoriel et $\|\cdot\|$ une norme sur E .

Exemple 1. 1. \mathbb{K} muni de $|\cdot|$ sont des espaces vectoriels normés.

2. Pour $p \geq 1$, on note $\|\cdot\|_p$ l'application définie pour tout $x = (x_1, \dots, x_n) \in \mathbb{K}^n$

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}.$$

Et si $p = \infty$,

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$

Alors l'espace \mathbb{K}^n muni de $\|\cdot\|_p$ est un espace vectoriel normé.

3. Plus généralement, si S est un ensemble dénombrable, sur \mathbb{K}^S , on définit les normes

$$\|x\|_p = \left(\sum_{s \in S} |x_s|^p \right)^{1/p} \quad \text{et} \quad \|x\|_\infty = \sup_{s \in S} |x_s|.$$

Alors l'ensemble $\ell_{\mathbb{K}}^p = \{x \in \mathbb{K}^S : \|x\|_p < \infty\}$, $p \in [1, \infty]$, muni de la norme $\|\cdot\|_p$ est un espace vectoriel normé.

4. Soit $(E, \|\cdot\|_E)$ et $(F, \|\cdot\|_F)$ deux espaces vectoriels normés et $A : E \rightarrow F$ un opérateur linéaire. On définit $\|\cdot\|_{E \rightarrow F}$ par

$$\|A\|_{E \rightarrow F} = \sup_{x \in E \setminus \{0\}} \frac{\|Ax\|_F}{\|x\|_E}.$$

On note $\mathcal{L}(E, F) = \{A : E \rightarrow F, A \text{ linéaire}, \|A\|_{E \rightarrow F} < \infty\}$ l'espace vectoriel des opérateurs linéaires continus de E dans F . On montre que $\|\cdot\|_{E \rightarrow F}$ est une norme sur $\mathcal{L}(E, F)$ appelée norme subordonnée.

Exercice 1. Montrer que les exemples donnés ci-dessus définissent bien des normes.

Exercice 2. Soit $(G, \|\cdot\|_G)$ est un troisième espace vectoriel normé. Montrer, pour tout $A \in \mathcal{L}(E, F)$ et $B \in \mathcal{L}(F, G)$, l'inégalité

$$\|BA\|_{E \rightarrow G} \leq \|B\|_{F \rightarrow G} \|A\|_{E \rightarrow F}.$$

1.2 Espaces métriques

1.2.1 Métrique, boule ouverte, boule fermée, parties bornées

Définition 1.2.1. Soit E un ensemble. Une distance (ou métrique) d sur E est une application $d : E \times E \rightarrow \mathbb{R}_+$ telle que :

1. $d(x, y) = 0$ si et seulement si $x = y$;
2. pour tout $x, y \in E$, $d(x, y) = d(y, x)$ (symétrie) ;
3. pour tout $x, y, z \in E$, $d(x, y) \leq d(x, z) + d(z, y)$ (inégalité triangulaire).

Un espace métrique est la donnée d'un couple (E, d) où E est un ensemble et d une distance sur E .

Proposition 1.2.2 (Deuxième inégalité triangulaire). Soit (E, d) un espace métrique, alors pour tout $x, y, z \in E$

$$|d(x, z) - d(y, z)| \leq d(x, y).$$

Démonstration. Nous avons par l'inégalité triangulaire et la propriété de symétrie

$$d(x, z) \leq d(x, y) + d(y, z) \quad \text{et} \quad d(y, z) \leq d(y, x) + d(x, z) = d(x, y) + d(x, z).$$

De ces deux inégalités, on déduit

$$-d(x, y) \leq d(x, z) - d(y, z) \leq d(x, y) \quad \implies \quad |d(x, z) - d(y, z)| \leq d(x, y).$$

□

Exemple 2. 1. Si $(E, \|\cdot\|)$ est un espace vectoriel normé alors (E, d) avec $d(x, y) = \|x - y\|$, $x, y \in E$, est un espace métrique.

2. \mathbb{R} muni de la métrique $d(x, y) = |\arctan x - \arctan y|$, $x, y \in E$, est espace métrique.
3. \mathbb{R}^2 muni de la métrique

$$\delta(x, y) = \begin{cases} \|x - y\|_2 & \text{si } 0, x \text{ et } y \text{ sont alignés,} \\ \|x\|_2 + \|y\|_2 & \text{sinon.} \end{cases}$$

4. Un ensemble E muni de la métrique discrète

$$d(x, y) = \begin{cases} 1 & \text{si } x \neq y, \\ 0 & \text{sinon.} \end{cases}$$

5. Soit $G = (V, E)$ un graphe fini simple non dirigé, i.e. V est un ensemble fini de nœuds et E est un ensemble de paires $\{x, y\} \subset V \times V$ appelées arêtes. Un chemin est une suite finie (x_1, x_2, \dots, x_n) de nœuds satisfaisant, pour tout $i = 1, \dots, n - 1$, $\{x_i, x_{i+1}\} \in E$. L'entier n est la longueur du chemin. On note $\Pi_{x,y}$ l'ensemble des chemins de longueur finie de x à y , i.e. $x_1 = x$ et $x_n = y$ avec les notations précédentes. Si $p \in \Pi_{x,y}$, on note $|p|$ la longueur de p . On définit pour tout $x, y \in V$

$$d(x, y) = \inf\{|p| : p \in \Pi_{x,y}\}.$$

Alors, (V, d) est un espace métrique. Un chemin $p \in \Pi_{x,y}$ tel que $|p| = d(x, y)$ est appelé géodésique. Le graphe G est dit complet si pour tout $x, y \in V$, $\{x, y\} \in E$. Dans ce cas, la métrique définie ci-dessus est la métrique discrète sur V .

Exercice 3. Vérifier que les exemples ci-dessus sont des espaces métriques.

Exercice 4. Montrer l'inégalité de Hölder et de Minkowski : soit $p \in [1, \infty)$, soit $a_i, b_i, i = 1, \dots, n$, des nombres réels ou complexes

1. Hölder : $|\sum_{i=1}^n a_i b_i| \leq (\sum_{i=1}^n |a_i|^p)^{1/p} (\sum_{i=1}^n |b_i|^q)^{1/q}$ avec $\frac{1}{p} + \frac{1}{q} = 1$;
2. Minkowski : $(\sum_{i=1}^n |a_i + b_i|^p)^{1/p} \leq (\sum_{i=1}^n |a_i|^p)^{1/p} + (\sum_{i=1}^n |b_i|^p)^{1/p}$.

Soit (E, d) un espace métrique. La boule ouverte de centre $a \in E$ et de rayon $r > 0$, notée $B(a, r)$ est définie par

$$B(a, r) = \{x \in E : d(a, x) < r\}.$$

La boule fermée sera notée $\overline{B}(a, r)$ et est définie par

$$\overline{B}(a, r) = \{x \in E : d(a, x) \leq r\}.$$

Si $A, B \subset E$ sont deux parties, la distance entre ces deux parties est donnée par

$$d(A, B) = \inf\{d(x, y) : x \in A, y \in B\}.$$

On utilise en général la convention $\inf \emptyset = \infty$ si bien que cette distance vaut l'infini si et seulement si l'une des deux parties est vide.

Proposition 1.2.3. Soit $A \subset E$ une partie non vide. Alors pour tout $x, y \in E$,

$$|d(x, A) - d(y, A)| \leq d(x, y).$$

Démonstration. C'est immédiat à partir de la proposition 1.2.2. □

Le diamètre d'une partie A de E est défini par $\text{Diam } A = \sup\{d(x, y) : x, y \in A\}$. Par convention, le diamètre d'une partie vide est égale à $-\infty$. Cette convention est cependant moins utile en pratique. Une partie $A \subset E$ est dite bornée si $\text{Diam } A < \infty$. En particulier, l'ensemble vide est borné.

1.2.2 Topologie des espaces métriques

Définition 1.2.4 (Topologie). Soit \mathbb{X} un ensemble. Une topologie sur \mathbb{X} est une famille de parties de \mathbb{X} , notée \mathcal{T} , satisfaisant

1. $\emptyset \in \mathcal{T}$ et $\mathbb{X} \in \mathcal{T}$;
2. pour toute famille $(O_i \in \mathcal{T})_{i \in I}$, la réunion $\cup_{i \in I} O_i \in \mathcal{T}$;
3. pour toute famille finie $O_1, \dots, O_n \in \mathcal{T}$, l'intersection $\cap_{i=1}^n O_i \in \mathcal{T}$.

Les éléments de \mathcal{T} sont appelés les ouverts.

Exemple 3. Soit \mathbb{X} un ensemble. Les familles $\mathcal{T}_1 = \{\emptyset, \mathbb{X}\}$ et $\mathcal{T}_2 = \mathcal{P}(\mathbb{X})$ (ensemble des parties de \mathbb{X}) sont des topologies sur \mathbb{X} appelée respectivement la topologie grossière et la topologie discrète.

Définition 1.2.5 (Ouvert). Soit (E, d) un espace métrique. Une partie $A \subset E$ est ouverte si pour tout $x \in A$, il existe $r > 0$ tel que $B(x, r) \subset A$.

Remarque 1. L'ensemble vide est ouvert !

Proposition 1.2.6. La boule ouverte est ouverte.

Démonstration. Soient $x \in E$, $r > 0$ et $y \in B(x, r)$. On pose $r_0 = d(x, y) < r$ et $\rho = r - r_0 > 0$, alors $B(y, \rho) \subset B(x, r)$. En effet, si $z \in B(y, \rho)$ alors $d(x, z) \leq d(x, y) + d(y, z) < r_0 + (r - r_0) = r$ et $z \in B(x, r)$. □

Proposition 1.2.7. Soit (E, d) un espace métrique. Une partie $A \subset E$ est ouverte si et seulement si elle est réunion de boule ouverte.

Démonstration. Une réunion arbitraire d'ouverts étant ouverte et la boule ouverte étant ouverte, la condition est évidemment suffisante. Réciproquement, pour tout $x \in A$, il existe $r_x > 0$ tel que $B(x, r_x) \subset A$. Alors, il est immédiat que $A = \cup_{x \in A} B(x, r_x)$. □

Proposition 1.2.8. Soit (E, d) un espace métrique. La famille $\mathcal{T} = \{O \subset E, O \text{ ouvert}\}$ définit une topologie sur E .

Démonstration. On vérifie facilement que $\emptyset, E \in \mathcal{T}$. Soit $x \in \cup_{i \in I} O_i$, alors il existe $i \in I$ tel que $x \in O_i$. Puisque que O_i est ouvert, il existe $r > 0$ tel que $B(x, r) \subset O_i \subset \cup_{i \in I} O_i$, d'où le résultat. Soit $x \in \cap_{i=1, \dots, n} O_i$, alors $x \in O_i$ pour tout $i = 1, \dots, n$. Pour chaque $i = 1, \dots, n$, il existe $r_i > 0$ tel que $B(x, r_i) \subset O_i$. Posons $r = \min r_i$, alors $B(x, r) \subset B(x, r_i)$ pour tout $i = 1, \dots, n$ et donc $B(x, r) \subset \cap_{i=1, \dots, n} O_i$. \square

Exercice 5. Soit E un ensemble. Montrer que la topologie associée à la métrique discrète est discrète.

Remarque 2. Il existe des topologies non métrisables qui peuvent être néanmoins intéressantes. Le contexte des espaces métriques est cependant très souvent suffisant mais les énoncés s'écrivent parfois plus facilement dans le langage de la topologie. Dans la suite, on jonglera avec les deux notions selon les cas.

Définition 1.2.9 (Voisinage ouvert, voisinage). Soient $(\mathbb{X}, \mathcal{T})$ un espace topologie et $A \subset \mathbb{X}$ non vide.

1. Un voisinage ouvert de A est un ouvert contenant A ;
2. Un voisinage de A est un ensemble contenant un voisinage ouvert de A .

Si $A = \{x\}$ on parle de voisinage ouvert (*resp.* de voisinage) de x . On notera $\mathcal{V}(x)$ l'ensemble des voisinages de x .

Proposition 1.2.10. Soit $(\mathbb{X}, \mathcal{T})$ un espace topologique. Un ensemble $V \subset \mathbb{X}$ est ouvert si et seulement si V est voisinage de chacun de ses points.

Démonstration. La condition est évidemment suffisante. Réciproquement, pour tout $x \in V$, il existe U_x ouvert tel que $x \in U_x \subset V$. Ainsi, $V = \cup_{x \in V} U_x$ ce qui montre que V est ouvert. \square

Proposition 1.2.11. Soient (E, d) un espace métrique et $x \in E$. Un ensemble V est un voisinage de x si et seulement si il existe $n \in \mathbb{N}^*$ tel que $B(x, 1/n) \subset V$.

Démonstration. Par définition, il existe un ouvert V' tel que $x \in V' \subset V$. Puisque V' est ouvert, il existe $r > 0$ tel que $B(x, r) \subset V'$, donc pour tout $n \in \mathbb{N}^*$ tel que $1/n < r$, on a $B(x, 1/n) \subset V' \subset V$. La réciproque est immédiate puisque $x \in B(x, 1/n) \subset V$. \square

Définition 1.2.12 (Base de voisinage). Soient $(\mathbb{X}, \mathcal{T})$ un espace topologique et $x \in \mathbb{X}$. Une collection \mathcal{B} est appelée base de voisinage de x si pour tout $V \in \mathcal{V}(x)$, il existe $U \in \mathcal{B}$ tel que $x \in U \subset V$.

La proposition 1.2.11 exprime que, dans le contexte des espaces métriques, $\{B(x, 1/n), n \geq 1\}$ est une base de voisinage de x . En particulier, un espace métrique est à base de voisinages dénombrable : chacun de ses points admet une base de voisinages dénombrable. C'est une des propriétés importantes que n'ont pas les topologies en générale. Une notion duale est celle de base d'ouverts.

Définition 1.2.13 (Base d'ouverts). Une famille d'ouverts \mathcal{O} est une base d'ouverts pour la topologie \mathcal{T} si tout $O \in \mathcal{T}$ est réunion d'éléments de \mathcal{O} .

Exemple 4. Soit (E, d) un espace métrique, alors tout ouvert $O \subset E$ s'écrit comme réunion de boules ouvertes :

$$O = \bigcup_{(x,r) \in \mathcal{S}} B(x, r) \quad \text{où} \quad \mathcal{S} = \{(x, r) \in E \times \mathbb{R}_*^+ : B(x, r) \subset O\}.$$

L'ensemble des boules ouvertes de E est donc une base d'ouverts de E .

Définition 1.2.14 (Intérieur). Soit $A \subset E$, l'intérieur de A est le plus grand ouvert contenu dans A noté $\text{Int } A$.

Définition 1.2.15. L'extérieur de $A \subset E$ est l'intérieur de A^c .

Une notion duale est celle de fermé.

Définition 1.2.16 (Fermé). Une partie $F \subset E$ est fermée si F^c est ouverte.

Proposition 1.2.17. Soit $(\mathbb{X}, \mathcal{T})$ un espace topologique. Alors,

1. \emptyset, E sont fermés ;
2. Si $(F_i)_{i \in I}$ est famille infinie de fermés, alors $\bigcap_{i \in I} F_i$ est fermée.
3. Si F_1, \dots, F_n sont des fermés, alors $\bigcup_{i=1}^n F_i$ est fermé.

Démonstration. Immédiat. □

Définition 1.2.18. L'adhérence d'une partie $A \subset E$ est le plus petit fermé contenant A , noté \overline{A} .

Remarque 3. Dans le cas d'un espace vectoriel normé sur \mathbb{K} , l'adhérence de la boule ouverte est la boule fermée. Dans le cas des espaces métriques, il se peut que l'adhérence de la boule ouverte ne soit pas la boule fermée. Pour voir cela, il suffit de considérer l'ensemble $E = \{0, 1\}$ muni de la métrique discrète. Dans ce cas, la boule ouverte $B(0, 1) = \{0\}$ et le plus petit fermé contenant la boule ouverte est $\{0\}$. En revanche la boule fermée $\overline{B}(0, 1) = \{0, 1\}$.

Proposition 1.2.19. Soient $(\mathbb{X}, \mathcal{T})$ un espace topologique et $A \subset \mathbb{X}$ une partie de \mathbb{X} . Alors

$$\overline{A} = \{x \in \mathbb{X}, \quad \forall V \in \mathcal{V}(x), \quad V \cap A \neq \emptyset\} = (\text{Ext } A)^{\text{c}}.$$

Démonstration. L'égalité suivante est immédiate :

$$(\text{Ext } A)^{\text{c}} = \left(\bigcup_{A^{\text{c}} \supset O \in \mathcal{T}} O \right)^{\text{c}} = \bigcap_{A \subset O^{\text{c}} \in \mathcal{T}} O^{\text{c}}.$$

En effet, il est facile de voir que la réunion sur des ouverts contenus dans A^{c} n'est autre que l'intérieur de A^{c} (c'est un ouvert contenu dans A^{c} et tous les ouverts contenu dans A^{c} sont inclus dans la réunion), soit l'extérieur de A . De même, l'intersection à droite est un fermé contenant A et tous les fermés contenant A contiennent cette intersection, c'est l'adhérence de A . Ceci montre $\overline{A} = (\text{Ext } A)^{\text{c}}$.

Montrons la seconde égalité qui peut se réécrire :

$$\{x \in \mathbb{X}, \quad \forall V \in \mathcal{V}(x), \quad V \cap A \neq \emptyset\}^{\text{c}} = \text{Ext } A.$$

Comme $\text{Ext } A$ est ouvert, il est voisinage de chacun de ses points $y \in \text{Ext } A$, or $\text{Ext } A \cap A = \emptyset$ puisque $\text{Ext } A \subset A^{\text{c}}$ et $A \cap A^{\text{c}} = \emptyset$. On a donc trouver un voisinage de $y \in \text{Ext } A$ qui n'intersecte pas A si bien que

$$\text{Ext } A \subset \{x \in \mathbb{X}, \quad \forall V \in \mathcal{V}(x), \quad V \cap A \neq \emptyset\}^{\text{c}}. \quad (1.1)$$

Inversement, si $x \in \mathbb{X}$ est tel qu'il existe $V \in \mathcal{V}(x)$ tel que $V \cap A = \emptyset$ alors il existe un ouvert U tel que $x \in U \subset V$ et $U \cap A = \emptyset$ d'où $U \subset A^{\text{c}}$ si bien que $U \subset \text{Ext } A$. Ceci montre l'inclusion inverse de (1.1). □

Proposition 1.2.20. Soient $(\mathbb{X}, \mathcal{T})$ un espace topologique et $A, B \subset \mathbb{X}$. Alors

1. $A \subset \overline{A}$,
2. $\overline{\overline{A}} = \overline{A}$,
3. $\overline{A \cup B} = \overline{A} \cup \overline{B}$,
4. $\overline{A \cap B} \subset \overline{A} \cap \overline{B}$.

Exercice 6. Montrer qu'en général l'inclusion du quatrième point est stricte.

Démonstration. Le premier point découle directement de la définition puisque \overline{A} est un fermé contenant A . Pour le deuxième point, il suffit de remarquer qu'un ensemble est fermé si et seulement si il est égale à son adhérence.

Montrons le point (3). On remarque que $A \cup B \subset \overline{A \cup B}$ par le point (1). Puisque $\overline{A \cup B}$ est fermé comme la réunion de deux fermés, c'est donc un fermé qui contient $A \cup B$, ainsi $\overline{A \cup B} \subset \overline{A \cup B}$. Réciproquement, nous avons $A \subset A \cup B \subset \overline{A \cup B}$, donc $\overline{A \cup B}$ est un fermé qui contient A . Il contient B également. Ainsi \overline{A} et \overline{B} sont tout deux contenus dans $\overline{A \cup B}$, leur réunion l'est donc aussi : $\overline{A} \cup \overline{B} \subset \overline{A \cup B}$.

Pour le point (4), nous avons $A \cap B \subset A \subset \overline{A}$ qui est fermé. De même, $A \cap B \subset B \subset \overline{B}$. Comme $\overline{A \cap B}$ est fermé, il vient que $\overline{A \cap B} \subset \overline{A} \cap \overline{B}$. □

Proposition 1.2.21. Soient $(\mathbb{X}, \mathcal{T})$ un espace topologique et $A, B \subset \mathbb{X}$. Alors

1. $\text{Int } A \subset A$,
2. $\text{Int } \text{Int } A = \text{Int } A$,
3. $\text{Int } (A \cap B) = \text{Int } A \cap \text{Int } B$,
4. $\text{Int } A \cup \text{Int } B \subset \text{Int } (A \cup B)$.

Exercice 7. Montrer que la dernière inclusion est stricte en général.

Démonstration. Le premier point découle directement de la définition. Pour le deuxième point, il suffit de remarquer qu'une partie est ouverte si et seulement si elle est égale à son intérieur.

Montrons (3). Puisque $\text{Int } A \subset A$ et $\text{Int } B \subset B$, il vient que l'ouvert $\text{Int } A \cap \text{Int } B$ est contenu dans $A \cap B$. Ainsi, $\text{Int } A \cap \text{Int } B \subset \text{Int } (A \cap B)$ par définition de l'intérieur. Réciproquement, $\text{Int } (A \cap B) \subset A \cap B \subset A$ mais il est aussi contenu dans B . Puisque $\text{Int } (A \cap B)$ est ouvert, il est contenu dans $\text{Int } A$ et dans $\text{Int } B$, il est donc contenu dans $\text{Int } A \cap \text{Int } B$.

Pour le point (4) on remarque que $\text{Int } A \subset A \subset A \cup B$ et $\text{Int } B \subset B \subset A \cup B$. Ainsi l'ouvert $\text{Int } A \cup \text{Int } B$ est contenu dans $A \cup B$ donc dans $\text{Int } (A \cup B)$. \square

Proposition 1.2.22. Soient $(\mathbb{X}, \mathcal{T})$ un espace topologique et $A \subset \mathbb{X}$. Alors

1. $\mathbb{X} \setminus \text{Int } A = \overline{\mathbb{X} \setminus A}$,
2. $\mathbb{X} \setminus \overline{A} = \text{Int } (\mathbb{X} \setminus A)$.

Démonstration. Pour le point (1), il suffit de remarquer que si un ouvert O est contenu dans A , alors O^c est un fermé qui contient A^c . Le complémentaire du plus grand ouvert contenu dans A correspond au plus petit fermé contenant A^c .

Le point (2) est immédiat par passage au complémentaire dans (i) appliqué à $B = \mathbb{X} \setminus A$. \square

Définition 1.2.23. La frontière d'une partie $A \subset E$, notée $\text{Fr } A$, est définie par $\text{Fr } A = \overline{A} \cap \overline{A^c}$.

Proposition 1.2.24. Soit $(\mathbb{X}, \mathcal{T})$ un espace topologique. Alors pour toute partie $A \subset \mathbb{X}$, le triplet $(\text{Int } A, \text{Fr } A, \text{Ext } A)$ forme une partition de \mathbb{X} .

Démonstration. Du fait de la définition de la frontière et de la proposition 1.2.22, on a

$$\text{Fr } A = \overline{A} \cap \overline{\mathbb{X} \setminus A} = \overline{A} \cap (\mathbb{X} \setminus \text{Int } A) = \overline{A} \setminus \text{Int } A.$$

Clairement $\mathbb{X} = \text{Int } A \cup \overline{A} \setminus \text{Int } A \cup \mathbb{X} \setminus \overline{A}$ est une réunion disjointe dont le deuxième ensemble est $\text{Fr } A$ et le troisième ensemble n'est autre que $\text{Ext } A$ puisque $\mathbb{X} \setminus \overline{A} = \text{Int } (\mathbb{X} \setminus A)$ par la proposition 1.2.22. \square

Définition 1.2.25 (Partie dense). Soit $(\mathbb{X}, \mathcal{T})$ un espace topologique. Une partie $D \subset \mathbb{X}$ est dite dense si $\overline{D} = \mathbb{X}$.

Proposition 1.2.26. Une partie D est dense dans un espace topologique $(\mathbb{X}, \mathcal{T})$ si et seulement si pour tout $x \in \mathbb{X}$ et tout voisinage V de x , $V \cap D \neq \emptyset$.

Démonstration. C'est une conséquence de la définition et de la proposition 1.2.19. \square

Remarque 4. Notons qu'on peut se restreindre aux voisinages ouverts.

Théorème 1.2.27 (Sous-groupe de $(\mathbb{R}, +)$). Soit G un sous-groupe de $(\mathbb{R}, +)$. Alors $G = a\mathbb{Z}$ pour un certain réel a , ou G est dense dans \mathbb{R} .

Démonstration. Soit G un sous-groupe de $(\mathbb{R}, +)$. Évacuons d'entrée le cas où G est le groupe trivial $\{0\}$. Alors il existe $g \in G \setminus \{0\}$ et ou bien $g > 0$ ou bien $g < 0$. Dans ce second cas, puisque G est un groupe, $-g \in G$ et $-g > 0$. Aussi, $G \cap \mathbb{R}_+^*$ est une partie non vide. Elle est également minorée et on note $a = \inf G \cap \mathbb{R}_+^*$ qui est positif car 0 est un minorant.

Supposons d'abord $a > 0$ et montrons que $a \in G$. Supposons au contraire que $a \notin G$. Nous avons $2a > a$ si bien que $2a$ n'est pas un minorant de $G \cap \mathbb{R}_+^*$. Il existe donc $b \in G$ tel que $a < b < 2a$. Mais b n'est pas plus un minorant de $G \cap \mathbb{R}_+^*$ et il existe donc de même $c \in G$ tel que $a < c < b < 2a$. Ainsi, $b - c \in G \cap \mathbb{R}_+^*$ et $b - c < a$. C'est une contradiction au fait a est la borne inférieure de $G \cap \mathbb{R}_+^*$. Par conséquent, $a \in G$ et donc $a\mathbb{Z} \subset G$. Il reste à montrer l'inclusion opposée.

Soit $g \in G$ et posons $n = \lfloor g/a \rfloor$ la partie entière de g/a . Par définition $n \leq g/a < n+1$ ou encore $na \leq g < a(n+1)$. Alors, $0 \leq g - na < a$. Puisque $a \in G$, $na \in G$ et donc $g - na \in G$. Supposons $g \neq na$, alors $0 < g - na < a$ et donc $g - na \in G \cap \mathbb{R}_+^*$ et $g - na < a$. C'est la même contradiction que précédemment, donc $g = na$. Finalement, $G \subset a\mathbb{Z}$.

Soit maintenant $a = 0$. Il s'agit de montrer que G est dense dans \mathbb{R} . Soit $x, y \in \mathbb{R}$ avec $x < y$. Puisque $a = 0$, il existe $g \in G$ tel que $0 < g < y - x$. Posons désormais $n = \lfloor x/g \rfloor + 1$. On obtient $(n-1)g \leq x < ng$. Alors :

$$x < ng = (n-1)g + g \leq x + g < x + (y-x) = y.$$

Autrement dit, pour tout $x, y \in \mathbb{R}$ tels que $x < y$, il existe $g \in G$ tel que $x < g < y$. Donc G est dense dans \mathbb{R} par la proposition 1.2.26. \square

1.2.3 Notion de limites

Convergence de suite

Définition 1.2.28 (Convergence dans les espaces topologiques). Soient $(\mathbb{X}, \mathcal{T})$ un espace topologique, $(x_n)_{n \geq 0}$ une suite de points de \mathbb{X} et $x \in \mathbb{X}$. On dit que la suite $(x_n)_{n \geq 0}$ converge vers x dans $(\mathbb{X}, \mathcal{T})$ si pour tout voisinage V de x , il existe $N \geq 0$ tel que si $n \geq N$ alors $x_n \in V$.

Exemple 5. Si on munit \mathbb{X} de la topologie grossière, alors toutes les suites sont convergentes. Pour la topologie discrète, les seules suites convergentes sont les suites constantes à partir d'un certain rang.

Dans le cas des espaces métriques, nous avons la définition suivante.

Proposition 1.2.29 (Convergence dans les espaces métriques). Soit $(x_n)_{n \geq 0}$ une suite à valeurs dans (E, d) et $x \in E$. La suite $(x_n)_{n \geq 0}$ converge vers x si et seulement si

$$\forall \varepsilon > 0, \quad \exists N \geq 0 : \quad n \geq N \implies d(x_n, x) < \varepsilon.$$

Démonstration. Exercice. \square

Définition 1.2.30 (Topologie séparée). Un espace topologique $(\mathbb{X}, \mathcal{T})$ est dite séparé si pour tout $x, y \in \mathbb{X}$, $x \neq y$, il existe V_x un voisinage de x et V_y un voisinage de y tel que $V_x \cap V_y = \emptyset$.

Lorsque la topologie sépare les points, le point vers lequel une suite converge est unique.

Proposition 1.2.31. Soit $(x_n)_{n \geq 0}$ qui converge vers x et y dans un espace topologique séparé $(\mathbb{X}, \mathcal{T})$. Alors $x = y$. Le point x est appelé limite de $(x_n)_{n \geq 0}$ et on note $\lim_{n \rightarrow \infty} x_n = x$.

Démonstration. Exercice. \square

Proposition 1.2.32. La topologie définie par une métrique sépare les points.

Démonstration. Immédiat. \square

Remarque 5. Les espaces métriques jouissent de deux propriétés remarquables : l'une d'elle est le fait que la topologie associée sépare les points ; l'autre est que chaque point admet une base de voisinage dénombrable.

Proposition 1.2.33. Soit \mathbb{X} un ensemble. Si \mathcal{T} et \mathcal{T}' sont deux topologies sur \mathbb{X} telle que $\mathcal{T} \subset \mathcal{T}'$, alors toute suite $(x_n)_{n \geq 0}$ qui converge vers x pour la topologie \mathcal{T}' converge vers x pour la topologie \mathcal{T} .

Démonstration. Immédiat. \square

Remarque 6. Cette proposition élémentaire est très régulièrement utilisée en probabilité mais aussi en analyse fonctionnelle : l'idée est qu'en supprimant des ouverts (ou des fermés) à une topologie, nous trouverons plus de suite convergente, la contrepartie étant que la convergence est plus faible : elle donne moins d'informations.

Proposition 1.2.34 (Caractérisation séquentielle des points adhérents). Soient (E, d) un espace métrique, $A \subset E$ et $x \in E$. Alors, $x \in \overline{A}$ si et seulement il existe une suite $(x_n)_{n \geq 0}$ de points de A qui converge vers x dans (E, d) .

Démonstration. Soit $x \in \overline{A}$, alors par la proposition 1.2.19, pour tout $n \geq 1$, la boule ouverte $B(x, 1/n) \cap A \neq \emptyset$, il suffit donc de choisir x_n dans cette intersection. De fait, $d(x, x_n) \rightarrow 0$, d'où $(x_n)_{n \geq 0}$ converge vers x .

Réciproquement, soit $(x_n)_{n \geq 0}$ une suite de points de A qui converge vers $x \in X$. Alors pour tout voisinage V de x , il existe $N \geq 0$ tel que pour tout $n \geq N$, $x_n \in V$. En particulier $V \cap A \neq \emptyset$ et $x \in \overline{A}$. \square

Remarque 7. Notons que la condition reste suffisante dans des espaces topologiques généraux.

Points d'accumulation, points isolés, valeurs d'adhérence

Nous avons déjà évoqué l'adhérence d'une partie A d'un espace topologie (X, \mathcal{T}) . Un point est dit adhérent si il est dans l'adhérence de A . Ci-dessous, nous définissons les notions de point d'accumulation, de point isolés et de valeur d'adhérence.

Définition 1.2.35 (Points d'accumulation, points isolés). Soit (X, \mathcal{T}) un espace topologique et soit $A \subset X$.

1. un point $x \in X$ est un point d'accumulation si pour tout voisinage V de x , $V \cap A \setminus \{x\} \neq \emptyset$;
2. un point $x \in A$ est isolé dans A si il existe un voisinage V de x tel que $V \cap A = \{x\}$.

Lorsque A est l'image d'une suite, c'est à dire $A = \{x_n : n \geq 0\}$, il existe une notion plus forte que celle de points adhérents : ce sont les valeurs d'adhérences.

Définition 1.2.36 (Valeur d'adhérence). Soient (X, \mathcal{T}) un espace topologique et $(x_n)_{n \geq 0}$ une suite de points de X . Un point $x \in X$ est une valeur d'adhérence (ou point limite) si pour tout voisinage V de x , $x_n \in V$ pour une infinité de $n \in \mathbb{N}$.

Exemple 6. — La suite $((-1)^n)_{n \geq 0}$ admet deux valeurs d'adhérences qui sont -1 et 1 .

- Soient $p \in \mathbb{Z}, q \in \mathbb{N}^*$, alors l'ensemble des valeurs d'adhérences de la suite $(\cos(2\pi np/q))_{n \geq 0}$ est $\{0, \cos(2\pi p/q), \dots, \cos(2\pi p(n-1)/q)\}$.
- Ces deux exemples font intervenir des suites périodiques. Mais, on peut considérer des exemples plus élaborés : si $\alpha \notin \mathbb{Q}$ alors l'ensemble des valeurs d'adhérence de la suite $(\cos(2\pi \alpha n))_{n \geq 0}$ est $[-1, 1]$.

On remarque d'abord en utilisant la parité de \cos que $\{\cos(2\pi \alpha n), n \geq 0\} = \cos(2\pi \alpha \mathbb{Z} + 2\pi \mathbb{Z})$. Comme la fonction $x \rightarrow \cos(x)$ est une surjection continue de \mathbb{R} dans $[-1, 1]$. Il suffit de montrer que $2\pi \alpha \mathbb{Z} + 2\pi \mathbb{Z}$ est dense dans \mathbb{R} . Or, $2\pi \alpha \mathbb{Z} + 2\pi \mathbb{Z}$ est un sous-groupe de \mathbb{R} . Par le théorème 1.2.27, il est soit dense soit de la forme $a\mathbb{Z}$ pour un certain $a \in \mathbb{R}$. Supposons qu'il soit de cette seconde forme, alors il existe $p, q \in \mathbb{Z}$ tel que $2\pi \alpha = pa$ et $2\pi = qa$ d'où $\alpha = p/q$ puisque $aq \neq 0$. C'est une contradiction avec $\alpha \notin \mathbb{Q}$ donc $2\pi \alpha \mathbb{Z} + 2\pi \mathbb{Z}$ est dense dans \mathbb{R} .

Le même genre de résultat est vrai pour la fonction \sin à l'aide d'un déphasage d'angle $\pi/2$.

Proposition 1.2.37. Soit $(x_n)_{n \geq 0}$ une suite de (E, d) . L'ensemble des valeurs d'adhérence de $(x_n)_{n \geq 0}$ est le fermé

$$F = \bigcap_{N \geq 0} \overline{\{x_n, n \geq N\}}.$$

Pour tout $a \in F$, il existe une suite n_k qui tend vers l'infini lorsque k tend vers l'infini telle que $\lim_{k \rightarrow \infty} x_{n_k} = a$. En particulier, une valeur d'adhérence est un point adhérent.

Démonstration. Notons $A_N = \{x_n : n \geq N\}$. Par la proposition 1.2.19

$$\overline{A_N} = \{x \in X : \forall \varepsilon > 0, B(x, \varepsilon) \cap A_N \neq \emptyset\} = \bigcap_{\varepsilon > 0} \{x \in X : B(x, \varepsilon) \cap A_N \neq \emptyset\}.$$

D'où

$$F = \bigcap_{N \geq 0} \bigcap_{\varepsilon > 0} \{x \in X : B(x, \varepsilon) \cap A_N \neq \emptyset\} = \bigcap_{\varepsilon > 0} \bigcap_{N \geq 0} \{x \in X : B(x, \varepsilon) \cap A_N \neq \emptyset\}.$$

C'est à dire, $x \in F$ si et seulement si pour tout $\varepsilon > 0$, tout $N \geq 0$, il existe $n \geq N$ tel que $x_n \in B(x, \varepsilon)$ si et seulement si x est une valeur d'adhérence.

Pour la deuxième partie de la proposition, si $a \in F$ alors pour $k \geq 1$, on peut trouver au moins x_{n_k} tel que $d(a, x_{n_k}) < 1/k$. D'où le résultat. \square

Limites et fonctions

Définition 1.2.38. Soient $(\mathbb{X}_1, \mathcal{T}_1)$, $(\mathbb{X}_2, \mathcal{T}_2)$ deux espaces topologiques. On dit que $f : \mathbb{X}_1 \rightarrow \mathbb{X}_2$ tend vers $b \in \mathbb{X}_2$ quand x tend vers $a \in \mathbb{X}_1$ si pour tout voisinage W de b dans \mathbb{X}_2 il existe un voisinage V de a dans \mathbb{X}_1 tel que $f(V) \subset W$.

Il arrive bien souvent qu'une fonction ne soit pas définie sur tout le domaine \mathbb{X}_1 mais plutôt sur un sous-ensemble strict $A \subset \mathbb{X}_1$. On peut alors chercher à définir une notion de limite vers un point en dehors de A . Typiquement, la fonction $x \rightarrow \sin(x)/x$ est définie sur $A = \mathbb{R} \setminus \{0\}$. Il est alors naturelle de chercher à définir la notion de limite lorsque x tend vers 0 ce qui n'est pas possible avec la définition donnée ci-dessus. De même, on s'intéresse souvent aux limites en $+\infty$ ou $-\infty$ sans pour autant que la fonction soit définie en ces points.

La définition suivante est celle considérée classiquement : il ne s'agit de rien d'autre que la définition ci-dessus appliquée à l'espace topologique induit (A, \mathcal{T}_A) — voir la proposition 1.2.46 — à ceci près qu'il faut ajouter l'hypothèse a est adhérent à A afin d'assurer que $V \cap A$ soit non vide — la notion serait alors triviale.

Définition 1.2.39. Soient $(\mathbb{X}_1, \mathcal{T}_1)$, $(\mathbb{X}_2, \mathcal{T}_2)$ deux espaces topologiques et $A \subset \mathbb{X}_1$. On dit qu'une fonction $f : A \rightarrow \mathbb{X}_2$ tend vers $b \in \mathbb{X}_2$ quand x tend vers $a \in \overline{A}$ dans \mathbb{X}_1 , $x \in A$, si pour tout voisinage W de b dans \mathbb{X}_2 , il existe un voisinage V de a dans \mathbb{X}_1 tel que $f(V \cap A) \subset W$.

Exemple 7. Il faut être vigilant et remarquer qu'il s'agit de la limite quand x tends vers a dans A et que cette limite dépend *a priori* de A comme l'illustre l'exemple suivant.

Soit $f : [0, 2] \rightarrow \mathbb{R}$ définie pour $x \in [0, 2)$ par $f(x) = x^2$ et $f(2) = 5$. Alors, $\lim_{x \rightarrow 2, x \in [0, 1] \cup \{2\}} f(x) = 5$ bien que $\lim_{x \rightarrow 2, x \in [0, 2]} f(x) = 4$.

En effet, soit W un voisinage de 5, alors on peut poser $V = B(2, 1/2)$ alors $f(V \cap A) = f(\{2\}) = \{5\} \subset W$. On montre de même la seconde limite. Les topologies étant séparées, la limite est unique et on observe l'importance du choix de A .

Exercice 8. Une suite $(x_n)_{n \geq 0}$ à valeurs dans un espace topologique $(\mathbb{X}, \mathcal{T})$ peut être vue comme une fonction $x : \mathbb{N} \rightarrow \mathbb{X}$. Montrer que la notion de convergence pour les fonctions coïncide avec celle des suites, en posant $a = \infty$ et $A = \mathbb{N}$ et en munissant $\mathbb{N} \cup \{\infty\}$ de la topologie dont les ouverts sont les singletons et les complémentaires de parties finies.

Si $(\mathbb{X}_2, \mathcal{T}_2)$ est séparé, la limite d'une fonction est unique et on note : $\lim_{x \rightarrow a, x \in A} f(x) = b$.

Proposition 1.2.40. Soient (E, d) et (E', d') deux espaces métriques, $A \subset E$ une partie de E , $f : A \rightarrow E'$ une application et $a \in \overline{A}$. On dit que f tend vers $b \in E'$ quand x tend vers a dans X , $x \in A$ si

$$\forall \varepsilon > 0, \quad \exists \delta > 0, \quad \forall x \in A : \quad d(x, a) < \delta \implies d(f(x), b) < \varepsilon.$$

Démonstration. Exercice. □

La proposition suivante est très utile en pratique. Par simplicité, on se restreint au cadre des espaces métriques même si certaines propriétés reste vraie dans un cadre plus général.

Proposition 1.2.41 (Caractérisation séquentielle). Soient (E, d) et (E', d') deux espaces métriques, $A \subset E$ une partie de E , $f : A \rightarrow E'$ une fonction, $a \in \overline{A}$ et $b \in E'$. Alors $b = \lim_{x \rightarrow a, x \in A} f(x)$ si et seulement si pour toute suite $(x_n)_{n \geq 0} \in A^{\mathbb{N}}$ telle que $\lim_{n \rightarrow \infty} x_n = a$ alors $\lim_{n \rightarrow \infty} f(x_n) = b$.

Démonstration. On se donne une suite $(x_n)_{n \geq 0}$ de points de A qui converge vers $a \in \overline{A}$ et on veut montrer que $(f(x_n))_{n \geq 0}$ converge vers $b \in E'$. Soit W un voisinage de b , puis que $\lim_{x \rightarrow a, x \in A} f(x) = b$, il existe V un voisinage de a tel que $f(A \cap V) \subset W$. Puisque V est un voisinage de a , il existe $N \geq 0$ tel que pour tout $n \geq N$, $x_n \in V$ et donc pour tout $n \geq N$, $f(x_n) \in W$ car $x_n \in V \cap A$.

Réciproquement, on suppose que f ne tend pas vers b quand $x \in A$ tend vers a . C'est à dire qu'il existe W un voisinage de b tel que pour tout voisinage de a , $f(V \cap A) \not\subset W$. En particulier, posons $V_n = B(a, 1/n)$, $n \geq 1$, alors il existe $x_n \in B(a, 1/n) \cap A$ tel que $f(x_n) \notin W$. Mais alors, $(x_n)_{n \geq 0}$ est une suite de points de A qui converge vers a tel que pour tout $n \geq 1$, $f(x_n) \notin W$, donc $(f(x_n))_{n \geq 0}$ ne peut converger vers b . □

1.2.4 Continuité

Définition 1.2.42 (Continuité ponctuelle, continuité). Soient $(\mathbb{X}, \mathcal{T})$, $(\mathbb{X}', \mathcal{T}')$ deux espaces topologiques et $f : \mathbb{X} \rightarrow \mathbb{X}'$ une fonction.

1. La fonction f est dite continue en $x \in \mathbb{X}$ si pour tout voisinage W de $f(x)$ il existe un voisinage V de x tel que $f(V) \subset W$;
2. La fonction f est dite continue si elle est continue en tout point $x \in \mathbb{X}$.

Remarque 8. La continuité d'une fonction f en $a \in \mathbb{X}$ est équivalente à l'égalité $\lim_{x \rightarrow a, x \in A} f(x) = f(a)$ avec $A = \mathbb{X}$.

Exercice 9. Donner une caractérisation séquentielle de la continuité.

Proposition 1.2.43. Soit (E, d) et (E', d') deux espaces métriques et f une application de (E, d) dans (E', d') . L'application f est dite continue en $x \in E$ si

$$\forall \varepsilon > 0, \quad \exists \delta > 0 : \quad \forall y \in E, \quad d(x, y) < \delta \implies d'(f(x), f(y)) < \varepsilon.$$

L'application f sera dite continue si f est continue pour tout $x \in E$.

Démonstration. C'est une condition suffisante. Soit W un voisinage de $f(x)$, alors il existe un ouvert O tel que $f(x) \in O \subset W$, et on peut trouver $\varepsilon > 0$ tel que $f(x) \in B(f(x), \varepsilon) \subset O \subset W$. Cela assure l'existence d'un $\delta > 0$ tel $f(B(x, \delta)) \subset B(f(x), \varepsilon) \subset W$. Or, $B(x, \delta)$ est un voisinage de x .

C'est une condition nécessaire. Soit $\varepsilon > 0$, comme $B(f(x), \varepsilon)$ est un voisinage de $f(x)$, il existe un voisinage V de x tel que $f(V) \subset B(f(x), \varepsilon)$. Or, par définition, il existe $\delta > 0$ tel que $x \in B(x, \delta) \subset V$ et on conclut $f(B(x, \delta)) \subset B(f(x), \varepsilon)$. \square

Proposition 1.2.44. Soient $(\mathbb{X}, \mathcal{T})$ et $(\mathbb{X}', \mathcal{T}')$ deux espaces topologiques. Les assertions suivantes sont équivalentes :

1. $f : \mathbb{X} \rightarrow \mathbb{X}'$ est continue ;
2. pour tout ouvert O de \mathbb{X}' , $f^{-1}(O)$ est un ouvert de \mathbb{X} ;
3. pour tout fermé F de \mathbb{X}' , $f^{-1}(F)$ est un fermé de \mathbb{X} ;
4. pour toute partie $A \subset \mathbb{X}$, $f(\overline{A}) \subset \overline{f(A)}$.

Démonstration. Montrons que (i) implique (ii). Clairement, $f^{-1}(\emptyset) = \emptyset$. Si O est un ouvert non vide de \mathbb{X}' , alors ou bien $f^{-1}(O) = \emptyset$ et $f^{-1}(O)$ est un ouvert de \mathbb{X} , ou bien $f^{-1}(O)$ est non vide et on peut choisir $x_0 \in f^{-1}(O)$, autrement dit $f(x_0) \in O$. Comme O est ouvert, c'est en particulier un voisinage de $f(x_0)$. La continuité de f donne l'existence d'un voisinage V de x_0 tel que $x_0 \in V \subset f^{-1}(O)$. Il existe donc un ouvert W_{x_0} de \mathbb{X} tel que $x_0 \in W_{x_0} \subset V \subset f^{-1}(O)$. On pose

$$U = \bigcup_{x \in f^{-1}(O)} W_x,$$

où W_x est un ouvert tel que $x \in W_x \subset f^{-1}(O)$. Ainsi, U est un ouvert contenu dans $f^{-1}(O)$ qui recouvre $f^{-1}(O)$: c'est exactement $f^{-1}(O)$.

Réciproquement, soit $x_0 \in \mathbb{X}$ et W un voisinage de $f(x_0)$, il existe donc O un ouvert de \mathbb{X}' tel que $f(x_0) \in O \subset W$. Par hypothèse, $f^{-1}(O)$ est un ouvert. De plus, $x_0 \in f^{-1}(\{f(x_0)\}) \subset f^{-1}(O)$, donc $f^{-1}(O)$ est un voisinage de x_0 .

Il est clair que (ii) est équivalent à (iii). Il reste à montrer que (i) est équivalent à (iv). On suppose f continue. Soit $A \subset \mathbb{X}$, $\overline{f(A)}$ est un fermé, donc $f^{-1}(\overline{f(A)})$ est un fermé par continuité. Puisque $f(A) \subset \overline{f(A)}$, $f^{-1}(\overline{f(A)})$ contient A , donc il contient \overline{A} . Ainsi, pour tout $x \in \overline{A}$, il existe $y \in \overline{f(A)}$ tel que $f(x) = y$. Autrement dit, $f(\overline{A}) \subset \overline{f(A)}$.

Réciproquement, soit F un fermé de \mathbb{X}' . On note $A = f^{-1}(F)$. Alors $f(A) = f(f^{-1}(F)) \subset F$. Par hypothèse, $f(\overline{A}) \subset \overline{f(A)} \subset \overline{F} = F$ puisque F est fermé. Donc, $\overline{A} = f^{-1}(\overline{f(A)}) \subset f^{-1}(F) = A$ par définition. Donc $\overline{A} = A$ et A est fermé. \square

Proposition 1.2.45. Soient $(\mathbb{X}_1, \mathcal{T}_1)$, $(\mathbb{X}_2, \mathcal{T}_2)$ et $(\mathbb{X}_3, \mathcal{T}_3)$ trois espaces topologiques. Soient $f : \mathbb{X}_1 \rightarrow \mathbb{X}_2$ continue en $x_0 \in \mathbb{X}_1$ et $g : \mathbb{X}_2 \rightarrow \mathbb{X}_3$ continue en $f(x_0)$. Alors $g \circ f$ est continue en x_0 .

Démonstration. Immédiat. \square

1.2.5 Topologies et opérations ensemblistes

Topologie induite

Proposition 1.2.46 (Topologie induite). Soient $(\mathbb{X}, \mathcal{T})$ un espace topologique et Y une partie de \mathbb{X} . Alors, l'ensemble de parties donné par

$$\mathcal{T}_Y = \{O \cap Y, O \in \mathcal{Y}\}$$

définit une topologie sur Y appelée topologie induite.

Démonstration. Immédiat. □

Proposition 1.2.47 (Sous-espace métrique). Soit (E, d) un espace métrique et $F \subset E$ une partie de E . L'application d restreinte à $F \times F$ définit encore une métrique appelée métrique induite sur F . L'espace (F, d) est appelé sous-espace métrique.

Démonstration. Immédiat. □

Proposition 1.2.48. La topologie définie par la métrique induite coïncide avec la topologie induite.

Démonstration. Soit O un ouvert de (F, d) , alors pour tout $x \in O$, il existe $\rho_x > 0$ tel que $B(x, \rho_x) \cap F \subset O$. Ainsi, $O = \cup_{x \in O} [B(x, \rho_x) \cap F] = (\cup_{x \in O} B(x, \rho_x)) \cap F$. De plus, pour tout $x \in O$, $B(x, \rho_x)$ est ouvert dans (E, d) d'où $O \in \mathcal{T}_F$.

Réciproquement, si $O \in \mathcal{T}_F$, alors il existe \tilde{O} un ouvert de (E, d) tel que $O = \tilde{O} \cap F$. Soit donc $x \in O = \tilde{O} \cap F$ alors il existe $\rho > 0$ tel que $B(x, \rho) \subset \tilde{O}$ et donc $B(x, \rho) \cap F \subset O$. Or $B(x, \rho) \cap F$ n'est rien d'autre que la boule ouverte centrée en $x \in F$ et de rayon $\rho > 0$ dans l'espace métrique (F, d) . Ainsi, O est un ouvert de (F, d) . □

De même que pour la notion de limite, une fonction définie sur une partie A d'un espace topologique est dite continue sur A si la restriction $f|_A$ de f à A est continue sur l'espace topologique induit (A, \mathcal{T}_A) . Le théorème suivant est alors une conséquence immédiate des définitions.

Proposition 1.2.49 (Prolongement par continuité). Soient (E, d) et (E', d') deux espaces métriques, $f : A \subset E \rightarrow E'$ une fonction et $a \in \bar{A}$. Alors, les assertions suivantes sont équivalentes

1. il existe une unique application continue $\tilde{f} : A \cup \{a\} \rightarrow E'$ qui coïncident avec f sur A ;
2. f est continue sur A et $\lim_{x \rightarrow a, x \in A} f(x)$ existe.

Démonstration. Exercice. □

Topologie initiale, topologie produit

Définition 1.2.50 (Comparaison de topologies). Soient \mathbb{X} un ensemble, \mathcal{T}_1 et \mathcal{T}_2 deux topologies sur \mathbb{X} . On dit que \mathcal{T}_2 est plus fine que \mathcal{T}_1 si $\mathcal{T}_1 \subset \mathcal{T}_2$. On dit aussi que \mathcal{T}_1 est moins fine que \mathcal{T}_2 .

Proposition 1.2.51 (Topologie engendrée). Soit \mathbb{X} un ensemble et \mathcal{O} un ensemble de parties. Alors il existe une topologie \mathcal{T} , contenant \mathcal{O} , qui est moins fine que toute autre topologie vérifiant cette propriété. La topologie \mathcal{T} est dite engendrée par \mathcal{O} .

Démonstration. On remarque que l'intersection d'une famille arbitraire de topologie est encore une topologie. On conclut en considérant la topologie définie comme l'intersection des topologies contenant \mathcal{O} : elle est non vide (la topologie discrète contient \mathcal{O}) et elle est moins fine que toute autre topologie contenant \mathcal{O} . □

Définition 1.2.52 (Topologie initiale). Soit \mathbb{X} un ensemble et $(f_i)_{i \in I}$ une famille d'applications chacune définie sur \mathbb{X} et à valeurs dans un espace topologique $(\mathbb{Y}_i, \mathcal{T}_i)$. La topologie la moins fine rendant les applications $f_i : \mathbb{X} \rightarrow \mathbb{Y}_i$ continues est appelée topologie initiale.

La topologie initiale est donc la topologie sur \mathbb{X} engendrée par $\mathcal{O} = \{f_i^{-1}(O), i \in I, O \in \mathcal{T}_i\}$. Cette définition s'applique directement à la notion d'espace produit.

Définition 1.2.53. Soit $((\mathbb{X}_i, \mathcal{T}_i))_{i \in I}$ une collection d'espaces topologique. On note $\mathbb{X} = \prod_{i \in I} \mathbb{X}_i$ et $p_i : \mathbb{X} \rightarrow \mathbb{X}_i$ la projection sur la coordonnée $i \in I$ qui à $x = (x_i)_{i \in I}$ associe $p_i(x) = x_i \in \mathbb{X}_i$. La topologie produit est la topologie initiale associée à la famille de projections $(p_i)_{i \in I}$.

Remarquons que si O est un ouvert de \mathbb{X}_{i_0} alors $p_{i_0}^{-1}(O) = \prod_{i \in I} O_i$ où $O_i = \mathbb{X}_i$ pour tout $i \in I \setminus \{i_0\}$. D'autre part, une intersection finie d'ouvert étant ouverte, cela mène à la définition de cylindre ouvert. Un cylindre ouvert est une partie $O = \prod_{i \in I} O_i$, $O_i \in \mathcal{T}_i$ pour tout $i \in I$, vérifiant $O_i = \mathbb{X}_i$ sauf pour un nombre fini de $i \in I$. On note \mathcal{T} la collection des parties de \mathbb{X} qui sont réunions de cylindre ouverts.

Proposition 1.2.54 (Topologie produit). *L'ensemble \mathcal{T} est la topologie produit sur \mathbb{X} .*

Démonstration. Il est immédiat que les projections p_i , $i \in I$, sont continues de $(\mathbb{X}, \mathcal{T})$ dans $(\mathbb{X}_i, \mathcal{T}_i)$ car, pour tout $i \in I$, $p_i^{-1}(O)$ est un cylindre ouvert pour tout $O \in \mathcal{T}_i$. Soit \mathcal{T}' une topologie sur \mathbb{X} telle que, pour tout $i \in I$, p_i est continue et considérons $O \in \mathcal{T}$. Il est immédiat que si C est un cylindre ouvert alors il existe un ensemble $I_0 \subset I$ fini et des ouverts $O_i \in \mathcal{T}_i$, $i \in I_0$, tels que $C = \bigcap_{i \in I_0} p_i^{-1}(O_i)$. Comme O est par définition une réunion de tels cylindres, il vient que O est réunion d'intersections finies d'images réciproques d'ouverts par les applications p_i donc $O \in \mathcal{T}'$ par continuité des $p_i : (\mathbb{X}, \mathcal{T}') \rightarrow (\mathbb{X}_i, \mathcal{T}_i)$. \square

Exemple 8. Soit $((E_n, d_n))_{n \in \mathbb{N}}$ une famille d'espaces métriques au plus dénombrable. Soit $(\alpha_n)_{n \geq 0}$ une suite de réels positifs telle que $\sum_{n \geq 0} \alpha_n < \infty$. On pose sur $E = \prod_{n \in \mathbb{N}} E_n$ la distance

$$d(x, y) = \sum_{n \geq 0} \alpha_n \frac{d_n(x_n, y_n)}{1 + d_n(x_n, y_n)}, \quad x = (x_n)_{n \geq 0}, y = (y_n)_{n \geq 0} \in E = \prod_{n \geq 0} E_n.$$

On définit également

$$\delta(x, y) = \sup_{n \geq 0} d_n(x_n, y_n) \wedge \frac{1}{1 + n}, \quad x = (x_n)_{n \geq 0}, y = (y_n)_{n \geq 0} \in E = \prod_{n \geq 0} E_n. \quad (1.2)$$

Exercice 10. Montrer que d et δ définissent des métriques sur E . Montrer que ces métriques définissent la même topologie sur E . Quelle est-elle ?

Proposition 1.2.55. *Soit $((\mathbb{X}_i, \mathcal{T}_i))_{i \in I}$ une collection d'espaces topologiques. On note $\mathbb{X} = \prod_{i \in I} \mathbb{X}_i$ et $p_i : \mathbb{X} \rightarrow \mathbb{X}_i$ la projection sur la coordonnée $i \in I$ qui à $x = (x_i)_{i \in I}$ associe $p_i(x) = x_i \in \mathbb{X}_i$. Une suite $(x_n)_{n \geq 0} \in \mathbb{X}^{\mathbb{N}}$ converge vers x dans \mathbb{X} muni de la topologie produit si et seulement si pour tout $i \in I$, la suite $(p_i(x_n))_{n \geq 0}$ converge vers x_i dans \mathbb{X}_i .*

Exemple 9. Soit $F = \mathbb{R}^{[0,1]}$ l'ensemble des fonctions de $[0, 1]$ dans \mathbb{R} . On munit \mathbb{R} de sa topologie usuelle associée à $|\cdot|$ et F de la topologie produit.

Par la proposition 1.2.55, une suite de fonctions $(f_n)_{n \geq 0}$ de F converge vers f si et seulement si, pour tout $x \in [0, 1]$, $f_n(x)$ converge vers $f(x)$. La topologie produit dans ce contexte n'est rien d'autre que la topologie de la convergence simple.

Cette topologie est séparée : si $x, y \in F$ sont tels que $x \neq y$ alors il existe $i \in [0, 1]$ tel que $x_i \neq y_i$. Puisque $(\mathbb{R}, |\cdot|)$ est un espace métrique, c'est un espace séparé : il existe un voisinage V_{x_i} de x_i et un voisinage V_{y_i} de y_i vérifiant $V_{x_i} \cap V_{y_i} = \emptyset$. Par continuité de la projection p_i , $p_i^{-1}(V_{x_i})$ et $p_i^{-1}(V_{y_i})$ sont des voisinages de x et y tels que $p_i^{-1}(V_{x_i}) \cap p_i^{-1}(V_{y_i}) = p_i^{-1}(V_{x_i} \cap V_{y_i}) = \emptyset$.

Notons que l'espace F muni de la topologie produit n'est par contre pas métrisable. Pour montrer ce fait, on cherche à contredire la caractérisation séquentielle des points adhérents.

On appelle fonction simple un élément $x \in F$ tel que $x_i = 0$ pour tout $i \in [0, 1]$ sauf peut-être un nombre fini. L'ensemble des fonctions simples est dense dans F : si $O = \prod_{i \in [0,1]} O_i$ est un cylindre ouvert non vide, alors seul un nombre fini d'ouverts O_{i_0}, \dots, O_{i_J} , tous non vides, ne sont pas \mathbb{R} tout entier. Soient $a_0 \in O_{i_0}, \dots, a_J \in O_{i_J}$ et posons $x_{i_j} = a_j$ pour tout $j = 0, \dots, J$ alors que $x_i = 0$ partout ailleurs. On constate que x est une fonction simple. Ainsi, l'ensemble des fonctions simples intersecte tous les cylindres ouverts et donc tous les ouverts : l'ensemble des fonctions simples est dense dans F pour la topologie produit. D'autre part, soient $y \in F$ limite d'une suite de fonctions simples $(y^{(n)})_{n \geq 0}$. Alors

$$A = \{i \in [0, 1] : y_i \neq 0\} \subset \bigcup_{n \geq 0, m \geq 0} \{i \in [0, 1] : |y_i^{(m)}| \geq 2^{-n}\}.$$

Ainsi, l'ensemble A est inclus dans une réunion dénombrable d'ensembles finis et est donc dénombrable. Ceci montre qu'une fonction non nulle sur un ensemble indénombrable ne peut être limite d'une suite de fonctions simples. La topologie de la convergence simple ne peut donc être métrisable.

Démonstration. Exercice. □

Proposition 1.2.56. Soient $((\mathbb{X}_i, \mathcal{T}_i)_{i \in I}$ une collection d'espaces topologique et $(\mathbb{Y}, \mathcal{U})$ un espace topologique. On note $\mathbb{X} = \prod_{i \in I} \mathbb{X}_i$ et $p_i : \mathbb{X} \rightarrow \mathbb{X}_i$ la projection sur la coordonnée $i \in I$ qui à $x = (x_i)_{i \in I}$ associe $p(x) = x_i \in \mathbb{X}_i$. Alors, \mathbb{X} muni de la topologie produit, une fonction $f : \mathbb{Y} \rightarrow \mathbb{X}$ est continue si et seulement si $p_i \circ f : \mathbb{Y} \rightarrow \mathbb{X}_i$ est continue pour tout $i \in I$.

Démonstration. Exercice. □

Topologie finale, topologie quotient

Proposition 1.2.57 (Topologie finale). Soient \mathbb{X} un ensemble, $((\mathbb{X}_i, \mathcal{T}_i)_{i \in I}$ une famille d'espace topologiques et pour chaque $i \in I$ une application $f_i : \mathbb{X}_i \rightarrow \mathbb{X}$. La topologie finale sur \mathbb{X} associée à la famille $(f_i)_{i \in I}$ est la topologie

$$\mathcal{T} = \{O \subset \mathbb{X} : \forall i \in I, f_i^{-1}(O) \in \mathcal{T}_i\}.$$

C'est la topologie la plus fine rendant les applications f_i continues.

Démonstration. C'est un exercice de montrer que \mathcal{T} est la topologie la plus fine rendant les applications f_i continues. □

Soit $(\mathbb{X}, \mathcal{T})$ un espace topologique et \mathcal{R} une relation d'équivalence sur \mathbb{X} . On note \mathbb{X}/\mathcal{R} l'ensemble quotient et $\pi : \mathbb{X} \rightarrow \mathbb{X}/\mathcal{R}$ la projection canonique.

Définition 1.2.58 (Topologie quotient). La topologie quotient sur \mathbb{X}/\mathcal{R} est la topologie finale associée à l'unique application π .

Proposition 1.2.59. Soient $(\mathbb{X}, \mathcal{T})$ et $(\mathbb{Y}, \mathcal{U})$ deux espaces topologiques, \mathcal{R} une relation d'équivalence sur \mathbb{X} et $f : \mathbb{X}/\mathcal{R} \rightarrow \mathbb{Y}$. Alors, si on munit \mathbb{X}/\mathcal{R} de la topologie quotient, l'application $f : \mathbb{X}/\mathcal{R} \rightarrow \mathbb{Y}$ est continue si et seulement si $f \circ \pi : \mathbb{X} \rightarrow \mathbb{Y}$ est continue.

Démonstration. Exercice. □

1.2.6 Compacité

Propriété de Borel-Lebesgue et théorème de Bolzano-Weierstrass

Définition 1.2.60 (Recouvrement ouvert). Soit $(\mathbb{X}, \mathcal{T})$ un espace topologique. Un recouvrement ouvert de \mathbb{X} est une famille $(O_i)_{i \in I}$ d'ouverts tels que $\mathbb{X} \subset \cup_{i \in I} O_i$. Un sous-recouvrement ouvert est un recouvrement ouvert donné par une famille $(O_j)_{j \in J}$ avec $J \subset I$. Un recouvrement ouvert est dit fini si I est lui-même fini.

Définition 1.2.61 (Compacité). Un espace topologique séparé $(\mathbb{X}, \mathcal{T})$ est dit compact s'il vérifie l'une des deux propriétés suivantes :

1. de tout recouvrement ouvert de \mathbb{X} , on peut extraire un sous-recouvrement fini de \mathbb{X} ;
2. de toute famille de fermés dont l'intersection est vide, on peut extraire une sous famille finie d'intersection vide.

Une partie $X \subset \mathbb{X}$ est compacte si, muni de la topologie induite, X est un espace topologique compact.

Remarque 9. Les deux propriétés ci-dessus, appelées propriétés de Borel-Lebesgue, sont évidemment équivalente par passage au complémentaire.

Exemple 10. Quelques exemples et contre-exemples :

- L'ensemble \emptyset est compact (pour n'importe quelle métrique) ;
- $(\mathbb{R}, |\cdot|)$ n'est pas compact ;
- un ensemble discret (c'est à dire muni de la métrique discrète) est compact si et seulement si il est fini.

Remarque 10. Remarquons que la compacité est une notion purement topologique. Par conséquent, si deux espaces topologiques sont homéomorphes — c.f. la définition 1.2.79 — alors ils sont simultanément compacts ou non compacts.

Proposition 1.2.62 (Propriétés des fermés emboîtés). Soit $(\mathbb{X}, \mathcal{T})$ compact et $(F_n)_{n \in \mathbb{N}}$ une suite décroissante de fermés (i.e. $F_{n+1} \subset F_n$ pour tout $n \geq 0$) d'intersection vide. Alors il existe $N \geq 0$ tel que $F_N = \emptyset$.

Remarque 11. La contraposée est particulièrement intéressante : si $(F_n)_{n \geq 0}$ est une famille décroissante de fermés non vides dans un espace topologique $(\mathbb{X}, \mathcal{T})$ compact, alors $\bigcap_{n \geq 0} F_n$ est (fermé) non vide. C'est la propriétés bien connues des segments emboîtés dans \mathbb{R} .

Démonstration. Soit $(F_n)_{n \geq 0}$ une suite décroissante de fermés non vides d'un espace topologique $(\mathbb{X}, \mathcal{T})$ compact. Par la seconde caractérisation de Borel-Lebesgue, il existe des indices n_0, \dots, n_k tels que $\bigcap_{\ell=0}^k F_{n_\ell} = \emptyset$. Soit alors n le plus grand de ces indices, alors par décroissance $\bigcap_{\ell=0}^k F_{n_\ell} = F_n$, d'où le résultat. \square

Le théorème de Bolzano-Weierstrass énoncé ci-dessous ne s'applique que dans le contexte des espaces métriques. La deuxième assertion, appelée propriété de Bolzano-Weierstrass, est parfois posée comme définition de la compacité.

Théorème 1.2.63 (Théorème de Bolzano-Weierstrass). Soit (E, d) un espace métrique et $A \subset E$. Alors les assertions suivantes sont équivalentes :

1. A est compact ;
2. (BW) de toute suite $(x_n)_{n \geq 0}$ de points de A on peut extraire une sous-suite $(x_{n_k})_{k \geq 0}$ convergente vers un point de A
3. tout ensemble infini $B \subset A$ admet au moins un point d'accumulation dans A .

Lemme 1.2.64 (Lemme de Lebesgue). Soient (E, d) un espace métrique et $(O_i)_{i \in I}$ un recouvrement ouvert de E . On suppose que (E, d) vérifie la propriété de Bolzano-Weierstrass : toute suite à valeur dans E admet une valeur d'adhérence dans E . Alors il existe $\rho > 0$ tel que pour tout $x \in E$ il existe $i = i(x) \in I$ tel que $B(x, \rho) \subset O_i$.

Preuve du lemme de Lebesgue. On suppose qu'un tel $\rho > 0$ n'existe pas. En particulier, pour tout $n \geq 1$, il existe un point $x_n \in E$ tel que pour tout $i \in I$, $B(x_n, 1/n) \cap O_i^c \neq \emptyset$. Par la propriété de Bolzano-Weierstrass, on peut trouver une sous-suite $(x_{n_k})_{k \geq 1}$ qui converge vers un point x de E .

Puisque $(O_i)_{i \in I}$ est un recouvrement de E , il existe $i \in I$ tel que $x \in O_i$. Or, O_i est ouvert donc on peut trouver $r > 0$ tel que $B(x, r) \subset O_i$. De plus, par convergence de $(x_{n_k})_{k \geq 1}$ il existe $K \geq 1$ tel que pour tout $k \geq K$, $x_{n_k} \in B(x, r)$. Choisissons $k \geq K$ tel que $1/n_k \leq r/2$ alors

$$B(x_{n_k}, 1/n_k) \subset B(x, r) \subset O_i.$$

C'est une contradiction. \square

Preuve du théorème de Bolzano-Weierstrass. On ne montre que l'équivalence entre les deux premiers points. L'équivalence entre les deux derniers points est immédiate.

Montrons que (1) implique (2). Soient A compact et $(x_n)_{n \geq 0}$ une suite de points de A . On pose, pour tout $n \geq 0$, $A_n = \{x_p : p > n\}$. Ainsi, $A_n \subset A$ et $A_{n+1} \subset A_n$. On note, pour tout $n \geq 0$, F_n la fermeture de A_n dans A si bien que $F_n \subset A$. Clairement, $F_{n+1} \subset F_n$. Par la propriété des fermés emboîtés dans les compacts, on obtient $\bigcap_{n \geq 0} F_n \neq \emptyset$. Soit $x \in \bigcap_{n \geq 0} F_n \subset A$, alors $x \in F_0 = \overline{A_0}$. Ainsi, $B(x, 1) \cap A_0 \neq \emptyset$ et il existe x_{n_0} tel que $d(x, x_{n_0}) < 1$. De même, $x \in F_{n_0} = \overline{A_{n_0}}$ si bien que $B(x, 1/2) \cap A_{n_0} \neq \emptyset$. On peut donc trouver $n_1 > n_0$ tel que $d(x, x_{n_1}) < 1/2$, et ainsi de suite. Nous construisons ainsi une suite $(x_{n_k})_{k \geq 0}$ extraite de $(x_n)_{n \geq 0}$ telle que $d(x, x_{n_k}) < \frac{1}{k+1}$. Ainsi, $x_{n_k} \rightarrow x \in A$.

Montrons que (2) implique (1). Soit $A \subset E$, alors puisque (E, d) est un espace métrique, E est séparé de même que ses parties. Montrons que A satisfait la première caractérisation de Borel-Lebesgue. Soit $(O_i)_{i \in I}$ un recouvrement d'ouvert (pour la topologie définie par (E, d)) de A . Soit $\rho > 0$ le rayon de Lebesgue, donné par le lemme 1.2.64 de Lebesgue, associé à ce recouvrement.

Par hypothèse, A vérifie la propriété de Bolzano-Weierstrass. Soit $x_1 \in A$, alors : ou bien $A \subset B(x_1, \rho)$ et c'est terminé puisque par le lemme de Lebesgue $B(x_1, \rho) \subset O_{i(x_1)}$, la caractérisation de Borel-Lebesgue est satisfaite ; sinon, il existe $x_2 \in A$ tel que $d(x_1, x_2) \geq \rho$. À nouveau : ou bien $A \subset B(x_1, \rho) \cup B(x_2, \rho)$, i.e. Borel-Lebesgue est satisfaite ; ou bien il existe $x_3 \in A$ tel que $d(x_i, x_j) \geq \rho$ dès $i \neq j$. Au besoin, on peut réitérer ce procédé indéfiniment, mais alors cela signifie que l'on a construit une suite $(x_n)_{n \geq 0}$ de point de A tel que $d(x_i, x_j) \geq \rho$ dès que $i \neq j$. C'est une contradiction avec la propriété de Bolzano-Weierstrass. \square

Quelques propriétés et conséquences de la compacité

Proposition 1.2.65. Soit $(\mathbb{X}, \mathcal{T})$ un espace topologique compact. Toute partie $F \subset \mathbb{X}$ fermée est compacte.

Démonstration. La partie $F \subset \mathbb{X}$ est séparée puisque \mathbb{X} l'est. Soit $(F_i)_{i \in I}$ une famille de fermés contenus dans F et d'intersection vide. Si F est fermé, $(F_i)_{i \in I} = (F \cap F_i)_{i \in I}$ est également une famille de fermés de \mathbb{X} d'intersection vide. On peut donc en extraire une sous-famille finie F_1, \dots, F_n tels que $\bigcap_{i=1}^n F_i$, d'où le résultat. \square

Proposition 1.2.66. Soient $(\mathbb{X}, \mathcal{T})$ un espace topologique séparé et $K \subset \mathbb{X}$ un compact. Alors K est fermé.

Démonstration. Il s'agit de montrer que $\mathbb{X} \setminus K$ est ouvert. Soit $x \in \mathbb{X} \setminus K$. Pour tout $y \in K$, par hypothèse de séparation, il existe un voisinage ouvert $O_{x,y}$ de x et O_y de y tel que $O_{x,y} \cap O_y = \emptyset$. Aussi $K \subset \bigcup_{y \in K} O_y$, puis par compacité, on peut trouver y_1, \dots, y_N tel que $K \subset \bigcup_{i=1}^N O_{y_i}$. On pose $O_x = \bigcap_{i=1}^N O_{x,y_i}$ qui est un voisinage ouvert de x vérifiant

$$\emptyset = O_x \cap \left(\bigcup_{i=1}^N O_{y_i} \right) \supset O_x \cap K.$$

D'où le résultat. \square

Proposition 1.2.67 (Théorème de Borel-Lebesgue). Tout segment $[a, b]$ est compact dans \mathbb{R} muni de sa topologie usuelle.

Démonstration. Soit $(O_i)_{i \in I}$ un recouvrement ouvert de $[a, b]$. On note

$$A = \{x \in [a, b] : [a, x] \text{ soit recouvert par un nombre fini de } O_i, i \in I\}.$$

Alors, A satisfait les trois assertions suivantes :

1. $A \subset [a, b]$;
2. $A \neq \emptyset$;
3. A est majoré par b .

Les premiers et troisièmes points sont des conséquences directes de la définition. Le second point, quant à lui, s'obtient en remarquant que $a \in A$ puisqu'il existe $i_a \in I$ tel que $a \in O_{i_a}$. On note $m = \sup A$. On va montrer que $m \in A$, puis que $m = b$.

Comme $m \in [a, b] \subset \bigcup_{i \in I} O_i$, il existe $i_m \in I$ tel que $m \in O_{i_m}$. Ainsi, on peut trouver $\varepsilon > 0$ tel que $(m - \varepsilon, m + \varepsilon) \subset O_{i_m}$ et par définition de la borne supérieure, il existe $x \in A \cap (m - \varepsilon, m + \varepsilon)$. Ensuite, en observant que $[a, m] = [a, x] \cup [x, m]$ et que $x \in A$, on obtient que le premier intervalle peut être recouvert par un nombre fini de O_i et le deuxième intervalle par l'ouvert O_{i_m} . Finalement, $m \in A$.

Supposons que $m < b$ et choisissons $\varepsilon > 0$ tel que $m + \varepsilon < b$ et $m \in (m - \varepsilon, m + \varepsilon) \subset O_{i_m}$. Soit $x \in (m, m + \varepsilon)$, puis on décompose comme précédemment l'intervalle $[a, x] = [a, m] \cup [m, x]$. On constate que $[m, x]$ est recouvert par O_{i_m} et que $[a, m]$ peut être recouvert par un nombre fini O_i puisque $m \in A$. Manifestement, $m < \sup A$, d'où $m = b$. \square

Corollaire 1.2.68. Soit E un \mathbb{K} -espace vectoriel normé de dimension finie. Les parties compactes de E sont les parties fermés bornés.

Démonstration. On considère seulement \mathbb{R} muni de la norme $|\cdot|$. L'extension aux K -espaces vectoriels normés de dimension finie est élémentaire.

Soit K une partie compacte de \mathbb{R} qui est séparé. Alors K est fermée par la proposition 1.2.66. De plus, on observe que $K \subset \bigcup_{m \geq 0} B(0, m)$. Par compacité, de ce recouvrement par des ouverts, on peut extraire un sous-recouvrement fini. Ainsi, il existe $m_0 \geq 0$ tel que $K \subset B(0, m_0)$ et K est une partie bornée.

Réciproquement, si K est fermé et borné, alors $K \subset [a, b]$ pour certains réels $a < b$. Les propositions 1.2.67 et 1.2.65 permettent de conclure. \square

Proposition 1.2.69. Soit $(\mathbb{X}, \mathcal{T})$ un espace topologique séparé. Alors

1. toute réunion finie de compacts est compacte ;
2. toute intersection de compacts est compacte.

Démonstration. Puisque $(\mathbb{X}, \mathcal{T})$ est séparé, il en va de même pour toutes ses parties.

Soient K_1, \dots, K_N des compacts et $(O_i)_{i \in I}$ un recouvrement ouvert de $\bigcup_{j=1}^N K_j$. C'est aussi un recouvrement d'ouvert de K_j pour tout $j = 1, \dots, N$. Pour chaque $j = 1, \dots, N$, on peut trouver un sous recouvrement fini $O_1^j, \dots, O_{N_j}^j$ de K_j . Alors,

$$\bigcup_{j=1}^N K_j \subset \bigcup_{j=1}^N \bigcup_{\ell=1}^{N_j} O_\ell^j,$$

si bien que $(O_\ell^j)_{j=1, \dots, N, \ell=1, \dots, N_j}$ est un recouvrement ouvert finie de l'union.

Soient $(K_i)_{i \in I}$ une famille de compacts et $(F_j)_{j \in J}$ des fermés contenus dans $\bigcap_{i \in I} K_i$ tel que $\bigcap_{j \in J} F_j = \emptyset$. Soit $i_0 \in I$, alors $(F_j)_{j \in J}$ est une famille de fermés contenus dans K_{i_0} , par compacité, on peut en extraire une sous-famille finie F_1, \dots, F_N de fermés tels que $\bigcap_{j=1}^N F_j = \emptyset$. Ainsi $\bigcap_{i \in I} K_i$ est compact. \square

Proposition 1.2.70. Soient $(\mathbb{X}_1, \mathcal{T}_1)$ un espace topologique, $(\mathbb{X}_2, \mathcal{T}_2)$ un espace topologique séparé et $f : \mathbb{X}_1 \rightarrow \mathbb{X}_2$ une application continue. Alors $f(K)$ est compacte pour tout compact $K \subset \mathbb{X}_1$.

Démonstration. Tout d'abord, $f(K)$ est séparée puisque \mathcal{T}_2 sépare les points. Soit $(O_i)_{i \in I}$ un recouvrement ouvert de $f(K)$, alors par continuité, $(f^{-1}(O_i))_{i \in I}$ est recouvrement ouvert de K . On peut extraire un recouvrement ouvert fini $f^{-1}(O_1), \dots, f^{-1}(O_N)$ de K . On montre que les ouverts O_1, \dots, O_N recouvre $f(K)$: soit $x \in f(K)$ tel que $x \in \bigcap_{i=1}^N O_i^c$, alors $f^{-1}(\{x\}) \subset \bigcap_{i=1}^N f^{-1}(O_i)^c$. On a ainsi trouvé $y \in K$ tel que $y \in (\bigcup_{i=1}^N f^{-1}(O_i))^c$. Contradiction. \square

Corollaire 1.2.71. Soit $f : (\mathbb{X}, \mathcal{T}) \rightarrow \mathbb{R}$ une application continue où $(\mathbb{X}, \mathcal{T})$ est un espace topologique compact non vide. Alors f est bornée et atteint ses bornes.

Démonstration. La fonction f est à valeurs dans \mathbb{R} un espace métrique donc séparé, il s'ensuit que $f(\mathbb{X})$ est séparé. Par la proposition précédente, $f(\mathbb{X})$ est donc compacte, or les compacts de \mathbb{R} sont les fermés bornés. Ainsi, f est bornée. De plus $\sup_{x \in \mathbb{X}} f(x)$ et $\inf_{x \in \mathbb{X}} f(x)$ sont dans $\overline{f(\mathbb{X})} = f(\mathbb{X})$, ainsi f atteint ses bornes. \square

Exercice 11. Soit $f : (E, d) \rightarrow \mathbb{R}$ une fonction continue strictement positive, c'est à dire $\{x \in E : f(x) \leq 0\} = \emptyset$. Montrer que si (E, d) est compact, il existe $\delta > 0$ tel que $f(x) \geq \delta$ pour tout $x \in E$.

Définition 1.2.72 (Espace topologique séparable). Un espace topologique $(\mathbb{X}, \mathcal{T})$ est dit séparable si il existe une partie dense au plus dénombrable.

Remarque 12. Il ne faut pas confondre "séparable" et "séparé".

Corollaire 1.2.73. Tout espace métrique compact est séparable.

Démonstration. Pour tout $n \geq 1$, la famille $\mathcal{O} = \{B(x, 1/n) : x \in \mathbb{X}\}$ est certainement un recouvrement ouvert, par compacité, on peut extraire un sous-recouvrement fini : il existe p_n points x_1, \dots, x_{p_n} tels que la famille finie d'ouverts $\{B(x_i, 1/n), i = 1, \dots, p_n\}$ soit un recouvrement. On note pour chaque $n \geq 1$, C_n l'ensemble des points x_i choisis. Alors, $D = \bigcup_{n \geq 1} C_n$ est dénombrable. Montrons qu'il est dense : soit $x \in \mathbb{X}$ alors pour tout $n \geq 1$, il existe $y_n \in C_n \subset D$ tel que $x \in B(y_n, 1/n)$. Autrement dit, $d(x, y_n) < 1/n$ et $(y_n)_{n \geq 1}$ converge vers x . \square

Théorème 1.2.74 (Théorème de Tychonoff). Soit $((\mathbb{X}_i, \mathcal{T}_i))_{i \in I}$ une famille d'espaces topologiques compacts, alors, muni de la topologie produit, $\prod_{i \in I} \mathbb{X}_i$ est compact.

Exemple 11. Soit K un compact de \mathbb{R} . L'ensemble $F = K^{[0,1]}$ des fonctions à valeurs dans un compact K , muni de la topologie de la convergence simple, est compact.

Démonstration. On va se contenter de montrer le théorème dans le cas d'une famille dénombrable $((E_n, d_n))_{n \geq 0}$ d'espaces métriques. Si la famille est non dénombrable, la preuve de ce théorème fait appel à l'axiome du choix. Le cas non métrique est également un peu pénible car on ne peut pas utiliser la caractérisation de Bolzano-Weierstrass.

On considère donc l'espace produit $E = \prod_{n \geq 0} E_n$ muni de la topologie produit (associée par exemple à la métrique δ définie par (1.2)) et on cherche à montrer la propriété de Bolzano-Weierstrass. Soit $(x_k)_{k \geq 0}$ une suite de points de E , pour chaque $k \geq 0$, on note $x_k^{(n)}$ la n -ième coordonnée de x_k .

Puisque E_0 est compact, on peut trouver $\phi_0 : \mathbb{N} \rightarrow \mathbb{N}$ strictement croissante telle que $x_{\phi_0(k)}^{(0)}$ converge vers $x^{(0)}$. De même, puisque E_1 est compact, la propriété de Bolzano-Weierstrass appliquée à $(x_{\phi_0(k)})_{k \geq 0}$ donne l'existence de $\phi_1 : \mathbb{N} \rightarrow \mathbb{N}$ telle que $x_{\phi_1(k)}^{(1)}$ converge vers $x^{(1)}$ mais également $x_{\phi_1(k)}^{(0)}$ converge vers $x^{(0)}$.

En répétant le procédé indéfiniment, nous construisons $\phi_n : \mathbb{N} \rightarrow \mathbb{N}$, $n \geq 0$, strictement croissante telle que pour tout $n \geq 0$ et tout $\ell = 0, \dots, n$, $x_{\phi_n(k)}^{(\ell)}$ converge vers $x^{(\ell)}$. On vérifie facilement que $(x_{\phi_n(n)}^{(\ell)})_{n \geq 0}$ converge pour tout $\ell \in \mathbb{N}$, c'est à dire $(x_{\phi_n(n)})_{n \geq 0}$ converge dans la topologie produit. \square

Remarque 13. Ce principe d'extraction successive s'appelle principe de la suite diagonale de Cantor.

Théorème de Heine et digression

Définition 1.2.75 (Continuité uniforme). Une fonction $f : (E, d) \rightarrow (E', d')$ est uniformément continue si

$$\forall \varepsilon > 0, \quad \exists \delta > 0 : \quad \forall x, y \in E, \quad d(x, y) < \delta \implies d'(f(x), f(y)) < \varepsilon.$$

Définition 1.2.76 (Application lipschitzienne). Une application $f : (E, d) \rightarrow (E', d')$ est dite lipschitzienne si il existe $K > 0$ tel que pour tout $x, y \in E$, $d'(f(x), f(y)) \leq Kd(x, y)$. Une telle fonction est dite K -lipschitzienne.

On remarque que si f est K -lipschitzienne, alors elle est K' -lipschitzienne pour toute $K' \geq K$. La plus petite constante K telle que f est K -lipschitzienne est appelée constante de Lipschitz.

Proposition 1.2.77. Soit $f : (E, d) \rightarrow (E', d')$. Alors,

1. si f est lipschitzienne, f est uniformément continue ;
2. si f est uniformément continue, f est continue.

Démonstration. Immédiat. \square

Exemple 12. Soit (E, d) un espace métrique et $A \subset E$ non vide. Alors, l'application qui à $x \in E$ associe à $d(x, A)$ dans \mathbb{R} muni de la métrique $|\cdot|$ est 1-lipschitzienne.

Définition 1.2.78. Une application $f : (E, d) \rightarrow (E', d')$ est une isométrie si pour tout $x, y \in E$, $d'(f(x), f(y)) = d(x, y)$.

Définition 1.2.79. Une application $f : (E, d) \rightarrow (E', d')$ est un homéomorphisme si f est bijective et f et f^{-1} sont continues. On dit que (E, d) et (E', d') sont homéomorphes.

Si l'application identité $\text{id} : (E, d) \rightarrow (E, d)$ est continue, on dit que d définit une topologie plus fine que d' . Si id est un homéomorphisme alors on dit que d et d' sont topologiquement équivalente. Si id et sa réciproque sont uniformément continue, on dit que d et d' sont uniformément équivalente. Enfin, si id et sa réciproque sont lipschitziennes, on dit que d et d' sont métriquement équivalente.

Proposition 1.2.80. Si d et d' sont métriquement équivalente alors d et d' sont uniformément équivalente.

Démonstration. Exercice. \square

Théorème 1.2.81 (Théorème de Heine). Soient (E, d) un espace métrique compact, (E', d') un espace métrique et $f : E \rightarrow E'$ une application continue. Alors f est uniformément continue sur E .

Démonstration. Exercice. \square

1.2.7 Espaces métriques complets

Si les notions de limites, de continuité, de compacité peuvent se définir dans le contexte général des espaces topologiques, la notion de suite de Cauchy est spécifique aux espaces métriques.

Définition 1.2.82 (Suite de Cauchy). Soit (E, d) un espace métrique. Une suite $(x_n)_{n \geq 0}$ à valeur dans E est dite de Cauchy si elle vérifie la propriété suivante, appelée propriété de Cauchy,

$$\forall \varepsilon > 0, \quad \exists N \in \mathbb{N} : \quad p, q \geq N \implies d(x_p, x_q) < \varepsilon.$$

Remarque 14. Une autre formulation pour une suite de Cauchy est la suivante : $(x_n)_{n \geq 0}$ est de Cauchy si et seulement si $\lim_{n \rightarrow \infty} \sup_{p \geq 0} d(x_{n+p}, x_n) = 0$.

Exemple 13. Toute suite convergente à valeurs dans un espace métrique vérifie la propriété de Cauchy.

Proposition 1.2.83. *Toute suite de Cauchy a au plus une valeur d'adhérence.*

Démonstration. Soient $x, y \in \mathbb{X}$ deux valeurs d'adhérences d'une suite de Cauchy $(x_n)_{n \geq 0}$. Alors, pour tout $n, m \in \mathbb{N}$

$$d(x, y) \leq d(x, x_n) + d(x_n, x_m) + d(x_m, y).$$

Soit $\varepsilon > 0$, il existe $N \geq 0$ tel que $n, m \geq N$ implique $d(x_n, x_m) < \varepsilon/3$. Puisque x, y sont des valeurs d'adhérence, il existe $n \geq N$ tel que $d(x, x_n) < \varepsilon/3$, de même il existe $m \geq N$ tel que $d(y, x_m) < \varepsilon/3$. Finalement, $d(x, y) < \varepsilon$. Comme $\varepsilon > 0$ peut être choisis arbitrairement petit, on obtient $d(x, y) = 0$ et $x = y$. \square

Définition 1.2.84 (Espace métrique complet). Un espace métrique (E, d) est complet si toute suite de Cauchy est convergente.

Exercice 12. Montrer que $(\mathbb{Q}, |\cdot|)$ n'est pas complet.

Proposition 1.2.85 (Propriété fondamentale des espaces de Baire). Soient (E, d) un espace métrique complet et $(F_n)_{n \geq 0}$ est une suite décroissante de fermés non vides dont le diamètre tend vers 0, alors $\bigcap_{n \geq 0} F_n$ est un singleton.

Démonstration. Pour chaque $n \geq 0$, on choisit un point $x_n \in F_n$. Soit $\varepsilon > 0$, il existe $N \geq 0$ tel que $n \geq N$ implique $\text{Diam } F_n < \varepsilon$, mais alors $n, m \geq N$ implique $d(x_n, x_m) \leq \text{Diam } F_n < \varepsilon$. Autrement dit, $(x_n)_{n \geq 0}$ est une suite de Cauchy donc admet une unique valeur d'adhérence $x \in \mathbb{X}$. Par la proposition 1.2.37,

$$\{x\} = \bigcap_{n \geq 0} \overline{\{x_m : m \geq n\}} \subset \bigcap_{n \geq 0} F_n.$$

Pour l'inclusion inverse, il suffit de voir que si $y \in \bigcap_{n \geq 0} F_n$ alors $d(x, y) \leq \text{Diam } F_n$ pour tout $n \geq 0$. D'où $x = y$. \square

Proposition 1.2.86. $(\mathbb{R}, |\cdot|)$ est complet.

Démonstration. Soit $(x_n)_{n \geq 0}$ une suite réelle vérifiant la propriété de Cauchy, alors $A_n = \{x_k : k \geq n\}$ est une suite décroissante de parties de \mathbb{R} . De plus, A_n est borné pour tout $n \geq 0$. En effet, pour tout $\varepsilon > 0$, il existe $N \geq 0$ tel que $|x_p - x_N| < \varepsilon$ dès que $p \geq N$, d'où

$$\sup_{k \geq 0} |x_k| \leq \max \left(\max_{i=0, \dots, N} |x_i|, \varepsilon + |x_N| \right) = r_\varepsilon, \quad A_n \in \overline{B}(0, r_\varepsilon).$$

On note $\alpha_n = \inf A_n$ et $\beta_n = \sup A_n$ alors $\text{Diam } A_n = \beta_n - \alpha_n$. Comme $(x_n)_{n \geq 0}$ est une suite de Cauchy, $\text{Diam } A_n \rightarrow 0$.

La suite $(\alpha_n)_{n \geq 0}$ est croissante majorée, $(\beta_n)_{n \geq 0}$ est décroissante minorée, elles sont convergentes. On note α et β leurs limites respectives. Comme $\text{Diam } A_n = \beta_n - \alpha_n$ tend vers 0, on a en fait $\alpha = \beta$. D'où $x_n \rightarrow \alpha = \beta$. \square

Remarque 15. C'est donc la propriété du supremum et de l'infimum dans \mathbb{R} qui permet de conclure.

Remarque 16. La notion de complétude dépend explicitement de la métrique, aussi la complétude est une notion métrique et non topologique. En particulier, on peut trouver des exemples de métriques définissant la même topologie sans être pour autant simultanément complète ou non complète.

Définition 1.2.87. Un \mathbb{K} -espace vectoriel normé E est un espace de Banach s'il est complet pour la métrique associée à la norme sur E .

Les démonstrations des trois propositions suivantes sont renvoyées en exercice de travaux dirigés.

Proposition 1.2.88. *Tout \mathbb{K} -espace vectoriel normé de dimension finie est un espace de Banach.*

Proposition 1.2.89. *L'espace $(\mathcal{C}_{\mathbb{K}}([0, 1]), \|\cdot\|_{\infty})$ est un espace de Banach.*

Proposition 1.2.90. *Soit $p \in [1, \infty]$ et \mathcal{S} un ensemble discret. Alors $\ell^p(\mathcal{S})$ est un espace de Banach.*

Définition 1.2.91. Soit $(u_n)_{n \geq 0}$ une suite d'éléments d'un espace vectoriel normé $(E, \|\cdot\|)$. On dit que la série de terme général u_n converge si la suite des sommes partielles converge au sens de la topologie induit par la norme. On dit qu'elle converge absolument si la série de terme général $\|u_n\|$ converge dans \mathbb{R} .

Proposition 1.2.92. *Un espace vectoriel normé $(E, \|\cdot\|)$ est un espace de Banach si et seulement si toute série absolument convergente converge.*

C'est précisément cette caractérisation que l'on utilisera dans le chapitre 6 pour montrer la complétude des espaces \mathbf{L}^p .

Démonstration. Supposons que E est un espace de Banach, alors on vérifie que la suite $(S_n)_{n \geq 0}$ des sommes partielles est de Cauchy, pour $q > p \geq n$:

$$\|S_q - S_p\| \leq \left\| \sum_{k=p+1}^q u_k \right\| \leq \sum_{k=p+1}^q \|u_k\| \leq \sum_{k=n+1}^{\infty} \|u_k\| \xrightarrow{n \rightarrow \infty} 0.$$

La propriété de complétude implique la convergence de la série.

Réciproquement, considérons une suite de Cauchy $(u_n)_{n \geq 0}$. On peut en extraire une sous-suite $(v_n)_{n \geq 0}$ telle que $\|v_{n+1} - v_n\| \leq 2^{-n}$. Ainsi, la série de terme générale $v_{n+1} - v_n$ est absolument convergente donc convergente par hypothèse. Autrement dit, la suite extraite $(v_n)_{n \geq 0}$ converge, or une suite de Cauchy admet au plus une valeur d'adhérence donc la suite $(x_n)_{n \geq 0}$ converge également. Ainsi, $(E, \|\cdot\|)$ est complet. \square

La proposition suivante montre que la complétude est une notion métrique en opposition à la compacité qui était une notion topologique.

Proposition 1.2.93. *Soit (E, d) un espace métrique complet. Alors, pour toute métrique δ métriquement équivalente à d sur E , (E, δ) est complet.*

Démonstration. Immédiat. \square

Proposition 1.2.94. *Tout produit dénombrable d'espaces métriques $((E_n, d_n))_{n \geq 0}$ complets est complet pour la métrique*

$$d(x, y) = \sum_{n=0}^{\infty} 2^{-n} \frac{d_n(x_n, y_n)}{1 + d_n(x_n, y_n)}, \quad (x, y) = ((x_n)_{n \geq 0}, (y_n)_{n \geq 0}) \in \prod_{n \geq 0} E_n.$$

Démonstration. Si $(x_k)_{k \geq 0}$ est une suite de Cauchy à valeurs dans $\prod_{n \geq 0} E_n$, alors pour tout $n \geq 0$, $(x_k^{(n)})_{k \geq 0}$ est de Cauchy dans (E_n, d_n) complet donc converge vers $x^{(n)}$. Il est immédiat que $(x_k)_{k \geq 0}$ converge vers $x = (x^{(n)})_{n \geq 0}$. \square

Proposition 1.2.95. *Tout espace métrique compact est complet.*

Démonstration. Une suite de Cauchy admet au plus une valeur d'adhérence. Or dans un espace métrique compact, par Bolzano-Weierstrass, toute suite admet au moins une valeur d'adhérence. Donc une suite de Cauchy dans un espace métrique compact admet exactement une valeur d'adhérence, elle est convergente. \square

Proposition 1.2.96. *Dans un espace métrique complet, les parties complètes sont exactement les parties fermées.*

Démonstration. Exercice. \square

Définition 1.2.97. Un espace métrique (E, d) est pré-compact si

$$\forall \varepsilon > 0, \quad \exists x_1, \dots, x_{N_\varepsilon} \in E : \quad E \subset \bigcup_{i=1}^{N_\varepsilon} B(x_i, \varepsilon).$$

Proposition 1.2.98. *Tout espace métrique pré-compact complet est compact.*

Démonstration. On va montrer la propriété de Bolzano-Weierstrass, on considère donc $(x_n)_{n \geq 0}$ une suite dans l'espace pré-compact (E, d) . Soit $\varepsilon = 1/2$, alors il existe $y_1 \in E$ tel que $B(y_1, 1/2)$ contient une infinité de points x_n , on note $(x_{\phi_1(n)})_{n \geq 0}$ cette suite extraite. Elle vérifie, pour tout $n, m \geq 0$, $d(x_{\phi_1(n)}, x_{\phi_1(m)}) \leq 1$.

On réitère le procédé avec $\varepsilon = 1/4$: on trouve une suite $(x_{\phi_2(n)})_{n \geq 0}$ de $(x_{\phi_1(n)})_{n \geq 0}$ vérifiant $d(x_{\phi_2(n)}, x_{\phi_2(m)}) \leq 1/2$ pour tout $n, m \geq 0$.

Ainsi, nous construisons successivement des sous-suites $(x_{\phi_k(n)})_{n \geq 0}$ satisfaisant, pour tout n, m , $d(x_{\phi_k(n)}, x_{\phi_k(m)}) < 1/k$. On considère la suite diagonale $(x_{\phi_n(n)})_{n \geq 0}$ extraite de $(x_n)_{n \geq 0}$ et les termes d'indices $n \geq k$ sont extraits de la suite $(x_{\phi_k(n)})_{n \geq 0}$ donc $d(x_{\phi_n(n)}, x_{\phi_m(m)}) < 1/k$ dès que $m, n \geq k$. Par conséquent, c'est une suite de Cauchy dans un espace complet donc elle converge. La propriété de Bolzano-Weierstrass est vérifiée. \square

Proposition 1.2.99. *Soient E, F deux \mathbb{K} -espace vectoriel normés. On suppose que F est un espace de Banach, alors $\mathcal{L}_{\mathbb{K}}(E, F)$ muni de la norme subordonnée est complet. En particulier, le dual topologique d'un \mathbb{K} -espace vectoriel normé est un espace de Banach.*

Cette proposition est démontrée en travaux dirigés.

Théorème 1.2.100 (Prolongement). *Soient (E, d) , (E', d') deux espaces métriques, $A \subset E$ une partie dense dans E , $f : A \rightarrow E'$ une application uniformément continue. Si (E', d') est complet alors f se prolonge de manière unique en une fonction $\tilde{f} : E \rightarrow E'$ (uniformément) continue sur E .*

Démonstration. Unicité : Si $f, g : E \rightarrow E'$ sont deux fonctions continues qui coïncident sur A , alors $f = g$. En effet, soit $x \in E$, il existe une suite $(x_n)_{n \geq 0}$ de points de A qui converge vers x . Pour tout $n \geq 0$, $f(x_n) = g(x_n)$. En faisant tendre n vers l'infini, la continuité de f et g implique $f(x) = g(x)$.

Existence : Soit $x \in E$, alors pour tout $n \geq 1$, $B(x, 1/n) \cap A \neq \emptyset$ par hypothèse de densité. On note $A_n = f(B(x, 1/n) \cap A) \subset E'$. Alors $A_{n+1} \subset A_n$ et $\text{Diam } A_n \rightarrow 0$ par uniforme continuité de f . La propriété fondamentale des espaces de Baire implique $\bigcap_{n \geq 0} A_n$ est réduite à un singleton, on note $\bigcap_{n \geq 0} A_n = \{\tilde{f}(x)\}$. Notons que $\bigcap_{n \geq 0} A_n \subset f(\{x\} \cap A)$ et donc si $x \in A$, on obtient $f(x) = \tilde{f}(x)$.

En fait, on a mieux, si $(x_\ell)_{\ell \geq 0}$ est une suite de points de A convergente vers x , alors pour tout $n \geq 0$, il existe ℓ_0 tel que pour tout $\ell \geq \ell_0$, $x_\ell \in B(x, 1/n) \cap A$ si bien que $f(x_\ell) \in A_n$. Donc $(f(x_\ell))_{\ell \geq 0}$ est de Cauchy convergente vers $\tilde{f}(x)$. Cela nous permet de montrer l'uniforme continuité de \tilde{f} . En effet, soient $x, y \in \mathbb{X}$, $(x_\ell)_{\ell \geq 0} \in A^{\mathbb{N}}$ et $(y_\ell)_{\ell \geq 0} \in A^{\mathbb{N}}$ deux suites qui convergent respectivement vers x et y , alors pour tout $\ell \geq 0$

$$d(\tilde{f}(x), \tilde{f}(y)) \leq d(\tilde{f}(x), f(x_\ell)) + d(f(x_\ell), f(y_\ell)) + d(f(y_\ell), \tilde{f}(y)).$$

Soit $\varepsilon > 0$. On peut trouver $L \geq 0$ tel que pour tout $\ell \geq L$,

$$d(\tilde{f}(x), f(x_\ell)) < \varepsilon/3 \quad \text{et} \quad d(\tilde{f}(y), f(y_\ell)) < \varepsilon/3.$$

De plus, on peut trouver $\eta > 0$ tel que $d(x_\ell, y_\ell) < 3\eta$ implique $d(f(x_\ell), f(y_\ell)) < \varepsilon/3$. Enfin, comme

$$d(x_\ell, y_\ell) \leq d(x_\ell, x) + d(x, y) + d(y, y_\ell),$$

choisissons $L' \geq 0$ tel que $\ell \geq L'$ implique $d(x_\ell, x) < \eta$ et $d(y_\ell, y) < \eta$, si bien que $d(x, y) < \eta$ implique $d(x_\ell, y_\ell) < 3\eta$. Finalement, pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que pour tout $\ell \geq L \vee L'$ et tout $x, y \in \mathbb{X}$ avec $d(x, y) < \eta$ alors

$$d(\tilde{f}(x), \tilde{f}(y)) \leq d(\tilde{f}(x), f(x_\ell)) + d(f(x_\ell), f(y_\ell)) + d(f(y_\ell), \tilde{f}(y)) \leq \varepsilon.$$

□

Théorème 1.2.101 (Complétion d'un espace métrique). *Tout espace métrique (E, d) est isométrique à un sous-espace dense d'un espace métrique complet $(\widehat{E}, \widehat{d})$ unique à isométrie près et appelé le complété de (E, d) .*

1.3 Espaces polonais

Nous nous bornons ici à définir ce qu'est un espace polonais qui est l'archétype des espaces que l'on croise souvent en théorie de la mesure.

Définition 1.3.1 (Espace polonais). Un espace métrique séparable complet est appelé espace polonais.

Chapitre 2

Tribus, applications mesurables et mesures

L'objet de ce chapitre est d'introduire les concepts fondamentaux de la théorie de la mesure. La théorie de la mesure donne un formalisme robuste aux notions de longueur, aire ou volume. Il est cependant important de noter que la théorie de la mesure ne se cantonne pas aux espaces euclidiens mais peut tout à fait s'appliquer à des espaces de fonctions ! Il ne s'agit pas là d'un érotisme mathématique mais bel et bien des fondations permettant la construction du mouvement brownien lequel apparaît comme la brique élémentaire dans de nombreux domaines applications (théorie cinétique des gaz, mathématiques financières...).

2.1 Tribus et Applications mesurables

2.1.1 Tribu

Dans toute la suite, \mathbb{X} désigne un ensemble. On notera $\mathcal{P}(\mathbb{X})$ l'ensemble des parties de \mathbb{X} .

Définition 2.1.1. Une tribu sur \mathbb{X} est un sous-ensemble non vide de $\mathcal{P}(\mathbb{X})$, noté \mathcal{X} , tel que

1. $\emptyset \in \mathcal{X}$;
2. $A \in \mathcal{X}$ implique $A^c \in \mathcal{X}$ (stable par passage au complémentaire) ;
3. si $(A_n)_{n \geq 1} \subset \mathcal{X}$ alors $\bigcup_{n \geq 1} A_n \in \mathcal{X}$ (stable par réunion dénombrable).

Proposition 2.1.2. Si \mathcal{X} est une tribu alors

1. $\mathbb{X} \in \mathcal{X}$,
2. $(A_n)_{n \geq 1} \subset \mathcal{X}$ implique $\bigcap_{n \geq 1} A_n \in \mathcal{X}$ (stable par intersection dénombrable),
3. $A, B \in \mathcal{X}$ alors $A \setminus B \in \mathcal{X}$,
4. $A, B \in \mathcal{X}$ alors $A \Delta B \in \mathcal{X}$.

Démonstration. Immédiat. □

Exemple 14. On peut mentionner deux tribus particulières la tribu la plus grossière $\mathcal{X} = \{\emptyset, \mathbb{X}\}$ et la tribu la plus fine $\mathcal{X} = \mathcal{P}(\mathbb{X})$. Lorsque \mathbb{X} est finie ou dénombrable, on choisira la plupart du temps la tribu la plus fine. Si \mathbb{X} est non dénombrable, le choix de la tribu n'est plus aussi naturel.

Exercice 13. Vérifier que la tribu grossière et la tribu la plus fine sont effectivement des tribus.

Définition 2.1.3. Un espace mesurable est la donnée d'un couple $(\mathbb{X}, \mathcal{X})$ constitué d'un ensemble \mathbb{X} et d'une tribu \mathcal{X} sur \mathbb{X} .

Proposition 2.1.4. Soit $(\mathcal{X}_i)_{i \in I}$ une famille de tribus sur \mathbb{X} alors $\mathcal{X} = \bigcap_{i \in I} \mathcal{X}_i$ est une tribu sur \mathbb{X} .

Remarque 17. Il s'agit bien d'une intersection quelconque, il n'est pas nécessaire qu'elle soit dénombrable !

Démonstration. Pour $i \in I$, \mathcal{X}_i est une tribu, donc $\emptyset \in \mathcal{X}_i$ pour tout $i \in I$ et il vient $\emptyset \in \mathcal{X}$. Soit $(A_j)_{j \geq 0}$ une famille dénombrable d'éléments de \mathcal{X} , alors pour tout $i \in I$, $\cup_{j \geq 0} A_j \in \mathcal{X}_i$ et donc $\cup_{j \geq 0} A_j \in \mathcal{X}$. La stabilité par passage au complémentaire se démontre de façon analogue. \square

Cette proposition permet de définir la notion de tribu engendrée.

Proposition 2.1.5 (Tribu engendrée). *Soit $\mathcal{C} \subset \mathcal{P}(\mathbb{X})$. Il existe une plus petite tribu (au sens de l'inclusion) contenant \mathcal{C} . Cette tribu est appelée tribu engendrée par \mathcal{C} et est notée $\sigma(\mathcal{C})$.*

Démonstration. Soit \mathfrak{S} l'ensemble des tribus sur \mathbb{X} qui contiennent \mathcal{C} . Alors

$$\mathcal{X} = \bigcap_{\mathcal{S} \in \mathfrak{S}} \mathcal{S} = \{A \subset \mathbb{X} : \forall \mathcal{S} \in \mathfrak{S}, A \in \mathcal{S}\}$$

est une tribu qui contient \mathcal{C} . Par définition, cette tribu est contenue dans toutes les tribus contenant \mathcal{C} . \square

Exemple 15. Un autre exemple de tribu classique est celle engendrée par un sous-ensemble $A \subset \mathbb{X}$, par définition

$$\sigma(\{A\}) = \{\emptyset, A, A^c, \mathbb{X}\}.$$

Proposition 2.1.6 (Image réciproque d'une tribu). *Soient \mathbb{X} et \mathbb{Y} deux ensembles, $f : \mathbb{X} \rightarrow \mathbb{Y}$ une application et \mathcal{Y} une tribu sur \mathbb{Y} . Alors*

$$f^{-1}(\mathcal{Y}) = \{f^{-1}(A), A \in \mathcal{Y}\}$$

est une tribu sur \mathbb{X} appelée tribu image réciproque de \mathcal{Y} par f .

Démonstration. Le résultat est conséquence directe du lemme ci-dessous dont la preuve est laissée en exercice.

Lemme 2.1.7. *Soit F un ensemble et $\mathcal{P}(F)$ l'ensemble de ses parties. Soient $A \in \mathcal{P}(F)$ et $(B_i)_{i \in I}$ une famille d'éléments de $\mathcal{P}(F)$. Alors pour toute application $f : E \rightarrow F$*

$$f^{-1}\left(\bigcup_{i \in I} B_i\right) = \bigcup_{i \in I} f^{-1}(B_i), \quad f^{-1}\left(\bigcap_{i \in I} B_i\right) = \bigcap_{i \in I} f^{-1}(B_i), \quad f^{-1}(A^c) = f^{-1}(A)^c.$$

\square

Proposition 2.1.8. *Soient f une application de \mathbb{X} dans \mathbb{Y} , \mathcal{C} un sous-ensemble de $\mathcal{P}(\mathbb{Y})$. Alors*

$$f^{-1}(\sigma(\mathcal{C})) = \sigma(f^{-1}(\mathcal{C})).$$

Démonstration. Comme $\mathcal{C} \subset \sigma(\mathcal{C})$, on a $f^{-1}(\mathcal{C}) \subset f^{-1}(\sigma(\mathcal{C}))$ qui est une tribu par la proposition précédente. Ainsi, $\sigma(f^{-1}(\mathcal{C}))$ est inclus dans $f^{-1}(\sigma(\mathcal{C}))$.

Montrons l'inclusion inverse. Notons \mathcal{A} l'ensemble des parties de $A \subset \mathbb{Y}$ telle $f^{-1}(A) \in \sigma(f^{-1}(\mathcal{C}))$. Alors \mathcal{A} est une tribu : $\emptyset \in \mathcal{A}$ et la stabilité de l'image inverse par réunion et passage au complémentaire permet de conclure. De plus \mathcal{A} contient \mathcal{C} , donc $\sigma(\mathcal{C})$. Il en résulte que $f^{-1}(\sigma(\mathcal{C})) \subset f^{-1}(\mathcal{A})$. Puis, par définition, $f^{-1}(\mathcal{A}) \subset \sigma(f^{-1}(\mathcal{C}))$, d'où le résultat $f^{-1}(\sigma(\mathcal{C})) \subset \sigma(f^{-1}(\mathcal{C}))$. \square

Proposition 2.1.9 (Tribu induite). *Soient $(\mathbb{X}, \mathcal{X})$ un espace mesurable et $B \subset \mathbb{X}$, l'ensemble $\mathcal{X}_B = \{A \cap B, A \in \mathcal{X}\}$ est une tribu sur B appelée tribu induite par \mathcal{X} sur B .*

Remarque 18. Notez que la partie B n'est pas supposée mesurable.

Démonstration. On vérifie les axiomes d'une tribu :

1. évidemment $\emptyset \in \mathcal{X}_B$ puisque $\emptyset = \emptyset \cap B$;
2. soit $A \in \mathcal{X}_B$, il existe par définition $\tilde{A} \in \mathcal{X}$ tel que $A = \tilde{A} \cap B$. Or le complémentaire de A dans B est

$$A^c \cap B = (\tilde{A}^c \cup B^c) \cap B = \tilde{A}^c \cap B.$$

Ainsi, $A^c \in \mathcal{X}_B$ puisque $\tilde{A}^c \in \mathcal{X}$.

3. Soit $(A_n)_{n \geq 1}$ des éléments de \mathcal{X}_B , alors pour chaque $n \geq 1$ on peut trouver $\tilde{A}_n \in \mathcal{X}$ tels que $A_n = \tilde{A}_n \cap B$. Alors

$$\bigcup_{n \geq 1} A_n = \bigcup_{n \geq 1} (\tilde{A}_n \cap B) = \left(\bigcup_{n \geq 1} \tilde{A}_n \right) \cap B.$$

Ceci montre la stabilité par réunion dénombrable puisque $\bigcup_{n \geq 1} \tilde{A}_n \in \mathcal{X}$. □

Définition 2.1.10 (Tribu produit). Soient $(\mathbb{X}, \mathcal{X})$ et $(\mathbb{Y}, \mathcal{Y})$ deux espaces mesurables. La tribu engendrée par les parties de $\mathbb{X} \times \mathbb{Y}$ s'écrivant comme $A \times B$ avec $A \in \mathcal{X}$ et $B \in \mathcal{Y}$ est appelée tribu produit et on la note $\mathcal{X} \otimes \mathcal{Y}$.

2.1.2 Tribu borélienne

Définition 2.1.11. Soit $(\mathbb{X}, \mathcal{T})$ un espace topologique. La tribu borélienne sur \mathbb{X} , notée $\mathcal{B}(\mathbb{X})$ est la tribu engendrée par les ouverts de $(\mathbb{X}, \mathcal{T})$. Autrement dit, $\mathcal{B}(\mathbb{X}) = \sigma(\mathcal{T})$. Les éléments de $\mathcal{B}(\mathbb{X})$ sont appelés les boréliens.

A priori, si (\mathbb{X}, d) est un espace métrique, la tribu engendrée par les boules ouvertes ne coïncident pas avec la tribu borélienne. On montre cependant que c'est vrai pour un espace métrique séparable en utilisant le lemme suivant.

Lemme 2.1.12. Soit (\mathbb{X}, d) un espace métrique séparable. Alors tout ouvert est réunion dénombrable de boules ouvertes.

Démonstration. Sous l'hypothèse de séparabilité, il existe une suite $(x_n)_{n \geq 0}$ dense dans \mathbb{X} . Soit O un ouvert et posons

$$I = \{(n, \rho) \in \mathbb{N} \times \mathbb{Q}_*^+ : B(x_n, \rho) \subset O\},$$

et montrons

$$O = \bigcup_{(n, \rho) \in I} B(x_n, \rho).$$

La réunion est par définition incluse dans O . Réciproquement, soit $x \in O$. Comme O est ouvert, il existe $r > 0$ tel $B(x, r) \subset O$. On a même pour tout $\rho \in \mathbb{Q} \cap (0, r)$, $B(x, \rho) \subset B(x, r) \subset O$. De plus, comme $(x_n)_{n \geq 0}$ est dense dans \mathbb{X} , il existe une sous-suite $(x_{n_k})_{k \geq 0}$ qui converge vers x . Autrement dit, il existe $K \geq 0$ tel que $k \geq K$ implique $x_{n_k} \in B(x, \rho/4)$. Par symétrie d'une métrique, $x \in B(x_{n_k}, \rho/4)$. Finalement, par l'inégalité triangulaire, pour tout $y \in B(x_{n_k}, \rho/4)$

$$d(x, y) \leq d(x, x_{n_k}) + d(x_{n_k}, y) \leq \rho/2 \implies x \in B(x_{n_k}, \rho/4) \subset B(x, \rho/2) \subset O.$$

□

Corollaire 2.1.13. Si (\mathbb{X}, d) est un espace métrique séparable, alors la tribu engendrée par les boules ouvertes coïncident avec la tribu borélienne.

Démonstration. En utilisant le lemme 2.1.12, la preuve est immédiate. □

On appliquera très souvent cette notion de tribu borélienne à \mathbb{R}^d muni de sa topologie usuelle, typiquement celle donnée par une métrique issue d'une norme.

Proposition 2.1.14. Sur \mathbb{R} muni de sa topologie usuelle, la tribu borélienne est engendrée par

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

Démonstration. 1. Notons \mathcal{E} l'ensemble des intervalles ouverts bornés de \mathbb{R} et \mathcal{O} les ouverts de \mathbb{R} . On a bien entendu $\mathcal{E} \subset \mathcal{O}$ si bien que $\sigma(\mathcal{E}) \subset \sigma(\mathcal{O})$.

Pour l'inclusion inverse, il suffit de remarquer que $(\mathbb{R}, |\cdot|)$ est séparable, ainsi tout ouvert est réunion dénombrable d'intervalles ouverts bornés. Ainsi $\mathcal{O} \subset \sigma(\mathcal{E})$ et donc $\sigma(\mathcal{O}) \subset \sigma(\mathcal{E})$.

2. Soit \mathcal{E}' la classe des intervalles de la forme $(-\infty, a)$. Encore une fois, $\sigma(\mathcal{E}') \subset \sigma(\mathcal{O})$ puisque $\mathcal{E}' \subset \mathcal{O}$. Pour l'inclusion inverse, il suffit de montrer que $\mathcal{E} \subset \sigma(\mathcal{E}')$, puisque par le point précédent, nous aurons $\sigma(\mathcal{E}) \subset \sigma(\mathcal{E}') \subset \sigma(\mathcal{O})$. Soit $(a, b) \in \mathcal{E}$. On a

$$(a, b) = (-\infty, b) \cap (a, \infty) = (-\infty, b) \cap (-\infty, a]^{\mathbb{C}}.$$

Puis, comme $(-\infty, a] = \bigcap_{n \geq 1} (-\infty, a + 1/n)$, nous avons montré que l'on peut écrire (a, b) comme l'intersection dénombrable d'éléments de \mathcal{E}' .

3. Ce point se démontre de la même manière que le précédent. □

Corollaire 2.1.15. *La tribu borélienne sur \mathbb{R}^d muni de sa topologie usuelle est engendrée par*

1. les pavés ouverts $\prod_{i=1}^d (a_i, b_i)$;
2. les pavés ouverts semi-infinis $\prod_{i=1}^d (-\infty, a_i)$;
3. les pavés fermés semi-infinis $\prod_{i=1}^d (-\infty, a_i]$.

Démonstration. C'est une application immédiate du lemme 2.1.12. □

2.1.3 La droite achevée

Pour diverses raisons, nous auront à considérer la compactification, généralement notée $\overline{\mathbb{R}}$, de la droite réelle \mathbb{R} . Pour se faire, il est possible de construire un homéomorphisme de \mathbb{R} dans $] - 1, 1[$. L'intervalle ouvert $] - 1, 1[$ se compactifie en $[-1, 1]$: c'est le plus petit compact de \mathbb{R} qui contient $] - 1, 1[$. Par exemple, les applications suivantes réalisent un tel homéomorphisme :

$$\begin{aligned} f : \mathbb{R} &\rightarrow] - 1, 1[& \text{et} & g : \mathbb{R} &\rightarrow] - 1, 1[\\ x &\rightarrow f(x) = \frac{x}{\sqrt{x^2+1}} & & x &\rightarrow g(x) = \frac{2}{\pi} \arctan(x). \end{aligned}$$

L'adhérence de l'intervalle ouvert $(-1, 1)$ dans la topologie \mathbb{R} est alors l'intervalle $[-1, 1]$. Il est alors possible de prolonger les homéomorphismes f et g à $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. Les points $-\infty$ et $+\infty$ sont alors les antécédents de -1 et par 1 par ces homéomorphismes. On note dans la suite \tilde{f} et \tilde{g} les prolongements de f et g à $\overline{\mathbb{R}}$.

Posons, pour tout $x, y \in \overline{\mathbb{R}}$,

$$\delta_{\tilde{f}}(x, y) = |\tilde{f}(x) - \tilde{f}(y)| \quad \text{et} \quad \delta_{\tilde{g}}(x, y) = |\tilde{g}(x) - \tilde{g}(y)|.$$

Ces deux applications sont des métriques sur $\overline{\mathbb{R}}$. L'application identité entre $(\overline{\mathbb{R}}, \delta_{\tilde{f}})$ et $(\overline{\mathbb{R}}, \delta_{\tilde{g}})$ est un homéomorphisme. Par ailleurs, la topologie résultant de la restrictions de ces métriques à \mathbb{R} coïncident avec la topologie usuelle sur \mathbb{R} associée à la valeur absolue. De plus, les espaces métriques $(\overline{\mathbb{R}}, \delta_{\tilde{f}})$ et $(\overline{\mathbb{R}}, \delta_{\tilde{g}})$ sont compacts. Ainsi, $\overline{\mathbb{R}}$ muni de l'une ou l'autre de ces métriques peut être vue comme une extension compacte de la droite réelle munie de la valeur absolue.

Notons qu'une base d'ouvert pour cette topologie est constituée des intervalles ouverts de la forme (a, b) , $(a, \infty]$ et $[-\infty, b)$ avec $a, b \in \mathbb{R}$. Une base de voisinage dénombrable de $+\infty$ (*resp.* $-\infty$) est données par $(n, +\infty]$ (*resp.* $[-\infty, -n)$), $n \geq 0$.

On démontre de façon analogue que la tribu borélienne de $\overline{\mathbb{R}}$ est engendrée par les classes $\{[-\infty, a), a \in \overline{\mathbb{R}}\}$ ou $\{[-\infty, a], a \in \overline{\mathbb{R}}\}$.

Enfin, il est à noter l'ordre total de \mathbb{R} peut également être étendu à $\overline{\mathbb{R}}$ puisque les homéomorphismes \tilde{f} et \tilde{g} sont monotones croissants. Les opérations algébriques tels que l'addition et la multiplication peuvent également être étendues dans une certaine mesure. Il subsiste néanmoins des indéterminations tels que $+\infty - \infty$. C'est une obstruction à la possibilité de définir une structure de groupe. *A priori*, les opérations $0 \times \pm\infty$ sont également indéterminées, néanmoins, par convention en théorie de la mesure, nous poserons $0 \times \pm\infty = 0$. Cette convention n'est pas source d'erreur et la raison apparaîtra plus clairement dans la suite.

2.1.4 Applications mesurables, applications boréliennes

Définition 2.1.16 (Applications mesurables). Soient $(\mathbb{X}, \mathcal{X})$ et $(\mathbb{Y}, \mathcal{Y})$ deux espaces mesurables. Une application $f : \mathbb{X} \rightarrow \mathbb{Y}$ est dite mesurable si

$$f^{-1}(\mathcal{Y}) \subset \mathcal{X},$$

ou de manière plus explicite

$$A \in \mathcal{E} \implies f^{-1}(A) \in \mathcal{X}.$$

Définition 2.1.17. Soient \mathbb{X}, \mathbb{Y} deux espaces topologiques. Une application mesurable de $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ dans $(\mathbb{Y}, \mathcal{B}(\mathbb{Y}))$ est dite borélienne.

Pour $A \subset \mathbb{X}$, on définit l'application indicatrice de A , notée $\mathbf{1}_A$, de \mathbb{X} dans $\{0, 1\}$ pour $x \in \mathbb{X}$ par

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

On munit $\{0, 1\}$ de la topologie discrète (l'ensemble des ouverts n'est rien d'autre que l'ensemble des parties de $\{0, 1\}$). La tribu borélienne correspondante est la tribu grossière. De fait, l'application $\mathbf{1}_A$ est mesurable dès que A est mesurable. Réciproquement, toute application de f de \mathbb{X} dans $\{0, 1\}$ s'écrit $\mathbf{1}_{\text{supp } f}$. De plus $\text{supp } f = \{x \in \mathbb{X} : f(x) = 1\}$ est mesurable dès que f est mesurable.

Proposition 2.1.18. Soient $(\mathbb{X}, \mathcal{X})$ et $(\mathbb{Y}, \mathcal{Y})$ deux espaces mesurables, f une application de \mathbb{X} dans \mathbb{Y} et \mathcal{B} un ensemble de parties sur \mathbb{Y} telle que $\sigma(\mathcal{B}) = \mathcal{Y}$. Alors f est mesurable si et seulement si l'image réciproque de tout élément de \mathcal{B} est dans \mathcal{X} .

Démonstration. La condition est évidemment nécessaire. Réciproquement, si \mathcal{X} contient l'image réciproque de \mathcal{B} , elle contient également la tribu engendrée par l'image réciproque de \mathcal{B} , i.e. $\sigma(f^{-1}(\mathcal{B}))$. Par la Proposition 2.1.8, $\sigma(f^{-1}(\mathcal{B})) = f^{-1}(\sigma(\mathcal{B})) = f^{-1}(\mathcal{Y})$. \square

Corollaire 2.1.19. Soient \mathbb{X}, \mathbb{Y} deux espaces topologiques munis de leurs tribus boréliennes. Toute application continue de \mathbb{X} dans \mathbb{Y} est mesurable.

Démonstration. La continuité implique que l'image inverse de tout ouvert est ouverte. D'où le résultat. \square

Exercice 14. Donner un exemple d'application borélienne non continue.

Proposition 2.1.20. Une application $f : (\mathbb{X}, \mathcal{X}) \rightarrow \mathbb{R}$ est mesurable si

1. $\forall a \in \mathbb{R}, \{x \in \mathbb{X} : f(x) \leq a\} \in \mathcal{X}$,
2. $\forall a \in \mathbb{R}, \{x \in \mathbb{X} : f(x) < a\} \in \mathcal{X}$,
3. $\forall a \in \mathbb{R}, \{x \in \mathbb{X} : f(x) \geq a\} \in \mathcal{X}$,
4. $\forall a \in \mathbb{R}, \{x \in \mathbb{X} : f(x) > a\} \in \mathcal{X}$.

Démonstration. C'est une application directe des propositions 2.1.14 et 2.1.18. \square

Propriétés de stabilité

Proposition 2.1.21 (Composition). Soient $(\mathbb{X}, \mathcal{X})$, $(\mathbb{Y}, \mathcal{Y})$ et $(\mathbb{Z}, \mathcal{Z})$ trois espaces mesurables, f une application mesurable de $(\mathbb{X}, \mathcal{X})$ dans $(\mathbb{Y}, \mathcal{Y})$ et g une application mesurable de $(\mathbb{Y}, \mathcal{Y})$ dans $(\mathbb{Z}, \mathcal{Z})$. Alors $f \circ g$ est mesurable de $(\mathbb{X}, \mathcal{X})$ dans $(\mathbb{Z}, \mathcal{Z})$.

Démonstration. Immédiat. \square

Proposition 2.1.22. Soient $(\mathbb{X}_1, \mathcal{X}_1)$ et $(\mathbb{X}_2, \mathcal{X}_2)$ deux espaces mesurables et p_1, p_2 les projections de $\mathbb{X}_1 \times \mathbb{X}_2$ sur \mathbb{X}_1 et \mathbb{X}_2 respectivement. On munit $\mathbb{X}_1 \times \mathbb{X}_2$ de la tribu produit $\mathcal{X}_1 \otimes \mathcal{X}_2$. Alors

1. les projections p_1 et p_2 sont mesurables ;

2. soient $(\mathbb{Y}, \mathcal{Y})$ un espace mesurable et $f : \mathbb{Y} \rightarrow \mathbb{X}_1 \times \mathbb{X}_2$ une application. Alors f est mesurable si et seulement si les composées $p_1 \circ f : \mathbb{Y} \rightarrow \mathbb{X}_1$ et $p_2 \circ f : \mathbb{Y} \rightarrow \mathbb{X}_2$ sont mesurables.

Démonstration. 1. Si $B_1 \in \mathcal{X}_1$, alors $p_1^{-1}(B_1) = B_1 \times \mathbb{X}_2 \in \mathcal{X}_1 \otimes \mathcal{X}_2$ et p_1 est mesurable. De la même manière p_2 est mesurable.

2. Si f est mesurable, par la proposition précédente, $p_1 \circ f$ et $p_2 \circ f$ sont mesurables. Réciproquement, supposons $p_1 \circ f$ et $p_2 \circ f$ mesurables. Alors pour tout $B_1 \in \mathcal{X}_1$, $f^{-1}(B_1 \times \mathbb{X}_2) = (p_1 \circ f)^{-1}(B_1)$ est dans la tribu \mathcal{Y} . De même, pour tout $B_2 \in \mathcal{X}_2$, $f^{-1}(\mathbb{X}_1 \times B_2) \in \mathcal{Y}$. Ainsi

$$f^{-1}(B_1 \times B_2) = f^{-1}((B_1 \times \mathbb{X}_2) \cap (\mathbb{X}_1 \times B_2)) = f^{-1}(B_1 \times \mathbb{X}_2) \cap f^{-1}(\mathbb{X}_1 \times B_2) \in \mathcal{Y}.$$

Comme $\mathcal{X}_1 \otimes \mathcal{X}_2$ est la tribu engendrée par les $B_1 \times B_2$ pour $B_1 \in \mathcal{X}_1$ et $B_2 \in \mathcal{X}_2$, on conclut à l'aide de la Proposition 2.1.8. □

Corollaire 2.1.23. *Pour qu'une application à valeurs complexes soit mesurable il faut et il suffit que sa partie réelle et sa partie imaginaire soient mesurables. Si f et g sont des applications mesurables de $(\mathbb{X}, \mathcal{X})$ à valeurs complexes, alors $f + g$, fg , $|f|$, ... sont mesurables.*

Démonstration. Il suffit d'appliquer les propositions 2.1.21 et 2.1.22 en remarquant que les applications $\mathbb{R}^2 \ni (x, y) \rightarrow x + y \in \mathbb{R}$, $\mathbb{R}^2 \ni (x, y) \rightarrow xy$ et $\mathbb{R} \ni x \rightarrow |x|$ sont continues donc mesurables. □

Définition 2.1.24. Soit $(x_n)_{n \geq 0}$ une suite à valeurs dans $\overline{\mathbb{R}}$. La plus grande (resp. la plus petite) valeur d'adhérence de $(x_n)_{n \geq 0}$ est notée $\limsup x_n$ (resp. $\liminf x_n$) et est définie par

$$\limsup x_n = \inf_{n \geq 0} \sup_{k \geq n} x_k \quad (\text{resp. } \liminf x_n = \sup_{n \geq 0} \inf_{k \geq n} x_k).$$

Remarque 19. On note parfois $\overline{\lim}$ et $\underline{\lim}$ en lieu et place de \limsup et \liminf .

Remarque 20. Les limites supérieures et inférieures sont a priori des éléments de $\overline{\mathbb{R}}$. Il est tout à fait possible d'avoir $\limsup x_n = \infty$ et $\liminf x_n = -\infty$, c'est le cas par exemple pour $x_n = (-1)^n n$. On a toujours $\liminf x_n \leq \limsup x_n$ et $(x_n)_{n \geq 0}$ converge si et seulement si $\liminf x_n \geq \limsup x_n$.

Pour une suite de fonctions $(f_n)_{n \geq 0}$ sur \mathbb{X} à valeurs dans $\overline{\mathbb{R}}$, on note $\limsup f_n$ et $\liminf f_n$ les fonctions qui à $x \in \mathbb{X}$ associe $\limsup f_n(x)$ et $\liminf f_n(x)$ respectivement.

Proposition 2.1.25 (Stabilité par passage à la limite). 1. Soit $(f_n)_{n \geq 0}$ une suite de fonctions mesurables sur $(\mathbb{X}, \mathcal{X})$ à valeur dans $\overline{\mathbb{R}}$. Les fonctions $\sup f_n$, $\inf f_n$, $\limsup f_n$ et $\liminf f_n$ sont mesurables.

2. Soit $(f_n)_{n \geq 0}$ une suite de fonctions mesurables à valeurs dans \mathbb{C} telle que pour tout $x \in E$, $\lim_n f_n(x) = f(x)$ existe. Alors f est mesurable.

Démonstration. 1. Par hypothèse, pour tout $a \in \mathbb{R}$, l'ensemble $\{f_n \leq a\}$ est dans \mathcal{X} . Or, $\{\sup f_n \leq a\} = \bigcap_{n \geq 0} \{f_n \leq a\}$. Par la Proposition 2.1.20, $\sup f_n$ est mesurable. Comme $\inf f_n = -\sup -f_n$, $\inf f_n$ est mesurable. Enfin, $\limsup f_n = \inf_{n \geq 0} \sup_{k \geq n} f_k$ et $\liminf f_n = \sup_{n \geq 0} \inf_{k \geq n} f_k$ sont mesurables par ce qui précède.

2. Il suffit de montrer que partie réelle et partie imaginaire de f est mesurable. Sans perte de généralité, on peut donc supposer seulement le cas réel. Dans ce cas, $f = \limsup f_n = \liminf f_n$ et est donc mesurable. □

Exercice 15. Soient f et g deux applications mesurables de $(\mathbb{X}, \mathcal{X})$ dans $\overline{\mathbb{R}}_+$ (muni de sa tribu borélienne). Montrer que $\{f \leq g\}$ et $\{f < g\}$ sont des ensembles mesurables.

2.1.5 Approximation des fonctions mesurables

Soit $(\mathbb{X}, \mathcal{X})$ un espace mesuré. On notera $m\mathcal{X}$ l'ensemble des fonctions mesurables et par $m\mathcal{X}_+$ l'ensemble des fonctions mesurables positives.

Définition 2.1.26 (Fonctions étagées). Une fonction mesurable sur $(\mathbb{X}, \mathcal{X})$ à valeurs dans \mathbb{C} est dite étagée si elle ne prend qu'un nombre fini de valeurs distinctes. On notera $m\mathcal{E}$ et $m\mathcal{E}_+$ respectivement l'ensemble des fonctions mesurables étagées bornées et mesurables étagées positives.

Soit f une fonction étagée et $\alpha_1, \dots, \alpha_n$ les n valeurs distinctes prises par f . Pour $i \in \{1, \dots, n\}$, on pose $A_i = \{f = \alpha_i\}$. Puisque f est mesurable, les ensembles A_i sont mesurables et f se réécrit

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

Réciproquement, toute combinaison linéaire finie à coefficients réels ou complexes de fonctions indicatrices d'ensembles mesurables est une fonction étagée. L'ensemble des fonctions étagées est \mathbb{K} -espace vectoriel (de dimension infinie).

Théorème 2.1.27. Soit f une fonction mesurable sur $(\mathbb{X}, \mathcal{X})$ à valeurs dans $\overline{\mathbb{R}}_+$. Alors, il existe une suite croissante $(f_n)_{n \geq 0}$ de fonctions étagées positives qui converge simplement vers f . De plus, la convergence est uniforme sur tout ensemble $B \in \mathcal{X}$ sur lequel f est bornée.

Démonstration. Pour $n \geq 0$ et $k = 0, 1, \dots, n2^n - 1$, posons (c.f. Figure 2.1)

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

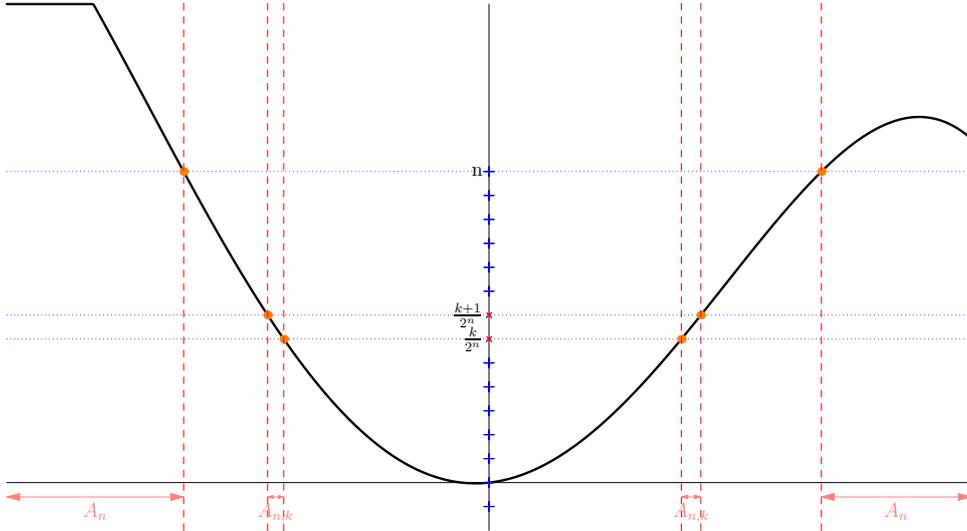


FIGURE 2.1 – Découpage dyadique tronqué d'une fonction mesurable f .

On définit alors la fonction f_n par :

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

Par définition, f_n est une fonction étagée positive telle que $f_n \leq f$. D'autre part, on vérifie que si $x \in A_{n,k}$,

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

D'autre part, si $x \in A_n$,

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

Ainsi, pour tout $n \geq 0$ et tout $x \in \mathbb{X}$, $f_n(x) \leq f_{n+1}(x)$: la suite $(f_n)_{n \geq 0}$ est croissante. De plus, $(A_n)_{n \geq 0}$ est une suite décroissante d'éléments de \mathcal{X} . Donc si $x \in A_{n_0}^{\mathbb{C}}$, alors pour tout $n \geq n_0$, $x \in A_n^{\mathbb{C}}$ ou encore pour tout $n \geq n_0$

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

Ceci implique que $(f_n(x))_{n \geq 0}$ converge vers $f(x)$. Ainsi, la suite $(f_n)_{n \geq 0}$ converge sur l'ensemble $\cup_{n \geq 0} A_n^{\mathbb{C}}$ qui n'est autre que $\{f < \infty\}$. Si $x \in \{f = \infty\}$, alors pour tout $n \geq 0$, $f_n(x) = n$ qui tend vers ∞ quand n tend vers ∞ . Soit à présent $B \in \mathcal{X}$ tel que f soit bornée sur B . Il existe n_1 tel que, pour tout $x \in B$, $f(x) < n_1$. Alors $B \cap A_{n_1} = \emptyset$ et ainsi

$$\forall n \geq n_1, \forall x \in B, 0 \leq f(x) - f_n(x) \leq \frac{1}{2^n}.$$

La convergence est donc bien uniforme sur B . □

Corollaire 2.1.28. *Toute fonction f mesurable sur $(\mathbb{X}, \mathcal{X})$ à valeurs dans \mathbb{R} (resp. \mathbb{R}^d , resp. \mathbb{C}) est limite simple d'une suite $(f_n)_{n \geq 0}$ de fonctions étagées à valeurs dans \mathbb{R} (resp. \mathbb{R}^d , resp. \mathbb{C}).*

Démonstration. Si f est à valeurs dans \mathbb{R} , on peut l'écrire $f = f^+ - f^-$ avec $f^+ = f \vee 0$ et $f^- = -f \wedge 0$. Comme f^+ et f^- sont mesurables à valeurs dans $\overline{\mathbb{R}}_+$, il existe des suites $(g_n)_{n \geq 0}$ et $(h_n)_{n \geq 0}$ de fonctions étagées positives tendant simplement vers f^+ et f^- respectivement. La suite $(f_n)_{n \geq 0}$, où $f_n = g_n - h_n$, est formée de fonctions étagées et converge simplement vers f . Si f est à valeurs dans \mathbb{R}^d , on raisonne composante par composante. De même, si f est à valeurs complexes, on l'écrira comme somme de ses parties réelle et imaginaire. □

2.2 Mesures positives

2.2.1 Définitions et propriétés élémentaires

Définition 2.2.1 (Mesure positive). Soit $(\mathbb{X}, \mathcal{X})$ un espace mesurable. Une mesure positive, ou simplement une mesure, sur $(\mathbb{X}, \mathcal{X})$ est une application $\mu : \mathcal{X} \rightarrow \overline{\mathbb{R}}_+$ satisfaisant

1. $\mu(\emptyset) = 0$,
2. si $(A_n)_{n \geq 0}$ est une suite d'éléments deux à deux disjoints (*i.e.* $A_n \cap A_m = \emptyset$ dès que $n \neq m$), alors

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

Cette deuxième propriété est appelée σ -additivité.

Une mesure positive vérifiant $\mu(\mathbb{X}) < \infty$ est dite finie. Si elle vérifie $\mu(\mathbb{X}) = 1$, c'est une mesure de probabilité. Enfin, si il existe une suite $(A_n)_{n \geq 0}$ d'éléments de \mathcal{X} telle que $\mathbb{X} = \cup_{n \geq 0} A_n$ et, pour tout $n \in \mathbb{N}$, $\mu(A_n) < \infty$, on dit que μ est σ -finie.

Définition 2.2.2 (Espace mesuré). Un espace mesuré est la donnée d'un triplet $(\mathbb{X}, \mathcal{X}, \mu)$ où $(\mathbb{X}, \mathcal{X})$ est un espace mesurable et μ est une mesure positive sur $(\mathbb{X}, \mathcal{X})$.

Lorsqu'il n'y aura pas d'ambiguïtés sur la tribu \mathcal{X} nous écrirons simplement (\mathbb{X}, μ) . Ce sera notamment le cas lorsque \mathbb{X} est discret ou lorsque \mathbb{X} est un espace topologique. Dans le premier cas, on munira l'ensemble discret presque exclusivement de la tribu la plus fine, celle qui contient toutes les parties de \mathbb{X} . Dans le second cas, on considérera en générale la tribu borélienne.

Proposition 2.2.3. *Soit $(\mathbb{X}, \mathcal{X}, \mu)$ un espace mesuré.*

1. Si $A_1, \dots, A_n \in \mathcal{X}$ sont deux à deux disjoints, alors

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

2. Soient $A, B \in \mathcal{X}$ tels que $A \subset B$, alors $\mu(A) \leq \mu(B)$. De plus, si $\mu(A) < \infty$, alors $\mu(B \setminus A) = \mu(B) - \mu(A)$.

3. Soient $A, B \in \mathcal{X}$, $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$.

Remarque 21. Une mesure μ est dite finiment additive si dans le deuxième point de la définition 2.2.1, la famille d'ensemble (resp. la réunion, la somme) dénombrable est remplacée par une famille finie.

Démonstration. 1. On pose $B_1 = A_1, \dots, B_n = A_n$ et pour tout $i \geq n+1$, $B_i = \emptyset$. Alors $(B_n)_{n \geq 0}$ est une famille dénombrable d'ensembles deux à deux disjoints. On obtient l'additivité finie à l'aide de la σ -additivité et du fait que $\mu(\emptyset) = 0$.

2. On écrit $B = A \cup (A^c \cap B)$, c'est la réunion de deux ensembles mesurables disjoints, d'où $\mu(B) = \mu(A) + \mu(A^c \cap B) \geq \mu(A)$. De plus $A^c \cap B = B \setminus A$ d'où, si $\mu(A) < \infty$, l'égalité $\mu(B \setminus A) = \mu(B) - \mu(A)$.

3. Comme $A \cap B \subset A$, si $\mu(A \cap B) = \infty$ alors $\mu(A) = \infty$ et l'égalité est vérifiée (en fait $\mu(B) = \infty$ également). Si $\mu(A \cap B) < \infty$, on peut écrire $A \cup B = A \setminus (A \cap B) \cup A \cap B \cup B \setminus (A \cap B)$ qui est une réunion disjointe. D'où

$$\mu(A \cup B) = \mu(A \setminus (A \cap B)) + \mu(A \cap B) + \mu(B \setminus (A \cap B)) = \mu(A) + \mu(B) - \mu(A \cap B).$$

□

La proposition suivante donne une définition équivalente d'une mesure positive.

Proposition 2.2.4. Une application $\mu : \mathcal{X} \rightarrow \overline{\mathbb{R}}_+$ est une mesure si et seulement si

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

Remarque 22. Cette définition équivalente a l'avantage de faire apparaître explicitement une propriété asymptotique des mesures (dans le point (3)). C'est cette propriété qui sera cruciale pour montrer le théorème de convergence monotone qui fait toute la puissance de l'intégrale de Lebesgue.

Démonstration. Ce sont des conditions suffisantes. En effet, par récurrence sur le point (ii), pour toute collection finie A_1, \dots, A_n d'ensembles mesurables deux à deux disjoints, on a

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

Si $(A_n)_{n \geq 1}$ est une collection d'ensembles mesurables deux à deux disjoints, en posant $B_k = \cup_{1 \leq n \leq k} A_n$, alors $\mu(B_k) = \sum_{n=1}^k \mu(A_n)$. De plus, $(B_k)_{k \geq 1}$ est une suite croissante telle que $\cup_{k \geq 1} B_k = \cup_{n \geq 1} A_n$. Par le point (3), on obtient

$$\mu(\cup_{n=1}^{\infty} A_n) = \mu(\cup_{k \geq 1} B_k) = \lim_{k \rightarrow \infty} \mu(B_k) = \lim_{n \rightarrow \infty} \sum_{k=1}^n \mu(A_k) = \sum_{k=1}^{\infty} \mu(A_k).$$

Réciproquement, supposons que μ soit une mesure. Soit $(B_n)_{n \geq 0}$ une suite croissante d'ensembles mesurables. Posons $A_0 = B_0$ et, pour tout $n \geq 1$, $A_n = B_n \setminus B_{n-1} \in \mathcal{X}$. Alors $(A_n)_{n \geq 0}$ est une suite d'ensembles mesurables deux à deux disjoints et, pour tout $n \geq 0$, $B_n = \cup_{k=0}^n A_k$. Il en résulte que

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

Ceci achève la preuve de la proposition. □

Proposition 2.2.5. Soit $(\mathbb{X}, \mathcal{X}, \mu)$ un espace mesuré.

1. Si $(B_n)_{n \geq 0}$ est une famille d'éléments de \mathcal{X} , alors $\mu(\cup_{n \geq 0} B_n) \leq \sum_{n \geq 0} \mu(B_n)$.
2. Si $(B_n)_{n \geq 0}$ est une suite décroissante de \mathcal{X} telle que $\mu(B_{n_0}) < \infty$ pour un certain $n_0 \geq 0$, alors la suite $(\mu(B_n))_{n \geq 0}$ converge en décroissant vers $\mu(\cap_{n \geq 0} B_n)$.

Remarque 23. Dans le deuxième point, l'existence d'un entier n_0 tel que $\mu(B_{n_0}) < \infty$ est nécessaire. En effet, si μ est la mesure de comptage sur \mathbb{N} et $B_n = \{n, n+1, \dots\}$ alors $\mu(B_n) = \infty$ pour tout $n \geq 0$ et $\cap_{n \geq 0} B_n = \emptyset$.

Démonstration. 1. On pose $A_0 = B_0$ et, pour tout $n \geq 1$, $A_n = B_n \setminus (\cup_{k=0}^{n-1} B_k) = B_n \cap (\cup_{k=0}^{n-1} B_k)^c$. Les ensembles $(A_n)_{n \geq 0}$ sont deux à deux disjoints et $\cup_{k=0}^n B_k = \cup_{k=0}^n A_k$. D'où, puisque $A_n \subset B_n$ pour tout $n \geq 0$.

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

2. Pour $k \geq n_0$, on pose $A_k = B_{n_0} \setminus B_k$. La suite $(A_k)_{k \geq n_0}$ est croissante et on a $\cup_{k \geq n_0} A_k = B_{n_0} \setminus (\cap_{k \geq n_0} B_k)$. Puisque $\cap_{k \geq n_0} B_k \subset B_{n_0}$ et $B_k \subset B_{n_0}$, on a

$$\mu(B_{n_0} \setminus (\cap_{k \geq n_0} B_k)) = \mu(B_{n_0}) - \mu(\cap_{k \geq n_0} B_k) \quad \text{et} \quad \mu(A_k) = \mu(B_{n_0}) - \mu(B_k),$$

d'où

$$\begin{aligned} \mu(B_{n_0}) - \mu(\cap_{k \geq n_0} B_k) &= \mu(\cup_{k \geq n_0} A_k) = \lim_{k \rightarrow \infty} \mu(A_k) \\ &= \lim_{k \rightarrow \infty} (\mu(B_{n_0}) - \mu(B_k)) = \mu(B_{n_0}) - \lim_{k \rightarrow \infty} \mu(B_k), \end{aligned}$$

et donc $\mu(\cap_{k \geq 1} B_k) = \lim_{k \rightarrow \infty} \mu(B_k)$. □

Définition 2.2.6. Soient $(\mathbb{X}, \mathcal{X}, \mu)$ un espace mesuré et $(A_n)_{n \geq 0}$ une famille d'éléments de \mathcal{X} . On définit la limite supérieure et inférieure de cette famille comme suit

$$\limsup A_n = \bigcap_{n \geq 0} \bigcup_{m \geq n} A_m \quad \text{et} \quad \liminf A_n = \bigcup_{n \geq 0} \bigcap_{m \geq n} A_m.$$

Intuitivement, si $x \in \limsup A_n$, cela signifie que x est dans une infinité de A_n . Si $x \in \liminf A_n$, cela signifie que x est dans tous les A_n à partir d'un certain rang $n \geq 0$. Notons également que

$$(\limsup A_n)^c = \liminf A_n^c.$$

Proposition 2.2.7 (Borel-Cantelli). Soient $(\mathbb{X}, \mathcal{X}, \mu)$ un espace mesuré et $(A_n)_{n \geq 0}$ une famille d'éléments de \mathcal{X} . Si, quitte à enlever un nombre fini de termes, la série des $(\mu(A_n))_{n \geq 0}$ est finie, alors $\mu(\limsup A_n) = 0$.

Démonstration. La suite $(\sum_{m \geq n} \mu(A_m))_{n \geq 0}$ est décroissante, de plus il existe $n_0 \geq 0$ tel que

$$\infty > \sum_{n \geq n_0} \mu(A_n) \geq \mu\left(\bigcup_{n \geq n_0} A_n\right).$$

D'où il vient que

$$\mu\left(\bigcap_{n \geq 0} \bigcup_{m \geq n} A_m\right) = \lim_{n \rightarrow \infty} \mu\left(\bigcup_{m \geq n} A_m\right) \leq \lim_{n \rightarrow \infty} \sum_{m \geq n} \mu(A_m) = 0.$$

□

2.2.2 Quelques exemples de mesures : mesures discrètes et mesure de Lebesgue

Définition 2.2.8 (Masse de Dirac). Soient $(\mathbb{X}, \mathcal{X})$ un espace mesuré et $a \in \mathbb{X}$. Posons pour tout $A \in \mathcal{X}$

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

L'application δ_a est une mesure de probabilité, appelée la mesure (ou masse) de Dirac au point a .

Exercice 16. Vérifier que la masse de Dirac en un point est bien une mesure.

Définition 2.2.9 (Mesure de Bernoulli). Soit $p \in (0, 1)$. Sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, la mesure de Bernoulli de paramètre p est définie par $\mu = (1-p)\delta_0 + p\delta_1$.

Remarque 24. La mesure de Bernoulli est ici définie sur $\mathbb{X} = \mathbb{R}$ mais on aurait pu choisir $\mathbb{X} = \{0, 1\}$, $\mathbb{X} = \mathbb{N}$ ou encore $\mathbb{X} = [0, 1] \dots$

Définition 2.2.10 (Mesures discrètes). Soit $(\mathbb{X}, \mathcal{X})$ un espace mesurable. Soient $(a_n)_{n \geq 0}$ une suite de points de \mathbb{X} et $(\alpha_n)_{n \geq 0}$ une suite à valeurs dans $\overline{\mathbb{R}}_+$. Pour tout $A \in \mathcal{X}$, on pose

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

L'application $\mu : \mathcal{X} \rightarrow \overline{\mathbb{R}}_+$ est une mesure positive. Tout point a_n tel que $\alpha_n > 0$ est appelé atome de μ .

Lemme 2.2.11. Soit $(a_{n,m})_{n,m \geq 0}$ une suite double de réels positifs telle que

- pour tout $n \geq 0$, $a_{n,m} \leq a_{n,m+1}$, et
- pour tout $m \geq 0$, $a_{n,m} \leq a_{n+1,m}$.

Alors

$$\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} a_{n,m} = \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} a_{n,m} \in \overline{\mathbb{R}}_+.$$

Démonstration. Immédiat. □

Exercice 17. Soit $(a_{k,n})_{k,n \in \mathbb{N}}$ est suite double de nombres positifs. Montrer que l'égalité suivante a lieu dans $\overline{\mathbb{R}}_+$

$$\sum_{k=0}^{\infty} \sum_{n=0}^{\infty} a_{k,n} = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} a_{k,n}.$$

En déduire que toute mesure discrète est en effet une mesure.

Exemple 16. Les 4 premières mesures sont des probabilités.

1. La mesure uniforme μ sur $\{1, \dots, n\}$ de paramètre $n \geq 1$ est définie par

$$\mu = \frac{1}{n} \sum_{k=1}^n \delta_k.$$

2. La mesure binomiale μ de paramètres $p \in (0, 1)$ et $n \geq 1$ est définie par

$$\mu = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \delta_k.$$

3. La mesure géométrique μ de paramètre $p \in (0, 1)$ est définie par

$$\mu = \sum_{n \geq 1} p(1-p)^{n-1} \delta_n.$$

4. La mesure de Poisson de paramètre $\lambda > 0$ est définie par

$$\mu = \sum_{n \geq 0} e^{-\lambda} \frac{\lambda^n}{n!}.$$

5. La mesure de comptage associée à une suite $(a_n)_{n \geq 0}$ de points de \mathbb{X} .

$$\mu = \sum_{n \geq 0} \delta_{a_n}.$$

Théorème 2.2.12 (Mesure de Lebesgue). *Il existe une unique mesure λ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que*

1. $\lambda([0, 1]) = 1$,
2. pour tout $a \in \mathbb{R}$ et tout $B \in \mathcal{B}(\mathbb{R})$, $\lambda(a + B) = \lambda(B)$.

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

La mesure de Lebesgue est donc l'unique mesure invariante par translation qui affecte une masse 1 à l'intervalle $[0, 1]$. C'est la mesure "qui correspond" à l'intégrale de Riemann.

La démonstration de ce théorème est loin d'être immédiate. C'est une conséquence du théorème de Carathéodory donné à la fin de ce chapitre. On peut d'ores et déjà donner le résultat suivant.

Proposition 2.2.13. *Pour tous $a < b$ réels,*

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

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

Démonstration. Si I est un intervalle non borné, alors $I = (-\infty, a)$, $I = (-\infty, a]$, $I = (a, \infty)$, $I = [a, \infty)$ ou $I = \mathbb{R}$. Traitons le premier cas par exemple. On note n_a le plus grand entier plus petit que a . Alors,

$$\bigcup_{k=-\infty}^{n_a} (k-1, k] \subset I,$$

ainsi, par la première partie de la proposition et croissance d'une mesure, on obtient

$$\lambda(I) \geq \lambda\left(\bigcup_{k=-\infty}^{n_a} (k-1, k]\right) = \sum_{k=-\infty}^{n_a} \lambda((k-1, k]) = \sum_{k=-\infty}^{n_a} 1 = \infty.$$

Posons $\alpha = \lambda(\{0\})$, alors par invariance par translation de la mesure de Lebesgue et croissance des mesures, il est facile de voir que $n\alpha = \lambda(\{1/k : 1 \leq k \leq n\}) \leq 1$, et ce pour tout $n \geq 1$ donc $\alpha = 0$. De même, pour tout $x \in \mathbb{R}$, $\lambda(\{x\}) = 0$. Cela permet de conclure pour les trois premières égalités.

Clairement,

$$(0, 1] = \left(0, \frac{1}{n}\right] \cup \left(\frac{2}{n}, \frac{3}{n}\right] \cup \dots \cup \left(\frac{n-1}{n}, 1\right].$$

Par additivité et invariance par translation, on obtient que $\lambda((0, 1/n]) = 1/n$ et même pour tout $k_1 \leq k_2$

$$\lambda\left(\left(\frac{k_1}{n}, \frac{k_2}{n}\right]\right) = \frac{k_2 - k_1}{n}.$$

De là, on peut passer à des bornes rationnelles en remarquant que si $r = p_1/q_1$ et $s = p_2/q_2$ alors $(r, s] = \left(\frac{p_1 q_2}{q_1 q_2}, \frac{p_2 q_1}{q_1 q_2}\right]$. En fait, si $a < b$ sont deux réels alors il existe $(r_n)_{n \geq 0}$ une suite décroissante de rationnels et $(s_n)_{n \geq 0}$ une suite croissante de rationnels qui convergent respectivement vers a et b , de sorte que

$$\bigcup_{n \geq 0} (r_n, s_n] = (a, b) \quad \text{et} \quad \lambda((a, b)) = \lim_{n \rightarrow \infty} \lambda((r_n, s_n]) = b - a,$$

en utilisant la continuité à gauche d'une mesure. □

Remarque 25. En remarquant $(0, 1) = \bigcup_{x \in (0, 1)} \{x\}$, il devient transparent qu'il est illusoire de considérer une forme d'additivité plus générale que la σ -additivité car si nous pouvons bien donner un sens **dans ce cas précis** à $\sum_{x \in (0, 1)} \lambda(\{x\})$ puisque tous les termes de cette somme sont nuls, il est bien entendu que $1 = \lambda((0, 1)) \neq \sum_{x \in (0, 1)} \lambda(\{x\})$.

2.2.3 Théorème des classes monotones, caractérisation des mesures et théorème de prolongement de Carathéodory

Théorème des classes monotones

Définition 2.2.14 (π -système). Une famille \mathcal{C} de parties de \mathbb{X} est un π -système si

1. $\mathcal{C} \neq \emptyset$,
2. si $A, B \in \mathcal{C}$ alors $A \cap B \in \mathcal{C}$.

La notion d'algèbre de Boole contient la notion de π -système. Le π -système est l'hypothèse minimale apparaissant dans le théorème de classe monotone, mais la notion d'algèbre de Boole est plus naturelle.

Définition 2.2.15 (Algèbre de Boole). Une algèbre de Boole sur \mathbb{X} est un ensemble de parties \mathcal{C} vérifiant

1. $\mathbb{X} \in \mathcal{C}$,
2. $A \in \mathcal{C}$ alors $A^c \in \mathcal{C}$,
3. $A, B \in \mathcal{C}$ alors $A \cup B \in \mathcal{C}$.

Exercice 18. Vérifier que, sur \mathbb{R} , l'ensemble des intervalles est un π -système. Vérifier que l'ensemble des réunions finies d'intervalles deux à deux disjoints est une algèbre de Boole.

Remarque 26. Une tribu sur \mathbb{X} est une algèbre sur \mathbb{X} stable par union dénombrable :

$$(A_n)_{n \geq 0} \in \mathcal{C}^{\mathbb{N}} \implies \bigcup_{n \geq 0} A_n \in \mathcal{C}.$$

Cette remarque explique la dénomination de σ -algèbre parfois employée qui est par ailleurs traduite en *σ -algebra* en anglais.

Définition 2.2.16 (λ -système ou classe monotone). Une famille Λ de parties de \mathbb{X} est appelée λ -système si

1. $\mathbb{X} \in \Lambda$,
2. si $(A_n)_{n \geq 0}$ est suite croissante d'éléments de Λ alors $\bigcup_{n \geq 0} A_n \in \Lambda$,
3. si $A, B \in \Lambda$ avec $A \subset B$ alors $B \setminus A \in \Lambda$.

Remarque 27. Une tribu est en particulier une classe monotone.

Lemme 2.2.17. Soit Λ un λ -système stable par intersection finie. Alors Λ est une tribu.

Démonstration. Les points (1) et (3) de la définition d'un λ -système implique qu'un λ -système est stable par passage au complémentaire. En particulier, $\emptyset \in \Lambda$. Il reste donc à montrer que Λ est stable par union dénombrable. Si $A, B \in \Lambda$ alors $A \cup B = (A^c \cap B^c)^c$ ainsi Λ est stable par union finie. Soit $(A_n)_{n \geq 0}$ une famille d'éléments de Λ alors

$$\bigcup_{n \geq 0} A_n = \bigcup_{p \geq 0} \bigcup_{k=0}^p A_k.$$

Comme $(\bigcup_{k=0}^p A_k)_{p \geq 0}$ est une suite croissante de Λ , on montre la stabilité de Λ par union dénombrable, ce qui termine la preuve du lemme. \square

Proposition 2.2.18. Si \mathcal{S} est un ensemble de parties de \mathbb{X} alors il existe un plus petit λ -système contenant \mathcal{S} noté $\Lambda(\mathcal{S})$.

Démonstration. C'est la même idée que pour les tribus engendrées : il faut remarquer que l'intersection de λ -systèmes est un λ -système. \square

Les λ -systèmes sont des classes beaucoup moins riches que les tribus. Le théorème suivant va permettre de caractériser les mesures en se restreignant à ces λ -système.

Théorème 2.2.19 (des classes monotones). Si \mathcal{S} est un π -système alors $\Lambda(\mathcal{S}) = \sigma(\mathcal{S})$.

Démonstration. Du fait du lemme 2.2.17, il suffit de montrer que $\Lambda(\mathcal{S})$ est stable par intersection finie. Soit $B \in \mathcal{S}$ fixé, posons

$$\Lambda_B = \{A \in \Lambda(\mathcal{S}) : A \cap B \in \Lambda(\mathcal{S})\}.$$

On vérifie que Λ_B est un λ -système :

- tout d'abord, puisque $\mathbb{X} \cap B = B \in \mathcal{S} \subset \Lambda(\mathcal{S})$, $\mathbb{X} \in \Lambda_B$;
- d'autre part, si $(A_n)_{n \geq 0}$ est une suite croissante d'éléments de Λ_B alors $A_n \cap B \in \Lambda(\mathcal{S})$ pour tout $n \geq 0$. Or

$$\left(\bigcup_{n \geq 0} A_n \right) \cap B = \bigcup_{n \geq 0} (A_n \cap B) \quad (2.1)$$

si bien que, puisque $\Lambda(\mathcal{S})$ est un λ -système et que $(A_n \cap B)_{n \geq 0}$ est une suite croissante d'éléments de $\Lambda(\mathcal{S})$, le membre de gauche est un élément de $\Lambda(\mathcal{S})$. Ainsi, $\bigcup_{n \geq 0} A_n \in \Lambda_B$;

- enfin, $(A_1 \setminus A_0) \cap B = A_1 \cap B \setminus A_0 \cap B \in \Lambda(\mathcal{S})$ car $A_0, A_1 \in \Lambda_B$ et $\Lambda(\mathcal{S})$ est un λ -système.

Par ailleurs, \mathcal{S} est un π -système, il est stable par intersection finie si bien que pour tout $A \in \mathcal{S}$, $A \cap B \in \mathcal{S} \subset \Lambda(\mathcal{S})$. Nous avons donc montré que Λ_B est un λ -système qui contient \mathcal{S} , par conséquent il contient $\Lambda(\mathcal{S})$. En particulier, on a montré que

$$\forall B \in \mathcal{S}, \quad \forall A \in \Lambda(\mathcal{S}), \quad A \cap B \in \Lambda(\mathcal{S}). \quad (2.2)$$

Soit maintenant $C \in \Lambda(\mathcal{S})$ et posons

$$\Lambda_C = \{A \in \Lambda(\mathcal{S}) : A \cap C \in \Lambda(\mathcal{S})\}$$

En procédant de la même façon que pour Λ_B , nous pouvons montrer que Λ_C est un λ -système. De plus il contient \mathcal{S} puisque si $A \in \mathcal{S}$ alors $A \cap C \in \Lambda(\mathcal{S})$ par (2.2). Le λ -système Λ_C contient donc $\Lambda(\mathcal{S})$. Mais par définition, $\Lambda_C \subset \Lambda(\mathcal{S})$ donc pour tout $C \in \Lambda(\mathcal{S})$, $\Lambda_C = \Lambda(\mathcal{S})$. Ceci implique en particulier que $\Lambda(\mathcal{S})$ est stable par intersection finie, c'est donc une tribu et $\sigma(\mathcal{S}) \subset \Lambda(\mathcal{S})$. Comme une tribu est un λ -système, l'inclusion précédente est en fait une égalité. \square

Théorème 2.2.20 (Caractérisation des mesures). *Soit \mathcal{S} un π -système.*

1. *Soient μ et ν deux mesures finies sur $\sigma(\mathcal{S})$ telles que*

- $\mu(\mathbb{X}) = \nu(\mathbb{X})$,
 - *pour tout $A \in \mathcal{S}$, $\mu(A) = \nu(A)$.*
- Alors $\mu = \nu$.*

2. *Soient μ et ν deux mesures sur $\sigma(\mathcal{S})$ telles que*

- *pour tout $A \in \mathcal{S}$, $\mu(A) = \nu(A)$,*
- *il existe une suite exhaustive croissante $(B_n)_{n \geq 0}$ d'éléments de \mathcal{S} telle que $\mu(B_n) = \nu(B_n) < \infty$ pour tout $n \in \mathbb{N}$.*

Alors $\mu = \nu$.

Remarque 28. Ainsi, si l'on veut vérifier que deux mesures de probabilités coïncident sur $(\mathbb{X}, \mathcal{X})$ il suffit de vérifier qu'elles coïncident sur un π -système (ou une algèbre de Boole) engendrant la tribu \mathcal{X} . Typiquement, pour les mesures boréliennes sur \mathbb{R} , il suffira de le vérifier pour les intervalles $(-\infty, a]$ pour tout $a \in \mathbb{R}$.

Démonstration. On commence par le cas où μ et ν des mesures finies. On pose

$$\Lambda = \{A \in \sigma(\mathcal{S}) : \mu(A) = \nu(A)\}.$$

Montrons que Λ est un λ -système. Par hypothèse, $\mathbb{X} \in \Lambda$. De plus Λ est stable par union finie. Soit $(A_n)_{n \geq 0}$ une suite croissante d'éléments de Λ . Par stabilité par union finie et continuité de la mesure

$$\mu \left(\bigcup_{n \geq 0} A_n \right) = \lim_{n \rightarrow \infty} \mu \left(\bigcup_{k=0}^n A_k \right) = \lim_{n \rightarrow \infty} \nu \left(\bigcup_{k=0}^n A_k \right) = \nu \left(\bigcup_{n \geq 0} A_n \right),$$

si bien que $\cup_{n \geq 0} A_n \in \Lambda$. Soient $A, B \in \Lambda$ avec $A \subset B$. Puisque $\mu(\mathbb{X})$ et $\nu(\mathbb{X})$ sont finis et que $B \in \Lambda$, on a $\mu(B), \nu(B) < \infty$. De plus,

$$\mu(B \setminus A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B \setminus A),$$

et donc $B \setminus A \in \Lambda$.

Ainsi Λ est un λ -système qui contient le π -système \mathcal{S} donc $\Lambda(\mathcal{S}) \subset \Lambda$. Par définition, $\Lambda \subset \sigma(\mathcal{S})$ et d'après le théorème des classes monotones, $\Lambda(\mathcal{S}) = \sigma(\mathcal{S})$. Finalement, on a montré

$$\Lambda \subset \sigma(\mathcal{S}) = \Lambda(\mathcal{S}) \subset \Lambda.$$

Supposons désormais que μ et ν soient seulement σ -finies et soit $(B_n)_{n \geq 0}$ une suite exhaustive satisfaisant aux hypothèses du théorème. Pour $n \geq 0$, on définit les mesures μ_n et ν_n sur $\sigma(\mathcal{S})$ par

$$\forall A \in \sigma(\mathcal{S}), \quad \mu_n(A) = \mu(A \cap B_n), \quad \text{et} \quad \nu_n(A) = \nu(A \cap B_n).$$

Les mesures μ_n et ν_n sont finies et coïncident sur \mathcal{S} (qui rappelons le est stable par intersections finies) donc sur $\sigma(\mathcal{S})$ par le premier point. Enfin, pour tout $A \in \sigma(\mathcal{S})$, puisque $(A \cap B_n)_{n \geq 0}$ est une suite croissante d'éléments de $\sigma(\mathcal{S})$,

$$\mu(A) = \lim_{n \rightarrow \infty} \mu_n(A) = \lim_{n \rightarrow \infty} \mu(A \cap B_n) = \lim_{n \rightarrow \infty} \nu(A \cap B_n) = \lim_{n \rightarrow \infty} \nu_n(A) = \nu(A).$$

Ceci achève la preuve. □

Théorème d'extension de Carathéodory

Définition 2.2.21 (Mesure sur une algèbre de Boole). Soit \mathcal{C} une algèbre de Boole sur \mathbb{X} . Une mesure sur \mathcal{C} est une application

$$\mu : \mathcal{C} \rightarrow \mathbb{R}_+ \cup \{\infty\}$$

satisfaisant

1. $\mu(\emptyset) = 0$,
2. μ est finiment additive : si $A, B \in \mathcal{C}$ et $A \cap B = \emptyset$ alors $\mu(A \cup B) = \mu(A) + \mu(B)$,
3. si $(A_n)_{n \geq 0}$ est une suite décroissante d'éléments de \mathcal{C} telle que

$$\mu(A_0) < \infty \quad \text{et} \quad \bigcap_{n \geq 0} A_n = \emptyset \quad \implies \quad \mu(A_n) \downarrow 0 \quad \text{quand} \quad n \rightarrow \infty.$$

Définition 2.2.22. Soit \mathcal{C} une algèbre de Boole sur \mathbb{X} et μ une mesure sur \mathcal{C} . On dit que

1. μ est finie si $\mu(\mathbb{X}) < \infty$;
2. μ est σ -finie s'il existe une suite exhaustive $(B_n)_{n \geq 0}$ d'éléments de \mathcal{C} avec $\mu(B_n) < \infty$ pour tout $n \geq 0$ et telle que pour tout $A \in \mathcal{C}$, $\mu(A) = \lim_{n \rightarrow \infty} \mu(B_n \cap A)$.

Proposition 2.2.23. Soit μ une mesure finie sur une algèbre de Boole \mathcal{C} . Alors

1. μ est σ -additive : si $(A_n)_{n \geq 0}$ est une suite d'éléments de \mathcal{C} deux à deux disjoints et $\cup_{n \geq 0} A_n \in \mathcal{C}$, alors

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

2. μ est continue à gauche : si $(A_n)_{n \geq 0}$ est une suite croissante d'éléments de \mathcal{C} et $\cup_{n \geq 0} A_n \in \mathcal{C}$, alors

$$\mu(A_n) \uparrow \mu \left(\bigcup_{n \geq 0} A_n \right) \quad \text{quand} \quad n \rightarrow \infty.$$

3. μ est σ -sous-additive : si $(A_n)_{n \geq 0}$ est une suite d'éléments de \mathcal{C} et $\cup_{n \geq 0} A_n \in \mathcal{C}$ alors

$$\mu \left(\bigcup_{n \geq 0} A_n \right) \leq \sum_{n \geq 0} \mu(A_n).$$

Démonstration. 1. Soit $(A_n)_{n \geq 0}$ une suite d'éléments de \mathcal{C} deux à deux disjoints tels que $\cup_{n \geq 0} A_n \in \mathcal{C}$.
On pose

$$B_n = \left(\bigcup_{n \geq 0} A_n \right) \setminus \left(\bigcup_{k=0}^{n-1} A_k \right).$$

Comme \mathcal{C} est une algèbre de Boole, il est stable par union finie et par passage au complémentaire si bien que $B_n \in \mathcal{C}$ pour tout $n \geq 0$. Les ensembles A_0, A_1, \dots, A_{n-1} et B_n sont par ailleurs deux à deux disjoints. Puisque μ est finiment additive (le point (2) définissant une mesure sur une algèbre de Boole), on obtient

$$\mu \left(\bigcup_{n \geq 0} A_n \right) = \mu(A_0) + \mu(A_1) + \dots + \mu(A_{n-1}) + \mu(B_n). \quad (2.3)$$

D'autre part, la suite $(B_n)_{n \geq 0}$ est décroissante, $\cap_{n \geq 0} B_n = \emptyset$ et comme μ est finie $\mu(B_0) < \infty$. Par conséquent, par le point (3) définissant une mesure sur une algèbre de Boole : $\lim_{n \rightarrow \infty} \mu(B_n) = 0$. En faisant tendre n vers l'infini dans (2.3), on obtient

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

2. Soit $(A_n)_{n \geq 0}$ une suite croissante d'éléments de \mathcal{C} . Posons $B_0 = A_0$ et pour tout $n \geq 1$, $B_n = A_n \setminus A_{n-1}$. Alors les B_n sont des éléments de \mathcal{C} deux à deux disjoints et tels que pour tout $n \geq 0$, $A_n = \cup_{k=0}^n B_k$. D'autre part $\cup_{n \geq 0} B_n = \cup_{n \geq 0} A_n \in \mathcal{C}$ par hypothèse. Le point précédent montre

$$\begin{aligned} \mu \left(\bigcup_{n \geq 0} A_n \right) &= \mu \left(\bigcup_{n \geq 0} B_n \right) = \sum_{n \geq 0} \mu(B_n) = \lim_{n \rightarrow \infty} \uparrow \sum_{k=0}^n \mu(B_k) \\ &= \lim_{n \rightarrow \infty} \uparrow \mu \left(\bigcup_{k=0}^n B_k \right) = \lim_{n \rightarrow \infty} \uparrow \mu(A_n). \end{aligned}$$

3. On pose $B_0 = A_0$ et pour tout $n \geq 1$,

$$B_n = A_n \setminus \bigcup_{k=0}^{n-1} A_k.$$

Les ensembles B_n sont des éléments de \mathcal{C} deux à deux disjoints et $B_n = \cup_{k=0}^n A_k$. De plus, $\cup_{n \geq 0} B_n = \cup_{n \geq 0} A_n$. D'où, puisque $B_n \subset A_n$ si bien que $\mu(A_n) = \mu(B_n) + \mu(A_n \setminus B_n)$,

$$\mu \left(\bigcup_{n \geq 0} A_n \right) = \mu \left(\bigcup_{n \geq 0} B_n \right) = \sum_{n \geq 0} \mu(B_n) \leq \sum_{n \geq 0} \mu(A_n).$$

□

Remarque 29. Les points 2 et 3 de la proposition 2.2.23 se montrent en fait de la même façon que pour les mesures sur une tribu. Cependant, une algèbre de Boole n'est pas stable par réunion dénombrable contrairement à une tribu. Il faut ainsi s'assurer que tous les ensembles que l'on mesure soient bien dans l'algèbre de Boole considérée, en dehors la mesure μ n'est *a priori* pas définie.

Théorème 2.2.24 (de prolongement de Carathéodory). *Soit \mathcal{C} une algèbre de Boole¹ et μ une mesure*

1. L'hypothèse selon laquelle \mathcal{C} est une algèbre de Boole est un peu forte. Le théorème reste vrai si \mathcal{C} est un anneau d'ensemble que l'on définit ci-dessous. Une algèbre de Boole est en particulier un anneau, la réciproque est fautive (considérer $\mathcal{R} = \{\emptyset\}$ par exemple). Étant donné que la plupart des familles d'ensembles génératrices que nous considérerons seront des algèbres, il apparaissait naturel d'énoncer le théorème sous cette forme.

Définition 2.2.25 (Anneaux). Une famille \mathcal{R} de parties de \mathbb{X} est anneau d'ensemble si

1. \mathcal{R} n'est pas vide;
2. \mathcal{R} est stable par différence ensembliste;
3. \mathcal{R} est stable par union finie.

σ -finie sur \mathcal{C} . Alors, il existe une unique mesure $\tilde{\mu}$ sur $\sigma(\mathcal{C})$ qui coïncide avec μ sur \mathcal{C} .

L'unicité dans le théorème de Carathéodory est une conséquence du théorème de caractérisation des mesures de 2.2.20. Il s'agit donc de montrer l'existence du prolongement. L'idée est d'étendre μ à une application, définie sur l'ensemble des parties $\mathcal{P}(\mathbb{X})$ de \mathbb{X} , appelée mesure extérieure et qui sera notée μ^* . En général, μ^* n'est pas une mesure parce que l'ensemble des parties $\mathcal{P}(\mathbb{X})$ est trop riche. La solution consiste à enlever les parties de \mathbb{X} pathologiques en construisant une tribu convenable, contenant l'algèbre de Boole \mathcal{C} , de sorte que, restreinte à cette tribu, la fonction d'ensembles μ^* soit une mesure.

Définition 2.2.26 (Mesure extérieure). Une mesure extérieure sur \mathbb{X} est une application $\mu^* : \mathcal{P}(\mathbb{X}) \rightarrow \overline{\mathbb{X}}_+$ vérifiant

1. $\mu^*(\emptyset) = 0$;
2. μ^* est croissante : si $A \subset B$ alors $\mu^*(A) \leq \mu^*(B)$;
3. μ^* est σ -sous-additive : si $(A_n)_{n \geq 0}$ est une famille de parties de \mathbb{X} alors

$$\mu^* \left(\bigcup_{n \geq 0} A_n \right) \leq \sum_{n \geq 0} \mu^*(A_n).$$

Nous aurons besoins des résultats suivants qui seront montrés en fin de section.

Lemme 2.2.27. Soit $B \in \mathcal{P}(\mathbb{X})$ et posons

$$\mu^*(B) = \inf \left\{ \sum_{n \geq 0} \mu(B_n) : (B_n)_{n \geq 0} \in \mathcal{C}^{\mathbb{N}}, B \subset \bigcup_{n \geq 0} B_n \right\}.$$

Alors μ^* est une mesure extérieure et μ^* coïncide avec μ sur \mathcal{C} .

Proposition 2.2.28. Soit

$$\mathcal{U} = \{A \in \mathcal{P}(\mathbb{X}) : \forall B \in \mathcal{P}(\mathbb{X}), \mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c)\}.$$

Alors,

$$\mathcal{U} = \{A \in \mathcal{P}(\mathbb{X}) : \forall B \in \mathcal{P}(\mathbb{X}), \mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c)\},$$

autrement dit, l'inégalité opposée est toujours vérifiée. De plus, \mathcal{U} est une tribu et, restreint à cette tribu, μ^* est une mesure sur \mathcal{U} .

Remarque 30. La tribu \mathcal{U} est en quelque sorte la plus grande tribu sur laquelle la mesure extérieure μ^* est une mesure.

Proposition 2.2.29. La tribu \mathcal{U} contient l'algèbre de Boole \mathcal{C} ainsi que $\sigma(\mathcal{C})$.

Preuve du théorème de Carathéodory 2.2.24. L'unicité dans le théorème de Carathéodory est une conséquence directe du théorème 2.2.20 caractérisant les mesures puisque une algèbre de Boole est en particulier un π -système.

Lorsque μ est finie, l'existence est une conséquence du lemme 2.2.27 ainsi que des propositions 2.2.28 et 2.2.29 : $\tilde{\mu}$ est simplement la restriction de la mesure extérieure μ^* à $\sigma(\mathcal{C}) \subset \mathcal{U}$.

Supposons désormais que μ est σ -finie. Dans ce cas, il existe $(E_n)_{n \geq 0}$ une suite croissante d'éléments de \mathcal{C} avec $\mu(E_n) < \infty$, $\mathbb{X} = \bigcup_{n \geq 0} E_n$ et pour tout $A \in \mathcal{C}$, $\mu(A) = \lim_{n \rightarrow \infty} \mu(E_n \cap A)$. Pour tout $n \geq 0$, on définit la mesure μ_n pour tout $A \in \mathcal{C}$ par $\mu_n(A) = \mu(A \cap E_n)$. Ainsi chaque mesure μ_n est finie sur l'algèbre de Boole \mathcal{C} donc se prolonge en une unique mesure $\tilde{\mu}_n$ sur $\sigma(\mathcal{C})$.

Comme $\tilde{\mu}_n$ et $\tilde{\mu}_{n+1}(\cdot \cap E_n)$ coïncident sur \mathcal{C} , elles coïncident sur $\sigma(\mathcal{C})$ par le théorème 2.2.20 de caractérisation des mesures finies. De plus, pour tout $A \in \sigma(\mathcal{C})$ et tout $n \geq 0$,

$$\tilde{\mu}_n(A) = \tilde{\mu}_{n+1}(A \cap E_n) \leq \tilde{\mu}_{n+1}(A).$$

On pose alors pour tout $A \in \sigma(\mathcal{C})$

$$\tilde{\mu}(A) = \lim_{n \rightarrow \infty} \tilde{\mu}_n(A) = \sup_{n \geq 0} \tilde{\mu}_n(A).$$

De fait, $\tilde{\mu}$ restreinte à \mathcal{C} coïncide avec μ . Montrons que $\tilde{\mu}$ est une mesure. Clairement $\tilde{\mu}(\emptyset) = 0$ puisque $\emptyset \in \mathcal{C}$, donc $\tilde{\mu}(\emptyset) = \mu(\emptyset) = 0$. Soit $(A_n)_{n \geq 0}$ une famille d'éléments de $\sigma(\mathcal{C})$ deux à deux disjoints. Alors, puisque $\tilde{\mu}_n$ est une mesure

$$\tilde{\mu}_n \left(\bigcup_{k \geq 0} A_k \right) \geq \tilde{\mu}_n \left(\bigcup_{k=0}^p A_k \right) = \sum_{k=0}^p \tilde{\mu}_n(A_k).$$

En faisant tendre n , puis p , vers l'infini, on obtient

$$\tilde{\mu} \left(\bigcup_{n \geq 0} A_n \right) \geq \sum_{n \geq 0} \tilde{\mu}(A_n).$$

En outre,

$$\tilde{\mu}_n \left(\bigcup_{p \geq 0} A_p \right) = \sum_{p \geq 0} \tilde{\mu}_n(A_p) \leq \sum_{p \geq 0} \tilde{\mu}(A_p).$$

En faisant tendre n vers l'infini, on obtient l'inégalité inverse ce qui achève la preuve du théorème. \square

Preuve du lemme 2.2.27. On commence par montrer que, restreint à \mathcal{C} , μ^* coïncide avec μ . Pour cela, on se donne $A \in \mathcal{C}$ que l'on peut écrire comme une réunion $A = \bigcup_{n \geq 0} B_n$ avec $B_0 = A$ et $B_n = \emptyset$ pour tout $n \geq 1$. Donc $\mu^*(A) \leq \mu(A) = \sum_{n \geq 0} \mu(B_n)$ en utilisant la σ -additivité de la mesure μ .

Soit $\varepsilon > 0$, alors on peut trouver une suite $(B_n)_{n \geq 0}$ d'éléments de \mathcal{C} tel que $A \subset \bigcup_{n \geq 0} B_n$ et

$$\sum_{n \geq 0} \mu(B_n) \leq \mu^*(A) + \varepsilon.$$

Aussi, on peut réécrire A comme suit

$$A = \left(\bigcup_{n \geq 0} B_n \right) \cap A = \bigcup_{n \geq 0} B_n \cap A.$$

Ainsi, la réunion à droite est un élément de \mathcal{C} . De plus, pour chaque $n \geq 0$, B_n est dans \mathcal{C} qui est stable par intersection finie. En utilisant la σ -sous-additivité de μ ainsi que sa croissance (points (i) et (ii) de la proposition 2.2.23), on obtient

$$\mu(A) \leq \sum_{n \geq 0} \mu(B_n \cap A) \leq \sum_{n \geq 0} \mu(B_n) \leq \mu^*(A) + \varepsilon.$$

On a donc montré que si $A \in \mathcal{C}$, alors $\mu(A) = \mu^*(A)$.

Montrons désormais que μ^* est une mesure extérieure. Tout d'abord, puisque $\emptyset \in \mathcal{C}$, nous avons $\mu^*(\emptyset) = \mu(\emptyset) = 0$. La croissance de μ^* provient de la croissance de μ . Il reste donc à montrer que μ^* est σ -sous-additive. Soit $(A_k)_{k \geq 0}$ un ensemble de parties de \mathbb{X} et $\varepsilon > 0$. Pour chaque $k \geq 0$, il existe une suite $(B_n^k)_{n \geq 0}$ d'éléments de \mathcal{C} tels que

$$A_k \subset \bigcup_{n \geq 0} B_n^k \quad \text{et} \quad \sum_{n \geq 0} \mu(B_n^k) \leq 2^{-k} \varepsilon + \mu^*(A_k).$$

Il est clair que $\bigcup_{k \geq 0} A_k \subset \bigcup_{k \geq 0} \bigcup_{n \geq 0} B_n^k$ qui est une réunion dénombrable, donc par définition de μ^*

$$\mu^* \left(\bigcup_{k \geq 0} A_k \right) \leq \sum_{k \geq 0} \sum_{n \geq 0} \mu(B_n^k) \leq \sum_{k \geq 0} \varepsilon 2^{-k} + \sum_{k \geq 0} \mu^*(A_k) = 2\varepsilon + \sum_{k \geq 0} \mu^*(A_k),$$

et ce pour tout $\varepsilon > 0$, d'où la σ -sous-additivité de μ^* . Ceci achève la preuve du lemme. \square

Preuve de la proposition 2.2.28. Soient $A, B \subset \mathbb{X}$, alors $B = (B \cap A) \cup (B \cap A^{\complement})$, par σ -sous-additivité de la mesure extérieure μ^* , $\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \cap A^{\complement})$, d'où l'égalité d'ensembles annoncée.

Montrons que \mathcal{U} est une tribu. Clairement, $\emptyset \in \mathcal{U}$ et si $A \in \mathcal{U}$ alors $A^{\complement} \in \mathcal{U}$. Reste à montrer la stabilité par union dénombrable. On commence par la stabilité par réunion finie. Soient $A_0, A_1 \in \mathcal{U}$, alors

$$\begin{aligned} \mu^*(B) &= \mu^*(B \cap A_0) + \mu^*(B \cap A_0^{\complement}) \\ &= \mu^*(B \cap A_0 \cap A_1) + \mu^*(B \cap A_0 \cap A_1^{\complement}) + \mu^*(B \cap A_0^{\complement} \cap A_1) + \mu^*(B \cap A_0^{\complement} \cap A_1^{\complement}). \end{aligned}$$

On remarque que $A_0^{\complement} \cap A_1^{\complement} = (A_0 \cup A_1)^{\complement}$ et $A_0 \cup A_1 = (A_0 \cap A_1) \cup (A_0 \cap A_1^{\complement}) \cup (A_0^{\complement} \cap A_1)$. En utilisant la σ -sous-additivité de μ^* appliqué à $B \cap (A_0 \cup A_1) = (B \cap A_0) \cup (B \cap A_1)$, il vient que

$$\mu^*(B) \geq \mu^*(B \cap (A_0 \cup A_1)) + \mu^*(B \cap (A_0 \cup A_1)^{\complement}).$$

Ainsi $A_0 \cup A_1 \in \mathcal{U}$. Nous avons donc montré que \mathcal{U} est stable par passage au complémentaire et unions finies. Pour montrer la stabilité par union dénombrable, nous aurons besoin du lemme suivant.

Lemme 2.2.30. *Si C_0, \dots, C_n sont des éléments de \mathcal{U} deux à deux disjoints alors pour tout $B \subset \mathbb{X}$*

$$\mu^*(B) \geq \sum_{k=0}^n \mu^*(B \cap C_k).$$

Démonstration. Soient $C_0, C_1 \in \mathcal{U}$ tels que $C_0 \cap C_1 = \emptyset$ si bien que $B \cap C_1 \subset B \cap C_0^{\complement}$ et

$$\mu^*(B) = \mu^*(B \cap C_0) + \mu^*(B \cap C_0^{\complement}) \geq \mu^*(B \cap C_0) + \mu^*(B \cap C_1),$$

en utilisant la croissance de μ^* . Par récurrence, on obtient l'inégalité du lemme pour des familles finies. \square

Soit $(A_n)_{n \geq 0}$ une famille d'éléments de \mathcal{U} . On pose $A'_0 = A_0$ et, pour tout $n \geq 1$, $A'_n = A_n \setminus \bigcup_{k=0}^{n-1} A_k$. Comme \mathcal{U} est stable par passage au complémentaire et par réunion finie, $A'_n \in \mathcal{U}$ pour tout $n \geq 0$. De plus, les ensembles de la famille $(A'_n)_{n \geq 0}$ sont deux à deux disjoints et $\bigcup_{n \geq 0} A_n = \bigcup_{n \geq 0} A'_n$.

Considérons $B \subset \mathbb{X}$, puisque pour tout $n \geq 0$, $\bigcup_{k=0}^n A'_k \in \mathcal{U}$, il vient que

$$\mu^*(B) = \mu^* \left(B \cap \bigcup_{k=0}^n A'_k \right) + \mu^* \left(B \cap \left(\bigcup_{k=0}^n A'_k \right)^{\complement} \right).$$

Pour le premier terme, on applique la sous-additivité finie de la mesure extérieure μ^* que l'on a montré plus haut, alors que pour le second terme on utilise la propriété de croissance. D'où

$$\mu^*(B) \geq \sum_{k=0}^n \mu^*(B \cap A'_k) + \mu^* \left(B \cap \left(\bigcup_{p \geq 0} A'_p \right)^{\complement} \right).$$

En faisant tendre n vers l'infini, on obtient

$$\begin{aligned} \mu^*(B) &\geq \sum_{p \geq 0} \mu^*(B \cap A'_p) + \mu^* \left(B \cap \left(\bigcup_{p \geq 0} A'_p \right)^{\complement} \right) \\ &\geq \mu^* \left(B \cap \left(\bigcup_{p \geq 0} A'_p \right) \right) + \mu^* \left(B \cap \left(\bigcup_{p \geq 0} A'_p \right)^{\complement} \right), \end{aligned} \tag{2.4}$$

en utilisant la σ -sous-additivité de μ^* . Ainsi, $\bigcup_{k \geq 0} A_k = \bigcup_{k \geq 0} A'_k \in \mathcal{U}$.

Il reste à montrer que μ^* est σ -additive sur \mathcal{U} . Si la suite $(A_n)_{n \geq 0}$ considérée au-dessus est constituée d'éléments deux à deux disjoints alors en fait $A_n = A'_n$ pour tout $n \geq 0$. En posant $B = \bigcup_{n \geq 0} A'_n$ dans (2.4), on trouve

$$\mu^* \left(\bigcup_{n \geq 0} A'_n \right) \geq \sum_{n \geq 0} \mu^*(A'_n).$$

Comme l'inégalité inverse est toujours vérifiée (c'est la σ -sous-additivité), on en déduit que μ^* est une mesure sur \mathcal{U} . \square

Preuve de la proposition 2.2.29. Soient $A \in \mathcal{C}$, $B \in \mathcal{P}(\mathbb{X})$. Alors pour tout $\varepsilon > 0$ il existe $(B_n)_{n \geq 0}$ une famille d'éléments de \mathcal{C} telle que

$$\sum_{n \geq 0} \mu^*(B_n) \leq \mu^*(B) + \varepsilon.$$

Puisque pour tout $n \geq 0$, $B_n \in \mathcal{C}$ et $A \in \mathcal{C}$, $B_n \cap A \in \mathcal{C}$, et que, de plus, μ^* coïncident sur \mathcal{C} avec μ qui est finiment additive, on a

$$\mu^*(B) + \varepsilon \geq \sum_{n \geq 0} \mu^*(B_n) \geq \sum_{n \geq 0} \mu^*(B_n \cap A) + \mu^*(B_n \cap A^c) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c),$$

par σ -sous-additivité. Lorsque $\varepsilon \rightarrow 0$, on obtient que $A \in \mathcal{U}$. Ceci montre que \mathcal{U} contient \mathcal{C} et $\sigma(\mathcal{C})$. \square

Construction effective de la mesure de Lebesgue et mesures de Stieltjes

Afin de rendre l'exposé de la construction de la mesure de Lebesgue plus lisible, nous introduisons la notion plus faible de semi-algèbre.

Définition 2.2.31 (Semi-algèbre). Une famille \mathcal{S} de parties de \mathbb{X} est une semi-algèbre si

1. $\emptyset \in \mathcal{S}$,
2. pour tout $A, B \in \mathcal{S}$, $A \cap B \in \mathcal{S}$,
3. pour tout $A \in \mathcal{S}$, il existe $n \geq 1$ et $A_1, \dots, A_n \in \mathcal{S}$ deux à deux disjoints tels que $A^c = \bigcup_{i=1}^n A_i$.

Exemple 17. L'ensemble des intervalles de \mathbb{R} constitue une semi-algèbre.

Proposition 2.2.32. Soit \mathcal{S} une semi-algèbre.

1. L'ensemble $\left\{ \bigcup_{i=1}^n A_i, A_i \in \mathcal{S}, \text{ deux à deux disjoints}, n \geq 1 \right\}$, notée $\mathcal{C}(\mathcal{S})$, est la plus petite algèbre de Boole contenant \mathcal{S} .
2. Soit $\mu : \mathcal{S} \rightarrow \overline{\mathbb{R}}_+$ une application vérifiant $\mu(\emptyset) = 0$ et finiment additive au sens suivant

$$\forall A, B \in \mathcal{S} : A \cup B \in \mathcal{S} \text{ et } A \cap B = \emptyset \implies \mu(A \cup B) = \mu(A) + \mu(B).$$

Alors μ admet un unique prolongement $\bar{\mu}$ à $\mathcal{C}(\mathcal{S})$ vérifiant la propriété d'additivité finie au sens d'une mesure sur une algèbre de Boole — c.f. le deuxième point de la définition 2.2.21.

Démonstration. 1. Par stabilité par réunion finie, toute algèbre de Boole contenant \mathcal{S} contient $\mathcal{C}(\mathcal{S})$. Reste à montrer que $\mathcal{C}(\mathcal{S})$ est une algèbre de Boole.

- $\emptyset \in \mathcal{S} \subset \mathcal{C}(\mathcal{S})$;
- $\mathcal{C}(\mathcal{S})$ est stable par intersection finie car, d'une part,

$$\left(\bigcup_{i=1}^n A_i \right) \cap \left(\bigcup_{j=1}^m B_j \right) = \bigcup_{i,j} A_i \cap B_j,$$

et d'autre part, les $A_i \cap B_j$ sont deux à deux disjoints dès que les A_i , $1 \leq i \leq n$, respectivement les B_j , $1 \leq j \leq m$, sont deux à deux disjoints; la stabilité par union finie de (\mathcal{S}) se déduit par passage au complémentaire que l'on montre ci-dessous;

- si $A = \bigcup_{i=1}^n A_i$ avec $A_i \in \mathcal{S}$ alors par hypothèse, pour chaque $i \in \{1, \dots, n\}$, $A_i^c = \bigcup_{k=1}^{m(i)} B_k^{(i)}$ où les $B_k^{(i)}$ sont des parties de \mathcal{S} deux à deux disjointes. Quitte à ajouter des $B_k^{(i)} = \emptyset$, on peut remplacer les $m(i)$ par $m = \max\{m(i) : 1 \leq i \leq n\}$. D'où

$$A^c = \bigcap_{i=1}^n A_i^c = \bigcap_{i=1}^n \left(\bigcup_{k=1}^m B_k^{(i)} \right) = \bigcup_{1 \leq k_1, \dots, k_n \leq m} \underbrace{\left(\bigcap_{i=1}^n B_{k_i}^{(i)} \right)}_{\in \mathcal{S}}.$$

Ainsi, $A^c \in \mathcal{C}(\mathcal{S})$ car les ensembles $\bigcap_{i=1}^n B_{k_i}^{(i)}$, $1 \leq k_1, \dots, k_n \leq m$, sont deux à deux disjoints.

2. Pour tout $A = \bigcup_{i=1}^n A_i \in \mathcal{C}(\mathcal{S})$ on pose

$$\bar{\mu}(A) = \sum_{i=1}^n \mu(A_i).$$

Rappelant que les $A_i \in \mathcal{S}$ sont deux à deux disjoints, cette définition est consistante : si $A = \bigcup_{j=1}^m A'_j$ est une autre représentation de A alors les $A_i \cap A'_j$ étant deux à deux disjoints, on peut réécrire $A = \bigcup_{i,j} A_i \cap A'_j$ et donc par additivité de μ sur \mathcal{S}

$$\sum_{i=1}^n \mu(A_i) = \sum_{i,j} \mu(A_i \cap A'_j) = \sum_{j=1}^m \mu(A'_j).$$

Finalement, $\bar{\mu}$ étant complètement déterminée par les valeurs de μ sur \mathcal{S} , $\bar{\mu}$ est unique alors que sa propriété d'additivité finie est évidente par définition. \square

Puisque $\bar{\mu}$ est finiment additive, on a facilement pour tout $A, B \in \mathcal{C}(\mathcal{S})$ tels que $A \subset B$ que $\bar{\mu}(B) = \bar{\mu}(B \cap A) + \bar{\mu}(B \cap A^c) \geq \mu(A)$, i.e. $\bar{\mu}$ est croissante. De même, $\bar{\mu}$ satisfait une formule du crible : pour tout $A, B \in \mathcal{C}(\mathcal{S})$, $\bar{\mu}(A \cup B) + \bar{\mu}(A \cap B) = \bar{\mu}(A) + \bar{\mu}(B)$. Enfin, $\bar{\mu}$ est finiment sous-additive :

$$\forall A_1, \dots, A_n \in \mathcal{C}(\mathcal{S}), \quad \bar{\mu}(A_1 \cup \dots \cup A_n) \leq \bar{\mu}(A_1) + \dots + \bar{\mu}(A_n).$$

Théorème 2.2.33 (Stieltjes). *Soit $F : \mathbb{R} \rightarrow \mathbb{R}$ une fonction croissante continue à droite. Il existe une unique mesure μ_F sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ appelée mesure de Stieltjes associée à F vérifiant*

$$\forall a, b \in \mathbb{R}, \quad \mu_F((a, b]) = F(b) - F(a).$$

La mesure de Lebesgue n'est rien d'autre que la mesure de Stieltjes associée à la fonction continue croissante $x \rightarrow x$. C'est alors un exercice de montrer le théorème 2.2.12.

Démonstration. Étape 1 :

On pose $\mathcal{S} = \{(a, b], (a, \infty), -\infty \leq a \leq b < \infty\}$. On vérifie facilement que \mathcal{S} est une semi-algèbre. On définit sur \mathcal{S} l'application $\ell = \ell_F$ par

$$\ell((a, b]) = F(b) - F(a), \quad \text{et} \quad \ell((a, \infty)) = F(+\infty) - F(a),$$

où $F(+\infty) = \lim_{t \rightarrow \infty} F(t)$ qui existe dans $\mathbb{R} \cup \{+\infty\}$ par croissance de F . On vérifie facilement que ℓ est finiment additive sur \mathcal{S} :

$$\ell((a, b] \cup (b, c]) = \ell((a, c]) = F(c) - F(a) = [F(c) - F(b)] + [F(b) - F(a)] = \ell((a, b]) + \ell((b, c]),$$

et de même pour $(a, b]$ et (b, ∞) — notons que la contrainte selon laquelle la réunion doit être dans la semi-algèbre réduit drastiquement les cas à vérifier. D'après la proposition 2.2.32, ℓ admet un unique prolongement $\bar{\ell}$ sur l'algèbre

$$\mathcal{C}(\mathcal{S}) = \{I_1 \cup \dots \cup I_n, I_k \in \mathcal{S}, n \geq 1\}$$

qui soit finiment additive sur l'algèbre de Boole. Autrement dit, afin d'appliquer le théorème 2.2.24 d'extension de Carathéodory, il est nécessaire de montrer que $\bar{\ell}$ satisfait le troisième axiome d'une mesure sur une algèbre de Boole — c.f. la définition 2.2.21 — et qu'elle est σ -finie. Il suffira alors de remarquer que $\mathcal{B}(\mathbb{R}) = \sigma(\{(a, \infty), a \in \mathbb{R}\})$ si bien que $\mathcal{B}(\mathbb{R}) \subset \sigma(\mathcal{S})$ et donc que $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C}(\mathcal{S}))$ pour conclure la preuve du théorème.

Étape 2 : on suppose que $F(\pm\infty) = \pm\infty$.

- Comme $F(\pm\infty) = \pm\infty$, il vient que, pour tout $A \in \mathcal{S}$, A est borné dans \mathbb{R} si et seulement si $\ell(A) < \infty$. Soit donc $(A_n)_{n \geq 0}$ une suite décroissante d'éléments de $\mathcal{C}(\mathcal{S})$ satisfaisant $\bar{\ell}(A_0) < \infty$ et $\bigcap_{n \geq 0} A_n = \emptyset$. Alors A_0 est borné comme réunion finie d'intervalles bornés. Pour tout $n \geq 0$, on écrit

$$A_n = I_1^{(n)} \cup \dots \cup I_{p_n}^{(n)}, \quad \text{où} \quad I_k^{(n)} = (\alpha_k^{(n)}, \beta_k^{(n)}] \quad (2.5)$$

sont deux à deux disjoints.

Si pour un certain n , $A_n = \emptyset$, le résultat est évident puisque $\bar{\ell}(\emptyset) = 0$. Sinon, tous les p_n sont non nuls, c'est à dire les réunions dans (2.5) contiennent au moins un éléments. Soit $\varepsilon > 0$, on construit pour tout n et tout k un intervalle $J_k^{(n)} = (\tilde{\alpha}_k^{(n)}, \beta_k^{(n)}]$ de façon que, d'un côté $\overline{J_k^{(n)}} \subset I_k^{(n)}$ et de l'autre

$$F(\tilde{\alpha}_k^{(n)}) - F(\alpha_k^{(n)}) \leq \frac{\varepsilon}{p_n 2^n}.$$

Ceci est toujours possible par continuité à droite de F . Chaque intervalle, éventuellement vide, $\overline{J_k^{(n)}}$ est compact. On pose alors pour tout $n \geq 0$, $A'_n = \bigcup_{k=0}^{p_n} J_k^{(n)}$ et il vient immédiatement que $A'_n \in \mathcal{C}(\mathcal{S})$, $A'_n \subset A_n$ et

$$\bar{\ell}(A_n \setminus A'_n) \leq \sum_{k=0}^{p_n} \frac{\varepsilon}{p_n 2^n} = \frac{\varepsilon}{2^n}.$$

Par construction, les $\overline{A'_n} = \bigcup_{k=0}^{p_n} \overline{J_k^{(n)}}$ sont compacts donc fermés dans le compact $\overline{A_0}$ et $\bigcap \overline{A'_n} \subset \bigcap A_n = \emptyset$. Par la proposition 1.2.62 du chapitre 1, on peut donc exhiber un entier $n_\varepsilon \geq 0$ tel que $\bigcap_{k=0}^{n_\varepsilon} A'_k \subset \bigcap_{k=0}^{n_\varepsilon} \overline{A'_k} = \emptyset$. Finalement,

$$\bar{\ell}(A_{n_\varepsilon}) = \bar{\ell} \left[\left(\bigcap_{k=0}^{n_\varepsilon} A_k \right) \setminus \left(\bigcap_{k=0}^{n_\varepsilon} A'_k \right) \right] \leq \sum_{k=0}^{n_\varepsilon} \bar{\ell}(A_k \setminus A'_k) \leq \varepsilon.$$

Ainsi, pour tout $n \geq n_\varepsilon$, $\bar{\ell}(A_{n_\varepsilon}) \leq \varepsilon$. Ceci termine la preuve de la convergence voulue.

- On montre maintenant que $\bar{\ell}$ est σ -finie. On pose pour cela $E_n = (-n, n]$. La suite $(E_n)_{n \geq 0}$ est croissante et exhaustive. Soit $A \in \mathcal{C}(\mathcal{S})$. De deux chose l'une, soit A est borné et pour tout n assez grand $A \subset E_n$ si bien que $\lim_{n \rightarrow \infty} \bar{\ell}(A \cap E_n) = \bar{\ell}(A)$; ou bien A est non borné et donc l'un des intervalles constituant A n'est pas borné, il suffit donc de vérifier la σ -finitude de $\bar{\ell}$ pour ces intervalles non bornés. Or, pour tout n assez grand $\bar{\ell}((a, \infty) \cap E_n) = \bar{\ell}((a, n]) = F(n) - F(a)$ qui tend vers $F(\infty)$, c'est à dire $\bar{\ell}((a, \infty))$.

Étape 3 :

Supposons que $F(\infty) \in \mathbb{R}$ et $F(-\infty) = -\infty$. Soit $\varepsilon > 0$, alors il existe un réel L_ε tel $F(\infty) - F(L_\varepsilon) \leq \varepsilon$. Alors, $A_n \cap (-\infty, L_\varepsilon]$ est suite décroissante (en n) d'éléments de $\mathcal{C}(\mathcal{S})$ bornés, d'intersection vide. D'autre part,

$$\bar{\ell}(A_n) = \bar{\ell}(A_n \cap (-\infty, L_\varepsilon]) + \bar{\ell}(A_n \cap (L_\varepsilon, \infty)).$$

D'après l'étape précédente,

$$\limsup_{n \rightarrow \infty} \bar{\ell}(A_n) \leq \bar{\ell}((L_\varepsilon, \infty)) \leq \varepsilon.$$

Ceci achève la preuve de la convergence voulue. La σ -finitude se montre de façon analogue. Les cas $F(-\infty) \in \mathbb{R}$ et $F(\infty) = \infty$ ainsi que $F(\pm\infty) \in \mathbb{R}$ se montrent également de manière similaire \square

2.2.4 Régularité des mesures, mesures de Borel et espaces polonais

Cette section fait une usage intensive de notions topologiques. On se reporte au chapitre 1 pour une introduction à toutes ces notions.

Définition 2.2.34 (Régularité d'une mesure). Soient (\mathbb{X}, d) un espace métrique et μ une mesure sur $\mathcal{B}(\mathbb{X})$. Alors μ est dite

1. extérieurement régulière si

$$\forall A \in \mathcal{B}(\mathbb{X}), \quad \mu(A) = \inf\{\mu(O), O \text{ ouvert}, A \subset O\};$$

2. intérieurement régulière si

$$\forall A \in \mathcal{B}(\mathbb{X}), \quad \mu(A) = \sup\{\mu(K), K \text{ compact}, K \subset A\};$$

3. régulière si elle est à la fois extérieurement et intérieurement régulière.

La proposition suivante, si elle ne conclut pas exactement à la régularité des mesures finie, donne un résultat proche.

Proposition 2.2.35. *Soit μ une mesure finie sur la tribu de Borel $\mathcal{B}(\mathbb{X})$ d'un espace métrique (\mathbb{X}, d) . Alors, pour tout $A \in \mathcal{B}(\mathbb{X})$ et tout $\varepsilon > 0$, il existe un ouvert O et un fermé F tels que*

$$F \subset A \subset O \quad \text{et} \quad \mu(O \setminus F) < \varepsilon.$$

Démonstration. On va montrer que

$$\mathcal{T} = \{A \in \mathcal{B}(\mathbb{X}) : \forall \varepsilon > 0, \exists O \text{ ouvert et } F \text{ fermé t.q. } F \subset A \subset O \text{ et } \mu(O \setminus F) < \varepsilon\}$$

est une tribu qui contient les ouverts et par conséquent la tribu borélienne.

Soit A un ouvert de \mathbb{X} et $\varepsilon > 0$. On pose $O = A$ et pour tout $\delta > 0$, $F_\delta = \{x \in \mathbb{X} : d(x, A^c) \geq \delta\}$. Puisque la fonction $x \rightarrow d(x, A^c)$ est continue, l'ensemble F_δ est fermé. On considère alors la réunion croissante

$$\bigcup_{p \geq 1} F_{1/p} = \{x \in \mathbb{X} : d(x, A^c) > 0\}.$$

Comme A est ouvert, la réunion contient A (c'est même la définition), mais comme F_δ est trivialement contenu dans A , la réunion est exactement égale à A .

Par continuité des mesures positives,

$$\mu(O) = \mu(A) = \mu\left(\bigcup_{p \geq 1} F_{1/p}\right) = \lim_{p \rightarrow \infty} \mu(F_{1/p}).$$

Comme μ est finie, $\lim_{p \rightarrow \infty} \mu(O \setminus F_{1/p}) = 0$ donc pour p assez grand, nous avons $\mu(O \setminus F_{1/p}) < \varepsilon$. Et de plus $F_{1/p} \subset A \subset O$. On a donc montré que \mathcal{T} contient les ouverts de \mathbb{X} .

Vérifions que \mathcal{T} est une tribu. Il est évident que \emptyset , qui est à la fois ouvert et fermé, est un élément de \mathcal{T} . De plus, si $A \in \mathcal{T}$, alors pour tout $\varepsilon > 0$, on peut trouver F fermé, O ouvert tel que $F \subset A \subset O$ et $\mu(O \setminus F) < \varepsilon$. Mais alors O^c est fermé, F^c est ouvert, $O^c \subset A^c \subset F^c$ et, comme $O \setminus F = F^c \setminus O^c$, on obtient $\mu(F^c \setminus O^c) < \varepsilon$.

Il reste donc à vérifier la stabilité par union dénombrable. Soient $(A_n)_{n \geq 1} \in \mathcal{T}^{\mathbb{N}^*}$ et $\varepsilon > 0$. Alors pour tout $n \geq 1$, on peut trouver un fermé $F_n \subset A_n$ et un ouvert $O_n \supset A_n$ tels que $\mu(O_n \setminus F_n) < \varepsilon 2^{-n-1}$. Or,

$$\bigcup_{n \geq 1} F_n \subset \bigcup_{n \geq 1} A_n \subset \bigcup_{n \geq 1} O_n,$$

et comme

$$\left(\bigcup_{n \geq 1} O_n\right) \setminus \left(\bigcup_{n \geq 1} F_n\right) = \left(\bigcup_{n \geq 1} O_n\right) \cap \left(\bigcup_{m \geq 1} F_m\right)^c = \left(\bigcup_{n \geq 1} \bigcap_{m \geq 1} O_n \cap F_m^c\right) \subset \bigcup_{n \geq 1} (O_n \setminus F_n),$$

il vient par σ -sous-additivité de μ

$$\mu\left(\left(\bigcup_{n \geq 1} O_n\right) \setminus \left(\bigcup_{n \geq 1} F_n\right)\right) \leq \sum_{n \geq 1} \mu(O_n \setminus F_n) = \varepsilon/2.$$

D'autre part, comme $\bigcup_{n \geq 1} F_n = \bigcup_{n \geq 1} \bigcup_{k=1}^n F_k$, il existe n_ε tel que, par continuité de la mesure μ ,

$$\mu\left(\bigcup_{n \geq 1} F_n\right) \leq \mu\left(\bigcup_{k=1}^{n_\varepsilon} F_k\right) + \varepsilon/2.$$

Posons alors, $O = \bigcup_{n \geq 1} O_n$ et $F = \bigcup_{k=1}^{n_\varepsilon} F_k$. Alors, $O \subset A$ est ouvert et $F \subset A$ est fermé. De plus, comme μ est finie

$$\mu(O \setminus F) = \mu(O) - \mu(F) = \mu\left(\bigcup_{n \geq 1} O_n\right) - \mu\left(\bigcup_{n \geq 1} F_n\right) + \mu\left(\bigcup_{n \geq 1} F_n\right) - \mu(F) \leq \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

□

Théorème 2.2.36. Soit μ une mesure σ -finie sur la tribu de Borel $\mathcal{B}(\mathbb{X})$ d'un espace métrique (\mathbb{X}, d) . Alors,

$$\forall A \in \mathcal{B}(\mathbb{X}), \quad \mu(A) = \sup\{\mu(F), F \text{ fermé}, F \subset A\}.$$

Si, de plus, $\mathbb{X} = \bigcup_{n \geq 1} \text{Int } E_n$ pour une famille croissante de boréliens $(E_n)_{n \geq 1}$ telle que $\mu(E_n) < \infty$ pour tout $n \geq 1$, alors la mesure est extérieurement régulière. Enfin, si l'on peut choisir les boréliens $(E_n)_{n \geq 1}$ compacts, la mesure μ est intérieurement régulière.

Exemple 18. La mesure de Lebesgue est régulière.

Démonstration. On montre les trois points dans l'ordre.

1. Supposons d'abord $\mu(A) < \infty$. Soit $\varepsilon > 0$. Nous pouvons réécrire sous la forme d'une réunion croissante $A = \bigcup_{n \geq 1} A \cap E_n$. Par continuité de μ , on peut trouver $n_\varepsilon \geq 1$ tel que $\mu(A) = \mu(A \cap E_{n_\varepsilon}) + \varepsilon/2$ (c'est à dire $\mu(A \cap E_{n_\varepsilon}^c) \leq \varepsilon/2$). On pose $\tilde{\mu}(\cdot) = \mu(\cdot \cap E_{n_\varepsilon})$. On se retrouve dans le cas d'une mesure $\tilde{\mu}$ finie, on peut appliquer la proposition précédente : il existe un fermé $F \subset A$ tel que $\tilde{\mu}(A \setminus F) < \varepsilon/2$. Ainsi,

$$\mu(A \setminus F) = \mu((A \setminus F) \cap E_{n_\varepsilon}) + \mu((A \setminus F) \cap E_{n_\varepsilon}^c) \leq \tilde{\mu}(A \setminus F) + \mu(A \cap E_{n_\varepsilon}^c) \leq \varepsilon.$$

On considère le cas $\mu(A) = \infty$. Toujours par continuité à gauche de μ , $\mu(A) = \lim_{n \rightarrow \infty} \mu(A \cap E_n)$, or d'après le cas $\mu(A) < \infty$, nous avons

$$\mu(A \cap E_n) = \sup\{\mu(F), F \subset A \cap E_n, F \text{ fermé}\} \leq \sup\{\mu(F), F \subset A, F \text{ fermé}\}.$$

D'où $\mu(A) \leq \sup\{\mu(F), F \subset A, F \text{ fermé}\}$. L'autre inégalité est immédiate.

2. On pose pour tout $n \geq 1$, $\mu_n = \mu_n(\cdot \cap E_n)$. Soit $A \in \mathcal{B}(\mathbb{X})$ et $\varepsilon > 0$. D'après la proposition précédente, il existe donc, pour tout $n \geq 1$, un ouvert O_n tel que $A \subset O_n$ et $\mu(O_n \setminus A) \leq \varepsilon 2^{-n}$, soit encore

$$A \subset O_n \quad \text{et} \quad \mu(O_n \cap E_n) \leq \mu(A \cap E_n) + \varepsilon 2^{-n}.$$

On montre la propriété P_n suivante par récurrence

$$P_n : \quad \mu \left(\bigcup_{k=1}^n (O_k \cap E_k) \right) \leq \mu(A \cap E_n) + \sum_{k=1}^n \frac{\varepsilon}{2^k}.$$

La propriété P_1 est immédiate avec ce qui a été fait un peu plus haut. On suppose P_n et on montre P_{n+1} . Comme $\mu \left(\bigcup_{k=1}^{n+1} (O_k \cap E_k) \right)$ est fini, on peut utiliser la formule de Poincaré (ou formule du crible), on obtient

$$\begin{aligned} \mu \left(\bigcup_{k=1}^{n+1} (O_k \cap E_k) \right) &= \mu(O_{n+1} \cap E_{n+1}) + \mu \left(\bigcup_{k=1}^n (O_k \cap E_k) \right) \\ &\quad - \mu \left((O_{n+1} \cap E_{n+1}) \cap \bigcup_{k=1}^n (O_k \cap E_k) \right). \end{aligned}$$

Puis, par hypothèse de récurrence,

$$\begin{aligned} \mu \left(\bigcup_{k=1}^{n+1} (O_k \cap E_k) \right) &\leq \mu(A \cap E_{n+1}) + \varepsilon 2^{-n-1} + \mu(A \cap E_n) + \sum_{k=1}^n \varepsilon 2^{-k} \\ &\quad - \mu \left((O_{n+1} \cap E_{n+1}) \cap \bigcup_{k=1}^n (O_k \cap E_k) \right). \end{aligned}$$

Or, on observe que

$$A \cap E_n \subset \bigcup_{k=1}^n (A \cap E_k) \subset \bigcup_{k=1}^n (O_k \cap E_k) \quad \text{et} \quad A \cap E_n \subset A \cap E_{n+1} \subset O_{n+1} \cap E_{n+1},$$

d'où

$$\mu(A \cap E_n) \leq \mu \left((O_{n+1} \cap E_{n+1}) \cap \bigcup_{k=1}^n (O_k \cap E_k) \right) < \infty,$$

ce qui montre P_{n+1} .

L'ouvert $\bigcup_{n \geq 1} (O_n \cap \text{Int } E_n)$ étant contenu dans $\bigcup_{n \geq 1} (O_n \cap E_n)$, le passage à la limite dans l'inégalité P_n implique

$$\mu \left(\bigcup_{n \geq 1} (O_n \cap \text{Int } E_n) \right) \leq \mu \left(\bigcup_{n \geq 1} (O_n \cap E_n) \right) \leq \mu(A) + \varepsilon.$$

Il reste à montrer que $\bigcup_{n \geq 1} (O_n \cap \text{Int } E_n)$ contient A . Ceci découle du fait que $\mathbb{X} = \bigcup_{n \geq 1} \text{Int } E_n$ puisque

$$A = \bigcup_{n \geq 1} (A \cap \text{Int } E_n) \subset \bigcup_{n \geq 1} (O_n \cap \text{Int } E_n).$$

3. On remarque que les ensembles $F \cap E_n$ sont compacts comme fermés dans des compacts et que $\mu(F) = \sup_{n \geq 1} \mu(F \cap E_n)$. □

Définition 2.2.37 (Mesure de Borel). Une mesure μ sur la tribu borélienne d'un espace métrique est appelée mesure de Borel si elle est finie sur les parties compacts.

Définition 2.2.38 (Espace localement compact). Un espace métrique (\mathbb{X}, d) est dit localement compact si tout point $x \in \mathbb{X}$ admet un voisinage compact, *i.e.* il existe un compact K tel que $x \in \text{Int } K$.

Théorème 2.2.39. *Sur un espace métrique localement compact et séparable, toute mesure de Borel μ est régulière.*

Remarque 31. Ce qu'on montre en substance c'est que dans un espace métrique séparable on peut trouver une suite exhaustive croissante d'ouverts relativement compact.

Démonstration. Par le théorème 2.2.36 précédant, il s'agit de construire une suite croissante $(L_n)_{n \geq 1}$ de compacts telle que $\mathbb{X} = \bigcup_{n \geq 1} \text{Int } L_n$.

Étape 1 : Soit (\mathbb{X}, d) un espace métrique localement compact et séparable, montrons l'existence d'une suite croissante exhaustive de compacts.

Comme (\mathbb{X}, d) est séparable, il existe une suite $(x_n)_{n \geq 0}$ dense dans \mathbb{X} . Soit $I = \{(n, r) \in \mathbb{N}^* \times \mathbb{Q}_+^* : \overline{B}(x_n, r) \text{ compact}\}$. Comme I est au plus dénombrable, on peut trouver une suite croissante d'ensembles finis I_p , $p \geq 1$, telle que $I = \bigcup_{p \geq 1} I_p$.

Soit $x \in \mathbb{X}$, alors x admet un voisinage compact K_x . Donc il existe $n \in \mathbb{N}^*$ et $r \in \mathbb{Q}_+^*$ tel que $x \in \overline{B}(x_n, r) \subset \text{Int } K_x$. Par conséquent, $\mathbb{X} = \bigcup_{(n,r) \in I} \overline{B}(x_n, r)$. On pose alors $K_p = \bigcup_{(n,r) \in I_p} \overline{B}(x_n, r)$. Il est immédiat que $\mathbb{X} = \bigcup_{p \geq 1} K_p$ et que les compacts K_p sont compacts comme réunion finie de compacts. Ceci achève l'étape 1.

Étape 2 : Construction des L_n par récurrence.

On pose $L_1 = K_1$ puis l'on suppose construits des compacts L_1, \dots, L_n tels que $K_k \subset L_k$, $1 \leq k \leq n$ et $L_{k-1} \subset \text{Int } L_k$, $2 \leq k \leq n$. L'ensemble $K_{n+1} \cup L_n$ est compact et, par locale compacité de \mathbb{X} , tout $x \in K_{n+1} \cup L_n$ a un voisinage compact V_x . Or, $x \in \text{Int } V_x$ par hypothèse donc la famille $(\text{Int } V_x)_{x \in K_{n+1} \cup L_n}$ est un recouvrement ouvert du compact $K_{n+1} \cup L_n$ dont on peut extraire un recouvrement fini $\text{Int } V_{x_1} \cup \dots \cup \text{Int } V_{x_p}$. On pose alors $L_{n+1} = V_{x_1} \cup \dots \cup V_{x_p}$. L'ensemble L_{n+1} ainsi construit est compact comme réunion finie de compacts, $K_{n+1} \subset L_{n+1}$ et $L_n \subset \text{Int } V_{x_1} \cup \dots \cup \text{Int } V_{x_p} \subset \text{Int } L_{n+1}$.

La suite de compacts ainsi construite vérifie finalement

$$\mathbb{X} = \bigcup_{n \geq 1} K_n \subset \bigcup_{n \geq 1} L_n \subset \bigcup_{n \geq 1} \text{Int } L_{n+1} \subset \bigcup_{n \geq 1} \text{Int } L_n \subset \mathbb{X}.$$

□

Théorème 2.2.40. Soit (\mathbb{X}, d) un espace polonais. Toute mesure μ finie sur $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ vérifie

1. pour tout $\varepsilon > 0$, il existe $K_\varepsilon \subset \mathbb{X}$ compact tel que $\mu(K_\varepsilon^c) \leq \varepsilon$.
2. μ est régulière.

Démonstration. On montre d'abord le point (i). Soient $(x_n)_{n \geq 0}$ une suite dense et $\varepsilon > 0$. Pour tout $p \geq 1$, il existe $n_p \in \mathbb{N}^*$ tel que

$$\mu \left(\left(\bigcup_{n=1}^{n_p} B(x_n, 1/p) \right)^c \right) \leq \varepsilon/2^p \quad \text{car} \quad \mathbb{X} = \bigcup_{n \geq 1} B(x_n, 1/p) \quad \text{et} \quad \mu(\mathbb{X}) < \infty.$$

On pose alors

$$K_\varepsilon = \overline{\bigcap_{p \geq 1} \bigcup_{n \leq n_p} B(x_n, 1/p)}.$$

Aussi, $K_\varepsilon \subset \overline{\bigcup_{n \leq n_p} B(x_n, 1/p)}$ pour tout $p \geq 1$, ainsi K_ε est pré-compacte. Or, K_ε est fermé dans un espace complet, il est lui-même complet. Donc, K_ε est compact. D'autre part,

$$\mu(K_\varepsilon) \leq \mu \left(\bigcup_{p \geq 1} \left(\bigcup_{n \leq n_p} B(x_n, 1/p) \right)^c \right) \leq \sum_{p \geq 1} \varepsilon/2^p = \varepsilon.$$

Le deuxième point de la proposition est une conséquence directe de la proposition 2.2.35 et du fait que $F \cap K$ est compact dès que F est fermé et K compact. \square

Chapitre 3

Intégrale au sens de Lebesgue

Dans ce chapitre est définie l'intégrale de Lebesgue contre une mesure sur un espace mesurable abstrait. Après avoir donné les propriétés essentielles de l'intégrale de Lebesgue, on s'attachera à donner des méthodes pratiques de calcul. On considérera notamment le cas des mesures discrètes et des mesures à densité. Il sera également évoqué le comportement de l'intégrale lorsque l'on transporte une mesure. Enfin, on étudiera le lien entre intégrale de Riemann et intégrale de Lebesgue.

3.1 Construction de l'intégrale de Lebesgue

La construction de l'intégrale de Lebesgue se fait en trois étapes. Tout d'abord nous la définissons pour les fonctions étagées positives, puis pour les fonctions positives en utilisant l'approximation monotone des fonctions positives par des fonctions étagées positives. Notons que la valeur de l'intégrale d'une fonction positive peut valoir l'infini, on dit qu'elle est à valeurs dans $\overline{\mathbb{R}}_+$. Toutes les propriétés usuelles telles que la linéarité ou la croissance de l'intégrale restent vraies pour l'intégrale des fonctions positives qui peut prendre des valeurs infinies ! La dernière étape consiste à définir l'intégrale pour des fonctions réelles en les écrivant comme la différence de leur partie négative et partie positive. Dans ce cas, il est nécessaire de faire une hypothèse d'intégrabilité car, contrairement au cas des fonctions positives, certaines formes indéterminées de type $+\infty - \infty$ peuvent apparaître. Enfin, les fonctions à valeurs complexes ou plus généralement à valeurs dans \mathbb{K}^n seront traitées.

3.1.1 Intégration des fonctions étagées positives

Définition 3.1.1. Soit f une fonction étagée positive prenant les valeurs distinctes $\alpha_1, \dots, \alpha_n$. On note $A_i = f^{-1}(\{\alpha_i\})$ pour tout $i = 1, \dots, n$. On appelle intégrale de f contre la mesure μ , et on note $\int f d\mu$, le nombre dans $\overline{\mathbb{R}}_+$ défini par

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

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

Proposition 3.1.2. L'intégrale de fonctions étagées positives vérifie les propriétés suivantes.

1. Si f et g sont deux fonctions étagées positives et $\lambda > 0$, alors

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

2. Si f et g sont deux fonctions étagées positives telles que $f \leq g$, alors

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

Démonstration. On montre le point (i) lorsque $\lambda = 1$. Le cas général s'en déduit immédiatement. On pose

$$f = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i} \quad \text{et} \quad g = \sum_{j=1}^m \beta_j \mathbf{1}_{B_j}$$

où $\alpha_1, \dots, \alpha_n$ (resp. β_1, \dots, β_m) sont distincts et les A_1, \dots, A_n (resp. B_1, \dots, B_m) sont des ensembles mesurables disjoints. On note $\gamma_1, \dots, \gamma_\ell$ les valeurs distinctes prises par $f + g$ et

$$C_k = (f + g)^{-1}(\gamma_k) = \bigcup_{(i,j) \in I_k} (A_i \cap B_j),$$

où $I_k = \{(i, j), \alpha_i + \beta_j = \gamma_k\}$. Puisque les ensembles $A_i \cap B_j$, $i = 1, \dots, n$, $j = 1, \dots, m$, sont deux à deux disjoints,

$$\mu(C_k) = \sum_{(i,j) \in I_k} \mu(A_i \cap B_j).$$

On calcule l'intégrale de $f + g$

$$\begin{aligned} \int f + g \, d\mu &= \sum_{k=1}^{\ell} \gamma_k \mu(C_k) = \sum_{k=1}^{\ell} \sum_{(i,j) \in I_k} (\alpha_i + \beta_j) \mu(A_i \cap B_j) \\ &= \sum_{i=1}^n \sum_{j=1}^m \alpha_i \mu(A_i \cap B_j) + \sum_{j=1}^m \sum_{i=1}^n \beta_j \mu(A_i \cap B_j) \\ &= \sum_{i=1}^n \alpha_i \mu(A_i) + \sum_{j=1}^m \beta_j \mu(B_j) \\ &= \int f \, d\mu + \int g \, d\mu. \end{aligned}$$

Pour le point (ii), on remarque que $g - f$ est une fonction étagée positive, son intégrale est positive, d'où, en utilisant le point (i)

$$\int f \, d\mu \leq \int f \, d\mu + \int g - f \, d\mu = \int f + g - f \, d\mu = \int g \, d\mu.$$

□

Remarque 32. Soit $f = \sum_i \alpha_i \mathbf{1}_{A_i}$ où les α_i ne sont pas nécessairement distincts — mais les A_i tout de même deux à deux disjoints. On a encore $\int f \, d\mu = \sum_i \alpha_i \mu(A_i)$.

3.1.2 Intégration des fonctions mesurables positives

Définition 3.1.3. Soit f une fonction mesurable à valeurs dans $\overline{\mathbb{R}}_+$. On appelle intégrale de f contre μ , et on note $\int f \, d\mu$ l'élément de $\overline{\mathbb{R}}_+$ défini par

$$\int f \, d\mu = \sup \left\{ \int u \, d\mu, \quad u \in m\mathcal{E}_+ : u \leq f \right\},$$

où $m\mathcal{E}_+$ désigne l'ensemble des fonctions étagées positives.

Remarque 33. Cette définition est consistante avec celle de l'intégrale d'une fonction étagées positives. Dans ce cas, le supremum est un maximum et on choisit $u = f$.

Proposition 3.1.4 (Croissance de l'intégrale). *Soient f, g des fonctions mesurables positives telles que $f \leq g$, alors*

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

Démonstration. C'est une conséquence immédiate de l'inclusion

$$\{u \in m\mathcal{E}_+ : u \leq f\} \subset \{u \in m\mathcal{E}_+ : u \leq g\}$$

et de la définition de l'intégrale. □

Théorème 3.1.5 (Théorème de convergence monotone de Beppo-Lévy). Soit $(f_n)_{n \geq 0}$ une suite monotone croissante de fonctions mesurables positives, i.e. $0 \leq f_n \leq f_{n+1}$ pour tout $n \geq 0$. Alors $f = \lim_{n \rightarrow \infty} f_n = \sup_{n \geq 0} f_n$ est mesurable positive et

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

Démonstration. D'après la proposition 2.1.25, on sait que la fonction supremum est mesurable. Comme $f_n \leq f$, on a $\int f_n d\mu \leq \int f d\mu$. La croissance de l'intégrale assure que la suite $(\int f_n d\mu)_{n \geq 0}$ est elle-même croissante et donc convergente dans $\overline{\mathbb{R}}_+$. On obtient donc

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

Démontrons l'inégalité opposée. Soit u une fonction étagée positive inférieure à f et $\lambda \in (0, 1)$. Posons,

$$E_n = \{x \in \mathbb{X} : f_n(x) \geq \lambda u(x)\}.$$

La suite $(E_n)_{n \geq 0}$ est une suite croissante d'ensembles mesurables. Soit $x \in \mathbb{X}$. Si $u(x) = 0$ alors $x \in E_n$ pour tout $n \geq 0$. Si $u(x) > 0$ alors

$$\lim_{n \rightarrow \infty} f_n(x) = f(x) \geq u(x) > \lambda u(x),$$

et ainsi $x \in E_n$ pour $n \geq 0$ assez grand de sorte que $\cup_{n \geq 0} E_n = \mathbb{X}$. D'autre part, par définition de E_n , $f_n \geq \lambda u \mathbf{1}_{E_n}$ et donc pour tout $n \geq 0$, par croissance de l'intégrale

$$\int f_n d\mu \geq \int \lambda u \mathbf{1}_{E_n} d\mu.$$

La fonction $\lambda u \mathbf{1}_{E_n}$ est étagée positive, on sait calculer son intégrale. Si $u = \sum_{i=1}^k \alpha_i \mathbf{1}_{A_i}$ alors

$$\int u d\mu = \sum_{i=1}^k \alpha_i \mu(A_i) \quad \text{et} \quad \int u \mathbf{1}_{E_n} d\mu = \sum_{i=1}^k \alpha_i \mu(A_i \cap E_n).$$

Or pour tout $i = 1, \dots, k$, $\mu(A_i \cap E_n)$ converge en croissant vers $\mu(A_i)$ donc $\int u \mathbf{1}_{E_n} d\mu$ converge vers $\int u d\mu$. On a donc établi que, pour tout $u \in m\mathcal{E}_+$ telle que $u \leq f$ et tout $\lambda \in (0, 1)$,

$$\lim_{n \rightarrow \infty} \int f_n d\mu \geq \lim_{n \rightarrow \infty} \lambda \int u \mathbf{1}_{E_n} d\mu = \lambda \int u d\mu.$$

En prenant le supremum sur $\lambda \in (0, 1)$, on obtient que l'intégrale de toute fonction étagée positive u majorée par f est inférieure à la limite des intégrales des fonctions f_n . Il en va de même pour l'intégrale de f :

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

ce qui est l'inégalité recherchée. □

Corollaire 3.1.6. Si $f, g \in m\mathcal{X}_+$, alors $\int (f + g) d\mu = \int f d\mu + \int g d\mu$.

Démonstration. D'après le théorème 2.1.27, il existe des suites $(f_n)_{n \geq 0}$ et $(g_n)_{n \geq 0}$ croissantes de fonctions étagées positives qui convergent simplement vers f et g respectivement. Alors $(f_n + g_n)_{n \geq 0}$ est une suite croissante de fonctions étagées positives qui converge simplement vers $f + g$. La linéarité de l'intégrale pour les fonctions étagées assure alors pour tout $n \geq 0$

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

Le théorème de convergence monotone permet de conclure. □

Corollaire 3.1.7. Soit $(f_n)_{n \geq 0}$ une suite de fonctions mesurables positives. Alors, l'égalité suivante a lieu dans $\overline{\mathbb{R}}_+$

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

Démonstration. Immédiat. □

3.1.3 Intégration des fonctions mesurables

Définition 3.1.8. Une application f de $(\mathbb{X}, \mathcal{X}, \mu)$ à valeurs dans \mathbb{K} est dite intégrable contre μ si elle est mesurable et $\int |f| d\mu < \infty$.

On notera $\mathcal{L}_{\mathbb{K}}^1(\mathbb{X}, \mathcal{X}, \mu)$, ou plus simplement $\mathcal{L}_{\mathbb{K}}^1(\mu)$ si il y a pas d'ambiguïtés, l'ensemble des fonctions intégrables à valeurs dans \mathbb{K} .

Proposition 3.1.9. Soit f une fonction mesurable à valeurs réelles. Alors f est intégrable si et seulement si f^+ et f^- le sont.

Démonstration. Il suffit de remarquer que $|f| = f^+ + f^- \geq 0$ et donc par linéarité de l'intégrale pour des fonctions mesurables positives

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

□

Définition 3.1.10. Soit $f \in \mathcal{L}_{\mathbb{R}}^1(\mu)$. On appelle intégrale de f contre μ et on note $\int f d\mu$ le nombre réel

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

Remarque 34. Remarquons que la définition a toujours un sens dans $\overline{\mathbb{R}}$ lorsque f^+ ou f^- est intégrable. Dans ce cas, il faut toutefois être attentif lorsque l'on calcule l'intégrale de la somme de deux fonctions, certaines indéterminations peuvent apparaître.

On note parfois lorsque l'on veut spécifier la variable muette

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

On rencontre parfois $\int f(x) d\mu(x)$ que nous éviterons d'employer du fait de la confusion possible avec les mesures de Stieltjes qui sont introduites par le Théorème 2.2.33.

Proposition 3.1.11. L'ensemble $\mathcal{L}_{\mathbb{R}}^1(\mu)$ est un espace vectoriel sur \mathbb{R} et l'application qui à f associe $\int f d\mu$ est une forme linéaire sur cet espace. De plus, on a

1. si $f \in \mathcal{L}_{\mathbb{R}}^1(\mu)$ et $f \geq 0$ alors $\int f d\mu \geq 0$;
2. si $f, g \in \mathcal{L}_{\mathbb{R}}^1(\mu)$ et $f \leq g$ alors $\int f d\mu \leq \int g d\mu$;
3. si $f \in \mathcal{L}_{\mathbb{R}}^1(\mu)$, alors $|\int f d\mu| \leq \int |f| d\mu$.

Démonstration. On sait déjà que l'ensemble des fonctions réelles mesurables est un espace vectoriel sur \mathbb{R} . De plus, si $f, g \in \mathcal{L}_{\mathbb{R}}^1(\mu)$ et $\lambda \in \mathbb{R}$, alors $|\lambda f + g| \leq |\lambda| |f| + |g|$. On en déduit, par la croissance des intégrales pour les fonctions positives

$$\int |\lambda f + g| d\mu \leq |\lambda| \int |f| d\mu + \int |g| d\mu < \infty.$$

La fonction mesurable identiquement nulle est évidemment intégrable, ainsi l'ensemble $\mathcal{L}_{\mathbb{R}}^1(\mu)$ est un espace vectoriel sur \mathbb{R} .

Soient $f, g \in \mathcal{L}_{\mathbb{R}}^1(\mu)$. On a

$$\begin{cases} f + g = (f + g)^+ - (f + g)^- \\ f + g = f^+ - f^- + g^+ - g^- \end{cases}$$

d'où l'égalité $(f + g)^+ + f^- + g^- = (f + g)^- + f^+ + g^+$. On intègre cette égalité en remarquant que tous les termes sont des fonctions mesurables positives. Il vient donc

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

Toutes ces quantités sont finies, donc on obtient

$$\int (f+g)^+ d\mu - \int (f+g)^- d\mu = \underbrace{\int f^+ d\mu - \int f^- d\mu}_{=\int f d\mu} + \underbrace{\int g^+ d\mu - \int g^- d\mu}_{=\int g d\mu}.$$

On montre de la même manière que

$$\int \lambda f d\mu = \lambda \int f d\mu.$$

Ceci montre la linéarité de l'intégrale.

Pour le point (1), il suffit de remarquer que si $f \in \mathcal{L}_{\mathbb{R}}^1(\mu)$ est positive alors $\int f d\mu = \int f^+ d\mu$ où f^+ est évidemment positive. Ainsi, la définition de l'intégrale de f coïncide avec celle d'une fonction mesurable positive et $\int f d\mu \geq 0$.

Le point (2) est une conséquence immédiate du point (1) appliqué à la fonction intégrable positive $g - f$.

Enfin, pour le point (3), on écrit simplement

$$\left| \int f d\mu \right| = \left| \int f^+ d\mu - \int f^- d\mu \right| \leq \int f^+ d\mu + \int f^- d\mu = \int |f| d\mu.$$

□

Proposition 3.1.12. Soit f une fonction mesurable à valeurs dans \mathbb{C} . Alors f est intégrable si et seulement si les parties réelle et imaginaire de f sont intégrables.

Démonstration. Il suffit d'intégrer les inégalités $|f| \leq |\operatorname{Re} f| + |\operatorname{Im} f| \leq 2|f|$.

□

Définition 3.1.13. Soit $f \in \mathcal{L}_{\mathbb{C}}^1(\mu)$. On appelle intégrale de f contre μ , et on note $\int f d\mu$, le nombre complexe

$$\int f d\mu = \int \operatorname{Re} f d\mu + i \int \operatorname{Im} f d\mu.$$

Proposition 3.1.14. L'ensemble $\mathcal{L}_{\mathbb{C}}^1$ est un \mathbb{C} -espace vectoriel et l'application qui à f associe $\int f d\mu$ est une forme linéaire sur cet espace. De plus,

$$\left| \int f d\mu \right| \leq \int |f| d\mu.$$

Démonstration. Le fait que l'intégrale d'une fonction intégrable à valeurs complexes définisse une forme linéaire se montre de la même manière que dans le cas réel.

Montrons la deuxième partie de la proposition. Soit $\alpha \in \mathbb{C}$ tel que $|\int f d\mu| = \alpha \int f d\mu$. On peut toujours choisir α de module 1 et, en utilisant le fait qu'un nombre réel est plus petit que sa valeur absolue,

$$\begin{aligned} \left| \int f d\mu \right| &= \int \alpha f d\mu = \int \operatorname{Re}(\alpha f) d\mu + i \underbrace{\int \operatorname{Im}(\alpha f) d\mu}_{=0} \\ &\leq \left| \int \operatorname{Re}(\alpha f) d\mu \right| \leq \int |\operatorname{Re}(\alpha f)| d\mu \leq \int |\alpha f| d\mu = \int |f| d\mu. \end{aligned}$$

□

Définition 3.1.15. Une application mesurable f à valeurs dans un \mathbb{K} -espace vectoriel normé $(E, \|\cdot\|)$ de dimension finie est dite intégrable si $\|f\| \in \mathcal{L}_{\mathbb{K}}^1(\mu)$, on note $f \in \mathcal{L}_E^1(\mu)$. De plus, si $\{e_1, \dots, e_d\}$ est une base de E , alors

$$\int f d\mu = \sum_{i=1}^d \left(\int f_i d\mu \right) e_i,$$

où $(f_i)_{i=1, \dots, d}$ sont les coordonnées de f dans la base $(e_i)_{i=1, \dots, d}$.

En général, cette notion d'intégrale vectorielle est utilisée dans le contexte $E = \mathbb{R}^d$ muni de n'importe quelle norme et on choisit la base canonique. Notons que si la valeur de l'intégrale ou, plus précisément, sa représentation vectorielle dépend effectivement de la base choisie, le morphisme linéaire exhibé dans la proposition précédente n'en dépend pas, de même que l'intégrabilité de f .

3.2 L'intégrale de Lebesgue en pratique

Au delà de l'aspect théorique de l'intégrale de Lebesgue, cette section s'intéresse à son aspect pratique. Les idées des quatre sous-sections suivantes sont très utiles en pratique dans le calcul des probabilités.

3.2.1 L'intégrale de Lebesgue contre des mesures discrètes

On considère un espace mesurable $(\mathbb{X}, \mathcal{X})$, une suite $(a_k)_{k \geq 0}$ de points de \mathbb{X} telle que $\{a_k\} \in \mathcal{X}$ et $(\alpha_k)_{k \geq 0}$ des réels positifs. On peut définir une mesure μ sur $(\mathbb{X}, \mathcal{X})$ par

$$\mu = \sum_{k \geq 0} \alpha_k \delta_{a_k}.$$

On souhaite comprendre ce que signifie $\int f d\mu$ pour $f \in \mathcal{L}_{\mathbb{K}}^1(\mu)$.

Proposition 3.2.1. *Soit μ définie comme ci-dessus.*

1. *Soit f une fonction mesurable de $(\mathbb{X}, \mathcal{X})$ dans $\overline{\mathbb{R}}_+$. Alors, dans $\overline{\mathbb{R}}_+$,*

$$\int f d\mu = \sum_{k=0}^{\infty} \alpha_k f(a_k).$$

2. *Une fonction f mesurable de $(\mathbb{X}, \mathcal{X})$ dans \mathbb{K} est intégrable si et seulement si $\sum_{k=0}^{\infty} \alpha_k |f(a_k)| < \infty$. Dans ce cas,*

$$\int f d\mu = \sum_{k=0}^{\infty} \alpha_k f(a_k).$$

De ce point de vue, une série numérique ou complexe n'est rien d'autre qu'une intégrale contre une mesure discrète. Aussi, tous les résultats s'appliquent en particulier aux séries numériques ou complexes. Notons cependant que la notion d'intégrabilité correspond à l'absolue convergence.

Démonstration. On commence par le point (1). On procède en trois étapes. Supposons d'abord que $f = \mathbf{1}_A$ avec $A \in \mathcal{X}$. Alors

$$\int f d\mu = \mu(A) = \sum_{k \geq 0} \alpha_k \mathbf{1}_A(a_k) = \sum_{k \geq 0} \alpha_k f(a_k).$$

Si f est à présent étagée positive, alors $f = \sum_{i=1}^n \beta_i \mathbf{1}_{A_i}$. Par linéarité de l'intégrale

$$\int f d\mu = \sum_{i=1}^n \beta_i \mu(A_i) = \sum_{i=1}^n \beta_i \sum_{k \geq 0} \alpha_k \mathbf{1}_{A_i}(a_k) = \sum_{k \geq 0} \alpha_k \sum_{i=1}^n \beta_i \mathbf{1}_{A_i}(a_k) = \sum_{k \geq 0} \alpha_k f(a_k).$$

Enfin, si f est mesurable positive, il existe une suite croissante $(f_n)_{n \geq 0}$ de fonctions étagées positives qui converge simplement vers f . Par le théorème de convergence monotone et le lemme 2.2.11 il vient que

$$\int f d\mu = \int \lim_{n \rightarrow \infty} f_n(x) d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu = \lim_{n \rightarrow \infty} \sum_{k \geq 0} \alpha_k f_n(a_k) = \sum_{k \geq 0} \alpha_k f(a_k).$$

Pour le point (2), soit f mesurable à valeur dans \mathbb{C} . Appliquons le point (i) à la fonction $|f|$: f est intégrable si et seulement si $\int |f| d\mu$ est finie si et seulement si $\sum_k \alpha_k |f(a_k)|$ est finie. Si tel est le cas, on écrit

$$f = (\operatorname{Re} f)^+ - (\operatorname{Re} f)^- + i(\operatorname{Im} f)^+ - i(\operatorname{Im} f)^-.$$

Les quatre fonctions sont mesurables positives et intégrables (puisque majorées par $|f|$). D'après le point (i), on obtient la relation annoncée. \square

Exercice 19. Exprimer $\int f d\mu$ lorsque μ est l'une des mesures discrètes des exemples 16 du chapitre 2.

3.2.2 Mesures à densité

Étant donné un espace mesuré $(\mathbb{X}, \mathcal{X}, \mu)$, on peut construire de nombreuses mesures à partir de μ comme le montre la proposition suivante.

Proposition 3.2.2. *Soit $(\mathbb{X}, \mathcal{X}, \mu)$ un espace mesuré et g une application mesurable positive sur $(\mathbb{X}, \mathcal{X})$. Soit ν l'application de \mathcal{X} dans \mathbb{R}_+ définie par*

$$\nu(A) = \int \mathbf{1}_A g \, d\mu = \int_A g \, d\mu.$$

Alors ν est une mesure sur $(\mathbb{X}, \mathcal{X})$.

Démonstration. On va utiliser la définition alternative d'une mesure donnée par la proposition 2.2.4. On a bien évidemment $\nu(\emptyset) = 0$ puisque $\mathbf{1}_\emptyset g = 0$. Soient $A, B \in \mathcal{X}$ disjoints, alors par linéarité de l'intégrale, $\nu(A \cup B) = \nu(A) + \nu(B)$. Soit $(B_n)_{n \geq 0}$ une suite croissante d'ensembles mesurables, alors, pour tout $n \geq 0$, $\mathbf{1}_{B_n} \leq \mathbf{1}_{B_{n+1}}$ et $\mathbf{1}_{B_n} g \leq \mathbf{1}_{B_{n+1}} g$ et

$$\lim_{n \rightarrow \infty} \mathbf{1}_{B_n} g = \mathbf{1}_{\cup_{n \geq 0} B_n} g.$$

Ainsi, par le théorème de convergence monotone

$$\nu(\cup_{n \geq 0} B_n) = \int \mathbf{1}_{\cup_{n \geq 0} B_n} g \, d\mu = \int \lim_{n \rightarrow \infty} \mathbf{1}_{B_n} g \, d\mu = \lim_{n \rightarrow \infty} \int \mathbf{1}_{B_n} g \, d\mu = \lim_{n \rightarrow \infty} \nu(B_n).$$

□

Remarque 35. Si on considère la définition initiale d'une mesure, alors il faut utiliser le corollaire du théorème de Beppo-Lévy permettant d'intervertir somme et intégrale.

Définition 3.2.3. La mesure ν est dite à densité g par rapport à μ . On note $\nu = g \cdot \mu$. On dit que g est la densité de ν par rapport à μ .

Exemple 19. Typiquement, les lois de probabilités à densité sont des mesures de probabilités ν qui sont à densité par rapport à la mesure de Lebesgue sur \mathbb{R} . Par exemple, la fonction g pour la loi normale centrée et réduite est définie pour tout $x \in \mathbb{R}$ par

$$g(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}.$$

Si $A \in \mathcal{B}(\mathbb{R})$ alors la mesure gaussienne de A est

$$\nu(A) = \int \mathbf{1}_A \frac{e^{-x^2/2}}{\sqrt{2\pi}} \lambda(dx).$$

En anticipant légèrement les résultats du paragraphe 3.2.4, si A est un intervalle (a, b) par exemple :

$$\nu((a, b)) = \int \mathbf{1}_{(a, b)} \frac{e^{-x^2/2}}{\sqrt{2\pi}} \lambda(dx) = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} \, dx.$$

Ce n'est rien d'autre que la probabilité qu'une variable aléatoire de loi normale centrée réduite prenne une valeur dans (a, b) .

Proposition 3.2.4 (Intégration par rapport à une mesure à densité). *En utilisant les notations de la proposition précédente*

1. Soit f une fonction mesurable positive sur $(\mathbb{X}, \mathcal{X})$. Alors, dans $\overline{\mathbb{R}}_+$,

$$\int f \, d\nu = \int (fg) \, d\mu.$$

2. Soit f une fonction mesurable à valeurs complexes sur $(\mathbb{X}, \mathcal{X})$. Alors f est intégrable pour ν si et seulement si fg est intégrable pour μ et on a alors

$$\int f d\nu = \int (fg) d\mu.$$

Démonstration. Pour montrer le point (i), on procède en trois étapes. Si $f = \mathbf{1}_A$ avec $A \in \mathcal{X}$, alors l'égalité est une conséquence immédiate de la définition de ν . Si f est étagée positive, elle se déduit de la linéarité de l'intégrale. Soient f mesurable positive et $(f_n)_{n \geq 0}$ une suite croissante de fonctions étagées positives qui converge simplement vers f . Le théorème de convergence monotone donne

$$\int f d\nu = \lim_{n \rightarrow \infty} \int f_n d\nu = \lim_{n \rightarrow \infty} \int f_n g d\mu = \int fg d\mu.$$

Pour le point (ii), on applique le point (i) à la fonction $|f|$: f est ν -intégrable si et seulement si et seulement si $\int |f|g d\mu$ est finie si et seulement si fg est μ -intégrable. Si tel est le cas, on écrit

$$f = (\operatorname{Re} f)^+ - (\operatorname{Re} f)^- + i(\operatorname{Im} f)^+ - i(\operatorname{Im} f)^-.$$

Les quatre fonctions sont mesurables positives et intégrables (puisque majorées par $|f|$). D'après le point (i), on obtient la relation annoncée. \square

3.2.3 Mesure image et théorème de transfert

Proposition 3.2.5 (Mesure image). Soient $(\mathbb{X}, \mathcal{X})$ et $(\mathbb{Y}, \mathcal{Y})$ deux espaces mesurables et ϕ une application mesurable de \mathbb{X} dans \mathbb{Y} . Soit μ une mesure sur $(\mathbb{X}, \mathcal{X})$. L'application ν qui à $B \in \mathcal{Y}$ associe $\nu(B) = \mu(\phi^{-1}(B))$ définit une mesure sur $(\mathbb{Y}, \mathcal{Y})$ appelée mesure image de μ par ϕ que l'on notera $\phi_*\mu$.

Démonstration. Nous avons $\nu(\emptyset) = 0$ puisque $\phi^{-1}(\emptyset) = \emptyset$. Soit $(A_n)_{n \geq 0}$ une famille d'ensembles \mathcal{Y} -mesurables deux à deux disjoints, alors $\phi^{-1}(A_i) \cap \phi^{-1}(A_j) = \phi^{-1}(A_i \cap A_j) = \emptyset$ dès que $i \neq j$. Ainsi, $(\phi^{-1}(A_n))_{n \geq 0}$ est une collection d'ensembles deux à deux disjoints qui sont \mathcal{X} -mesurables puisque ϕ est mesurable et

$$\nu\left(\bigcup_{n \geq 0} A_n\right) = \mu\left(\phi^{-1}\left(\bigcup_{n \geq 0} A_n\right)\right) = \mu\left(\bigcup_{n \geq 0} \phi^{-1}(A_n)\right) = \sum_{n \geq 0} \mu(\phi^{-1}(A_n)) = \sum_{n \geq 0} \nu(A_n).$$

\square

Théorème 3.2.6 (Théorème de transfert). À l'aide des mêmes notations,

1. soit f une fonction mesurable positive définie sur $(\mathbb{Y}, \mathcal{Y})$. Alors dans $\overline{\mathbb{R}}_+$,

$$\int f d\phi_*\mu = \int f \circ \phi d\mu. \quad (3.1)$$

2. Soit f une fonction à valeurs complexes définies sur $(\mathbb{X}, \mathcal{X})$. Alors f est intégrable par rapport à $\phi_*\mu$ si et seulement si $f \circ \phi$ est intégrable par rapport à μ . Dans ce cas,

$$\int f d\phi_*\mu = \int f \circ \phi d\mu.$$

Démonstration. Si f est mesurable positive, alors f est limite monotone de fonctions étagées positives. Par convergence monotone, il suffit donc de vérifier l'égalité (3.1) pour les fonctions étagées positives. Si g est une telle fonction, alors elle s'écrit pour $A_1, \dots, A_n \in \mathcal{E}$

$$g = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}, \quad \alpha_1, \dots, \alpha_n \geq 0.$$

Par définition, l'intégrale de g se calcule comme suit

$$\int_E g d\phi_*\mu = \sum_{i=1}^n \alpha_i \phi_*\mu(A_i) = \sum_{i=1}^n \alpha_i \mu(\phi^{-1}A_i) = \int_\Omega g \circ \phi d\nu,$$

en remarquant que $\mathbf{1}_{\phi^{-1}A_i} = \mathbf{1}_{A_i} \circ \phi$.

Pour des fonctions f mesurables à valeurs réelles, au vu de l'égalité précédente, il est clair que $f \circ \phi$ est μ -intégrable si et seulement si f est $\phi_*\mu$ -intégrable. De plus, en écrivant $f = f^+ - f^-$, on a $f \circ \phi = (f \circ \phi)^+ - (f \circ \phi)^-$ et le résultat suit immédiatement. Le cas des fonctions à valeurs complexes se montrent comme d'habitude en décomposant en partie réelle et imaginaire. \square

3.2.4 Intégrale de Riemann et intégrale de Lebesgue

En substance, cette sous section permet de montrer que les fonctions intégrables au sens de Riemann sont intégrables au sens de Lebesgue et que les intégrales coïncident dans ce cas là. Dans cette partie, on va également généraliser le théorème fondamental du calcul intégral.

Intégration sur un intervalle compact

Soient f une fonction réelle bornée sur $[a, b]$ et $\sigma : a = x_0 < x_1 < \dots < x_{n+1} = b$ une subdivision de $[a, b]$. Le nombre $\delta(\sigma) = \max\{x_k - x_{k-1}, 1 \leq k \leq n+1\}$ est appelé pas de la subdivision σ . On pose

$$m_k = \inf\{f(t), t \in [x_k, x_{k+1}]\} \quad \text{et} \quad M_k = \sup\{f(t), t \in [x_k, x_{k+1}]\}.$$

Les sommes de Darboux associées à la subdivision σ sont

$$s(\sigma) = \sum_{k=1}^n m_k(x_{k+1} - x_k) \quad \text{et} \quad S(\sigma) = \sum_{k=1}^n M_k(x_{k+1} - x_k).$$

Définition 3.2.7. On dit qu'une fonction réelle f sur un intervalle $[a, b]$ est intégrable au sens de Riemann s'il existe un nombre réel I tel que les sommes $s(\sigma)$ et $S(\sigma)$ tendent vers I quand $\delta(\sigma)$ tend vers 0 :

$$\forall \varepsilon > 0, \exists \eta > 0, \quad \forall \sigma : \delta(\sigma) < \eta \implies |s(\sigma) - I| + |S(\sigma) - I| < \varepsilon.$$

Le nombre I est alors appelé l'intégrale de Riemann de f sur $[a, b]$ et on le note $\int_a^b f(t) dt$.

Considérons à nouveau la subdivision σ et, pour chaque $k = 1, \dots, n+1$, choisissons $\xi_k \in [x_{k-1}, x_k]$. La somme de Riemann définie par σ et $\xi = (\xi_1, \dots, \xi_n)$ est par définition

$$S(\sigma, \xi) = \sum_{k=1}^n f(\xi_k)(x_k - x_{k-1}).$$

Il est alors facile de voir que si f est intégrable au sens de Riemann, les sommes de Riemann convergent vers $\int_a^b f(t) dt$ lorsque $\delta(\sigma)$ tend vers 0, uniformément par rapport au choix de ξ . Plus précisément,

$$\forall \varepsilon > 0, \quad \exists \eta > 0, \quad \forall \sigma : \delta(\sigma) < \eta, \quad \forall \xi \text{ associé à } \sigma, \quad \left| S(\sigma, \xi) - \int_a^b f(t) dt \right| < \varepsilon.$$

Théorème 3.2.8 (Théorème fondamental du calcul intégrale). *Tout fonction continue par morceaux sur $[a, b]$ est intégrable au sens de Riemann. De plus, si f est continue, la fonction $x \rightarrow F(x) = \int_a^x f(t) dt$ est dérivable sur $[a, b]$ de dérivée $F' = f$.*

Démonstration. Exercice. \square

Intégrale généralisée

Soit $f : [a, b) \rightarrow \mathbb{R}$, où b peut valoir $+\infty$, localement intégrable au sens de Riemann : c'est à dire $f\mathbf{1}_{[a, c]}$ est Riemann intégrable pour intervalle compacte $[a, c] \subset [a, b)$.

On dit que f admet une intégrale généralisée sur $[a, b)$ si la fonction $x \rightarrow \int_a^x f(t) dt$ admet une limite lorsque x tend vers b , $x < b$. On pose alors

$$\int_a^b f(t) dt = \lim_{x \rightarrow b, x < b} \int_a^x f(t) dt.$$

Dans ce cas, on dit aussi que l'intégrale est convergente. On dira que l'intégrale généralisée est absolument convergente si $\int_a^b |f(t)| dt$ est convergente. On rappelle que l'absolue convergence implique la convergence mais que la réciproque est fautive (penser à l'exemple classique $\int_0^\infty \frac{\sin(t)}{t} dt$).

Comparaison de l'intégrale de Riemann et de l'intégrale de Lebesgue pour une fonction bornée sur un intervalle compact

Proposition 3.2.9. Soit f une fonction continue sur $[a, b]$. Alors si λ désigne la mesure de Lebesgue sur \mathbb{R} , $f\mathbf{1}_{[a,b]} \in \mathcal{L}_{\mathbb{R}}^1(\lambda)$ et

$$\int_{\mathbb{R}} f\mathbf{1}_{[a,b]} d\lambda = \int_a^b f(t) dt.$$

Démonstration. Il est immédiat que $f\mathbf{1}_{[a,b]}$ est borélienne. De plus, comme f est continue sur le compact $[a, b]$, elle est bornée sur $[a, b]$. Nous obtenons, en posant $M = \sup_{t \in [a,b]} |f(t)|$, que $|f\mathbf{1}_{[a,b]}| \leq M\mathbf{1}_{[a,b]}$ qui est manifestement Lebesgue intégrable. De même, pour tout $x \in [a, b]$, $f\mathbf{1}_{[a,x]}$ est Lebesgue intégrable. Posons donc $F(x) = \int f\mathbf{1}_{[a,x]} d\lambda$ et montrons que F est dérivable sur $[a, b]$ de dérivée f . Soit $x_0 \in [a, b]$ et $h > 0$. On calcule

$$\mathbf{1}_{[a, x_0+h]}f = \mathbf{1}_{[a, x_0]}f + \mathbf{1}_{(x_0, x_0+h]}f \implies \frac{F(x_0+h) - F(x_0)}{h} = \frac{1}{h} \int \mathbf{1}_{(x_0, x_0+h]}f d\lambda,$$

d'où

$$\frac{F(x_0+h) - F(x_0)}{h} - f(x_0) = \frac{1}{h} \int \mathbf{1}_{(x_0, x_0+h]}(f - f(x_0)) d\lambda$$

Soit $\varepsilon > 0$. Puisque f est continue en x_0 , il existe $\eta > 0$ tel que pour tout x satisfaisant $|x - x_0| \leq \eta$ implique $|f(x) - f(x_0)| \leq \varepsilon$. Ainsi si $h \in (0, \eta)$ alors

$$\left| \frac{F(x_0+h) - F(x_0)}{h} - f(x_0) \right| = \frac{1}{h} \int \varepsilon \mathbf{1}_{(x_0, x_0+h]} d\lambda = \varepsilon.$$

Le cas $h < 0$ se traite de façon analogue. Donc F est dérivable sur $[a, b]$ de dérivée f . Or $F(a) = 0$ car $\lambda(\{a\}) = 0$, d'où $F(x) = \int_a^x f(t) dt$ pour tout $x \in [a, b]$ et notamment

$$F(b) = \int_a^b f(t) dt = \int_{\mathbb{R}} f\mathbf{1}_{[a,b]} d\lambda. \quad (3.2)$$

□

Remarque 36. Il faut bien noter qu'*a priori* l'intégrale de Lebesgue de $f\mathbf{1}_{[a,b]}$, notée $\int f d\lambda$, et l'intégrale de Riemann de f sur $[a, b]$, notée $\int_a^b f(t) dt$, sont deux objets différents, elles sont construites de façon radicalement différentes. La proposition 3.2.9 ci-dessus, ainsi que le raffinement donné par le théorème 3.2.10 ci-dessous, permet de conclure que pour une grande classe de fonction les deux intégrales coïncident. Il est très commode de faire la confusion entre les notations $\int f\mathbf{1}_{[a,b]} d\lambda$, $\int f(x)\mathbf{1}_{[a,b]}(x) \lambda(dx)$ et $\int_a^b f(x) dx$: sauf mention contraire, ces notations désigneront toujours l'intégrale au sens de Lebesgue. Cet abus n'apporte pas de problème particulier en pratique.

Théorème 3.2.10 (Critère de Lebesgue). Une fonction $f : [a, b] \rightarrow \mathbb{R}$ bornée est intégrable au sens de Riemann si et seulement si il existe $N \subset [a, b]$ de mesure de Lebesgue nulle tel que f est continue en tout $x \in [a, b] \setminus N$. Dans ce cas, il y a coïncidence entre les deux intégrales

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

Intégrale de Riemann généralisée et intégrale de Lebesgue

La proposition suivante est très utile pour alléger le traitement des intégrales généralisées sous la condition d'absolue convergence. Les intégrales généralisées simplement convergentes devront toutefois être traitées de façon plus classique.

Proposition 3.2.11. Soit $f : [a, b] \rightarrow \mathbb{R}$ une fonction continue. Alors $f\mathbf{1}_{[a,b]} \in \mathcal{L}_{\mathbb{R}}^1(\lambda)$ si et seulement si $\int_a^b f(t) dt$ est absolument convergente et, dans ce cas, on a

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

Remarque 37. En pratique cela autorise à écrire des choses comme ceci :

$$\int_0^\infty \frac{e^{-x}}{\sqrt{x}} dx = 2 \int_0^\infty e^{-y^2} dy,$$

là où, dans le contexte de l'intégrale de Riemann, nous devrions écrire pour être tout à fait rigoureux

$$\int_0^\infty \frac{e^{-x}}{\sqrt{x}} dx = \lim_{A \rightarrow 0, B \rightarrow \infty} \int_A^B e^{-x} \sqrt{x} dx = \lim_{A \rightarrow 0, B \rightarrow \infty} \int_{\sqrt{A}}^{\sqrt{B}} 2e^{-y^2} dy,$$

en justifiant tous les passages à la limite.

De même pour une intégration par parties, on peut écrire les bornes infinies directement, sous la condition bien entendu que l'intégrale généralisée est absolument convergente.

Démonstration. Supposons d'abord f positive. Soit $(b_n)_{n \geq 0}$ une suite croissante de points de $[a, b)$ qui converge vers b . Pour tout $n \geq 0$,

$$\int f \mathbf{1}_{[a, b_n]} d\lambda = \int_a^{b_n} f(t) dt.$$

Le théorème de convergence monotone (pour l'intégrale de Lebesgue), on obtient

$$\int f \mathbf{1}_{[a, b)} d\lambda = \lim_{n \rightarrow \infty} \int f \mathbf{1}_{[a, b_n]} d\lambda = \lim_{n \rightarrow \infty} \int_a^{b_n} f(t) dt \in \overline{\mathbb{R}}_+.$$

Or, par définition, $f \mathbf{1}_{[a, b)}$ est Lebesgue intégrable si et seulement si cette limite est finie donc si et seulement si f est Riemann intégrable. De plus, ces deux intégrables coïncident.

Dans le cas général, on sait que f est Lebesgue intégrable si et seulement si $|f|$ l'est, donc si et seulement si $\int_a^b f(t) dt$ est absolument convergente. Si tel est le cas, nous écrivons $f = f^+ - f^-$. On a $f^+ \leq |f|$ et $f^- \leq |f|$ si bien que f^+, f^- sont positives et intégrables aussi bien dans le sens de Lebesgue que dans le sens de Riemann. Or,

$$\int f^+ \mathbf{1}_{[a, b)} d\lambda = \int_a^b f^+(t) dt \quad \text{et} \quad \int f^- \mathbf{1}_{[a, b)} d\lambda = \int_a^b f^-(t) dt,$$

et la linéarité de l'intégrale permet de conclure. □

Chapitre 4

Théorèmes limites

4.1 Lemme de Fatou

Lors de la construction de l'intégrale, nous avons établi un théorème limite fondamental : le théorème de Beppo-Lévy également appelé le théorème de convergence monotone.

Théorème 4.1.1 (Théorème de Beppo-Lévy). *Soit $(f_n)_{n \geq 0}$ une suite monotone croissante de fonctions mesurables positives, i.e. $0 \leq f_n \leq f_{n+1}$ pour tout $n \geq 0$. Alors $f = \lim_{n \rightarrow \infty} f_n = \sup_{n \geq 0} f_n$ est mesurable positive et*

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

Théorème 4.1.2 (Lemme de Fatou). *Si $(f_n)_{n \geq 0}$ est une suite de fonctions mesurables positives, alors*

$$\int \liminf_{n \rightarrow \infty} f_n d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu.$$

Démonstration. Posons $g = \liminf_{n \rightarrow \infty} f_n$, cette fonction est mesurable et prend ses valeurs dans $\overline{\mathbb{R}}_+$. On pose également $g_n = \inf_{k \geq n} f_k$ pour tout $n \geq 0$. Par définition, $g = \lim_{n \rightarrow \infty} g_n$. De plus, $(g_n)_{n \geq 0}$ est croissante. Le théorème de croissance monotone assure donc

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

D'autre part, pour tout $n \geq 0$, $g_n \leq f_n$ et par suite $\int g_n d\mu \leq \int f_n d\mu$. En particulier, pour tout $n \geq 0$, il vient que

$$\int g_n d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu.$$

Le second membre de l'inégalité ne dépend plus de n , d'où en passant à la limite dans le premier membre (cette limite existe par le théorème de Beppo-Lévy), on obtient

$$\int \liminf_{n \rightarrow \infty} f_n d\mu = \lim_{n \rightarrow \infty} \int g_n d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu.$$

□

4.2 Ensembles et fonctions mesurables négligeables

Définition 4.2.1. Soit $(\mathbb{X}, \mathcal{X}, \mu)$ un espace mesuré.

1. On dit qu'une partie N de \mathbb{X} est négligeable pour μ s'il existe $A \in \mathcal{X}$ tel que $N \subset A$ et $\mu(A) = 0$.
2. On dit que la σ -algèbre \mathcal{X} est complète pour μ si toute partie négligeable pour μ appartient à \mathcal{X} .

Il est toujours possible d'ajouter les ensembles négligeables à une tribu non complète pour la rendre complète. Nous supposons désormais que les tribus considérées sont complètes.

Définition 4.2.2. Soit $(\mathbb{X}, \mathcal{X}, \mu)$ un espace mesuré. On dit qu'une propriété \mathcal{P} sur \mathbb{X} est vraie presque partout (en abrégé p.p. ou μ -p.p.) si l'ensemble des points de \mathbb{X} où elle est fautive est négligeable.

Une fonction définie sur \mathbb{X} à valeurs réelles ou complexes est dite μ -négligeable si $\{f \neq 0\}$ est négligeable.

Deux fonctions f et g définies sur \mathbb{X} et à valeurs dans un même espace mesurable \mathbb{Y} sont dites égales presque partout si $\{f \neq g\}$ est négligeable.

On dit qu'une suite $(f_n)_{n \geq 0}$ de fonctions définies sur \mathbb{X} à valeurs dans un espace topologique (séparé) converge vers f presque partout si il existe un ensemble négligeable N tel que pour tout $x \notin N$, $\lim_n f_n(x) = f(x)$.

Lemme 4.2.3 (Inégalité de Markov). *Soit f une fonction mesurable positive sur $(\mathbb{X}, \mathcal{X})$. Alors pour tout $\lambda > 0$, on a*

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

Démonstration. Par positivité de f , pour tout $\lambda > 0$, $\lambda \mathbf{1}_{f \geq \lambda} \leq f$. Par croissance de l'intégrale, on obtient le résultat. \square

Proposition 4.2.4. *Si $f \in \mathcal{L}^1_{\mathbb{R}}(\mu)$, alors f est finie μ -p.p..*

Démonstration. Par l'inégalité de Markov, $\mu(\{|f| \geq n\}) \leq \frac{1}{n} \int |f| \, d\mu$. La suite $A_n = \{|f| > n\}$ est décroissante et $\mu(A_1) \leq \int |f| \, d\mu < \infty$. Par la continuité à droite de la mesure μ :

$$\mu(\{|f| = \infty\}) = \mu\left(\bigcap_{n \geq 1} A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n) = 0.$$

\square

Exercice 20. Montrer que la réciproque est fautive. Donner au moins un exemple dans le cas d'une mesure μ finie.

Proposition 4.2.5. *Soit f une fonction mesurable sur $(\mathbb{X}, \mathcal{X})$ à valeurs complexes. Alors f est négligeable si et seulement si $\int |f| \, d\mu = 0$.*

Démonstration. L'inégalité de Markov encore implique que $\mu(\{|f| > \frac{1}{n}\}) \leq n \int |f| \, d\mu = 0$ pour tout $n \geq 1$. Or, la suite $A_n = \{|f| > \frac{1}{n}\}$ est croissante, par la continuité à gauche de μ on obtient

$$\mu(\{|f| > 0\}) = \mu\left(\bigcup_{n \geq 1} A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n) = 0.$$

Réciproquement, soit $n \geq 1$ alors $|f|$ est limite monotone de $|f| \wedge n$, le théorème de convergence monotone implique

$$\int |f| \, d\mu = \lim_{n \rightarrow \infty} \int |f| \wedge n \, d\mu = \lim_{n \rightarrow \infty} \int |f| \wedge n \mathbf{1}_N \, d\mu + \int |f| \wedge n \mathbf{1}_{N^c} \, d\mu \leq n\mu(N^c) = 0,$$

où $N = \{|f| = 0\}$. \square

Proposition 4.2.6. *Soit $(\mathbb{X}, \mathcal{X}, \mu)$ un espace mesuré.*

1. *Soient f et g deux fonctions mesurables positives telles $f \leq g$ presque partout. Alors $\int f \, d\mu \leq \int g \, d\mu$.*
2. *Soient f et g deux fonctions mesurables positives telles que $f = g$ presque partout. Alors $\int f \, d\mu = \int g \, d\mu$.*
3. *Soient f et g deux fonctions mesurables complexes telles que $f = g$ presque partout. Alors f est intégrable si et seulement si g est intégrable et, dans ce cas, $\int f \, d\mu = \int g \, d\mu$.*

Démonstration. 1. Il suffit d'appliquer la proposition 4.2.5 à la fonction $(f - g)^+$. Celle-ci est nulle presque-partout, $|(f - g)^+| = (f - g)^+$ et donc

$$0 = \int (f - g)^+ \, d\mu = \int (f - g) \, d\mu + \int (f - g)^- \, d\mu \geq \int (f - g) \, d\mu.$$

2. C'est une conséquence d'une double application du point précédant car $f = g$ presque-partout si et seulement si $f \leq g$ presque partout et $g \leq f$ presque partout.
3. Il suffit de poser $h = f - g$ alors $h = 0$ presque partout et donc

$$0 = \int |h| d\mu \geq \left| \int h d\mu \right|.$$

□

4.3 Théorème de convergence dominée

Théorème 4.3.1 (Théorème de convergence dominée). *Soit $(f_n)_{n \geq 0}$ une suite de fonctions mesurables sur $(\mathbb{X}, \mathcal{X})$ à valeurs dans \mathbb{R} ou \mathbb{C} telle que :*

1. $(f_n)_{n \geq 0}$ converge μ -presque partout vers une fonction f mesurable,
2. il existe une fonction $g \in \mathcal{L}^1_{\mathbb{R}}(\mu)$ positive telle que pour tout $n \geq 0$, $|f_n| \leq g$ μ -presque partout.

Alors les fonctions $(f_n)_{n \geq 0}$ et f sont intégrables et

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu \quad \text{et} \quad \lim_{n \rightarrow \infty} \int |f_n - f| d\mu = 0.$$

Exemple 20. Considérons par exemple la suite de fonctions $(f_n)_{n \geq 1}$ définie pour tout $n \geq 1$ et tout $x \in \mathbb{R}$ par $f_n(x) = \frac{\sin(x)^n}{x(1+x)} \mathbf{1}_{[0, \infty)}(x)$. On vérifie que pour tout $x \in \mathbb{R} \setminus (\pi/2 + \pi\mathbb{Z})$:

$$\lim_{n \rightarrow \infty} f_n(x) = \lim_{n \rightarrow \infty} \frac{\sin(x)^n}{x(1+x)} \mathbf{1}_{[0, \infty)}(x) = 0.$$

Comme $\lambda(\pi/2 + \pi\mathbb{Z}) = 0$, la suite $(f_n)_{n \geq 1}$ converge presque partout vers la fonction nulle. D'autre part, en utilisant l'inégalité $\sin(x) \leq x$ pour tout $x \geq 0$, on obtient :

$$|f_n(x)| \leq \frac{x^{n-1}}{1+x} \mathbf{1}_{[0,1]}(x) + \frac{1}{x(1+x)} \mathbf{1}_{[1, \infty)}(x) \leq \frac{1}{2} \mathbf{1}_{[0,1]}(x) + \frac{1}{x(1+x)} \mathbf{1}_{[1, \infty)}(x) \leq g(x).$$

Les deux termes sont mesurables positifs, l'intégrale de Lebesgue a donc un sens dans $\overline{\mathbb{R}}_+$. L'intégrale contre la mesure de Lebesgue du premier terme vaut $\frac{1}{2}$ alors que le second terme est un $O(x^{-2} \mathbf{1}_{[1, \infty)})$ qui est intégrable également. Le théorème de convergence dominée implique donc

$$\lim_{n \rightarrow \infty} \int_0^{\infty} \frac{\sin(x)^n}{x(1+x)} dx = 0.$$

Dans la pratique, on peut se permettre d'aller un peu plus vite, l'idée étant toujours d'utiliser les indicatrices pour décomposer le domaine et majorer uniformément sur chaque sous-domaine par la fonction adéquate.

Notons que l'on ne s'occupe absolument pas des points pathologiques pour lesquelles le sinus vaut 1 ou -1 . D'autre part, ces points pathologiques sont exactement ceux qui empêchent la convergence uniforme de la suite $(f_n)_{n \geq 1}$. Avec des techniques de type Riemann, il faudrait procéder autrement en enlevant des petits voisinages ouverts autour de ces points pathologiques puis justifier le passage à la limite. C'est possible mais bien plus pénible.

Remarque 38. Remarquons que nous ne disons rien sur la limite de la suite de fonction contrairement au théorème de convergence dominée que l'on énonce dans le cadre Riemann. Ceci est dû au fait que, dans le contexte de la théorie de la mesure, la limite est automatiquement mesurable et la condition de domination implique qu'elle est intégrable. Dans le cadre riemannien, la limite de fonction Riemann intégrables n'est pas nécessairement Riemann intégrable.

Démonstration. Supposons tout d'abord que la convergence de $(f_n)_{n \geq 0}$ vers f ait lieu partout et que l'inégalité du deuxième point est vraie pour tout $x \in \mathbb{X}$. Posons $g_n = 2g - |f_n - f|$. Alors $(g_n)_{n \geq 0}$ est une suite de fonctions mesurables positives et d'après le lemme de Fatou,

$$2 \int g d\mu = \int \liminf_{n \rightarrow \infty} g_n d\mu \leq \liminf_{n \rightarrow \infty} \int g_n d\mu = 2 \int g d\mu - \limsup_{n \rightarrow \infty} \int |f_n - f| d\mu.$$

Puisque $\int g \, d\mu < \infty$, on déduit que $\limsup_{n \rightarrow \infty} \int |f_n - f| \, d\mu \leq 0$. On en déduit donc que

$$\lim_{n \rightarrow \infty} \int |f_n - f| \, d\mu = 0 \quad \text{et} \quad \lim_{n \rightarrow \infty} \int f_n \, d\mu = \int f \, d\mu.$$

Passons à présent au cas général. Par définition, il existe $N \in \mathcal{X}$ tel que, si $x \notin N$, $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ et $\mu(N) = 0$. Il existe également des ensembles $N_n \in \mathcal{X}$, $n \geq 0$, tel que, si $x \notin N_n$, $|f_n(x)| \leq g(x)$ et $\mu(N_n) = 0$. Posons $M = N \cup (\cup_{n \geq 0} N_n) \in \mathcal{X}$. On a encore $\mu(M) = 0$. On pose $h_n = f_n \mathbf{1}_{M^c}$ et $h = f \mathbf{1}_{M^c}$. Alors, pour tout $x \in \mathbb{X}$ et tout $n \geq 0$,

$$\lim_{m \rightarrow \infty} h_m(x) = h(x) \quad \text{et} \quad |h_n(x)| \leq g(x).$$

La première partie de la preuve assure donc que $\lim_{n \rightarrow \infty} \int |h_n - h| \, d\mu = 0$. Pour conclure, il suffit de remarquer que $h_n = f_n$ μ -p.p. et $h = f$ μ -p.p. si bien que $|h_n - h| = |f_n - f|$ μ -p.p. et donc leurs intégrales sont égales. \square

Remarque 39. En utilisant le même principe de preuve, on peut montrer un théorème de convergence monotone presque-partout ou un lemme de Fatou presque-partout.

Corollaire 4.3.2. Soit $(f_n)_{n \geq 0}$ une suite de fonctions mesurables sur $(\mathbb{X}, \mathcal{X})$ à valeurs sur \mathbb{R} ou \mathbb{C} telle que

$$\sum_{n \geq 0} \int |f_n| \, d\mu < \infty.$$

Alors les fonctions $(f_n)_{n \geq 0}$ sont intégrables, la série $\sum_n f_n$ converge μ -p.p. et il existe $f \in \mathcal{L}_{\mathbb{K}}^1(\mu)$ telle que

$$f = \sum_{n=0}^{\infty} f_n \quad \mu\text{-p.p.}, \quad \lim_{n \rightarrow \infty} \int \left| f - \sum_{k=0}^n f_k \right| \, d\mu = 0, \quad \int f \, d\mu = \sum_{n=0}^{\infty} \int f_n \, d\mu.$$

Démonstration. Le théorème de convergence monotone pour les séries à termes positifs implique

$$\int \sum_{n \geq 0} |f_n| \, d\mu = \sum_{n \geq 0} \int |f_n| \, d\mu.$$

Par la proposition 4.2.4, on déduit que $\sum_{n \geq 0} f_n$ est absolument convergente. L'application du théorème de convergence dominée à la suite des sommes partielles achève la preuve du corollaire. \square

4.4 Intégrale à paramètres

On termine ce chapitre par des cas particuliers d'interversion de limites, à savoir continuité et dérivation sous le signe intégral.

Théorème 4.4.1 (Continuité d'une intégrale à paramètre). Soient $(\mathbb{X}, \mathcal{X}, \mu)$ un espace mesuré, (E, d) un espace métrique et f une fonction définie sur $\mathbb{X} \times E$ à valeurs réelles ou complexes. On suppose que

1. pour μ -presque tout $x \in \mathbb{X}$, la fonction $\alpha \rightarrow f(x, \alpha)$ est continue sur E ;
2. pour tout $\alpha \in E$, la fonction $x \rightarrow f(x, \alpha)$ est mesurable sur $(\mathbb{X}, \mathcal{X})$;
3. il existe une fonction g sur $(\mathbb{X}, \mathcal{X})$ mesurable, positive et intégrable telle que pour tout $\alpha \in E$, $|f(x, \alpha)| \leq g(x)$ μ -presque partout.

Alors $F : \alpha \rightarrow \int_{\mathbb{X}} f(x, \alpha) \, \mu(dx)$ est définie et continue sur E .

Démonstration. Pour tout $\alpha \in E$, la fonction $x \rightarrow f(x, \alpha)$ est intégrable par rapport à μ donc F est bien définie sur E . Soit $\alpha \in E$ et montrons que F est continue au point α . Pour cela, on va utiliser la caractérisation séquentielle de la continuité. Soit donc $(\alpha_n)_{n \geq 0}$ une suite de E convergente vers α . Notons pour tout $x \in E$, $f_n(x) = f(x, \alpha_n)$, puis on applique le théorème de convergence dominée. On obtient que $F(\alpha_n)$ converge vers $F(\alpha)$. \square

Théorème 4.4.2 (Dérivabilité d'une intégrale à paramètre). Soient $(\mathbb{X}, \mathcal{X}, \mu)$ un espace mesuré, I un intervalle ouvert de \mathbb{R} et f une fonction définie sur $\mathbb{X} \times I$ à valeurs réelles ou complexes. On suppose que

1. pour μ -presque tout $x \in \mathbb{X}$, la fonction $\alpha \rightarrow f(x, \alpha)$ est dérivable sur I ;
2. pour tout $\alpha \in I$, la fonction $x \rightarrow f(x, \alpha)$ est μ -intégrable ;
3. il existe une fonction g sur $(\mathbb{X}, \mathcal{X})$ intégrable et positive telle que pour μ -presque tout $x \in E$

$$\forall \alpha \in I, \quad \left| \frac{\partial f}{\partial \alpha}(x, \alpha) \right| \leq g(x).$$

Alors pour tout $\alpha \in I$, la fonction $x \rightarrow \frac{\partial f}{\partial \alpha}$ est intégrable. De plus, la fonction $F : \alpha \rightarrow \int f(x, \alpha) \mu(dx)$ est dérivable sur I et

$$\forall \alpha \in I, \quad F'(\alpha) = \int \frac{\partial f}{\partial \alpha}(x, \alpha) \mu(dx).$$

Démonstration. Par hypothèse, il existe un ensemble de mesure nulle $N \in \mathcal{X}$ tel que si $x \notin N$, la dérivée $\frac{\partial f}{\partial \alpha}(x, \alpha)$ existe pour tout point $\alpha \in I$ et

$$\left| \frac{\partial f}{\partial \alpha}(x, \alpha) \right| \leq g(x).$$

Il en résulte que $x \rightarrow \frac{\partial f}{\partial \alpha}(x, \alpha)$ est μ -intégrable pour tout $\alpha \in I$. Étudions la dérivabilité de F en $\alpha \in I$. Soit $(\alpha_n)_{n \geq 0}$ une suite de I qui converge vers α mais telle que $\alpha_n \neq \alpha$ pour tout $n \geq 0$. Le théorème des accroissements finis implique pour tout $x \notin N$ que

$$|f(x, \alpha_n) - f(x, \alpha)| \leq |\alpha_n - \alpha| \sup_{\alpha \in I} \left| \frac{\partial f}{\partial \alpha}(x, \alpha) \right| \leq |\alpha_n - \alpha| g(x).$$

On introduit la suite $(h_n)_{n \geq 0}$ où la fonction h_n est définie sur \mathbb{X} par

$$h_n(x) = \frac{f(x, \alpha_n) - f(x, \alpha)}{\alpha_n - \alpha}.$$

Cette suite converge simplement sur $\mathbb{X} \setminus N$ vers la fonction $x \rightarrow \frac{\partial f}{\partial \alpha}(x, \alpha)$. De plus $(h_n)_{n \geq 0}$ est uniformément bornée en valeur absolue par g , d'où par le théorème de convergence dominée

$$\int \frac{\partial f}{\partial \alpha}(x, \alpha) d\mu = \lim_{n \rightarrow \infty} \int \frac{f(x, \alpha_n) - f(x, \alpha)}{\alpha_n - \alpha} d\mu = \lim_{n \rightarrow \infty} \frac{F(\alpha_n) - F(\alpha)}{\alpha_n - \alpha}.$$

Il en résulte que F est dérivable en α de dérivée

$$F'(\alpha) = \int \frac{\partial f}{\partial \alpha}(x, \alpha) d\mu.$$

□

Chapitre 5

Mesure produit

Dans ce chapitre, on souhaite construire une mesure m sur un produit d'espaces mesurables $(E \times F, \mathcal{X} \otimes \mathcal{Y})$ tel que $m(A \times B) = \mu(A)\nu(B)$ où μ et ν sont des mesures prescrites sur \mathcal{X} et \mathcal{Y} .

En fait, une telle mesure m existe et est unique sous la condition de σ -finitude de μ et ν . L'unicité de la mesure produit découlera du théorème de caractérisation des mesures σ -finies. Pour l'existence, on donnera une preuve "constructive" au sens où l'on ne fera pas usage du théorème d'extension de Carathéodory. La raison pour laquelle on préférera cette preuve directe est qu'elle nous permettra, au delà de l'existence de la mesure produit, de montrer les théorèmes de Tonelli et Fubini qui permettent de ramener un calcul d'intégrale multiple en autant d'intégrales simples successives

5.1 Mesure produit

Soient $(\mathbb{X}, \mathcal{X}, \mu)$ et $(\mathbb{Y}, \mathcal{Y}, \nu)$ deux espaces σ -finis. On dispose déjà d'une tribu naturelle sur $\mathbb{X} \times \mathbb{Y}$ construite à partir de \mathcal{X} et \mathcal{Y} , c'est la tribu produit $\mathcal{X} \otimes \mathcal{Y}$, i.e la tribu engendrée par les pavés $A \times B$ où $A \in \mathcal{X}$ et $B \in \mathcal{Y}$.

Soit $C \in \mathcal{X} \otimes \mathcal{Y}$. On note $C_x = \{y \in \mathbb{Y} : (x, y) \in C\}$ la section verticale et $C^y = \{x \in \mathbb{X} : (x, y) \in C\}$ la section horizontale.

Lemme 5.1.1. *Soit $C \in \mathcal{X} \otimes \mathcal{Y}$. Alors pour tout $x \in \mathbb{X}$ et tout $y \in \mathbb{Y}$, $C_x \in \mathcal{Y}$ et $C^y \in \mathcal{X}$.*

Remarque 40. Si C et D sont des éléments de $\mathcal{X} \otimes \mathcal{Y}$ alors pour tout $x \in \mathbb{X}$

$$(C_x)^c = (C^c)_x, \quad C_x \cup D_x = (C \cup D)_x \quad \text{et} \quad C_x \cap D_x = (C \cap D)_x.$$

Il en va de même pour les unions et intersections dénombrables.

Démonstration. Soit \mathcal{C} l'ensemble des parties $C \in \mathcal{X} \otimes \mathcal{Y}$ telles que, pour tout $x \in \mathbb{X}$ et tout $y \in \mathbb{Y}$, $C_x \in \mathcal{Y}$ et $C^y \in \mathcal{X}$. Alors \mathcal{C} est clairement une tribu. Soit $C = A \times B$ un rectangle, alors pour tout $x \in \mathbb{X}, y \in \mathbb{Y}$,

$$C_x = \begin{cases} B & \text{si } x \in A \\ \emptyset & \text{si } x \notin A \end{cases} \in \mathcal{Y} \quad \text{et} \quad C^y = \begin{cases} A & \text{si } y \in B \\ \emptyset & \text{si } y \notin B \end{cases} \in \mathcal{X}.$$

Ainsi, \mathcal{C} est une tribu qui contient les rectangles, elle contient donc la tribu produit $\mathcal{X} \otimes \mathcal{Y}$. □

Théorème 5.1.2. *Soient $(\mathbb{X}, \mathcal{X}, \mu)$ et $(\mathbb{Y}, \mathcal{Y}, \nu)$ deux espaces mesurés σ -finis.*

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

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

avec la convention $0 \times \infty = 0$. Cette mesure est σ -finie. On la note généralement $\mu \otimes \nu$ et on l'appelle mesure produit de μ et ν .

2. Pour tout $C \in \mathcal{X} \otimes \mathcal{Y}$, les applications $x \rightarrow \nu(C_x)$ et $y \rightarrow \mu(C^y)$ sont respectivement \mathcal{X} -mesurable et \mathcal{Y} -mesurable et

$$\mu \otimes \nu(C) = \int_{\mathbb{X}} \nu(C_x) \mu(dx) = \int_{\mathbb{Y}} \mu(C^y) \nu(dy). \quad (5.1)$$

Démonstration. L'unicité de la mesure produit est une conséquence du théorème 2.2.20 de caractérisation des mesures. Supposons qu'il existe une autre mesure produit m' , alors pour tout $A \in \mathcal{X}$ et $B \in \mathcal{Y}$, on a

$$m'(A \times B) = \mu(A)\nu(B) = m(A \times B).$$

L'algèbre de Boole engendrée par les rectangles est constituée des réunions finies de rectangles disjoints, donc m' et m coïncident sur l'algèbre de Boole engendrée par les rectangles.

Les mesures μ et ν sont σ -finies, ainsi il existe $(X_n)_{n \geq 0}$ et $(Y_n)_{n \geq 0}$ des suites croissantes d'ensembles mesurables dans \mathcal{X} et \mathcal{Y} respectivement tels que $\mu(X_n) < \infty$ et $\nu(Y_n) < \infty$ pour tout $n \geq 0$, $\mathbb{X} = \cup_{n \geq 0} X_n$ et $\mathbb{Y} = \cup_{n \geq 0} Y_n$. La suite $(X_n \times Y_n)_{n \geq 0}$ est elle-même croissante exhaustive et satisfait

$$m(X_n \times Y_n) = \mu(X_n)\nu(Y_n) = m'(X_n \times Y_n) < \infty.$$

Les mesures m et m' sont donc σ -finies et donc coïncident sur la tribu engendrée par les rectangles, c'est à dire la tribu produit.

Pour l'existence de la mesure produit, nous allons considérer la fonction d'ensembles suivante

$$\forall C \in \mathcal{X} \otimes \mathcal{Y}, \quad m(C) = \int_{\mathbb{X}} \nu(C_x) \mu(dx). \quad (5.2)$$

Pour que cette application soit bien définie, il faut tout d'abord montrer le lemme suivant qui correspond à la première partie du deuxième point du théorème.

Lemme 5.1.3. Si $C \in \mathcal{X} \otimes \mathcal{Y}$, l'application $x \rightarrow \nu(C_x)$ est mesurable sur $(\mathbb{X}, \mathcal{X})$ et l'application $y \rightarrow \mu(C^y)$ est mesurable sur $(\mathbb{Y}, \mathcal{Y})$.

Démonstration. Il suffit de montrer la première assertion. Supposons dans un premier temps que ν est finie. Soit \mathcal{C} l'ensemble des parties $C \in \mathcal{X} \otimes \mathcal{Y}$ telles que $x \rightarrow \nu(C_x)$ soit mesurable. Nous allons montrer que \mathcal{C} est un λ -système contenant l'algèbre de Boole, notée \mathcal{B} , engendrée par les rectangles. En effet, comme le plus petit λ -système contenant \mathcal{B} (qui est stable par intersections finies) est la tribu engendrée par \mathcal{B} (mais aussi par les rectangles), c'est donc que $\mathcal{C} = \mathcal{X} \otimes \mathcal{Y}$.

Étape 1 : $\mathcal{B} \subset \mathcal{C}$.

Si $C = A \times B \in \mathcal{X} \otimes \mathcal{Y}$, alors $\nu(C_x) = \mathbf{1}_A(x)\nu(B)$ et donc $C \in \mathcal{C}$. Si $C = \cup_{i=1}^n C^i$ où les $(C^i)_{1 \leq i \leq n}$ sont des rectangles mesurables deux à deux disjoints, on a $\nu(C_x) = \sum_i \nu(C_x^i)$ et $x \rightarrow \nu(C_x)$ est mesurable en tant que somme de fonctions mesurables. D'où $\mathcal{B} \subset \mathcal{C}$.

Étape 2 : \mathcal{C} est un λ -système.

Il est clair que $C = \mathbb{X} \times \mathbb{Y} \in \mathcal{C}$ car alors $C_x = \mathbb{Y}$ et donc $x \rightarrow \nu(C_x) = \nu(\mathbb{Y})$ est mesurable. Soit $(C^n)_{n \geq 0}$ une suite croissante d'éléments de \mathcal{C} et C sa réunion. Pour tout $x \in \mathbb{X}$, le théorème de convergence monotone appliqué à la suite croissante $(\mathbf{1}_{C_x^n})_{n \geq 0}$ implique que $(\nu(C_x^n))_{n \geq 0}$ converge vers $\nu(C_x)$. Donc $x \rightarrow \nu(C_x)$ est mesurable en tant que limite simple d'une suite de fonctions mesurables. Enfin, si C et D sont dans \mathcal{C} avec $C \subset D$, alors $(D \setminus C)_x = D_x \setminus C_x$ et comme ν est supposée finie, $x \rightarrow \nu((D \setminus C)_x) = \nu(D_x) - \nu(C_x)$ est mesurable comme différence de fonctions mesurables.

Si ν est seulement σ -finie, soit $(Y_n)_{n \geq 0}$ une suite croissante exhaustive d'éléments de \mathcal{Y} telle que $\nu(Y_n) < \infty$ pour tout $n \geq 0$. Soit $C \in \mathcal{X} \otimes \mathcal{Y}$ et, pour tout $n \geq 0$, posons $C^n = C \cap (\mathbb{X} \times Y_n)$. D'après la première partie de la démonstration l'application $x \rightarrow \nu(C_x^n)$ est mesurable. Par convergence monotone, il en est de même pour $x \rightarrow \nu(C_x)$. \square

Désormais, on a justifié que la quantité donnée dans (5.2) est bien définie. Montrons qu'il s'agit d'une mesure. Il est clair que $m(\emptyset) = 0$. Soit $(C^n)_{n \geq 0}$ une suite d'éléments de $\mathcal{X} \otimes \mathcal{Y}$ deux à deux disjoints et C leur réunion. On a $C_x = \cup_n C_x^n$ avec $(C_x^n)_{n \geq 0}$ deux à deux disjoints dans \mathcal{Y} , d'où $\nu(C_x) = \sum_n \nu(C_x^n)$.

Encore une fois, le théorème de convergence monotone appliqué aux sommes partielles de la séries de fonctions $\sum_n \mathbf{1}_{C_x^n}$ implique

$$\int_{\mathbb{X}} \nu(C_x) \mu(dx) = \sum_{n \geq 0} \int_{\mathbb{X}} \nu(C_x^n) \mu(dx) = \sum_{n \geq 0} m(C^n).$$

Ainsi, m est une mesure et il reste à vérifier qu'elle affecte la mesure souhaitée aux rectangles. Si $C = A \times B \in \mathcal{X} \otimes \mathcal{Y}$, on a

$$m(C) = \int_{\mathbb{X}} \nu(C_x) \mu(dx) = \int_{\mathbb{X}} \mathbf{1}_A(x) \nu(B) \mu(dx) = \mu(A) \nu(B).$$

De même, on montre que $C \rightarrow \int_{\mathbb{Y}} \mu(C^y) \nu(dy)$ définit une mesure sur $\mathcal{X} \otimes \mathcal{Y}$ qui coïncide avec m sur les rectangles. Par unicité, cette mesure est égale à m et l'on obtient la relation (5.1). \square

Si $(\mathbb{X}_i, \mathcal{X}_i, \mu_i)_{1 \leq i \leq N}$ sont N espaces mesurés σ -finis, on peut vouloir définir une mesure produit π sur $(\prod_{i=1}^N \mathbb{X}_i, \otimes \mathcal{X}_i)$ telle que, pour tout $A = A_1 \times \dots \times A_N$, $A_i \in \mathcal{X}_i$, $1 \leq i \leq N$, $\pi(A) = \mu_1(A_1) \dots \mu_N(A_N)$. Pour se faire, on peut procéder par étapes : pour le cas $N = 3$, on peut commencer par construire $\mu_1 \otimes \mu_2$, puis $(\mu_1 \otimes \mu_2) \otimes \mu_3$. On peut cependant procéder différemment et construire $\mu_2 \otimes \mu_3$, puis $\mu_1 \otimes (\mu_2 \otimes \mu_3)$. Ces deux constructions définissent-elles une même mesure ? La réponse est heureusement oui.

Proposition 5.1.4. *Le produit tensoriel de mesure est associatif.*

Exercice 21. Démontrer la proposition 5.1.4. On pourra pour cela utiliser le théorème de caractérisation des mesures.

Si \mathbb{X} et \mathbb{Y} sont des espaces topologiques, on peut les munir de leurs tribus boréliennes. Sur le produit $\mathbb{X} \times \mathbb{Y}$, on peut donner *a priori* plusieurs structures mesurables : soit on munit le produit de la tribu produit des tribus boréliennes, c'est à dire $\mathcal{B}(\mathbb{X}) \otimes \mathcal{B}(\mathbb{Y})$; soit on munit le produit de la tribu borélienne issue de la topologie produit. Ces deux tribus sont-elles identiques ?

Proposition 5.1.5. *Soient \mathbb{X} et \mathbb{Y} deux espaces métriques séparables. Alors $\mathcal{B}(\mathbb{X}) \otimes \mathcal{B}(\mathbb{Y}) = \mathcal{B}(\mathbb{X} \times \mathbb{Y})$ où $\mathbb{X} \times \mathbb{Y}$ est muni de la topologie produit.*

Exercice 22. Démontrer la proposition 5.1.5. On pourra montrer que tout ouvert est réunion dénombrable de pavés ouverts.

Théorème 5.1.6 (Mesure de Lebesgue sur \mathbb{R}^d). *Il existe une unique mesure λ_d sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ telle que, pour tout produit d'intervalles $I_1 \times \dots \times I_d$, $\lambda_d(I_1 \times \dots \times I_d)$ soit égal au produit des longueurs des intervalles $(I_j)_{j=1, \dots, d}$. De plus, λ_d est le produit tensoriel répété d fois de la mesure de Lebesgue λ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on note $\lambda_d = \lambda^{\otimes d}$. Cette mesure est appelée mesure de Lebesgue sur \mathbb{R}^d . Enfin, la mesure λ_d est l'unique mesure sur $\mathcal{B}(\mathbb{R}^d)$ telle que*

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

5.2 Théorèmes de Fubini-Tonelli et de Fubini-Lebesgue

On remarque que l'égalité (5.1) s'écrit encore

$$\int_{\mathbb{X} \times \mathbb{Y}} \mathbf{1}_C d\mu \otimes \nu = \int_{\mathbb{X}} \left[\int_{\mathbb{Y}} \mathbf{1}_C(x, y) \nu(dy) \right] \mu(dx) = \int_{\mathbb{Y}} \left[\int_{\mathbb{X}} \mathbf{1}_C(x, y) \mu(dx) \right] \nu(dy). \quad (5.3)$$

Ainsi, calculer l'intégrale de la fonction indicatrice d'un élément de la tribu produit revient à intégrer l'intégrale des sections, l'ordre d'intégration ne jouant aucun rôle.

Le théorème de Fubini-Tonelli, qu'on appellera plus simplement théorème de Tonelli, montre que ce fait reste vrai pour les fonctions mesurables positives, ce qui ne devrait pas nous étonner au vu de la construction de l'intégrale de Lebesgue.

Théorème 5.2.1 (Fubini-Tonelli). *Soit f une fonction mesurable de $(\mathbb{X} \times \mathbb{Y}, \mathcal{X} \otimes \mathcal{Y})$ dans $\overline{\mathbb{R}}_+$ et soient μ et ν deux mesures σ -finies respectivement sur $(\mathbb{X}, \mathcal{X})$ et $(\mathbb{Y}, \mathcal{Y})$. Alors,*

1. les fonctions partout définies $x \rightarrow \int_{\mathbb{Y}} f(x, y) \nu(dy)$ et $y \rightarrow \int_{\mathbb{X}} f(x, y) \mu(dx)$ sont respectivement \mathcal{X} et \mathcal{Y} -mesurables.
2. les égalités suivantes ont lieu dans $\overline{\mathbb{R}_+}$:

$$\int_{\mathbb{X} \times \mathbb{Y}} f \, d\mu \otimes \nu = \int_{\mathbb{X}} \left[\int_{\mathbb{Y}} f(x, y) \nu(dy) \right] \mu(dx) = \int_{\mathbb{Y}} \left[\int_{\mathbb{X}} f(x, y) \mu(dx) \right] \nu(dy). \quad (5.4)$$

Démonstration. Nous montrons dans un même temps le point (1) pour la première fonction et la première égalité du point (2). La stratégie de preuve est similaire à ce qu'elle était pour la construction de l'intégrale, elle se fait en trois étapes (indicatrices, fonction étagées positives et mesurables positives). À chaque étape, on doit montrer

1. pour tout $x \in \mathbb{X}$, $y \rightarrow f(x, y)$ est \mathcal{Y} -mesurable et positive,
2. $x \rightarrow \int_{\mathbb{Y}} f(x, y) \nu(dy)$ est \mathcal{X} -mesurable et positive,
3. la relation (5.4) est vérifiée par f .

Étape 1 : Si f est l'indicatrice d'un élément C de $\mathcal{X} \otimes \mathcal{Y}$, alors le point (1) est assuré par le lemme 5.1.1 puisque $y \rightarrow f(x, y)$ est en fait l'application $y \rightarrow \mathbf{1}_{C_x}(y)$; le point (2) est assuré par le lemme 5.1.3 et l'égalité (5.3) n'est rien d'autre que l'égalité (5.1) montrée dans le théorème 5.1.2.

Étape 2 : Si f est une fonction étagée positive, le résultat découle de la linéarité de l'intégrale et de la stabilité de la mesurabilité par combinaison linéaire.

Étape 3 : Si f est mesurable positive, il existe une suite $(f_n)_{n \geq 0}$ croissante de fonctions étagées positives qui converge simplement vers f . Donc, pour tout $x \in \mathbb{X}$, $(y \rightarrow f_n(x, y))_{n \geq 0}$ est suite de fonctions mesurables positives qui converge vers $y \rightarrow f(x, y)$. Le théorème de convergence monotone assure

$$\int_{\mathbb{Y}} f(x, y) \nu(dy) = \int_{\mathbb{Y}} f_n(x, y) \nu(dy) = \lim_{n \rightarrow \infty} \int_{\mathbb{Y}} f_n(x, y) \nu(dy).$$

Pour chaque $n \geq 0$, par l'étape 2, la fonction $x \rightarrow \int_{\mathbb{Y}} f_n(x, y) \nu(dy)$ est mesurable positive. Donc, la fonction $x \rightarrow \int_{\mathbb{Y}} f(x, y) \nu(dy)$ est mesurable positive comme limite de fonctions mesurables positives. De plus,

$$\begin{aligned} \int_{\mathbb{X} \times \mathbb{Y}} f \, d\mu \otimes \nu &\stackrel{\text{CM}}{=} \lim_{n \rightarrow \infty} \int_{\mathbb{X} \times \mathbb{Y}} f_n \, d\mu \otimes \mu \stackrel{\text{étape 2}}{=} \lim_{n \rightarrow \infty} \int_{\mathbb{X}} \left[\int_{\mathbb{Y}} f_n(x, y) \nu(dy) \right] \mu(dx) \\ &\stackrel{\text{CM}}{=} \int_{\mathbb{X}} \lim_{n \rightarrow \infty} \left[\int_{\mathbb{Y}} f_n(x, y) \nu(dy) \right] \mu(dx) \stackrel{\text{CM}}{=} \int_{\mathbb{X}} \left[\int_{\mathbb{Y}} f(x, y) \nu(dy) \right] \mu(dx), \end{aligned}$$

ce qui achève la preuve. □

Corollaire 5.2.2. Soit f une fonction mesurable sur $(\mathbb{X} \times \mathbb{Y}, \mathcal{X} \otimes \mathcal{Y})$ à valeurs complexes. Alors f est intégrable pour la mesure $\mu \otimes \nu$ si et seulement si l'une des deux conditions suivantes est satisfaite

$$\int_{\mathbb{X}} \left[\int_{\mathbb{Y}} |f(x, y)| \nu(dy) \right] \mu(dx) < \infty \quad \text{ou} \quad \int_{\mathbb{Y}} \left[\int_{\mathbb{X}} |f(x, y)| \mu(dx) \right] \nu(dy) < \infty.$$

Démonstration. C'est l'application du théorème de Tonelli à la fonction positive $|f|$. □

Théorème 5.2.3 (Fubini-Lebesgue). Soit f une fonction intégrable sur $(\mathbb{X} \times \mathbb{Y}, \mathcal{X} \otimes \mathcal{Y}, \mu \otimes \nu)$. Alors,

1. pour presque tout $x \in \mathbb{X}$, la fonction $y \mapsto f(x, y)$ est dans $\mathcal{L}^1(\nu)$; de plus la fonction $x \mapsto \int_{\mathbb{Y}} f(x, y) \nu(dy)$, définie μ -p.p., est μ -intégrable.
2. pour presque tout $y \in \mathbb{Y}$, la fonction $x \mapsto f(x, y)$ est dans $\mathcal{L}^1(\mu)$; de plus la fonction $y \mapsto \int_{\mathbb{X}} f(x, y) \mu(dx)$, définie ν -p.p., est ν -intégrable.
3. Enfin,

$$\int_{\mathbb{X} \times \mathbb{Y}} f \, d\mu \otimes \nu = \int_{\mathbb{X}} \left[\int_{\mathbb{Y}} f(x, y) \nu(dy) \right] \mu(dx) = \int_{\mathbb{Y}} \left[\int_{\mathbb{X}} f(x, y) \mu(dx) \right] \nu(dy).$$

Démonstration. On montre le point (1) et la première égalité de (3) pour une fonction à valeur dans \mathbb{R} . D'après le théorème de Tonelli et l'hypothèse d'intégrabilité, on a

$$\infty > \int_{\mathbb{X} \times \mathbb{Y}} |f| \, d\mu \otimes \nu = \int_{\mathbb{X}} \left[\int_{\mathbb{Y}} |f(x, y)| \, \nu(dy) \right] \mu(dx).$$

L'inégalité de Markov implique que l'application $x \rightarrow \int_{\mathbb{Y}} |f(x, y)| \, \nu(dy)$ est μ -presque partout finie, on note N l'ensemble négligeable sur lequel elle est infinie. Si $x \notin N$, l'application $y \rightarrow f(x, y)$ est ν -intégrable. On décompose f en la différence de la partie positive et de la partie négative : $f = f^+ - f^-$. Si $x \notin N$, les applications $y \rightarrow f^+(x, y)$ et $y \rightarrow f^-(x, y)$ sont ν -intégrables et on a

$$\forall x \in N^c, \quad \int f(x, y) \, \nu(dy) = \int f^+(x, y) \, \nu(dy) - \int f^-(x, y) \, \nu(dy).$$

D'après le théorème de Tonelli, les fonctions $x \rightarrow \int f^\pm(x, y) \, \nu(dy)$ sont mesurables sur $\mathbb{X} \setminus N$ muni de la tribu induite et

$$\int_{\mathbb{X} \setminus N} \left[\int_{\mathbb{Y}} f^\pm(x, y) \, \nu(dy) \right] \mu(dx) = \int_{\mathbb{X}} \left[\int_{\mathbb{Y}} f^\pm(x, y) \, \nu(dy) \right] \mu(dx) = \int_{\mathbb{X} \times \mathbb{Y}} f^\pm \, d\mu \otimes \nu < \infty.$$

Par conséquent, l'application $x \rightarrow \int f(x, y) \, \nu(dy)$ définie sur $\mathbb{X} \setminus N$ est intégrable comme combinaison linéaire de deux fonctions intégrables. On a enfin

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

Ceci termine la preuve du théorème. □

5.3 La mesure produit en application

Exemple 21 (Normalisation de la gaussienne). On se propose de montrer que

$$I = \int_0^\infty e^{-x^2/2} \, dx = \sqrt{\frac{\pi}{2}}. \quad (5.5)$$

On définit f sur \mathbb{R}_+^2 par $f(x, y) = y \exp(y^2(1 + x^2)/2)$. La fonction f est continue donc mesurable, de plus elle est positive sur \mathbb{R}_+^2 donc le théorème de Tonelli s'applique. Or, d'une part,

$$\int_{\mathbb{R}_+} f(x, y) \, dy = \left[\frac{\exp(-y^2(1 + x^2)/2)}{1 + x^2} \right]_0^\infty = \frac{1}{1 + x^2}.$$

Donc

$$\int_{\mathbb{R}_+} \left(\int_{\mathbb{R}_+} f(x, y) \, dy \right) dx = \int_{\mathbb{R}_+} \frac{1}{1 + x^2} = \frac{\pi}{2}.$$

D'autre part, pour $y > 0$, à l'aide du changement de variable $u = xy$

$$\int_{\mathbb{R}_+} f(x, y) \, dx = e^{-y^2/2} \int_{\mathbb{R}_+} e^{-(xy)^2/2} y \, dx = e^{-y^2/2} \int_{\mathbb{R}_+} e^{-u^2/2} \, du = I e^{-y^2/2}.$$

En intégrant par rapport à la variable y , on obtient

$$\int_{\mathbb{R}_+} \left(\int_{\mathbb{R}_+} f(x, y) \, dx \right) dy = I^2.$$

Le théorème de Tonelli implique $I^2 = \pi/2$ et par positivité de l'intégrande dans (5.5) on obtient le résultat.

Remarque 41. Dans le calcul, nous avons supposé, pour que le changement de variable soit inversible, que $y > 0$. Il n'est pas nécessaire de considérer le cas $y = 0$ puisque

$$\int_{\mathbb{R}_+} f(x, y) dx \mathbf{1}_{\mathbb{R}_+}(y) = \int_{\mathbb{R}_+} f(x, y) dx \mathbf{1}_{\mathbb{R}_+^*}(y), \quad \text{p.p.}$$

Si dans l'exemple ci-dessus, on a profité de la positivité de la fonction f pour appliquer le théorème de Tonelli, on considère ci-dessous un exemple de fonction à intégrer qui n'est pas de signe constant. Dans ce cas, on commence par étudier l'intégrabilité de la valeur absolue de la fonction à intégrer en utilisant le théorème de Tonelli. Une fois l'intégrabilité assurée, on applique le théorème de Fubini.

Exemple 22. On veut calculer

$$\int_{\mathbb{R}_+^2} \sin(xy) \exp\{-(x+y)\} dx dy.$$

On trouve facilement une majoration de la valeur absolue

$$\int_{\mathbb{R}_+^2} |\sin(xy)| \exp\{-(x+y)\} dx dy \leq \int_{\mathbb{R}_+^2} \exp\{-(x+y)\} dx dy.$$

Dans l'intégrale de droite, on peut appliquer le théorème de Tonelli, d'où

$$\int_{\mathbb{R}_+^2} \exp\{-(x+y)\} dx dy = \left(\int_{\mathbb{R}_+} e^{-x} dx \right) \left(\int_{\mathbb{R}_+} e^{-y} dy \right) = 1.$$

On en déduit l'intégrabilité de la fonction de départ. Par le théorème de Fubini, on obtient (c'est un exemple, on peut bien sûr intervertir le rôle de x et y si le calcul est facilité)

$$\int_{\mathbb{R}_+^2} \sin(xy) \exp\{-(x+y)\} dx dy = \int_{\mathbb{R}_+} \left[\int_{\mathbb{R}_+} \sin(xy) \exp\{-(x+y)\} dx \right] dy$$

En intégrant par parties deux fois (on intègre l'exponentielle), on montre que

$$J(y) = \int_{\mathbb{R}_+} \sin(xy) e^{-x} dx = \frac{y}{1+y^2}.$$

Ainsi, on obtient finalement

$$\int_{\mathbb{R}_+^2} \sin(xy) \exp\{-(x+y)\} dx dy = \int_{\mathbb{R}_+} \frac{e^{-y} y}{1+y^2} dy.$$

Exemple 23. On veut calculer

$$\int_{\mathbb{R}_+^2} \sin(y) \exp\{-(x+y)\} dx dy.$$

On trouve facilement une majoration de la valeur absolue

$$\int_{\mathbb{R}_+^2} |\sin(y)| \exp\{-(x+y)\} dx dy \leq \int_{\mathbb{R}_+^2} \exp\{-(x+y)\} dx dy.$$

Dans l'intégrale de droite, on peut appliquer le théorème de Tonelli, d'où

$$\int_{\mathbb{R}_+^2} \exp\{-(x+y)\} dx dy = \left(\int_{\mathbb{R}_+} e^{-x} dx \right) \left(\int_{\mathbb{R}_+} e^{-y} dy \right) = 1.$$

On en déduit l'intégrabilité de la fonction de départ. Par le théorème de Fubini, on obtient

$$\int_{\mathbb{R}_+^2} \sin(y) \exp\{-(x+y)\} dx dy = \left(\int_{\mathbb{R}_+} \sin(y) e^{-y} dy \right) \left(\int_{\mathbb{R}_+} e^{-x} dx \right).$$

De plus, en intégrant par partie deux fois, on obtient

$$\int_{\mathbb{R}_+} \sin(y)e^{-y} dy = 1 - \int_{\mathbb{R}_+} \sin(y)e^{-y} dy = \frac{1}{2}.$$

L'intégrale voulue initialement vaut donc $1/2$.

Exemple 24. Soit $f(x, y) = xy^2$ et Δ le domaine intérieur au triangle ABC avec $A = (0, -1)$, $B = (1, 3)$ et $C = (0, 1)$. La fonction f est continue sur Δ compact donc f est bornée disons par $M \geq 0$. On voit facilement par croissance de l'intégrale

$$\int_{\Delta} |f(x, y)| dx dy \leq M\lambda(\Delta) < \infty.$$

On peut donc appliquer le théorème de Fubini. Un calcul simple donne

$$\int_{\mathbb{R}^2} \mathbf{1}_{\Delta}(x, y) f(x, y) dx dy = \int_{\mathbb{R}^2} \mathbf{1}_{[-1, 1]}(x) \mathbf{1}_{[0, \frac{3}{2}(x+1)]}(y) f(x, y) dx dy = \int_{-1}^1 x \int_0^{\frac{3}{2}(x+1)} y^2 dy dx.$$

Ainsi,

$$\int_{\mathbb{R}^2} \mathbf{1}_{\Delta}(x, y) f(x, y) dx dy = \frac{9}{8} \int_{-1}^1 \left[\frac{x^5}{5} + 3\frac{x^4}{4} + x^3 + \frac{x^2}{2} \right] dx = \frac{27}{10}.$$

5.4 Mesure image et changement de variables

On commence par rappeler le théorème de transfert.

Théorème 5.4.1 (Théorème de transfert). Soient $\phi : (\Omega, \mathcal{F}, \nu) \rightarrow (\mathbb{X}, \mathcal{E})$. Soit $f : (\mathbb{X}, \mathcal{E}) \rightarrow \overline{\mathbb{R}}_+$ mesurable. Alors,

$$\int_{\mathbb{X}} f d\phi_*\nu = \int_{\Omega} f \circ \phi d\nu. \quad (5.6)$$

Si f est à valeurs complexes, alors f est $\phi_*\nu$ -intégrable si et seulement si $f \circ \phi$ est ν -intégrable et on a l'égalité 5.6 dans \mathbb{C} .

Le théorème de transfert donne donc une formule de changement de variable théorique. Cependant, cela reste peu exploitable en pratique puisque la mesure $\phi_*\nu$ n'est pas explicite.

Définition 5.4.2 (Mesures absolument continues). Soit $(\mathbb{X}, \mathcal{X})$ un espace mesurable et μ et ν deux mesures sur $(\mathbb{X}, \mathcal{X})$. La mesure ν est dite absolument continue par rapport à μ si, pour tout $A \in \mathcal{X}$, $\mu(A) = 0$ implique $\nu(A) = 0$. On note $\nu \prec \mu$.

Proposition 5.4.3. Soit $(\mathbb{X}, \mathcal{X}, \mu)$ un espace mesuré et f une fonction mesurable positive sur $(\mathbb{X}, \mathcal{X})$. Si ν est à densité f par rapport à μ , alors ν est absolument continue par rapport à μ .

Démonstration. Si $A \in \mathcal{X}$ est tel que $\mu(A) = 0$ alors $f\mathbf{1}_A$ est nulle μ -presque partout et

$$\nu(A) = \int f\mathbf{1}_A d\mu = 0.$$

□

Le théorème de Radon-Nikodým énoncé et démontré au chapitre 6 établit une réciproque à cette proposition dans le cas σ -fini : si $\nu \prec \mu$ alors ν est à densité par rapport à μ .

Lorsque ν est absolument continue par rapport à la mesure de Lebesgue et ϕ un C^1 -difféomorphisme, on peut donner une expression explicite de $\phi_*\nu$. Le lemme suivant, utile pour démontrer le théorème de changement de variable général, donne une telle expression pour le cas où ϕ est un automorphisme linéaire.

Lemme 5.4.4. Soient A une matrice inversible de taille $d \times d$ et $b \in \mathbb{R}^d$ un vecteur. Soit $f : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}_+$ une fonction mesurable positive. Alors

$$\int_{\mathbb{R}^d} f(Ax + b) \lambda(dx) = \frac{1}{|\det A|} \int_{\mathbb{R}^d} f(x) d\lambda(x).$$

Remarque 42. Ce lemme est aussi valable pour une fonction mesurable à valeurs dans \mathbb{C} sous condition d'intégrabilité.

Remarque 43. En d'autres termes, le lemme établit que la mesure image de la mesure de Lebesgue λ_d par l'application affine $\phi : x \rightarrow Ax + b$ est donnée par

$$\phi_* \lambda_d = |\det A|^{-1} \lambda_d.$$

Démonstration. Puisque la mesure de Lebesgue est invariante par translation et que A est linéaire, on peut supposer d'abord $b = 0$. Il s'agit d'identifier la mesure $\nu = \phi_* \lambda_d$. On commence par montrer qu'elle est proportionnelle à la mesure de Lebesgue λ_d .

Soient $a \in \mathbb{R}^d$ et $B \in \mathcal{B}(\mathbb{R}^d)$, alors

$$\nu(a + B) = \lambda(A^{-1}a + A^{-1}B) = \lambda(A^{-1}B) = \nu(B).$$

Ainsi, ν est invariante par translation. D'autre part, $\nu([0, 1]^d) \neq 0$ car

$$\nu(\mathbb{R}^d) = \lambda(\mathbb{R}^d) = \infty \quad \text{et} \quad \nu(\mathbb{R}^d) \leq \sum_{n \in \mathbb{Z}^d} \nu(n + [0, 1]^d) = \sum_{n \in \mathbb{Z}^d} \nu([0, 1]^d).$$

Enfin, on montre que $\nu([0, 1]^d) = \lambda(A^{-1}[0, 1]^d) < \infty$ car $A^{-1}[0, 1]^d$ est compact. Ceci montre que la mesure $\mu = \nu/\nu([0, 1]^d)$ est invariante par translation et vérifie $\mu([0, 1]^d) = 1$, il s'agit donc de la mesure de Lebesgue. Aussi il existe une constante $c(A) \in \mathbb{R}_+^*$ telle que $\nu = c\lambda_d$. Il s'agit de montrer que $c(A) = |\det A|^{-1}$. Si A est inversible, la méthode du pivot de Gauss consiste à multiplier à gauche par des matrices élémentaires M_1, \dots, M_k de sorte que $M_k \cdots M_1 A = I$ et donc que $A^{-1} = M_k \cdots M_1$. Les matrices élémentaires permettent de permuter des lignes (type 1), multiplier une ligne par un scalaire non nul (type 2), ajouter une ligne à une autre ligne (type 3). On remarque que

$$A^{-1}B = M_k \cdots M_1 B \quad \text{et donc} \quad c(A) = c(M_k^{-1}) \cdots c(M_1^{-1}).$$

La preuve sera terminée si l'on montre que $c(M) = |\det M|^{-1}$ dans le cas où M est une matrice de type 1, 2 et 3. Il faut en effet juste remarquer que si M est de type 1 (*resp.* de type 2) alors M^{-1} est de type 1 (*resp.* de type 2) alors que si M est de type 3, M^{-1} est le produit d'une matrice de type 2 et de type 3. Or,

1. si A est matrice de type 1, alors

$$\nu(A) = \lambda_d(A^{-1}[0, 1]^d) = \lambda_d([0, 1]^d) = 1 = |\det A|^{-1};$$

2. si A est une matrice de type 2, $\alpha \in \mathbb{R}_+^*$, alors

$$\nu(A) = \lambda_d(A^{-1}[0, 1]^d) = \lambda([0, 1]^{d-1} \times [0, \alpha^{-1}]) = |\alpha|^{-1} = |\det A|^{-1};$$

Si $\alpha \in \mathbb{R}_-^*$, par le même raisonnement, on obtient le même résultat.

3. si A est une matrice de type 3, alors, à une permutation de la base de \mathbb{R}^d près, c'est à dire à des permutations de lignes près,

$$A = \begin{pmatrix} P & 0 \\ 0 & I \end{pmatrix} \quad \text{avec} \quad P = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{ou} \quad P = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

De plus, on calcule $A[0, 1]^d = D \times [0, 1]^{d-2}$ avec D le parallélogramme de \mathbb{R}^2 dont les sommets ont pour coordonnées $(0, 0)$, $(0, 1)$, $(1, 1)$ et $(1, 2)$ pour le premier cas et $(0, 0)$, $(1, 0)$, $(1, 1)$ et $(2, 1)$ dans le second cas. Dans chacun des cas, ces parallélogrammes sont la réunion de deux triangles isocèles rectangle de côté 1. De fait,

$$\nu([0, 1]^d) = \lambda_d(A^{-1}[0, 1]^d) = \lambda_2(D) \lambda_{d-2}([0, 1]^{d-2}) = 1 = |\det(A)|^{-1}.$$

□

Définition 5.4.5. Soient U un ouvert de \mathbb{R}^p et $\phi : U \subset \mathbb{R}^p \rightarrow \mathbb{R}^q$ une application C^1 , i.e. les dérivées partielles de ϕ existes et sont continues sur U . On note $\phi = (\phi_1, \dots, \phi_q)$. Le jacobien de ϕ au point $a \in U$, noté $\text{Jac}_\phi(a)$, est une application linéaire de \mathbb{R}^p dans \mathbb{R}^q dont la matrice dans la base canonique est donnée par

$$\text{Jac}_\phi(a) = \begin{pmatrix} \partial_1 \phi_1(a) & \cdots & \partial_p \phi_1(a) \\ \vdots & & \vdots \\ \partial_1 \phi_q(a) & \cdots & \partial_p \phi_q(a) \end{pmatrix}$$

Remarque 44. Rappelons que la matrice jacobienne d'une application ϕ de classe C^1 est la matrice représentant le morphisme linéaire de \mathbb{R}^p dans \mathbb{R}^q au sens suivant :

$$\phi(a + h) = \phi(a) + \text{Jac}_\phi(a)h + o(\|h\|)$$

avec $h \in \mathbb{R}^p$ suffisamment petit de sorte que $a+h \in U$. Cette remarque donne un moyen mnémotechnique pour se rappeler de la forme de la jacobienne lorsque dimension de départ et dimension à l'arrivée diffèrent.

Avant d'énoncer le théorème de changement de variables, rappelons deux théorèmes importants du calcul différentiel : le théorème d'inversion locale et la caractérisation des difféomorphismes. Le lecteur trouvera la démonstration de ces deux théorèmes dans [Car67].

Théorème 5.4.6 (Inversion locale). *Soit D un ouvert de \mathbb{R}^d et $\phi : D \rightarrow \mathbb{R}^d$ une application de classe C^1 sur D . Si $a \in D$ est tel que $\text{Jac}_\phi(a)$ est inversible, alors il existe un voisinage ouvert V_a de a dans D tel que $\phi|_{V_a}$ soit un difféomorphisme de V_a sur son image ouverte $\phi(V_a)$.*

Théorème 5.4.7. *Soit D un ouvert de \mathbb{R}^d . La fonction $\phi : D \rightarrow \mathbb{R}^d$ est C^1 -difféomorphisme sur son image $\Delta = \phi(D)$ si et seulement si elle vérifie*

1. ϕ est injective sur D ,
2. ϕ est de classe C^1 sur D ,
3. Jac_ϕ est inversible en tout point $a \in D$.

Dans ce cas, Δ est un ouvert de \mathbb{R}^d et, pour tout $a \in \Delta$, $\text{Jac}_{\phi^{-1}}(a) = \text{Jac}_\phi(\phi^{-1}(a))^{-1}$.

Théorème 5.4.8. *Soient D, Δ deux ouverts non vides de \mathbb{R}^d et ϕ un C^1 -difféomorphisme de D dans Δ . Si $f : \Delta \rightarrow \mathbb{R}$ est mesurable positive alors,*

$$\int_{\Delta} f(u) du = \int_D f(\phi(v)) |\det \text{Jac}_\phi(v)| dv. \quad (5.7)$$

Si $f : \Delta \rightarrow \mathbb{C}$ est mesurable, alors $f \circ \phi |\det \text{Jac}_\phi|$ est intégrable si et seulement si f est intégrable et l'égalité (5.7) a lieu dans \mathbb{C} .

Remarque 45. Formellement, si $u = \phi(v)$ est un changement de variable alors

$$du = d\phi(v) = |\text{Jac}_\phi(v)| dv.$$

Démonstration. Le changement de variable pour un automorphisme linéaire n'est rien d'autre que le lemme 5.4.4. La preuve dans le cadre générale est longue et fastidieuse (voir [BP04] par exemple), on peut néanmoins en donner une heuristique :

1. recouvrir le domaine D par des pavés de diamètre $< \delta$ petit ;
2. approcher la fonction ϕ sur chaque pavé par son jacobien ;
3. appliquer le théorème de transfert au jacobien à l'aide du lemme 5.4.4 ;
4. faire tendre δ vers 0 en contrôlant uniformément les restes.

□

Exemple 25 (Densité gaussienne). On cherche à montrer que $I = \int_{\mathbb{R}^2} e^{-x^2/2} dx dy = \sqrt{2\pi}$. Pour cela, une application du théorème de Tonelli implique

$$\int_{\mathbb{R}^2} \underbrace{e^{-\frac{x^2+y^2}{2}}}_{=f(x,y)} dx dy = \int_{\mathbb{R}} e^{-x^2/2} dx \int_{\mathbb{R}} e^{-y^2/2} dy = I^2$$

D'autre part, on fait le changement de variable $(x, y) = \phi(\rho, \theta) = (\rho \cos(\theta), \rho \sin(\theta))$. L'application ϕ est un C^1 -difféomorphisme de $\mathbb{R}_+^* \times [0, 2\pi)$ dans $\mathbb{R}^2 \setminus \{0\}$. De plus,

$$\text{Jac}_\phi(\rho, \theta) = \begin{pmatrix} \cos(\theta) & -\rho \sin(\theta) \\ \sin(\theta) & \rho \cos(\theta) \end{pmatrix} \implies |\det \text{Jac}_\phi(\rho, \theta)| = |\rho| = \rho.$$

La formule du changement variable donne

$$\int_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dx dy = \int_{\mathbb{R}^2 \setminus \{0\}} e^{-\frac{x^2+y^2}{2}} dx dy = \int_{\mathbb{R}_+^* \times [0, 2\pi)} \rho e^{-\rho^2/2} d\rho d\theta = 2\pi \int_0^\infty \rho e^{-\rho^2/2} d\rho,$$

par le théorème de Tonelli encore une fois. Finalement, $I^2 = 2\pi$.

Remarque 46. Si il est primordial que ϕ soit un C^1 -difféomorphisme pour éviter les effets de type courbe de Peano, il est parfois possible d'enlever des points au domaine pour obtenir un tel C^1 -difféomorphisme. Typiquement, dans l'exemple ci-dessus, ϕ est un C^1 -difféomorphisme de $\mathbb{R}_+^* \times [0, 2\pi)$ dans $\mathbb{R}^2 \setminus \{0\}$ et non de $\mathbb{R}_+^* \times [0, 2\pi)$ dans \mathbb{R}^2 . Ce problème est levé en remarquant que $f \mathbf{1}_{\mathbb{R}^2} = f \mathbf{1}_{\mathbb{R}^2 \setminus \{0\}}$ presque partout.

Chapitre 6

Espaces \mathcal{L}^p et L^p

Dans ce chapitre, \mathbb{K} désigne indifféremment le corps des réels ou des complexes.

6.1 Généralités

Définition 6.1.1. Pour tout réel $p > 0$, on définit

$$\mathcal{L}_{\mathbb{K}}^p(\mathbb{X}, \mathcal{X}, \mu) = \left\{ f : (\mathbb{X}, \mathcal{X}) \rightarrow (\mathbb{K}, \mathcal{B}(\mathbb{K})) \text{ mesurable} : \int |f|^p d\mu < \infty \right\}.$$

Si il n'y a pas d'ambiguïtés, on notera plus simplement $\mathcal{L}_{\mathbb{K}}^p(\mu)$ voir \mathcal{L}^p .

Remarque 47. Si m désigne la mesure de comptage sur $(\mathbb{S}, \mathcal{P}(\mathbb{S}))$ où \mathbb{S} est un ensemble dénombrable, alors

$$\mathcal{L}_{\mathbb{K}}^p(m) = \ell_{\mathbb{K}}^p(\mathbb{S}) = \left\{ (a_s)_{s \in \mathbb{S}} : \sum_{s \in \mathbb{S}} |a_s|^p < \infty \right\}.$$

Proposition 6.1.2. Pour tout $p > 0$, $\mathcal{L}_{\mathbb{K}}^p(\mu)$ est un \mathbb{K} -espace vectoriel.

Démonstration. On vérifie que $\mathcal{L}_{\mathbb{K}}^p(\mu)$ est un sous espace vectoriel du \mathbb{K} -espace vectoriel des fonctions mesurables sur $(\mathbb{X}, \mathcal{X})$ à valeurs dans \mathbb{K} . Il est immédiat que la fonction nulle est dans $\mathcal{L}_{\mathbb{K}}^p(\mu)$. Soient $\lambda \in \mathbb{K}$ et $f, g \in \mathcal{L}_{\mathbb{K}}^p(\mu)$. Les majorations

$$|\lambda f + g|^p \leq (|\lambda||f| + |g|)^p \leq (2 \max\{|\lambda||f|, |g|\})^p \leq 2^p |\lambda|^p |f|^p + 2^p |g|^p$$

assurent que $\lambda f + g \in \mathcal{L}_{\mathbb{K}}^p(\mu)$. □

Proposition 6.1.3. Si $\mu(\mathbb{X}) < \infty$, alors, pour tout $p \in (0, q]$, $\mathcal{L}_{\mathbb{K}}^q(\mu) \subset \mathcal{L}_{\mathbb{K}}^p(\mu)$.

Démonstration. Si $p \in (0, q]$, alors $|f|^p \leq |f|^q \mathbf{1}_{|f|>1} + \mathbf{1}_{|f|\leq 1}$. Ainsi, dès que $f \in \mathcal{L}_{\mathbb{K}}^q(\mu)$, on obtient

$$\int |f|^p d\mu \leq \int |f|^q d\mu + \mu(\{|f| \leq 1\}) < \infty.$$

□

Exercice 23. Montrer que les inclusions sont strictes en général. Montrer que l'hypothèse de mesure finie est primordiale.

Proposition 6.1.4. Pour tout $p \in (0, q]$, $\ell_{\mathbb{K}}^p(\mathbb{S}) \subset \ell_{\mathbb{K}}^q(\mathbb{S})$.

Démonstration. Soient $p \in (0, q]$ et $(a_s)_{s \in \mathbb{S}}$ alors $S = \{s \in \mathbb{S} : |a_s| > 1\}^{\complement}$ est fini. Comme pour tout $s \in S^{\complement}$, $|a_s|^q \leq |a_s|^p$, on obtient que $\sum_{s \in \mathbb{S}} |a_s|^q < \infty$. □

Définition 6.1.5 (*p*-Norme). Soient $f : (\mathbb{X}, \mathcal{X}, \mu) \rightarrow \mathbb{K}$ et $p \geq 1$. On définit la quantité

$$\|f\|_p = \left(\int_{\mathbb{X}} |f|^p \right)^{1/p} \in \overline{\mathbb{R}}_+,$$

avec la convention $(+\infty)^{1/p} = +\infty$.

Pour $p = \infty$, on définit la quantité

$$\|f\|_\infty = \inf\{r \geq 0 : \mu(\{|f| > r\}) = 0\} = \text{ess sup}|f|.$$

Remarque 48. On peut aussi considérer le cas $p \in (0, 1)$ mais alors l'appellation "*p*-norme" est abusive car l'inégalité triangulaire n'est plus valide. On parle alors de quasi-norme que l'on éludera pour notre part.

6.2 Inégalités de Hölder et de Minkowski

Deux réels $p, q \in [1, \infty]$ sont dits conjugués si $\frac{1}{p} + \frac{1}{q} = 1$ avec la convention $1/\infty = 0$.

Théorème 6.2.1 (Inégalité de Hölder). Soient $f, g : (\mathbb{X}, \mathcal{X}, \mu) \rightarrow \mathbb{K}$ et $p, q \in [1, \infty]$ deux indices conjugués.

1. Si f, g sont positives, alors dans $\overline{\mathbb{R}}_+$

$$0 \leq \int fg \, d\mu \leq \|f\|_p \|g\|_q.$$

En outre, si $p, q \in (1, \infty)$ et $\|f\|_p$ et $\|g\|_q$ sont finis, l'inégalité est une égalité si et seulement il existe $\alpha, \beta > 0$ tels que $\alpha f^p = \beta g^q$ μ -presque partout.

2. Si $f \in \mathcal{L}_{\mathbb{K}}^p(\mu)$ et $g \in \mathcal{L}_{\mathbb{K}}^q(\mu)$, alors $fg \in \mathcal{L}_{\mathbb{K}}^1(\mu)$ et

$$\|fg\|_1 \leq \|f\|_p \|g\|_q.$$

En outre, si $p, q \in (1, \infty)$, l'inégalité est une égalité si et seulement si il existe $\alpha, \beta > 0$ tels que $\alpha|f|^p = \beta|g|^q$ μ -presque partout.

Démonstration. Si $p = 1$ et $q = \infty$ (ou l'inverse en échangeant les rôles de p et q), il suffit de remarquer que $0 \leq fg \leq f\|g\|_\infty$ μ -p.p. et d'intégrer cette inégalité.

Si $p, q \in (1, \infty)$, on commence par établir une inégalité utile dans la suite. Soient $\alpha \in (0, 1)$ et $x \in \mathbb{R}_+$, on pose $\phi_\alpha(x) = x^\alpha - \alpha x$. La fonction ϕ_α est dérivable sur \mathbb{R}_+^* et $\phi'_\alpha(x) = \alpha(x^{\alpha-1} - 1)$. D'où, $\phi'_\alpha < 0$ sur $(1, \infty)$ et $\phi'_\alpha > 0$ sur $(0, 1)$. Donc pour tout $x \in \mathbb{R}_+$, $\phi_\alpha(x) \leq \phi_\alpha(1)$ avec égalité si et seulement si $x = 1$. Aussi, $x^\alpha \leq \alpha x + 1 - \alpha$ avec égalité si et seulement si $x = 1$. En posant $x = u/v$ avec $u \geq 0$ et $v > 0$, il vient

$$u^\alpha v^{1-\alpha} \leq \alpha u + (1 - \alpha)v \quad \text{avec égalité si et seulement si } u = v. \quad (6.1)$$

Cette inégalité est encore vraie pour $u, v \in \mathbb{R}_+$.

On revient à la preuve du premier point. Si $\|f\|_p$ ou $\|g\|_q$ est nulle alors f ou g est nulle μ -presque partout et il en va de même pour fg . L'inégalité est alors triviale. De même si $\|f\|_p$ ou $\|g\|_q$ vaut ∞ . On suppose donc que ces deux quantités sont strictement positives et finies. On pose

$$\alpha = \frac{1}{p}, \quad \text{d'où } 1 - \alpha = \frac{1}{q}, \quad u = \frac{f^p}{\|f\|_p^p} \quad \text{et} \quad v = \frac{g^q}{\|g\|_q^q}.$$

D'après l'inégalité (6.1),

$$\frac{fg}{\|f\|_p \|g\|_q} \leq \frac{1}{p} \frac{f^p}{\|f\|_p^p} + \frac{1}{q} \frac{g^q}{\|g\|_q^q}.$$

En intégrant de chaque côté de l'inégalité contre μ , il vient

$$0 \leq \int fg \, d\mu \leq \|f\|_p \|g\|_q \left(\frac{1}{p} \int \frac{f^p}{\|f\|_p^p} \, d\mu + \frac{1}{q} \int \frac{g^q}{\|g\|_q^q} \, d\mu \right) = \|f\|_p \|g\|_q.$$

L'égalité a lieu si et seulement si $f/\|f\|_p = g/\|g\|_q$ μ -presque partout. Le deuxième point du théorème est immédiat. \square

Corollaire 6.2.2. Si μ est une mesure de probabilité, l'application $r \rightarrow \|f\|_r$ est croissante.

Théorème 6.2.3 (Inégalité de Minkowski). Si $p \in [1, \infty]$, alors, pour tout $f, g \in \mathcal{L}_{\mathbb{K}}^p(\mu)$,

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

L'égalité a lieu si et seulement si

- $f = 0$ μ -presque partout ou $g = \alpha f$ μ -presque partout, pour $\alpha \geq 0$ si $p > 1$
- $f = 0$ μ -presque partout ou $f\bar{g} \geq 0$ μ -presque partout si $p = 1$.

Démonstration. Si $\|f + g\|_p = 0$, l'inégalité est triviale. Sinon, on intègre par rapport à μ l'inégalité

$$|f + g|^p \leq |f||f + g|^{p-1} + |g||f + g|^{p-1} \quad \text{avec la convention } x^0 = 1 \text{ pour } x \geq 0.$$

On obtient alors

$$\|f + g\|_p^p \leq \int |f||f + g|^{p-1} d\mu + \int |g||f + g|^{p-1} d\mu.$$

Si $p = 1$, l'inégalité est établie. Sinon, puisque $(p - 1)q = p$, l'inégalité de Hölder assure que

$$\int |f||f + g|^{p-1} d\mu \leq \|f\|_p \left(\int |f + g|^{(p-1)q} d\mu \right)^{1/q} = \|f\|_p \|f + g\|_p^{p/q}.$$

Ainsi,

$$\|f + g\|_p^p \leq (\|f\|_p + \|g\|_p) \|f + g\|_p^{p/q}.$$

Il ne reste plus qu'à simplifier par $\|f + g\|_p^{p/q}$ qui est strictement positif et à remarquer $p - p/q = 1$ pour obtenir l'inégalité souhaitée.

L'inégalité pour $p = \infty$ est une conséquence immédiate de l'inégalité triangulaire pour la valeur absolue. \square

Remarque 49. L'inégalité de Minkowski n'est rien d'autre que l'inégalité triangulaire pour la p -norme. L'homogénéité est immédiate. Ainsi, $\|\cdot\|_p$ est une semi-norme. Pour que ce soit une norme, il faudrait que $\|f\|_p = 0$ implique $f = 0$, or f n'est nulle que μ -presque partout.

Il existe une façon simple de construire un espace vectoriel normé à partir de \mathcal{L}^p et $\|\cdot\|_p$: il suffit de considérer l'espace quotient \mathcal{L}^p/\sim où \sim est la relation d'équivalence d'égalité μ -presque partout : $f \sim g$ si et seulement si $f = g$ μ -presque partout.

Définition 6.2.4 (Espaces \mathbf{L}^p). L'espace $\mathbf{L}_{\mathbb{K}}^p(\mu)$ est défini comme l'espace $\mathcal{L}_{\mathbb{K}}^p(\mu)$ modulo l'égalité μ -presque partout : $\mathbf{L}_{\mathbb{K}}^p(\mu) = \mathcal{L}_{\mathbb{K}}^p(\mu)/\sim$. L'espace $\mathbf{L}_{\mathbb{K}}^p(\mu)$ muni de l'application $\|\cdot\|_p$ est un \mathbb{K} -espace vectoriel normé.

Théorème 6.2.5. Pour tout $p \geq 1$, l'espace vectoriel normé $(\mathbf{L}_{\mathbb{K}}^p(\mu), \|\cdot\|_p)$ est un espace de Banach.

Démonstration. Fixons $p \in [1, \infty)$ et considérons une suite de Cauchy $(f_n)_{n \geq 0}$ d'éléments dans $\mathbf{L}_{\mathbb{K}}^p(\mu)$. On peut trouver une sous-suite $(f_{n_k})_{k \geq 0}$ telle que pour tout $k \geq 0$: $\|f_{n_{k+1}} - f_{n_k}\|_p \leq 2^{-k}$. Pour prouver que $(f_n)_{n \geq 0}$ converge, il suffit de montrer que $(f_{n_k})_{k \geq 0}$ converge car les suites de Cauchy ont au plus une valeur d'adhérence (voir la proposition 1.2.83). Pour cela, posons

$$g = \sum_{k \geq 0} |f_{n_{k+1}} - f_{n_k}|.$$

La fonction g est mesurable et par convergence monotone ainsi que l'inégalité de Minkowski :

$$\|g\|_p \leq \sum_{k \geq 0} \|f_{n_{k+1}} - f_{n_k}\|_p < \infty.$$

La fonction g est donc finie μ -presque partout comme toute fonction dans $\mathbf{L}_{\mathbb{K}}^p(\mu)$. Ainsi, hors d'un certain ensemble négligeable, la série numérique $g_n(x) = \sum_{k \geq 0} (f_{n_{k+1}} - f_{n_k})(x)$ est absolument convergente, donc

convergente. Ainsi, en dehors de cet ensemble négligeable, la suite $(g_n)_{n \geq 0}$ converge donc simplement vers une certaine fonction f qui est, de ce fait, mesurable. On conclut en remarquant que f vérifie

$$\|f - f_{n_k}\|_p \leq \sum_{\ell \geq k} \|f_{n_{\ell+1}} - f_{n_\ell}\|_p \leq 2^{-k+1},$$

si bien que f est dans $\mathbf{L}_{\mathbb{K}}^p(\mu)$ et que $(f_{n_k})_{k \geq 0}$ ainsi que $(f_n)_{n \geq 0}$ converge vers f dans cet espace.

Considérons le cas $p = \infty$ et soit $(f_n)_{n \geq 0}$ une suite de Cauchy. Soit $\varepsilon > 0$, alors il existe $N \geq 0$ tel que pour tout $n, m \geq N$, $\|f_n - f_m\|_\infty \leq \varepsilon$. Cela implique l'existence d'ensemble $N_{n,m}$ tel que $\mu(N_{n,m}) = 0$ et si $x \notin N_{n,m}$, $|f_n(x) - f_m(x)| \leq \varepsilon$. Posons $N = \cup_{n,m \geq N} N_{n,m}$, alors $\mu(N) = 0$ et pour tout $n, m \geq 0$ et tout $x \in N^c$, $|f_n(x) - f_m(x)| \leq \varepsilon$. Ainsi, pour tout $x \in N^c$, $(f_n(x))_{n \geq 0}$ est une suite de Cauchy dans \mathbb{K} complet, elle converge donc vers un réel $f(x)$. Or,

$$|f(x)| \leq |f(x) - f_n(x)| + |f_n(x)| \leq \varepsilon + \sup_{n \geq N} \|f_n\|_\infty, \quad n \geq N, x \in N^c.$$

Pour conclure, il suffit de remarquer que $|\|f_n\|_\infty - \|f_m\|_\infty| \leq \varepsilon$ ainsi $(\|f_n\|_\infty)_{n \geq 0}$ est convergente donc bornée. \square

Remarque 50. On a montré au passage le fait suivant : si $(f_n)_{n \geq 0}$ est une suite de Cauchy dans $\mathbf{L}_{\mathbb{K}}^p(\mu)$, $p \in [1, \infty)$, alors il existe une suite extraite qui converge μ -presque partout.

6.3 Théorème de Radon-Nikodym

6.3.1 Un peu d'espace de Hilbert

Dans toute la suite $\mathbb{K} = \mathbb{R}$ ou \mathbb{C} .

Définition 6.3.1. Un produit scalaire sur un \mathbb{K} -espace vectoriel E est une application $\langle \cdot, \cdot \rangle : E \times E \rightarrow \mathbb{K}$ satisfaisant

1. pour tout $y \in E$, l'application de E dans \mathbb{K} qui à $x \in E$ associe $\langle x, y \rangle$ est linéaire ;
2. — si $\mathbb{K} = \mathbb{R}$ alors $\langle x, y \rangle = \langle y, x \rangle$ pour tout $x, y \in E$;
— si $\mathbb{K} = \mathbb{C}$ alors $\langle x, y \rangle = \overline{\langle y, x \rangle}$ pour tout $x, y \in E$;
3. pour tout $x \in E$, $\langle x, x \rangle \in \mathbb{R}_+$;
4. $\langle x, x \rangle = 0$ si et seulement si $x = 0$.

Autrement dit, un produit scalaire est une forme bilinéaire (ou hermitienne si $\mathbb{K} = \mathbb{C}$) symétrique définie positive.

Un espace vectoriel E muni d'un produit scalaire est appelée espace préhilbertien.

Proposition 6.3.2 (Inégalité de Cauchy-Schwarz). *Un $(E, \langle \cdot, \cdot \rangle)$ un espace préhilbertien. Alors pour tout $x, y \in E$, $|\langle x, y \rangle|^2 \leq \langle x, x \rangle \langle y, y \rangle$.*

Démonstration. Soit $u \in \mathbb{K}$ unitaire ($|u| = 1$) tel que $u\langle x, y \rangle = |\langle x, y \rangle|$. Alors, par définition, pour tout $t \in \mathbb{R}$, $\langle ux + ty, ux + ty \rangle \geq 0$. Or

$$\langle ux + ty, ux + ty \rangle = \langle ux, ux \rangle + 2t \operatorname{Re} \langle ux, y \rangle + t^2 \langle y, y \rangle = \langle x, x \rangle + 2t |\langle x, y \rangle| + t^2 \langle y, y \rangle.$$

Ce polynôme du second degré est positif pour tout $t \in \mathbb{R}$ si bien que son discriminant est négatif ou nul. C'est à dire $|\langle x, y \rangle|^2 \leq \langle x, x \rangle \langle y, y \rangle$. \square

Corollaire 6.3.3. *Soit $(E, \langle \cdot, \cdot \rangle)$ un espace préhilbertien alors $\|x\| = \sqrt{\langle x, x \rangle}$ est une norme sur E .*

Démonstration. Notons tout d'abord que $\langle x, x \rangle \geq 0$ si bien que la racine est correctement définie, en particulier $\|x\| \in \mathbb{R}_+$. On vérifie les trois axiomes d'une norme :

- Soit $\lambda \in \mathbb{K}$, $\|\lambda x\| = \sqrt{\langle \lambda x, \lambda x \rangle} = \sqrt{\lambda \overline{\lambda} \langle x, x \rangle} = |\lambda| \|x\|$.
- Soit $x \in E$, $\|x\| = 0$ si et seulement si $\langle x, x \rangle = 0$ si et seulement si $x = 0$.

— Enfin, pour tout $x, y \in E$:

$$\begin{aligned} \langle x + y, x + y \rangle &= \langle x, x \rangle + \langle y, y \rangle + \langle x, y \rangle + \overline{\langle x, y \rangle} \leq \langle x, x \rangle + \langle y, y \rangle + 2|\langle x, y \rangle| \\ &\leq \left(\sqrt{\langle x, x \rangle} + \sqrt{\langle y, y \rangle} \right)^2. \end{aligned}$$

□

Proposition 6.3.4 (Identité du parallélogramme). *Soient $x, y \in E$ préhilbertien, alors*

$$\left\| \frac{x + y}{2} \right\|^2 + \left\| \frac{x - y}{2} \right\|^2 = \frac{1}{2} (\|x\|^2 + \|y\|^2).$$

Démonstration. Exercice. □

Définition 6.3.5. Un espace préhilbertien complet pour la norme issue du produit scalaire est appelé espace de Hilbert.

Exemple 26. On munit $\mathbf{L}_\mu^2(\mathbb{R}^d)$, l'espace des fonctions de carré intégrable sur \mathbb{R}^d muni d'une mesure μ , du produit scalaire

$$\langle f, g \rangle = \int_{\mathbb{R}^d} f(x) \overline{g(x)} \mu(dx), \quad f, g \in \mathbf{L}_\mu^2(\mathbb{R}^d).$$

La norme associée est la norme \mathbf{L}^2 usuelle et, muni de cette norme, on sait que $\mathbf{L}_\mu^2(\mathbb{R}^d)$ est complet. Ainsi, muni du produit scalaire défini ci-dessus, c'est un espace de Hilbert.

Définition 6.3.6. Si E est un espace préhilbertien, alors x est dit orthogonal à y et on note $x \perp y$ si $\langle x, y \rangle = 0$. Si A est une partie de E , l'orthogonal de A noté A^\perp est défini par

$$A^\perp = \{x \in E : \forall y \in A, \quad x \perp y\}$$

Proposition 6.3.7. 1. Si $A \subset B \subset E$ alors $B^\perp \subset A^\perp$;

2. A^\perp est un s.e.v. fermé de E ;

Démonstration. 1. Soit $x \in B^\perp$. Alors pour tout $y \in B$, $\langle x, y \rangle = 0$. Or $A \subset B$ donc pour tout $y \in A$, $\langle x, y \rangle = 0$. Ainsi, $B^\perp \subset A^\perp$.

2. On observe $A^\perp = \bigcap_{y \in A} \{x \in E : \langle x, y \rangle = 0\}$. Or, pour tout $y \in E$, l'application $x \rightarrow \langle x, y \rangle$ est continue. Par conséquent, $\{x \in E : \langle x, y \rangle = 0\}$ est un fermé comme l'image réciproque de $\{0\}$ par une application continue. D'où A^\perp est fermé. □

Théorème 6.3.8 (Théorème de Pythagore). *Soient $x, y \in E$ tels que $x \perp y$, alors $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.*

Exemple 27. Dans $\mathbf{L}_\mu^2([0, 2\pi]^d)$ où μ est une probabilité, les fonctions $e_n : x \rightarrow e^{i\langle z, x \rangle}$, $z \in \mathbb{Z}^d$, sont deux à deux orthogonales.

Démonstration. Soient $x, y \in E$ tels que $x \perp y$ alors

$$\|x + y\|^2 = \langle x + y, x + y \rangle = \langle x, x \rangle + \langle y, y \rangle + 2\langle x, y \rangle = \|x\|^2 + \|y\|^2.$$

□

Définition 6.3.9. Soit $(E, \|\cdot\|)$ un espace vectoriel normé. Une partie $C \subset E$ est dite convexe si pour tout $x, y \in C$ le segment $[x, y] = \{tx + (1 - t)y : t \in [0, 1]\}$ est incluse dans C .

Exemple 28. Dans un espace vectoriel normé, les boules ouvertes, les boules fermés, les sous-espace vectoriels sont convexes.

Théorème 6.3.10 (Projection sur un convexe). *Soient $(E, \langle \cdot, \cdot \rangle)$ un espace de Hilbert et C un sous-ensemble convexe fermé non vide de E . Alors, pour tout $x \in E$, il existe un unique $y \in C$ tel que $\|x - y\| = d(x, C)$. On note $P_C(x) = y$ le projeté de x sur C . Le projeté $P_C(x)$ est aussi l'unique élément $y \in C$ tel que*

$$\operatorname{Re} \langle x - y, z - y \rangle \leq 0, \quad \forall z \in C.$$

Le projecteur P_C est 1-lipschitzien

Démonstration. On commence par l'existence. Soit $x \in E$, on choisit une suite $(y_n)_{n \geq 0}$ d'éléments de C telle que $\|x - y_n\|$ converge vers $d(x, C) = \delta$ et plus précisément $\delta^2 \leq \|x - y_n\|^2 \leq \delta^2 + 1/n$. Alors $(y_n)_{n \geq 0}$ est une suite de Cauchy. En effet, par l'identité du parallélogramme,

$$\left\| \frac{x - y_n}{2} + \frac{x - y_m}{2} \right\|^2 + \left\| \frac{x - y_n}{2} - \frac{x - y_m}{2} \right\|^2 = \frac{1}{2} (\|x - y_n\|^2 + \|x - y_m\|^2),$$

et donc

$$\|y_n - y_m\|^2 = 2 \left(\underbrace{\|x - y_n\|^2}_{\leq \delta^2 + 1/n} + \underbrace{\|x - y_m\|^2}_{\leq \delta^2 + 1/m} \right) - 4 \underbrace{\left\| x - \frac{y_n + y_m}{2} \right\|^2}_{\geq \delta^2}.$$

Puisque C est un fermé dans un espace complet, il est complet et cela assure l'existence d'un élément $y \in C$ tel que $y_n \rightarrow y$. De plus, $\|x - y\| = \delta = d(x, C)$ par continuité.

Pour l'unicité, on suppose qu'il existe $y \neq z$ vérifiant l'égalité $\|x - y\| = \|x - z\| = d(x, C)$. En remplaçant, y_n et y_m par y et z dans le calcul précédant, on constate que

$$\|y - z\|^2 = 2(\|x - y\|^2 + \|x - z\|^2) - 4\left\| x - \frac{y + z}{2} \right\|^2 \leq 0,$$

et $z = y$.

On peut donc poser $y = P_C(x)$. Montrons que pour tout $z \in C$, $\operatorname{Re} \langle x - y, z - y \rangle \leq 0$. Soit $z \in C$ et $t \in (0, 1]$, par convexité, $(1 - t)y + tz \in C$ et donc, puisque $t \in \mathbb{R}$

$$0 \leq \|x - y\|^2 \leq \|x - [(1 - t)y + tz]\|^2 = \|(x - y) - t(z - y)\|^2 = \|x - y\|^2 - 2t \operatorname{Re} \langle x - y, z - y \rangle + t^2 \|y - z\|^2.$$

Il vient, pour tout $t \in (0, 1]$,

$$2t \operatorname{Re} \langle x - y, z - y \rangle \leq t^2 \|y - z\|^2,$$

d'où le résultat.

Réciproquement, si $y \in C$ vérifie, pour tout $z \in C$, $\operatorname{Re} \langle x - y, z - y \rangle \leq 0$, alors pour tout $z \in C$

$$\|x - z\|^2 = \|(x - y) - (z - y)\|^2 = \|x - y\|^2 - 2 \operatorname{Re} \langle x - y, z - y \rangle + \|z - y\|^2 \geq \|x - y\|^2.$$

Reste à montrer que P_C est 1-lipschitzien. Soient $x, y \in E$ alors

$$\begin{aligned} \operatorname{Re} \langle x - y, P_C(x) - P_C(y) \rangle &= \operatorname{Re} \langle (x - P_C(x)) + (P_C(x) - P_C(y)) + (P_C(y) - y), P_C(x) - P_C(y) \rangle \\ &= \operatorname{Re} \langle x - P_C(x), P_C(x) - P_C(y) \rangle + \|P_C(x) - P_C(y)\|^2 \\ &\quad + \operatorname{Re} \langle P_C(y) - y, P_C(x) - P_C(y) \rangle \geq \|P_C(x) - P_C(y)\|^2, \end{aligned} \quad (6.2)$$

en utilisant la caractérisation du projeté démontrée au dessus impliquant que le premier et troisième termes sont négatifs. L'inégalité de Cauchy-Schwarz appliqué au membre de gauche de (6.2) implique

$$\|P_C(x) - P_C(y)\|^2 \leq \operatorname{Re} \langle x - y, P_C(x) - P_C(y) \rangle \leq \|x - y\| \|P_C(x) - P_C(y)\|. \quad (6.3)$$

De deux choses l'une, ou bien $P_C(x) = P_C(y)$ et $0 = \|P_C(x) - P_C(y)\| \leq \|x - y\|$ quoiqu'il arrive; ou bien $P_C(x) \neq P_C(y)$ et en simplifiant (6.3) on obtient $\|P_C(x) - P_C(y)\| \leq \|x - y\|$. \square

Remarque 51. Si E est seulement préhilbertien, le résultat reste valable si C est complet pour la norme induite par le produit scalaire.

Théorème 6.3.11 (Projeté sur un s.e.v. fermé). *Soit F un s.e.v. fermé d'un espace de Hilbert E . Alors le projecteur P_F est linéaire et si $x \in E$, $P_F(x)$ est l'unique élément $y \in F$ tel que $x - y \in F^\perp$.*

Démonstration. Par le théorème 6.3.10, $P_F(x)$ est l'unique élément $y \in E$ tel que

$$\left\{ \begin{array}{l} y \in F \\ \operatorname{Re} \langle x - y, z - y \rangle \leq 0 \quad \forall z \in F \end{array} \right. \iff \left\{ \begin{array}{l} y \in F \\ \operatorname{Re} \langle x - y, \lambda z \rangle \leq 0 \quad \forall z \in F \quad \forall \lambda \in \mathbb{C} \end{array} \right.$$

car $z - y \in F$. Or si $\operatorname{Re} \lambda \langle x - y, z \rangle \leq 0$ pour tout $\lambda \in \mathbb{C}$ alors pour $\lambda = \overline{\langle x - y, z \rangle}$, on obtient que $|\langle x - y, z \rangle|^2 \leq 0$ d'où $\langle x - y, z \rangle = 0$.

Il reste à montrer que P_F est linéaire : soient $x_1, x_2 \in E$ et $\lambda \in \mathbb{K}$. Notons $y_1 = P_F(x_1) \in F$ et $y_2 = P_F(x_2) \in F$. Or

$$(x_1 + \lambda x_2) - (y_1 + \lambda y_2) = \underbrace{(x_1 - y_1)}_{\in F^\perp} + \lambda \underbrace{(x_2 - y_2)}_{\in F^\perp}. \quad (6.4)$$

Ainsi $y_1 + \lambda y_2$ est un élément de F tel que $(x_1 + \lambda x_2) - (y_1 + \lambda y_2) \in F^\perp$, par unicité du projeté, $y_1 + \lambda y_2 = P_F(x_1 + \lambda x_2)$. Ceci montre que $P_F(x_1) + \lambda P_F(x_2) = P_F(x_1 + \lambda x_2)$. \square

Corollaire 6.3.12. 1. Tout s.e.v. fermé F de E admet un supplémentaire orthogonal, i.e. $E = F \oplus F^\perp$ et l'identité satisfait $I = P_F + (I - P_F)$ avec $P_F(I - P_F) = (I - P_F)P_F = 0$ et P_F le projecteur linéaire sur F ;

2. Pour tout s.e.v. F de E , $E = \overline{F} \oplus F^\perp$;

3. Un s.e.v. F est dense dans E si et seulement si $F^\perp = \{0\}$;

4. Pour tout s.e.v. F de E , $(F^\perp)^\perp = \overline{F}$.

Exemple 29. Le s.e.v. $F = \{e_n : n \in \mathbb{Z}^d\} \subset \mathbf{L}_\mu^2([0, 2\pi]^d)$ est dense. C'est donc une base orthonormée. L'analyse de Fourier dans le contexte \mathbf{L}^2 consiste en fait à décomposer les fonctions le long d'une base orthonormée. En exercice, on pourra retrouver le théorème de Parseval qui est une extension du théorème de Pythagore. Attention toutefois, les convergences des séries ont lieu dans \mathbf{L}^2 et ne préjuge en rien de la convergence dans des topologies autres (convergence simple par exemple).

Démonstration. 1. Il est immédiat par le théorème 6.3.11 que

$$x = \underbrace{P_F(x)}_{\in F} + \underbrace{x - P_F(x)}_{\in F^\perp}. \quad (6.5)$$

De plus, si $x \in F \cap F^\perp$ alors $x \perp x$ donc $\langle x, x \rangle = 0$ d'où $x = 0$. Ceci montre que $E = F \oplus F^\perp$. L'équation (6.5) implique que $I = P_F + (I - P_F)$ et de plus que $\operatorname{Im} P_F = F$ et $\operatorname{Im} (I - P_F) = F^\perp$ d'où $P_F(I - P_F) = (I - P_F)P_F = 0$.

2. On applique le point précédent à \overline{F} en remarquant que $(\overline{F})^\perp = F^\perp$ par la proposition 6.3.7.

3. Le s.e.v. F est dense dans E si et seulement si $E = \overline{F}$ si et seulement si $F^\perp = \{0\}$.

4. Soient $x \in F$ et $y \in F^\perp$ alors $\langle x, y \rangle = 0$ donc $x \in (F^\perp)^\perp$ par définition et $F \subset (F^\perp)^\perp$. Par la proposition 6.3.7, $(F^\perp)^\perp$ est fermé ainsi $W = (F^\perp)^\perp$ est un espace de Hilbert pour le produit hermitien restreint à W . L'orthogonale de V dans W est $V^\perp \cap W = V^\perp \cap (V^\perp)^\perp = \{0\}$. Autrement dit, V est dense dans W , i.e. $\overline{V} = (V^\perp)^\perp$. \square

Remarque 52. Tous ces résultats restent valides dans un espace E préhilbertien pour autant que \overline{F} est complet.

6.3.2 Lemme de Fréchet-Riesz

Définition 6.3.13. Soit E un \mathbb{K} -espace vectoriel. Une forme linéaire est une application linéaire de E dans \mathbb{K} . Si E est muni d'une norme, une forme linéaire sur E est dite continue si elle est continue de $(E, \|\cdot\|)$ dans $(\mathbb{K}, |\cdot|)$.

Théorème 6.3.14 (Lemme de Fréchet-Riesz). Soient $(E, \langle \cdot, \cdot \rangle)$ un espace de Hilbert et ϕ une forme linéaire continue sur E . Alors il existe un unique $y \in E$ tel que $\phi(x) = \langle x, y \rangle$ pour tout $x \in E$. De plus, $\|\phi\|_{\mathcal{L}(E, \mathbb{K})} = \|y\|_E$. Autrement dit, l'application qui à $y \in E$ associe la forme linéaire continue $E \ni x \rightarrow \langle x, y \rangle \in \mathbb{K}$ est une isométrie surjective — une isométrie est en effet toujours injective.

Démonstration. Notons $\phi_y = \langle \cdot, y \rangle$. Cette application est linéaire continue et plus précisément, pour tout $x \in E$, $|\phi_y(x)| \leq \|x\| \|y\|$ et $\phi_y(y) = \|y\|^2$. On en déduit $\|\phi_y\| = \|y\|$. Ceci montre que $y \rightarrow \phi_y$ est une isométrie, reste à montrer qu'elle est surjective.

Considérons ϕ une forme linéaire continue. Si $\phi = 0$, alors $y = 0$ convient (et c'est la seule!). Supposons donc $\phi \neq 0$. Par continuité, $\text{Ker } \phi = \phi^{-1}(\{0\})$ est un *s.e.v.* fermé et par le théorème du rang (infini dimensionnel) il est de codimension 1. Ainsi, nous avons la décomposition en somme directe $E = \text{Ker } \phi \oplus (\text{Ker } \phi)^\perp$. Puisque $\phi \neq 0$, il existe $e \in (\text{Ker } \phi)^\perp$ non nul et de norme 1. On pose $y = \overline{\phi(e)}e \neq 0$. Puis pour $x \in E$ et $x_0 + x_1$ sa décomposition en somme directe. Alors

- $\phi(x_0) = 0 = \langle x_0, y \rangle = \overline{\phi_y}(x_0)$,
- $\phi_y(e) = \langle e, y \rangle = \langle e, \overline{\phi(e)}e \rangle = \phi(e)\|e\|^2 = \phi(e)$, et donc $\phi_y(x_1) = \phi(x_1)$ car $x_1 \in (\text{Ker } \phi)^\perp$ qui est de dimension 1.

Finalement, pour tout $x \in E$, $\phi_y(x) = \phi(x)$. □

6.3.3 Théorème de Radon-Nikodym, cas des mesures positives

On commence par rappeler la définition d'une mesure à densité.

Définition 6.3.15 (Mesure à densité). Soit $(\mathbb{X}, \mathcal{X}, \mu)$ un espace mesuré. Une mesure ν sur $(\mathbb{X}, \mathcal{X})$ est à densité par rapport à μ s'il existe $f : \mathbb{X} \rightarrow [0, \infty]$ mesurable telle que $\nu(A) = \int \mathbf{1}_A f d\mu$ pour tout $A \in \mathcal{X}$. On appelle f la densité de ν par rapport à μ . On note $\nu = f \cdot \mu$ ou $d\nu = f d\mu$.

Définition 6.3.16 (Mesure absolument continue). Soient $(\mathbb{X}, \mathcal{X})$ un espace mesurable et μ, ν deux mesures sur $(\mathbb{X}, \mathcal{X})$. On dit que ν est absolument continue par rapport à μ si pour tout $A \in \mathcal{X}$, $\mu(A) = 0$ implique $\nu(A) = 0$. On note $\nu \prec \mu$.

Proposition 6.3.17. Soit $(\mathbb{X}, \mathcal{X}, \mu)$ un espace mesuré. Si ν est une mesure sur $(\mathbb{X}, \mathcal{X})$ à densité par rapport à μ , alors ν est absolument continue par rapport à μ

Démonstration. Il existe $f : \mathbb{X} \rightarrow [0, \infty]$ mesurable telle que $\nu = f \cdot \mu$. Soit $A \in \mathcal{X}$ tel que $\mu(A) = 0$. Ainsi $\mathbf{1}_A = 0$ μ -presque partout, d'où $\mathbf{1}_A f = 0$ μ -presque partout. Donc,

$$\nu(A) = \int \mathbf{1}_A f d\mu = 0 \implies \nu \prec \mu.$$

□

Le théorème de Radon-Nikodym établit la réciproque : si $\nu \prec \mu$ alors ν est à densité par rapport à μ .

Théorème 6.3.18 (Radon-Nikodym). Soit $(\mathbb{X}, \mathcal{X}, \mu)$ un espace mesuré. On suppose que μ est σ -finie. Soit ν une mesure sur $(\mathbb{X}, \mathcal{X})$, alors les assertions suivantes sont équivalentes :

1. ν est finie et est absolument continue par rapport à μ ;
2. ν est à densité intégrable par rapport à μ , c'est à dire qu'il existe $f \in \mathbf{L}^1(\mu)$ positive telle que $\nu(A) = \int \mathbf{1}_A f d\mu$.

De plus f est unique, on l'appelle dérivée de Radon-Nikodym de ν par rapport à μ , on la note $f = d\nu/d\mu$.

Démonstration. Si on suppose que ν est à densité f positive et intégrable par rapport à μ alors ν est clairement finie et est absolument continue par la proposition précédente

En ce qui concerne l'unicité de la densité, si g est une densité de ν par rapport à μ , on pose $A_n = \{f \geq g + 1/n\}$ et il vient par l'inégalité de Markov que

$$\nu(A_n) = \int_{A_n} f d\mu = \int_{A_n} g d\mu \implies 0 = \int_{A_n} (f - g) d\mu \geq \mu(A_n)/n.$$

D'où $\mu(A_n) = 0$ et $\mu(\cup_{n \geq 1} A_n) = 0$ si bien que $f \leq g$ μ -p.p.. De la même manière, on montre que $g \leq f$ μ -p.p.. Enfin, $f = g$ dans $\mathbf{L}^1(\mu)$.

Soit $\rho = \nu + \mu$, c'est à dire, $\rho(A) = \nu(A) + \mu(A)$ pour tout $A \in \mathcal{X}$. Alors ρ est σ -finie et pour tout f mesurable positive

$$\int f d\rho = \int f d\nu + \int f d\mu.$$

De plus, $f \in \mathcal{L}^p(\rho)$ si et seulement si $f \in \mathcal{L}^p(\nu) \cap \mathcal{L}^p(\mu)$. Enfin, comme $\rho(A) = 0$ si et seulement si $\nu(A) = \mu(A) = 0$, il vient que $f \in \mathbf{L}^p(\rho)$ si et seulement si $f \in \mathbf{L}^p(\nu) \cap \mathbf{L}^p(\mu)$.

Étape 1 : Lemme de Riesz-Fisher

Soit $g \in \mathbf{L}^2(\rho)$, alors $g \in \mathbf{L}^2(\nu)$ puis par l'inégalité de Cauchy-Schwartz et le fait que ν soit finie

$$\int |g| d\nu \leq \|g\|_2 \sqrt{\nu(\mathbb{X})}.$$

On peut donc définir l'application $T : \mathbf{L}^2(\rho) \rightarrow \mathbb{R}$ définie par $Tg = \int g d\nu$. C'est une forme linéaire continue (sa norme est $\sqrt{\nu(\mathbb{X})}$). D'après le lemme de Riesz-Fisher, il existe un unique $\phi \in \mathbf{L}^2(\rho)$ tel que pour tout $g \in \mathbf{L}^2(\rho)$:

$$T(g) = \int g d\nu = \int g\phi d\rho = \int g\phi d\nu + \int g\phi d\mu.$$

Remarquons que la fonction ϕ est nulle si et seulement si $\nu = 0$, mais dans ce cas le résultat est évident. Heuristiquement, en posant $g = \mathbf{1}_A(1 - \phi)^{-1}$ on a

$$\nu(A) = \int \mathbf{1}_A d\nu = \int \mathbf{1}_A \frac{\phi}{1 - \phi} d\mu.$$

Il s'agit de justifier cette heuristique.

Étape 2 : Bornes sur ϕ

1. Montrons que $\phi \geq 0$ μ -p.p. et donc ν -p.p.. Soit $(E_n)_{n \geq 0} \in \mathcal{X}^{\mathbb{N}}$ une suite croissante d'ensembles tels que $\mu(E_n) < \infty$ et $\mathbb{X} = \cup_{n \geq 0} E_n$, puis notons $B_n = \{\phi < 0\} \cap E_n$. Clairement, $g = \mathbf{1}_{B_n} \in \mathbf{L}^2(\mu) \cap \mathbf{L}^2(\nu)$ et donc $g \in \mathbf{L}^2(\rho)$. La représentation de Riesz-Fisher donne

$$\int \mathbf{1}_{B_n}(1 - \phi) d\nu = \int \mathbf{1}_{B_n}\phi d\mu.$$

On remarque que l'intégrande à gauche est positive alors que celle de droite est négative si bien que les deux intégrales doivent être nulles. Ainsi :

$$\mathbf{1}_{B_n}\phi = 0, \mu - \text{p.p.} \quad \text{et} \quad \mathbf{1}_{B_n}(1 - \phi) = 0, \nu - \text{p.p.}$$

Or par définition de B_n , $\mathbf{1}_{B_n}\phi < 0$ et $\mathbf{1}_{B_n}(1 - \phi) > 1$ partout. En particulier, il vient que $\mu(B_n) = 0$ et $\nu(B_n) = 0$. Ceci montre que $\phi \geq 0$ μ -p.p. et ν -p.p.. De plus, la positivité de ϕ ν -p.p. ne requiert pas d'hypothèses particulières.

2. Montrons que $\phi < 1$ μ -p.p. et ν -p.p.. Cette fois-ci, on pose

$$C_n = \{\phi \geq 1\} \cap E_n,$$

et on obtient par la représentation de Riesz-Fisher

$$\int \mathbf{1}_{C_n}(1 - \phi) d\nu = \int \mathbf{1}_{C_n}\phi d\mu.$$

Là-encore, l'intégrande à gauche est négative alors que celle de droite est positive. On en déduit $0 = \mathbf{1}_{C_n}\phi \geq \mathbf{1}_{C_n}$ μ -p.p. donc $\mu(C_n) = 0$. En particulier, puisque $\nu \prec \mu$, on a aussi $\nu(C_n) = 0$.

Contrairement au point précédent, l'hypothèse d'absolue continuité est essentielle, c'est ici qu'elle apparaît.

Quitte à modifier ϕ sur un ensemble ρ -négligeable, on peut considérer que $\phi(x) \in [0, 1)$ pour tout $x \in \mathbb{X}$. On pose alors $f = \phi/(1 - \phi)$ qui est mesurable positive.

Étape 3 : Montrons que $d\nu = f d\mu$. Pour cela, on utilise à nouveau l'égalité provenant de la représentation de Riesz-Fisher pour tout $g \in \mathbf{L}^2(\rho)$

$$\int g(1 - \phi) d\nu = \int g\phi d\mu. \tag{6.6}$$

Soit $A \in \mathcal{X}$ tel que $\mu(A) < \infty$ et posons $g = \mathbf{1}_A$. Alors $g \in \mathbf{L}^2(\rho)$ et

$$\int \mathbf{1}_A(1 - \phi) d\nu = \int \mathbf{1}_A\phi d\mu. \tag{6.7}$$

Si $\mu(A) = \infty$, on pose $A_n = A \cap E_n$ et par convergence monotone l'égalité (6.7) est toujours vérifiée, éventuellement dans $\overline{\mathbb{R}}_+$.

Ce qui est vrai pour les indicatrices reste vraie pour les fonctions étagées positives en utilisant la linéarité de l'intégrale pour les fonctions positives. C'est également vrai pour les fonctions mesurables positives en utilisant le théorème de convergence monotone. Ainsi, l'égalité 6.6 déduite du lemme de Riesz-Fisher est satisfaite pour toute fonction g mesurable positive. En particulier, pour tout $A \in \mathcal{X}$, la fonction $g = \mathbf{1}_A/(1 - \phi)$ est mesurable positive et on obtient

$$\int \frac{\mathbf{1}_A}{1 - \phi} (1 - \phi) d\nu = \int \mathbf{1}_A \frac{\phi}{1 - \phi} d\mu \iff \nu(A) = \int \mathbf{1}_A f d\mu.$$

Comme ν est supposée finie, la fonction f est intégrable positive. \square

Définition 6.3.19 (Mesures étrangères). Deux mesures μ et ν sont dites étrangères s'il existe un ensemble $E \in \mathcal{X}$ tel que $\mu(E^c) + \nu(E) = 0$. On dit que μ est concentrée sur E et ν sur E^c .

6.3.4 Théorème de Radon-Nikodym, cas des mesures signées

Définition 6.3.20 (Mesure signée). Soit $(\mathbb{X}, \mathcal{X})$ un espace mesurable. Une mesure signée μ est une application σ -additive de \mathcal{X} dans \mathbb{R} , i.e. pour toute suite $(A_n)_{n \geq 0}$ d'ensembles mesurables deux à deux disjoints, $\mu(\cup_{n \geq 0} A_n) = \sum_{n \geq 0} \mu(A_n)$.

Si μ est une mesure signée, on peut lui associer sa variation totale, notée $|\mu|$, définie pour tout $A \in \mathcal{X}$

$$|\mu(A)| = \sup \left\{ \sum_{n \geq 0} |\mu(E_n)| : i \neq j \Rightarrow E_i \cap E_j = \emptyset, \bigcup_{n \geq 0} E_n = A \right\}.$$

Remarque 53. Une mesure signée μ vérifie en particulier, par définition, $\mu(\mathbb{X}) < \infty$. Il n'est de même pas nécessaire de supposer $\mu(\emptyset) = 0$ pour les mesures signées. En effet, $\emptyset = \cup_{n \geq 0} \emptyset$ qui est une réunion d'ensembles deux à deux disjoints. Si $\mu(\emptyset) \neq 0$, on aurait $\sum_{n \geq 0} \mu(\emptyset) = \pm\infty$ ce qui est exclu.

Proposition 6.3.21. Soit μ une mesure signée sur un espace mesurable $(\mathbb{X}, \mathcal{X})$. Alors,

1. $|\mu|$ est une mesure positive,
2. $|\mu|$ est une mesure finie,
3. Pour tout $A \in \mathcal{X}$, $|\mu(A)| \leq |\mu|(A)$.

Démonstration. 1. Suivant la remarque précédente, nous avons déjà que $\mu(\emptyset) = 0$ d'où il vient facilement que $|\mu|(\emptyset) = 0$. Soit $(A_n)_{n \geq 0}$ une famille d'ensembles mesurables deux à deux disjoints. On note $A = \cup_{n \geq 0} A_n$ et on se donne une partition $(E_k)_{k \geq 0}$ de A . On remarque que $E_k = \cup_{n \geq 0} E_k \cap A_n$ qui est une réunion disjointe. C'est à dire $(E_k \cap A_n)_{n \geq 0}$ est une partition de E_k pour tout $k \geq 0$. D'où

$$\begin{aligned} \sum_{k \geq 0} |\mu(E_k)| &= \sum_{k \geq 0} \left| \sum_{n \geq 0} \mu(E_k \cap A_n) \right| \\ &\leq \sum_{n \geq 0} \sum_{k \geq 0} |\mu(A_n \cap E_k)| \leq |\mu|(A_n), \end{aligned}$$

en intervertissant les sommes et en remarquant que $(E_k \cap A_n)_{k \geq 0}$ est une partition de A_n . Donc, $|\mu|(A) \leq \sum_{n \geq 0} |\mu|(A_n)$.

Réciproquement, on considère pour tout $n \geq 0$ une partition $(E_{n,k})_{k \geq 0}$ de A_n . Manifestement, $(E_{n,k})_{n,k \geq 0}$ est une partition de A . Donc,

$$\sum_{n \geq 0} \left(\sum_{k \geq 0} |\mu(E_{n,k})| \right) = \sum_{k,n \geq 0} |\mu(E_{n,k})| \leq |\mu|(A).$$

En passant à la borne supérieure pour tout $n \geq 0$, on obtient l'inégalité inverse.

2. Ce deuxième résultat nécessite deux lemmes.

Lemme 6.3.22. *Pour $x_1, \dots, x_N \in \mathbb{R}$, on pose $S = \sum_{k=1}^N |x_k|$. Alors il existe $I \subset \{1, 2, \dots, N\}$ tel que*

$$\left| \sum_{i \in I} x_i \right| \geq S/2.$$

Démonstration. Immédiat. □

Lemme 6.3.23. *Si $E \in \mathcal{X}$ est tel que $|\mu|(E) = \infty$ alors il existe une partition de E en deux ensembles mesurables A et B tels que $|\mu|(A) \geq 1$ et $|\mu|(B) = \infty$.*

Démonstration. Nous avons $\mu(E) \in \mathbb{R}$ et donc $|\mu(E)| < \infty$. Posons $M = 2(1 + |\mu(E)|)$. Si $|\mu|(E) = \infty$ alors il existe E_1, E_2, \dots, E_N disjoints et contenu dans E tels que

$$\sum_{k=1}^N |\mu(E_k)| \geq M.$$

Par le lemme 6.3.22, il existe $I \subset \{1, \dots, N\}$ tel que

$$\left| \sum_{k \in I} \mu(E_k) \right| \geq M/2.$$

On pose alors $A = \cup_{k \in I} E_k$, et nous avons

$$|\mu(A)| = \left| \sum_{k \in I} \mu(E_k) \right| \geq M/2 \geq 1.$$

D'autre part, en posant $B = E \setminus A$, on obtient $\mu(E) = \mu(A) + \mu(B)$ et donc $|\mu(B)| \geq |\mu(E)| - |\mu(A)| \geq M/2 - |\mu(E)| = 1$. De plus, $|\mu|$ est σ -additive donc

$$\infty = |\mu|(E) = |\mu|(A) + |\mu|(B),$$

d'où $|\mu|(A) = \infty$ ou $|\mu|(B) = \infty$. C'est le résultat du lemme quitte à modifier le rôle de A et B . □

Si $|\mu|(\mathbb{X}) = \infty$, on construit par récurrence grâce au lemme 6.3.23 une suite $(A_n)_{n \geq 0}$ d'ensembles mesurables deux à deux disjoints tel que $|\mu(A_n)| \geq 1$ pour tout $n \geq 0$. Or

$$\sum_{n \geq 0} \mu(A_n) = \mu \left(\bigcup_{n \geq 0} A_n \right) \in \mathbb{R}$$

est une série convergente, donc $\mu(A_n)$ tend vers 0 quand n tend vers l'infini. C'est une contradiction.

3. Ce point est immédiat en remarquant que $\{A, \emptyset, \emptyset, \dots\}$ est une partition de A donc vérifie en particulier $|\mu(A)| \leq |\mu|(A)$. □

Si μ est une mesure signée sur un espace mesurable $(\mathbb{X}, \mathcal{X})$ alors, en posant,

$$\mu^+ = \frac{|\mu| + \mu}{2} \quad \text{et} \quad \mu^- = \frac{|\mu| - \mu}{2}.$$

on vérifie que $\mu = \mu^+ - \mu^-$. De plus, μ^+ et μ^- sont des mesures positives.

Théorème 6.3.24 (Décomposition de Hahn). *Si μ est une mesure signée sur $(\mathbb{X}, \mathcal{X})$, alors il existe une partition de \mathbb{X} en deux ensembles N et P tels que*

— pour tout $A \subset P$ mesurable, $\mu(A) \geq 0$,

— et pour tout $B \subset N$ mesurable, $\mu(B) \leq 0$.

De plus, μ^+ et μ^- sont caractérisées par $\mu^+(\cdot) = \mu(\cdot \cap P)$ et $\mu^- = -\mu(\cdot \cap N)$. En particulier, μ^+ et μ^- sont étrangères.

Démonstration. Admis. □

Théorème 6.3.25 (Théorème de Radon-Nikodym). *Soit $(\mathbb{X}, \mathcal{X}, \mu)$ un espace mesuré avec μ une mesure positive σ -finie. Soit ν une mesure signée sur $(\mathbb{X}, \mathcal{X})$, alors les assertions suivantes sont équivalentes.*

1. ν est absolument continue par rapport à μ ;
2. ν est à densité intégrable par rapport à μ , i.e. il existe une unique $f \in \mathbf{L}^1(\mu)$ telle que $\nu(A) = \int \mathbf{1}_A f \, d\mu$ pour tout $A \in \mathcal{X}$.

Démonstration. Si ν est à densité intégrable alors ν est absolument continue par rapport à μ , c'est immédiat.

Réciproquement, supposons ν est absolument continue par rapport à μ , et commençons par montrer que $|\nu|$ est absolument continue par rapport à μ . En effet, soit $A \in \mathcal{X}$ tel que $\mu(A) = 0$ et soit $(E_n)_{n \geq 0}$ une partition de A . Pour tout $n \geq 0$, $E_n \subset A$ donc $\mu(E_n) = 0$ et donc $\nu(E_n) = 0$. Puis,

$$\sum_{n \geq 0} |\nu(E_n)| = 0 \implies |\nu|(A) = 0.$$

Les mesures ν^+ et ν^- sont donc aussi absolument continue par rapport à μ . Elles sont également finies. On applique donc le théorème de Radon-Nikodym pour les mesures positives qui nous assure l'existence de deux fonctions f^+ et f^- positives intégrables telles que $d\nu^\pm = f^\pm \, d\mu$. Ainsi, $f = f^+ - f^- \in \mathbf{L}^1(\mu)$ et par linéarité de l'intégrale pour les fonctions intégrables, on obtient $d\nu = f \, d\mu$. □

Définition 6.3.26. On note $\mathcal{M}(\mathbb{X}, \mathcal{X}, \mathbb{R}) = \{\mu \text{ mesures signées sur } (\mathbb{X}, \mathcal{X})\}$. C'est un \mathbb{R} -espace vectoriel. On peut le munir de la norme de la variation totale définie par $\|\mu\|_{VT} = |\mu|(\mathbb{X})$.

Théorème 6.3.27 (Cohn). *L'espace vectoriel normé $(\mathcal{M}(\mathbb{X}, \mathcal{X}, \mathbb{R}), \|\cdot\|_{VT})$ est un espace de Banach.*

Remarque 54. Les mesures signées apparaissent comme une généralisation naturelle des mesures positives. Nous ne l'avons pas évoqué ici, mais il existe également une notion de mesure à valeurs complexes. Ces deux généralisations, bien qu'utile en théorie de la mesure, ne sont plus interprétables en termes de mesure d'aire ou de volume.

6.4 Approximation dans les espaces \mathbf{L}^p , $p \in [1, \infty)$

L'espace \mathbf{L}^∞ est particulièrement gros et possède peu de bonnes propriétés. C'est le cas notamment en théorie de l'approximation : hormis le paragraphe 6.4.1, les autres ne concernent que le cas p fini.

Lemme 6.4.1. *Si $(f_n)_{n \geq 0}$ et $(g_n)_{n \geq 0}$ converge respectivement vers f et g dans $\mathbf{L}^p_{\mathbb{K}}(\mu)$ et que $(\lambda_n)_{n \geq 0}$ converge vers λ , alors $(\lambda_n f_n + g_n)_{n \geq 0}$ converge dans $\mathbf{L}^p_{\mathbb{K}}(\mu)$ vers $\lambda f + g$.*

Démonstration. Le lemme se déduit de l'inégalité suivante

$$\|\lambda f + g - \lambda_n f_n - g_n\|_p = \|\lambda f - \lambda_n f + \lambda_n f + g - \lambda_n f_n - g_n\|_p \leq |\lambda - \lambda_n| \|f\|_p + |\lambda_n| \|f - f_n\|_p + \|g - g_n\|_p.$$

□

6.4.1 Approximation par des fonctions étagées mesurables

Proposition 6.4.2. *L'espace vectoriel des fonctions étagées intégrables est dense dans l'espace $\mathbf{L}^p_{\mathbb{K}}(\mu)$ pour la norme $\|\cdot\|_p$, $p \in [1, \infty]$.*

Démonstration. Soit $f \in \mathbf{L}^p_{\mathbb{K}}(\mu)$. Si f est à valeurs dans \mathbb{K} , on a la décomposition $f = \operatorname{Re} f^+ - \operatorname{Re} f^- + i \operatorname{Im} f^+ - i \operatorname{Im} f^-$. Aussi, on peut supposer f positive sans perte de généralité.

On sait qu'il existe une suite $(f_n)_{n \geq 0}$ monotone croissante de fonctions étagées positives vérifiant $0 \leq f_n \leq f$ pour tout $n \geq 0$ et $\lim_{n \rightarrow \infty} f_n = f$ μ -presque partout. Alors $|f - f_n|^p$ converge vers 0 μ -presque partout et $|f - f_n|^p \leq 2f^p$. Par le théorème de convergence dominée, on obtient le résultat. □

Remarquons enfin que, si ϕ est étagée, ϕ est intégrable si et seulement si ϕ^p est intégrable si et seulement si $\mu(\{\phi \neq 0\}) < \infty$.

6.4.2 Approximation par des fonctions continues à support compact

Comme annoncé, on se restreint ici aux espaces \mathbf{L}^p avec p fini. Dans cette partie, sauf mention contraire, on suppose de plus que (\mathbb{X}, d) est un espace métrique et $(\mathbb{X}, \mathcal{X}, \mu)$ est un espace mesuré où $\mathcal{X} = \mathcal{B}(\mathbb{X})$ est la tribu borélienne. Les théorèmes ci-dessous s'appliquent typiquement pour $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$.

Définition 6.4.3 (Fonctions en escalier). Une fonction mesurable sur \mathbb{X} est dite en escalier si il existe des ouverts $O_i, i = 1, \dots, n$, deux à deux disjoints et des $\alpha_i \in \mathbb{K}, i = 1, \dots, n$ tels que

$$f = \sum_{i=1}^n \alpha_i \mathbf{1}_{O_i}.$$

Le support (topologique) d'une fonction f est l'adhérence de l'ensemble $\{f \neq 0\}$, on note $\text{supp } f = \overline{\{f \neq 0\}}$.

Proposition 6.4.4. Soient $(\mathbb{X}, \mathcal{T})$ un espace topologique et μ une mesure extérieurement régulière sur $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$. Alors les fonctions en escaliers intégrables sont denses dans $\mathbf{L}^p(\mu), p \in [1, \infty)$.

Démonstration. Les fonctions en escaliers sont évidemment mesurables et dans $\mathbf{L}^p(\mu)$ si et seulement si elles sont intégrables.

Soit $A \in \mathcal{B}(\mathbb{X})$ tel que $\mathbf{1}_A \in \mathbf{L}^p(\mu)$ (i.e. $\mu(A) < \infty$). Par régularité de la mesure μ , on peut trouver une suite $(O_n)_{n \geq 0}$ contenant A tel que $\mu(O_n)$ tende vers $\mu(A)$. En particulier, pour tout n assez grand, $\mu(O_n) < \infty$ et donc $\mathbf{1}_{O_n} \in \mathbf{L}^p(\mu)$. Puis, on calcule

$$\|\mathbf{1}_A - \mathbf{1}_{O_n}\|_p^p = \mu(O_n) - \mu(A) \rightarrow 0, \quad \text{quand } n \rightarrow \infty.$$

□

Remarque 55. La preuve met en évidence l'obstruction pour le cas $p = \infty$. Puisque $\|\mathbf{1}_{O_n} - \mathbf{1}_A\|_\infty = 1$ sauf si $\mu(O_n \setminus A) = 0$ à partir d'un certain rang, c'est à dire A est, à ensemble de mesure nulle près, ouvert.

Proposition 6.4.5. Soient (\mathbb{X}, d) un espace métrique et μ est une mesure sur $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$. On suppose qu'il existe une suite croissante $(K_n)_{n \geq 0}$ de compacts de μ -mesure finie telle que $\mathbb{X} = \bigcup_{n \geq 0} \text{Int } K_n$. Alors μ est une mesure de Borel régulière et les fonctions en escaliers à support compact sont denses dans $\mathbf{L}^p(\mu), p \in [1, \infty)$.

Démonstration. La régularité de μ provient du théorème 2.2.36. Soit K un compact, alors on peut extraire du recouvrement d'ouvert $(\text{Int } K_n)_{n \geq 0}$ un sous-recouvrement fini. Il existe donc n_0 tel que $K \subset K_{n_0}$ et $\mu(K) < \infty$, c'est une mesure de Borel.

Soit O un ouvert tel que $\mu(O) < \infty$, on pose $O_n = O \cap \text{Int } K_n$ alors par continuité à gauche $\mu(O_n)$ converge vers $\mu(O)$ et donc $\mathbf{1}_{O_n}$ converge vers $\mathbf{1}_O$ dans $\mathbf{L}^p(\mu)$. De plus, $O_n \subset K_n$. Le résultat découle du lemme 6.4.1 et de la proposition 6.4.4 □

Proposition 6.4.6. Soient (\mathbb{X}, d) un espace métrique et μ est une mesure sur $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$. On suppose qu'il existe une suite croissante $(K_n)_{n \geq 0}$ de compacts de μ -mesure finie telle que $\mathbb{X} = \bigcup_{n \geq 0} \text{Int } K_n$. Alors μ est une mesure de Borel régulière et les fonctions continues à support compact sont denses dans $\mathbf{L}^p(\mu), p \in [1, \infty)$.

Démonstration. Du fait de la proposition 6.4.5 et le lemme 6.4.1, on doit montrer que pour tout O ouvert de \mathbb{X} relativement compact, $\mathbf{1}_O \in \mathbf{L}^p_{\mathbb{K}}(\mu)$ est limite dans \mathbf{L}^p de fonctions continues à support compact. Si O est vide, alors la suite de fonctions constante égale à la fonction nulle convient. Supposons donc O non vide relativement compact, alors la régularité de μ implique pour tout $\varepsilon > 0$ l'existence d'un compact $K \subset O$ tel que $\mu(O \setminus K) < \varepsilon$. On pose

$$\forall x \in \mathbb{X}, \quad f_\varepsilon(x) = \frac{d(x, O^c)}{d(x, O^c) + d(x, K)}.$$

C'est une fonction qui vaut 1 sur K et 0 sur O^c , elle est à support dans \overline{O} qui est compact. La fonction f_ε est continue. En effet, pour toute partie $A \subset \mathbb{X}$ non vide, $x \rightarrow d(x, A)$ est continue (même lipschitzienne) et $d(x, O^c) + d(x, K) = 0$ si et seulement si $x \in O^c \cap K = \emptyset$ puisque K et O^c sont fermés. Enfin, $\|f_\varepsilon - \mathbf{1}_O\|_p \leq \mu(O \setminus K) < \varepsilon$. □

Remarque 56. Les hypothèses des propositions 6.4.5 et 6.4.6 sont satisfaites si (\mathbb{X}, d) est un espace métrique séparable et μ une mesure σ -finie sur $\mathcal{B}(\mathbb{X})$ d'après la démonstration du théorème 2.2.39.

Remarque 57. Il pourrait sembler de prime abord que la métrisabilité de l'espace est essentielle pour exhiber la fonction f_ε . En fait l'existence d'une telle fonction est assurée dans un cadre plus général.

Lemme 6.4.7. *Soit $(\mathbb{X}, \mathcal{T})$ un espace topologique séparé localement compact. Pour tout ouvert U et tout compact $K \subset U$, il existe une fonction continue $f : \mathbb{X} \rightarrow [0, 1]$ tel que*

$$\forall x \in K, \quad f(x) = 1 \quad \text{et} \quad \forall x \in U^c, \quad f(x) = 0.$$

Proposition 6.4.8. *Soient $(\mathbb{X}, \mathcal{T})$ un espace topologique séparé localement compact et μ une mesure sur $\mathcal{B}(\mathbb{X})$ régulière extérieurement. Alors l'espace des fonctions continues intégrables est dense dans $\mathbf{L}^p(\mu)$.*

Démonstration. C'est une conséquence des lemmes 6.4.1 et 6.4.7 ainsi que de la proposition 6.4.4 \square

Exercice 24. Soient (\mathbb{X}, d) un espace métrique et μ une mesure sur $\mathcal{B}(\mathbb{X})$ extérieurement régulière. Montrer que pour tout $p \in [1, \infty)$ l'espace des fonctions Lipschitz bornées et intégrables est dense dans $\mathbf{L}^p(\mu)$.

Exercice 25. En utilisant la proposition 6.4.4, montrer que si (\mathbb{X}, d) un espace métrique séparable et μ une mesure extérieurement régulière, alors l'espace $\mathbf{L}^p(\mathbb{X}, \mu)$ est séparable.

Proposition 6.4.9. *Soit $(\mathbb{X}, \mathcal{X}, \mu)$ un espace mesuré. On définit l'application $\rho : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_+$ pour $A, B \in \mathcal{X}$ par $\rho(A, B) = \mu(A \Delta B)$. Muni de ρ , l'ensemble \mathcal{X} est un espace métrique. La tribu \mathcal{X} est dite séparable si l'espace métrique (\mathcal{X}, ρ) est lui-même séparable. Une tribu \mathcal{X} est séparable si et seulement si il existe une famille $(B_n)_{n \geq 0} \in \mathcal{X}^{\mathbb{N}}$ telle que $\mathcal{X} = \sigma(B_n)_\mu$.*

Proposition 6.4.10. *Soit $(\mathbb{X}, \mathcal{X}, \mu)$ un espace mesuré. Si la tribu \mathcal{X} est séparable alors, pour tout $p \in [0, \infty)$, l'espace $\mathbf{L}^p(\mathbb{X}, \mathcal{X}, \mu)$ est séparable.*

Exercice 26. Montrer que ces résultats sont faux pour $p = \infty$.

6.4.3 Convolution

Définition et propriétés élémentaires

Dans cette partie, on se place exclusivement sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$. La convolution se généralise très bien aux groupes topologiques localement compact, mais cela nous emmènerait trop loin par rapport à l'objectif de ce cours. Dans tout ce qui suit, on utilise effectivement de manière cruciale la structure de groupe additif de \mathbb{R}^d ainsi que l'invariance par translation de la mesure de Lebesgue (que l'on appelle mesure de Haar pour les groupes topologiques localement compact).

On se permettra de noter plus simplement l'espace $\mathbf{L}^p(\mathbb{R}^d, \lambda_d)$ par \mathbf{L}^p .

Proposition 6.4.11. *Soient $f, g \in \mathbf{L}^1$. La fonction $\phi : x \rightarrow \int_{\mathbb{R}^d} f(x-y)g(y) dy$ est définie μ -presque partout, mesurable et intégrable par rapport à λ_d . De plus, $\|\phi\|_1 \leq \|f\|_1 \|g\|_1$.*

La fonction ϕ s'appelle la convolée de f et g et est notée $f * g$.

Démonstration. On pose $\psi(x, y) = f(x-y)g(y)$. La fonction ψ est mesurable sur $\mathbb{R}^d \times \mathbb{R}^d$. Puis on calcule en utilisant Tonelli puis l'invariance par translation de λ_d

$$\begin{aligned} \int |\psi(x, y)| dx dy &= \int |f(x-y)| |g(y)| dx dy \\ &= \int |g(y)| \left(\int |f(x-y)| dx \right) dy \\ &= \int |f(x)| dx \int |g(y)| dy = \|f\|_1 \|g\|_1 < \infty. \end{aligned}$$

Ainsi ψ est intégrable et le théorème de Fubini implique que $x \rightarrow \phi(x)$ est mesurable et intégrable. Puis par un calcul très similaire on obtient la majoration de $\|\phi\|_1$. \square

On définit ainsi une loi de composition interne sur \mathbf{L}^1 qui à $(f, g) \in \mathbf{L}^1 \times \mathbf{L}^1$ associe le produit de convolution $f * g \in \mathbf{L}^1$.

Proposition 6.4.12. *Le produit de convolution est commutatif, associatif, distributif par rapport à l'addition et homogène par multiplication par un scalaire. Il n'admet pas d'éléments neutre.*

L'espace $\mathbf{L}^1(+, *)$ une algèbre de Banach sans unité.

Démonstration. On commence par la commutativité. Soient $f, g \in \mathbf{L}^1$, alors

$$f * g(x) = \int f(x-y)g(y) dy = \int f(u)g(x-u) du = g * f(x),$$

en faisant le changement de variable affine $y = x - u$. Cette égalité est satisfaite pour tout $x \in \mathbb{R}^d$ pour lequel $g * f$ et $f * g$ sont définies, c'est à dire pour presque tout $x \in \mathbb{R}^d$.

La distributivité et l'homogénéité reflètent la linéarité de l'intégrale. Pour l'associativité, on utilise le théorème de Fubini : soient $f, g, h \in \mathbf{L}^1$ et calculons, d'une part,

$$\begin{aligned} [(f * g) * h](x) &= \int (f * g)(x-y)h(y) dy \\ &= \int \left(\int f(x-y-z)g(z) dz \right) h(y) dy \\ &= \int \int f(x-y-z)g(z)h(y) dydz \end{aligned}$$

et d'autre part

$$\begin{aligned} [f * (g * h)](x) &= \int f(x-u)g * h(y) dy \\ &= \int f(x-u) \int g(u-v)h(v) dvdu \\ &= \int \int f(x-u)g(u-v)h(v) dudv. \end{aligned}$$

On conclut en effectuant dans la deuxième expression, le changement de variable affine $u = y + z$ et $v = y$.

On termine la proposition en montrant qu'il n'existe pas d'élément neutre pour le produit de convolution. Pour cela, supposons au contraire qu'il existe $g \in \mathbf{L}^1$ tel que pour tout $f \in \mathbf{L}^1$ on ait $g * f = f$. Pour $n \in \mathbb{N}$, on définit f_n par $f_n(x) = e^{-n\|x\|^2}$. Les fonctions f_n sont continues et dans \mathbf{L}^1 . Le produit de convolution entre f_n est g est donné par

$$f_n * g(x) = \int e^{-n\|x-y\|^2} g(y) dy.$$

L'intégrande est continue en tout $x \in \mathbb{R}^d$ et est bornée en valeur absolue par $y \rightarrow |g(y)|$ qui est intégrable. Ainsi, $f_n * g$ est continue sur \mathbb{R}^d .

D'autre part, $f_n * g = f_n$ presque partout, mais en fait partout par continuité. Donc, en particulier, $f_n(0) = f_n * g(0)$ c'est à dire

$$1 = \int g(y)e^{-n\|y\|^2} dy.$$

Puis le théorème de convergence dominée appliquée à l'intégrale à droite donne la contradiction cherchée. \square

Exercice 27. Deux fonctions continues égales presque partout sont égales.

Approximation de l'identité

Si notre algèbre de convolution est sans unité, il existe une notion d'approximation de l'unité.

Définition 6.4.13 (Approximation de l'identité). Une suite $(\phi_n)_{n \geq 0}$ d'éléments de \mathbf{L}^1 est une approximation de l'unité si

1. pour tout $n \geq 0$, $\int \phi_n d\lambda_d = 1$;

2. $\sup_{n \geq 0} \int |\phi_n| d\lambda_d < \infty$, on dit que $(\phi_n)_{n \geq 0}$ est bornée dans \mathbf{L}^1 ;
3. pour tout $\varepsilon > 0$, $\lim_{n \rightarrow \infty} \int_{\|x\| > \varepsilon} |\phi_n| d\lambda_d = 0$.

Théorème 6.4.14. *Si $(\phi_n)_{n \geq 0}$ est une suite d'approximations de l'unité, alors $\phi_n * f$ converge vers f dans \mathbf{L}^1 .*

Exemple 30 (de suite d'approximations de l'unité). Soit $\phi \in \mathbf{L}^1$ telle que $\int \phi d\lambda_d = 1$. Pour tout $n \geq 1$, on définit ϕ_n par $\phi_n(x) = n^d \phi(nx)$. Alors $(\phi_n)_{n \geq 1}$ est une suite d'approximation de l'unité. En effet,

- $\int \phi_n(x) dx = \int n^d \phi(nx) dx = \int \phi(x) dx = 1$,
- $\sup_{n \geq 1} \int |\phi_n| d\lambda_d \leq \|\phi\|_1$,
- pour tout $\varepsilon > 0$,

$$\int_{\|x\| > \varepsilon} |\phi_n(x)| dx = \int_{\|x\| > n\varepsilon} |\phi(x)| dx \xrightarrow{n \rightarrow \infty} 0$$

par convergence dominée.

Remarque 58. La convolution d'une fonction par une approximation de l'unité a pour effet de prendre des moyennes de f localement autour de chaque point x . Cela a pour effet de régulariser la fonction f et la régularité obtenue est celle de l'approximation de l'unité. Par exemple, dans l'exemple précédent, si on choisit ϕ infiniment dérivable à support compact, alors f peut être approchée par $f * \phi_n$ dans \mathbf{L}^1 qui est également infiniment dérivable à support compact. La convolution permet de montrer de nombreux résultats de densité dans \mathbf{L}^1 . Néanmoins, rappelons que nous utilisons de manière cruciale la structure de \mathbb{R}^d

Avant de montrer le théorème, montrons le lemme suivant.

Lemme 6.4.15. *Soit $p \in [1, \infty)$. Si $y \in \mathbb{R}^d$ et $f \in \mathbf{L}^p$, on définit $\tau_y f \in \mathbf{L}^p$, la translatée de f , par $\tau_y f(x) = f(x + y)$. Alors, $\lim_{y \rightarrow 0} \|\tau_y f - f\|_p = 0$.*

Démonstration. Supposons d'abord f continue à support compact. Si $f(x+y) - f(x) \neq 0$, alors $x \in \text{supp } f$ ou $x \in (\text{supp } f) - y$. Il existe donc un compact K de \mathbb{R}^d tel que si $\|y\| \leq 1$ alors $\text{supp } (\tau_y f - f) \subset K$. Ainsi, si $\|y\| \leq 1$, nous avons

$$\|\tau_y f - f\|_p = \left(\int_K |f(x+y) - f(x)|^p dx \right)^{1/p}.$$

Puisque f est continue sur K , f est uniformément continue :

$$\forall \varepsilon > 0, \quad \exists \eta \in (0, 1), \quad \|y\| \leq \eta \implies \forall x \in K, |f(x+y) - f(x)| < \varepsilon.$$

Aussi, si $\|y\| \leq \eta$, $\|\tau_y f - f\|_p \leq (\lambda(K)\varepsilon^p)^{1/p} = \lambda(K)^{1/p}\varepsilon$.

Maintenant, si $f \in \mathbf{L}^p$, alors on peut trouver une suite $(f_n)_{n \geq 0}$ de fonctions continue à support compact convergeant vers f dans \mathbf{L}^p . De plus, pour tout $n \geq 0$ et tout $y \in \mathbb{R}^d$, nous avons

$$\begin{aligned} \|\tau_y f - f\|_p &= \|\tau_y f - \tau_y f_n + \tau_y f_n - f_n + f_n - f\|_p \\ &\leq \|\tau_y f - \tau_y f_n\|_p + \|\tau_y f_n - f_n\|_p + \|f_n - f\|_p \\ &\leq 2\|f - f_n\|_p + \|\tau_y f_n - f_n\|_p. \end{aligned}$$

Soit $\varepsilon > 0$, il existe $n_0 \geq 0$ tel que $\|f - f_{n_0}\|_p \leq \varepsilon/4$. D'après ce qui précède, on peut trouver $\eta > 0$ tel que $\|y\| \leq \eta$ implique $\|\tau_y f_{n_0} - f_{n_0}\|_p \leq \varepsilon/2$. Ce qui conclut la preuve du lemme. \square

Revenons à la preuve du théorème.

Démonstration. On cherche à montrer la convergence vers zéro de la quantité $(*) = \|\phi_n * f - f\|_1$. On utilise d'abord le fait que $\int \phi_n(x) dx = 1$, on obtient

$$\|\phi_n * f - f\|_1 = \int |\phi_n * f(x) - f(x)| dx = \int \left| \int f(x-y)\phi_n(y) dy - \int f(x)\phi_n(y) dy \right| dx.$$

Des majorations standards ainsi que le théorème de Tonelli implique

$$\|\phi_n * f - f\|_1 \leq \int |\phi_n(y)| \left(\int |\tau_{-y} f(x) - f(x)| dx \right) dy \leq \int |\phi_n(y)| \|\tau_{-y} f - f\|_1 dy.$$

Soit $\varepsilon > 0$. On écrit $\mathbf{1} = \mathbf{1}_{\|y\| \leq \varepsilon} + \mathbf{1}_{\|y\| > \varepsilon}$. Nous avons alors d'une part, par l'inégalité de Hölder

$$\int_{\|y\| \leq \varepsilon} |\phi_n(y)| \|\tau_{-y}f - f\|_1 dy \leq \sup_{\|y\| \leq \varepsilon} \|\tau_y f - f\|_1 \times \sup_{n \geq 0} \int |\phi_n(y)| dy = \mathcal{O} \left(\sup_{\|y\| \leq \varepsilon} \|\tau_y f - f\|_1 \right).$$

D'autre part,

$$\int_{\|y\| > \varepsilon} |\phi_n(y)| \|\tau_{-y}f - f\|_1 dy \leq 2\|f\|_1 \int_{\|y\| > \varepsilon} |\phi_n(y)| dy.$$

Ceci implique que

$$\limsup_{n \rightarrow \infty} \|\phi_n * f - f\|_1 = \mathcal{O} \left(\sup_{\|y\| \leq \varepsilon} \|\tau_y f - f\|_1 \right).$$

Comme $\varepsilon > 0$ peut être choisi arbitrairement petit, cela termine la preuve du théorème. □

Régularisation par convolution

Le théorème suivant et surtout son corollaire n'est qu'un exemple parmi tant d'autre de l'intérêt de l'effet régularisant de la convolution.

Théorème 6.4.16. *On fixe $k \in \mathbb{N} \cup \{\infty\}$. Soient $g \in \mathbf{L}^1$ et $f \in C_b^k(\mathbb{R}^d)$, i.e. f est k fois continûment dérivable et toutes ses dérivées partielles jusqu'à l'ordre k sont bornées. Alors $g * f$ a un sens et $g * f \in C_n^k(\mathbb{R}^d)$. Pour $\alpha \in \mathbb{N}^d$, on note $\partial_\alpha = \partial_1^{\alpha_1} \dots \partial_d^{\alpha_d}$. Alors pour tout $\alpha \in \mathbb{N}^d$, $|\alpha| \leq k$, $\partial_\alpha(f * g) = (\partial_\alpha f) * g$.*

Démonstration. Soient $g \in \mathbf{L}^1$ et $f \in C_b^k(\mathbb{R}^d)$. Par l'inégalité de Hölder, $f * g$ est bien définie. L'application qui à $(x, y) \rightarrow f(x - y)g(y)$ est mesurable en y et de classe C^k en la variable x . Soit $\alpha \in \mathbb{N}^d$ tel que $|\alpha| \leq k$, alors $|\partial_\alpha f(x - y)g(y)| = \mathcal{O}(|g(y)|)$, ce qui implique que $x \rightarrow \int f(x - y)g(y) dy$ est de classe C_b^k . De plus,

$$\forall \alpha \in \mathbb{N}^d, |\alpha| \leq k, \quad \partial_\alpha g * f(x) = \int \partial_\alpha f(x - y)g(y) dy.$$

D'où le résultat. □

Corollaire 6.4.17. *L'espaces $C_c^\infty(\mathbb{R}^d) \cap \mathbf{L}^1$ (ainsi que tous les espaces $C_c^k(\mathbb{R}^d)$) est dense dans \mathbf{L}^1 .*

Exercice 28. Montrer le corollaire. On pourra utiliser comme fonction de base la fonction ϕ suivante :

$$\phi(x) = \begin{cases} \exp \left[-\frac{1}{1 - \|x\|^2} \right] & \text{si } \|x\| < 1, \\ 0 & \text{sinon.} \end{cases}$$

Deuxième partie

Probabilités générales

Chapitre 7

Variables aléatoires réelles et vecteurs aléatoires

Dans ce chapitre et les suivants, on se donne un espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$. Nous allons définir les notions de variables aléatoires et étudier plus précisément les notions de variables aléatoires réelles et vecteurs aléatoires.

Le formalisme des probabilités est le même que celui de la théorie de la mesure : une variable aléatoire n'est rien d'autre qu'une application mesurable. Toujours en termes de terminologie, en probabilité, les ensembles mesurables $A \in \mathcal{F}$ s'appellent traditionnellement des événements.

7.1 Variables aléatoires

Définition 7.1.1 (Variable aléatoire). Soient $(\Omega, \mathcal{F}, \mathbf{P})$ un espace probabilisé et (E, \mathcal{E}) un espace mesurable. Une variable aléatoire à valeurs dans E est une application mesurable $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{E})$.

Exemple 31 (Pile ou face). On pose $\Omega = \{0, 1\}$ muni de la tribu $\mathcal{P}(\Omega)$ et on pose $\mathbf{P} = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. On pose également $E = \{p, f\}$ muni de la tribu $\mathcal{P}(E)$. Alors l'application $X : \Omega \rightarrow E$ définie par $X(0) = p$ et $X(1) = f$ est une variable aléatoire. Elle modélise l'expérience aléatoire du pile ou face.

Exemple 32 (Le dé à 6 faces). On pose $\Omega = \{1, 2, \dots, 6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ et \mathbf{P} la mesure uniforme sur Ω . De même, on pose $(E, \mathcal{E}) = (\Omega, \mathcal{F})$. Alors l'application $X : \Omega \rightarrow E$ qui à $\omega \in \Omega$ associe $X(\omega) = \omega$ est une variable aléatoire. Elle modélise l'expérience le lancé d'un dé équilibré.

Une variable aléatoire n'est donc rien de plus qu'une application mesurable entre deux espaces mesurables. Traditionnellement, on utilise des lettres capitales $X, Y, Z \dots$ pour désigner de telles variables aléatoires.

Définition 7.1.2 (Loi d'une variable aléatoire). Soit X une variable aléatoire à valeurs dans E . La loi de X , notée \mathbf{P}_X , est la mesure image par X de la probabilité \mathbf{P} . Autrement dit, c'est la mesure de probabilité sur (E, \mathcal{E}) définie pour tout $A \in \mathcal{E}$ par $\mathbf{P}_X(A) = \mathbf{P}(X^{-1}(A))$.

En termes de notations, on préférera souvent écrire $\mathbf{P}(X \in A)$ en lieu et place de la notation un peu lourde $\mathbf{P}(X^{-1}(A))$.

Une grande partie du travail en théorie des probabilités consiste à caractériser la loi d'une variable aléatoire X , c'est à dire, stricto sensu, renseigner la valeur de $\mathbf{P}(X \in A) \in [0, 1]$ pour tout $A \in \mathcal{E}$. Remarquons que si X est à valeurs dans E fini et si $\mathcal{E} = \mathcal{P}(E)$ alors $\text{card } \mathcal{E} = 2^{\text{card } E}$. Il paraît alors vite inenvisageable d'énumérer les valeurs de $\mathbf{P}(X \in A)$ pour tout $A \in \mathcal{E}$. Dans la suite, on explicitera d'autres méthodes permettant de caractériser la loi de X de manière plus compacte sans faire cette énumération fastidieuse. À titre d'illustration, il est clair que dans le cas E fini ci-dessus il est suffisant de renseigner $\mathbf{P}(X = k) = \mathbf{P}(X \in \{k\})$ pour tout $k \in E$ puisque si $A \in \mathcal{P}(E)$, alors

$$\mathbf{P}(X \in A) = \mathbf{P}\left(X \in \bigcup_{k \in A} \{k\}\right) = \sum_{k \in A} \mathbf{P}(X = k),$$

du fait de la σ -additivité de la mesure \mathbf{P} .

Pour l'exemple 31, la loi de X est même complètement caractérisée par $\mathbf{P}(X = f)$ car $\mathbf{P}(X = p) = 1 - \mathbf{P}(X = f)$.

Le choix du triplet probabiliste n'est pas unique mais il n'est pas complètement arbitraire non plus. En particulier, il faut choisir un espace suffisamment gros pour exprimer les variables aléatoires modélisant le problème idoine. Par exemple, considérons le problème du pile ou face. Outre le choix fait dans l'exemple 31, nous aurions pu poser $(\Omega, \mathcal{F}, \mathbf{P}) = ([0, 1], \mathcal{B}([0, 1]), \lambda)$ et définir $Y : \Omega \rightarrow E = \{p, f\}$ par

$$Y(\omega) = \begin{cases} p & \text{si } \omega \in [0, 1/2) \\ f & \text{si } \omega \in [1/2, 1] \end{cases}.$$

Remarquons finalement que $\mathbf{P}(Y = f) = 1 - \mathbf{P}(Y = p) = \lambda([0, 1/2)) = 1/2$. Autrement dit, pour ces deux choix de triplets probabilistes, les variables aléatoires X et Y ont même loi.

La proposition suivante, qui découle directement de la définition de la loi d'une variable aléatoire, donne une façon de caractériser la loi d'une variable aléatoire.

Proposition 7.1.3. *Soient $(\Omega, \mathcal{F}, \mathbf{P})$ et $(\Omega', \mathcal{F}', \mathbf{P}')$ deux espaces probabilisés et (E, \mathcal{E}) un espace mesurable. Alors deux variables aléatoires X et Y à valeurs dans E ont même loi si et seulement si pour toute fonction borélienne $f : (E, \mathcal{E}) \rightarrow \mathbb{R}$ bornée*

$$\int_{\Omega} f(X) d\mathbf{P} = \int_{\Omega'} f(Y) d\mathbf{P}'.$$

Remarque 59. En probabilité, on note

$$\mathbf{E}(f(X)) = \int_{\Omega} f(X) d\mathbf{P} = \int_E f(x) \mathbf{P}_X(dx),$$

où la seconde inégalité est obtenue par le théorème 3.2.6 de transfert.

Démonstration. C'est une condition suffisante car pour $A \in \mathcal{E}$, l'égalité appliquée à $f = \mathbf{1}_A$ implique $\mathbf{P}_X(A) = \mathbf{P}_Y(A)$.

Réciproquement, par hypothèse, pour $f = \mathbf{1}_A$, à l'aide du théorème de transfert, on a

$$\int_E \mathbf{1}_A(z) \mathbf{P}_X(dz) = \int_E \mathbf{1}_A(z) \mathbf{P}_Y(dz).$$

La preuve s'achève en remarquant que l'égalité est satisfaite pour les fonctions étagées positives, les fonctions boréliennes positives et enfin les fonctions boréliennes bornées. \square

Remarque 60. On remarque que X et Y peuvent être définies sur deux espaces probabilisés différents, ce qui reflète encore une fois l'idée que le choix du triplet probabiliste $(\Omega, \mathcal{F}, \mathbf{P})$ n'est pas unique.

Supposons que l'espace (E, \mathcal{E}) soit muni d'une mesure σ -finie μ telle que $\mathbf{P}_X \prec \mu$ alors le théorème 6.3 de Radon-Nikodym garantit l'existence d'une fonction $f : (E, \mathcal{E}) \rightarrow \mathbb{R}$ mesurable positive telle que

$$\mathbf{P}_X(A) = \mathbf{P}(X \in A) = \int \mathbf{1}_A f d\mu. \quad (7.1)$$

Remarquons que $\mathbf{P}_X(E) = 1$ si bien que $\int f d\mu = 1$.

Une variable aléatoire X dont la loi \mathbf{P}_X satisfait l'équation (7.1) pour une fonction f mesurable positive et telle que $\int f d\mu = 1$ sera dite à densité de probabilité par rapport à μ ou plus simplement à densité si il n'y a aucune ambiguïté. La fonction f sera appelée la densité de X par rapport à μ .

Lorsqu'une variable aléatoire X est à densité par rapport à une mesure de référence μ fixée, cette densité caractérise la loi de X .

Proposition 7.1.4. *Soient $(\Omega, \mathcal{F}, \mathbf{P})$ et $(\Omega', \mathcal{F}', \mathbf{P}')$ deux espaces probabilisés et (E, \mathcal{E}, μ) un espace mesuré σ -fini. Alors deux variables aléatoires X et Y à valeurs dans E et à densité f_X et f_Y respectivement par rapport à μ ont même loi, i.e. $\mathbf{P}_X = \mathbf{P}'_Y$, si et seulement si $f_X = f_Y$ μ -p.p..*

Démonstration. C'est évidemment une condition suffisante car pour tout $A \in \mathcal{E}$

$$\mathbf{P}_X(A) = \int_E f_X \mathbf{1}_A d\mu = \int_E f_Y \mathbf{1}_A d\mu = \mathbf{P}'_Y(A).$$

C'est également une condition nécessaire puisque $\mathbf{P}_X = \mathbf{P}'_Y$. Ces deux mesures (finies) sont absolument continue par rapport à μ , qui est σ -finie, et l'unicité de la densité dans le théorème de Radon-Nikodym implique que $f_X = d\mathbf{P}_X/d\mu$ et $f_Y = d\mathbf{P}'_Y/d\mu$ sont égales μ -p.p. \square

Remarque 61. Lorsque E est dénombrable (muni de la tribu idoine), la mesure de comptage est une mesure de référence de choix. Dans ce cas, la proposition ci-dessus dit que X et Y ont même loi si et seulement si $\mathbf{P}(X = x) = \mathbf{P}'(Y = x)$ pour tout $x \in E$.

7.2 Variables aléatoires réelles

On s'intéresse ici aux variables aléatoires à valeurs dans \mathbb{R} qui est muni de la tribu borélienne.

Définition 7.2.1. Une variable aléatoire réelle est une application mesurable X de $(\Omega, \mathcal{F}, \mathbf{P})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

On se réfère à l'annexe 13.7.2 pour un tableau de quelques lois usuelles.

7.2.1 Intégration des variables aléatoires réelles

Moment d'ordre 1, Moment d'ordre p

Tout ce qui suit est la transposition des définitions de l'analyse fonctionnelle au contexte des variables aléatoires ainsi que des applications du théorème de transfert.

Définition 7.2.2. Une *v.a.r* est dite intégrable si

$$\int |X| d\mathbf{P} = \int |x| \mathbf{P}_X(dx) < \infty.$$

Définition 7.2.3. Soit X une *v.a.r* positive ou intégrable. L'espérance de X , notée $\mathbf{E}(X)$ est définie par

$$\mathbf{E}(X) = \int X d\mathbf{P} = \int x \mathbf{P}_X(dx).$$

Soit X une variable aléatoire à valeurs dans \mathbb{N} , alors \mathbf{P}_X est absolument continue par rapport à la mesure de comptage sur \mathbb{N} et

$$\mathbf{P}_X = \sum_{k \in \mathbb{N}} \mathbf{P}(X = k) \delta_k.$$

Ainsi,

$$\mathbf{E}(X) = \int_{\Omega} X d\mathbf{P} = \sum_{k \geq 0} k \mathbf{P}(X = k).$$

Si X est une variable de densité f par rapport à la mesure de Lebesgue, cette fois-ci on calcule l'espérance par

$$\mathbf{E}(X) = \int_{\Omega} X d\mathbf{P} = \int_{\mathbb{R}} x f(x) dx.$$

Remarque 62. Souvent, la mesure de référence ne sera pas spécifiée et devra être comprise implicitement à l'aide du contexte. Concrètement, ce sera la mesure de comptage sur les espaces discrets et la mesure de Lebesgue sur \mathbb{R} . Cela conduira à parler de *v.a.* discrètes et de *v.a.r.* à densité.

Exemple 33. Soient X, Y et Z des *v.a.r.* de loi uniforme standard $\mathcal{U}_{[0,1]}$, exponentielle $\mathcal{E}(\lambda)$ et de Cauchy $\mathcal{C}(1)$ respectivement. Alors, X et Y sont intégrables. En effet,

$$\mathbf{E}(|X|) = \int_{\mathbb{R}} \mathbf{1}_{[0,1]}(x) |x| dx = 1/2 < \infty \quad \text{et} \quad \mathbf{E}(|Y|) = \int_{\mathbb{R}} \mathbf{1}_{\mathbb{R}^+}(x) \lambda e^{-\lambda x} |x| dx < \infty.$$

De plus,

$$\mathbf{E}(X) = 1/2 \quad \text{et} \quad \mathbf{E}(Y) = 1/\lambda.$$

Quant à la variable de Cauchy Z , on remarque que

$$\mathbf{E}(|Z|) = \int_{\mathbb{R}} \frac{|x| dx}{\pi(1+x^2)} = \infty.$$

Définition 7.2.4. Une *v.a.r.* X admet un moment d'ordre p , $p \in [1, \infty)$, et on note $X \in \mathbf{L}^p$, si $\int |X|^p d\mathbf{P} < \infty$. Dans ce cas, le moment d'ordre p de X est défini par

$$\mathbf{E}(X^p) = \int X^p d\mathbf{P}.$$

Définition 7.2.5. Soit X une *v.a.r.* admettant un moment d'ordre 2. La variance de X , notée $\mathbf{V}(X)$, est définie par

$$\mathbf{V}(X) = \mathbf{E}[X - \mathbf{E}(X)]^2$$

Si l'espérance est un paramètre dit de position, la variance est un paramètre de dispersion. Il existe d'autre paramètre en statistique (médiane, quantile, intervalle interquartile etc), cependant moyenne et variance restent centraux eût égard notamment à la loi des grands nombres et le théorème central limite que l'on démontrera au chapitre 12.

Proposition 7.2.6. Soit X une *v.a.r.* admettant un moment d'ordre 2, alors

$$\mathbf{V}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2.$$

Covariance et coefficient de corrélation linéaire

Définition 7.2.7. Soient X, Y deux *v.a.r.* admettant un moment d'ordre 2, on appelle covariance entre X et Y la quantité $\text{Cov}(X, Y)$ définie par

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

Le coefficient de corrélation linéaire entre les variables X et Y , noté $\rho(X, Y)$, est donné par

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\mathbf{V}(X)\mathbf{V}(Y)}}$$

Proposition 7.2.8. Soient X et Y deux *v.a.r.* admettant un moment d'ordre 2. La covariance est une application bilinéaire symétrique vérifiant $\mathbf{V}(X) = \text{Cov}(X, X)$. De plus, $\rho(X, Y) \in [-1, 1]$ et $\rho(X, X) = 1$. Enfin,

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

Inégalité de Markov et inégalité de Bienaymé-Tchebychev

Proposition 7.2.9 (Inégalité de Markov). Soit X une *v.a.r.* positive. Alors pour tout $\lambda > 0$

$$\mathbf{P}(X > \lambda) \leq \frac{\mathbf{E}(X)}{\lambda}.$$

Proposition 7.2.10 (Inégalité de Bienaymé-Tchebychev). Soit X une *v.a.r.* admettant un moment d'ordre 2. Alors pour $\lambda > 0$

$$\mathbf{P}(|X - \mathbf{E}(X)| > \lambda) \leq \frac{\mathbf{V}(X)}{\lambda^2}.$$

Démonstration. Remarquer que

$$\mathbf{P}(|X - \mathbf{E}(X)| > \lambda) = \mathbf{P}(|X - \mathbf{E}(X)|^2 > \lambda^2)$$

et appliquer l'inégalité de Markov. □

Proposition 7.2.11. Soit X une v.a.r. et $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction convexe telles que X et $\varphi(X)$ soient intégrables. Alors $\varphi(\mathbf{E}(X)) \leq \mathbf{E}(\varphi(X))$.

Lemme 7.2.12 (des trois cordes). Soit I un intervalle et $\varphi : I \rightarrow \mathbb{R}$ une fonction convexe. Alors pour tout $a, b, c \in I$ tels que $a < c < b$

$$\frac{\varphi(c) - \varphi(a)}{c - a} \leq \frac{\varphi(b) - \varphi(a)}{b - a} \leq \frac{\varphi(b) - \varphi(c)}{b - c}.$$

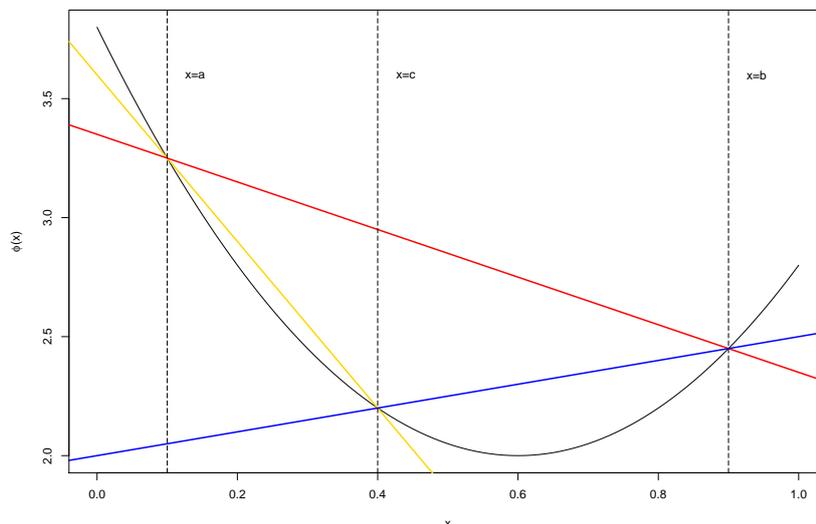


FIGURE 7.1 – La corde “rouge” a une pente plus grande que la corde “jaune” et plus petite que la corde “bleu”.

Démonstration. Il est assez facile de voir que ces inégalités sont en fait équivalentes à l’unique inégalité

$$\varphi(c) \leq \frac{b - c}{b - a} \varphi(a) + \frac{c - a}{b - a} \varphi(b).$$

En posant $t = \frac{b-c}{b-a}$ on obtient facilement que $c = ta + (1 - t)b$ — remarquons que $a \neq b$. L’inégalité ci-dessus n’est alors qu’une traduction de la propriété de convexité. \square

Remarque 63. Remarquons que les inégalités des trois cordes caractérise la convexité.

Démonstration. Soit $a \in \mathbb{R}$ et définissons la fonction $\tau_a : x \rightarrow \frac{\varphi(x) - \varphi(a)}{x - a}$. Notons que τ_a est bien définie sauf peut-être en a . D’après le lemme 7.2.12 des trois cordes appliqués aux trois cas $a < x < y$, $x < a < y$ et $x < y < a$, la fonction τ_a est croissante. Puisque $\tau_a(2a)$ et $\tau_a(a/2)$ sont des réels, il vient que τ_a admet une limite en a à gauche $\varphi'_g(a) > -\infty$ et une limite à droite $\varphi'_d(a) < \infty$. Il est également claire que $\varphi'_d(a) \geq \varphi'_g(a)$ par croissance de τ_a .

Par définition de φ'_d , pour tout $x > a$,

$$\frac{\varphi(x) - \varphi(a)}{x - a} \geq \varphi'_d(a) \iff \varphi(x) \geq \varphi(a) + \varphi'_d(a)(x - a).$$

D’autre part, pour tout $x < a$:

$$\frac{\varphi(x) - \varphi(a)}{x - a} \leq \varphi'_g(a) \leq \varphi'_d(a) \iff \varphi(x) \geq \varphi(a) + \varphi'_d(a)(x - a).$$

En posant $a = \mathbf{E}(X)$ et $x = X$, on obtient par croissance de l’intégrale

$$\varphi(X) \leq \varphi(\mathbf{E}(X)) + \varphi'_d(\mathbf{E}(X))(X - \mathbf{E}(X)) \implies \mathbf{E}(\varphi(X)) \geq \varphi(\mathbf{E}(X)).$$

\square

7.2.2 Caractérisation de la loi d'une v.a.r.

Dans cette partie, on explicite différente façon de caractériser la loi d'une v.a.r..

Fonction de répartition

Définition 7.2.13 (Fonction de répartition). Soit X une v.a.r., on appelle fonction de répartition de X la fonction $F_X : \mathbb{R} \rightarrow [0, 1]$ définie pour $t \in \mathbb{R}$ par

$$F_X(t) = \mathbf{P}_X((-\infty, t]) = \mathbf{P}(X \in (-\infty, t]) = \mathbf{P}(X \leq t).$$

Proposition 7.2.14. La fonction de répartition F_X d'une v.a.r X est :

1. croissante à valeurs dans $[0, 1]$;
2. continue à droite ;
3. $\lim_{t \rightarrow -\infty} F_X(t) = 0$ et $\lim_{t \rightarrow \infty} F_X(t) = 1$.

Réciproquement, pour toute fonction F vérifiant les propriétés 1,2 et 3 ci-dessus il existe une v.a.r. X telle que $F_X = F$.

Remarque 64. Toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ croissante admet une limite à gauche et une limite à droite. En particulier, une fonction de répartition est limitée à gauche.

Remarque 65. La réciproque de cette proposition signifie la chose suivante : si on se donne une fonction F vérifiant les points 1,2 et 3, alors il existe un espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$ et une variable aléatoire X de $(\Omega, \mathcal{F}, \mathbf{P})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ tel que $\mathbf{P}(X \leq t) = F(t)$ pour tout $t \in \mathbb{R}$.

Démonstration. Vérifions les trois points.

1. Tout d'abord, puisque $F_X(t) = \mathbf{P}_X((-\infty, t])$ est une probabilité, F_X est à valeurs dans $[0, 1]$. La croissance de F_X est une conséquence de la croissance des mesures : si $A \subset B$, alors $\mathbf{P}_X(A) \leq \mathbf{P}_X(B)$.
2. La continuité à droite provient de la continuité à droite des mesures. En effet, soient $t \in \mathbb{R}$ et $(t_n)_{n \geq 0}$ une suite de réels tels que $t_n > t$ et $\lim_{n \rightarrow \infty} t_n = t$. La fonction F_X est croissante donc quitte à considérer une sous-suite, on peut supposer $(t_n)_{n \geq 0}$ décroissante. Posons $A_n = (-\infty, t_n]$. La suite d'ensembles mesurables $(A_n)_{n \geq 0}$ est décroissante et $\mathbf{P}_X(A_0) < \infty$ puisque \mathbf{P}_X est une probabilité. Ainsi,

$$\mathbf{P}_X((-\infty, t]) = \mathbf{P}_X \left(\bigcap_{n \geq 0} A_n \right) = \lim_{n \rightarrow \infty} \mathbf{P}_X(A_n) = \lim_{n \rightarrow \infty} \mathbf{P}_X((-\infty, t_n]).$$

Cela montre la continuité à droite de F_X .

3. On peut considérer A_n avec $t_n = -n$ pour tout $n \geq 0$. Alors,

$$\lim_{t \rightarrow -\infty} F_X(t) = \mathbf{P}_X \left(\bigcap_{n \geq 0} A_n \right) = 0.$$

Pour la limite en ∞ , on peut poser $B_n = (-\infty, n]$ et utiliser la continuité à gauche des mesures.

La réciproque est un corollaire du théorème de Stieltjes 2.2.33. □

Exemple 34. La fonction de répartition X donnant la valeur numérique de la face d'une dé équilibré à six faces est donnée pour $t \in \mathbb{R}$ par

$$F_X(t) = \left[\frac{1}{6} \mathbf{1}_{[1,2)}(t) + \frac{2}{6} \mathbf{1}_{[2,3)}(t) + \frac{3}{6} \mathbf{1}_{[3,4)}(t) + \frac{4}{6} \mathbf{1}_{[4,5)}(t) + \frac{5}{6} \mathbf{1}_{[5,6)}(t) + \mathbf{1}_{[6,\infty)}(t) \right].$$

Proposition 7.2.15. La fonction de répartition caractérise la loi d'une v.a.r. : si X et Y sont deux v.a.r., alors $F_X = F_Y$ si et seulement si X et Y ont même loi.

Démonstration. C'est une conséquence directe du théorème 2.2.20 car

$$\mathcal{S} = \{(-\infty, a], a \in \mathbb{R}\}$$

est un π -système (non vide et stable par intersection finie). \square

Lemme 7.2.16. *Soit h une fonction croissante de \mathbb{R} dans \mathbb{R} . Alors h admet un nombre dénombrable de discontinuités.*

Les limites à gauche et à droite de h au point $t \in \mathbb{R}$ sont communément notées $h(t^-)$ et $h(t^+)$. En particulier, h est continue à droite (*resp.* à gauche, *resp.* continue) si et seulement si $h(t) = h(t^+)$ (*resp.* $h(t) = h(t^-)$, *resp.* $h(t) = h(t^+) = h(t^-)$).

Démonstration. L'ensemble des points de discontinuités de h s'écrit

$$\{t \in \mathbb{R} : h(t^+) - h(t^-) > 0\} = \bigcup_{M \in \mathbb{N}} \bigcup_{n \geq 1} \left\{ t \in [-M, M] : h(t^+) - h(t^-) > \frac{1}{n} \right\}.$$

Or $\sup_{t \in [-M, M]} h(t^+) - h(t^-) \leq h(M+1) - h(-M-1) = K < \infty$, donc le nombre de discontinuités de $[-M, M]$ plus grande que $\frac{1}{n}$ est majoré par Kn . Ainsi l'ensemble des points de discontinuités de h est réunion dénombrable d'ensembles finis, il est dénombrable. \square

En particulier, l'ensemble $D_F = \{t \in \mathbb{R} : F(t) - F(t^-) > 0\}$ des points de discontinuités de la fonction de répartition F est dénombrable.

Définition 7.2.17 (Variables aléatoires discrètes, continues). 1. Une *v.a.r.* X est dite discrète si il existe un ensemble $A \subset \mathbb{R}$ au plus dénombrable tel que $\mathbf{P}(X \in A) = 1$.

2. Une *v.a.r.* X est dite continue ou diffuse si pour tout $a \in \mathbb{R}$, $\mathbf{P}(X = a) = 0$.

Ces deux propriétés peuvent être caractérisée à l'aide de la fonction de répartition.

Proposition 7.2.18. 1. Une *v.a.r.* X est discrète si et seulement si la somme des sauts de F_X vaut 1, i.e.

$$\sum_{t \in D_{F_X}} F(t) - F(t^-) = 1.$$

2. Une *v.a.r.* X est continue si et seulement si la fonction de répartition F_X est continue.

Démonstration. Soit X une *v.a.r.* et posons $A = \{s \in \mathbb{R} : \mathbf{P}(X = s) > 0\}$. Il se trouve que $\mathbb{P}(X = s) = F(s) - F(s^-)$ pour tout $s \in \mathbb{R}$, aussi A n'est rien d'autre que l'ensemble de discontinuité de F_X . De là, la *v.a.r.* X est continue si et seulement si $\mathbf{P}(X \in A) = 0$; elle est discrète si et seulement si $\mathbf{P}(X \in A) = 1$. \square

Densité de probabilité

Une *v.a.r.* X est dite à densité si elle est à densité par rapport à la mesure de Lebesgue. Pour une *v.a.r.* X à densité f , sa fonction de répartition F_X se calcule, par définition d'une mesure à densité, pour $t \in \mathbb{R}$ par

$$F_X(t) = \mathbf{P}_X((-\infty, t]) = \int \mathbf{1}_{(-\infty, t]} f \, d\lambda = \int_{-\infty}^t f(x) \, dx.$$

Proposition 7.2.19. *Soit X une *v.a.r.* à densité f . On note F sa fonction de répartition. Alors F est continue sur \mathbb{R} et est dérivable presque-partout. Sa dérivée est presque partout égale à f .*

Démonstration. Puisque F est une fonction de répartition, F est continue à droite. Il reste donc à montrer que F est continue à gauche. Soit $t \in \mathbb{R}$ et $(t_n)_{n \geq 0}$ une suite de réels qui converge vers t et telle que $t_n \leq x$ pour tout $n \geq 0$. Alors, on vérifie facilement que $f \mathbf{1}_{(-\infty, t_n]}$ converge simplement vers $f \mathbf{1}_{(-\infty, t)}$ donc vers $f \mathbf{1}_{(-\infty, t]}$ presque partout (par rapport à la mesure de Lebesgue). De plus, $0 \leq f \mathbf{1}_{(-\infty, t_n]} \leq f$ qui est intégrable (c'est une densité de probabilité). Le théorème de convergence dominée implique

$$\lim_{n \rightarrow \infty} F(t_n) = \lim_{n \rightarrow \infty} \int f \mathbf{1}_{(-\infty, t_n]} \, d\lambda = \int f \mathbf{1}_{(-\infty, t]} \, d\lambda = F(t).$$

La seconde partie de la proposition n'est en fait rien d'autre que le théorème de différentiation de Lebesgue. Sa preuve est un peu plus fine et fait intervenir l'inégalité maximale de Hardy-Littlewood. On pourra se référer à [Rud87] pour la preuve complète. \square

Exemple 35. La loi d'une variable aléatoire X exponentielle de paramètre $\lambda > 0$, notée $\mathcal{E}(\lambda)$ est caractérisée par sa densité définie par

$$f_X(x) = \mathbf{1}_{\mathbb{R}_+}(x) \lambda e^{-\lambda x}, \quad x \in \mathbb{R}.$$

Sa fonction de répartition est donc

$$F_X(t) = \int_{-\infty}^t f_X(x) dx = \mathbf{1}_{\mathbb{R}_+}(t) [1 - e^{-\lambda t}], \quad t \in \mathbb{R}.$$

Proposition 7.2.20. *La loi d'une v.a.r. X à densité est caractérisée par sa densité de probabilité : si X et Y ont pour densité f_X et f_Y respectivement alors $f_X = f_Y$ presque partout si et seulement si X et Y ont même loi.*

Démonstration. Ce résultat est un corollaire de la proposition 7.1.4. \square

Puisqu'une fonction de répartition est dérivable presque partout, on pourrait penser de prime abord que la dérivée (définie seulement presque partout néanmoins) est une densité de probabilité. Ceci n'est pas vrai en général comme le montre l'exemple suivant.

Exemple 36. Soit F la fonction réelle définie par :

$$F(t) = \begin{cases} 0, & t < 0 \\ 1, & t \geq 0. \end{cases}$$

On vérifie facilement que c'est une fonction de répartition. Elle est dérivable presque partout (en fait partout sauf en 0) et la dérivée est la fonction nulle presque partout (sauf en 0 également). Il est bien évident que cette fonction dérivée n'est pas une densité de probabilité.

La fonction de répartition de l'exemple 36 est celle de la masse de Dirac en 0. Autrement dit, une v.a.r. X admettant la fonction de répartition F de l'exemple 36 satisfait $\mathbf{P}(X = 0) = 1$, c'est une variable purement discrète.

L'exemple suivant est encore plus fin : on construit une fonction de répartition continue dont la v.a.r. correspondante n'est pas à densité.

Exemple 37 (Escalier de Cantor). L'escalier du diable, ou l'escalier de Cantor se construit comme une limite uniforme de fonctions continues. Plus précisément, on définit $(f_n)_{n \geq 0}$ une suite de fonction continue sur $[0, 1]$ par récurrence — c.f. Figure 7.2 :

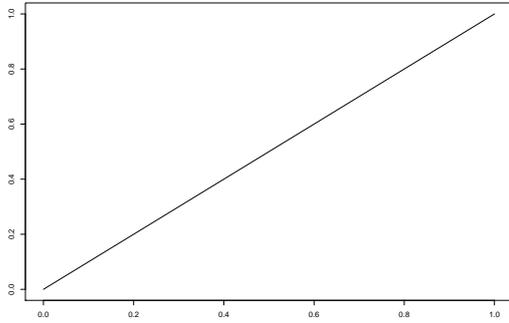
- $f_0(x) = x, x \in [0, 1]$;
- on construit f_{n+1} à partir de f_n en remplaçant f_n sur chaque intervalle d'intérieur non vide $[u, v]$ qui ne contient pas de plateaux de f_n par une fonction affine par morceaux qui est constante égale à $\frac{f_n(u)+f_n(v)}{2}$ sur le tiers central de $[u, v]$.

Par construction, f_{n+1} et f_n ne diffèrent que sur les intervalles non vides $[u, v]$ ne contenant pas de plateaux de f_n . Un tel intervalle sépare les extrémités de deux plateaux successifs de f_n dont la hauteur est de 2^{-n} (on divise par 2 étape par étape). Ainsi, pour tout $x \in [0, 1]$ et tout $n \in \mathbb{N}$,

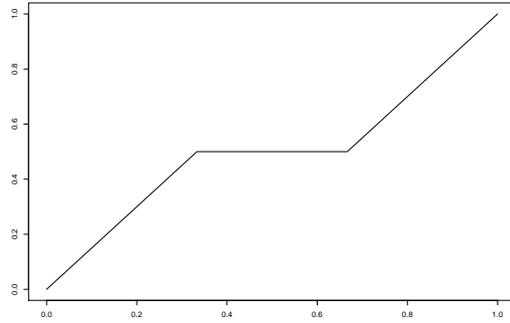
$$|f_{n+1}(x) - f_n(x)| \leq 2^{-n}.$$

Il vient que la séries $\sum (f_{n+1} - f_n)$ converge donc uniformément. Ainsi, la suite $(f_n)_{n \geq 0}$ converge uniformément vers une fonction f continue monotone croissante. La fonction f est appelée escalier du diable.

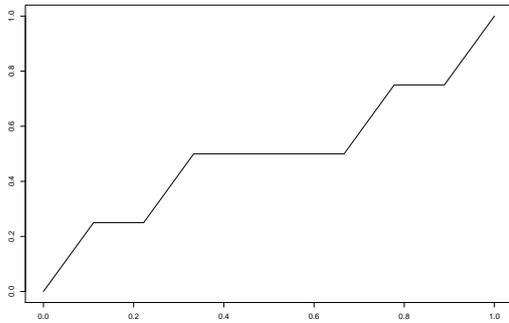
Soit $F : \mathbb{R} \rightarrow \mathbb{R}$ supposée nulle sur $(-\infty, 0]$, constante égale à 1 sur $[1, \infty)$ et égale à l'escalier du diable, c'est à dire f , sur $[0, 1]$. Du fait de ce que l'on vient de montrer, F est une fonction de répartition.



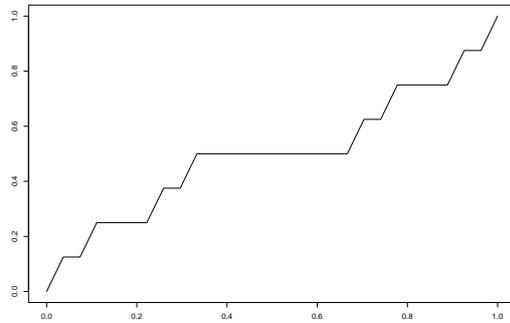
(a) Étape 0.



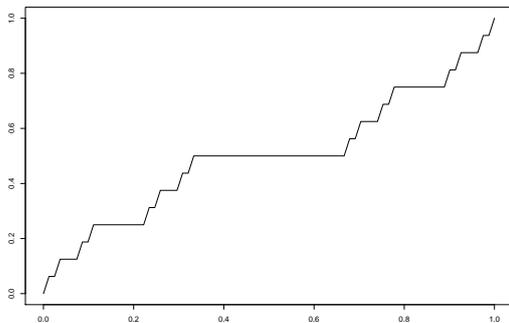
(b) Étape 1.



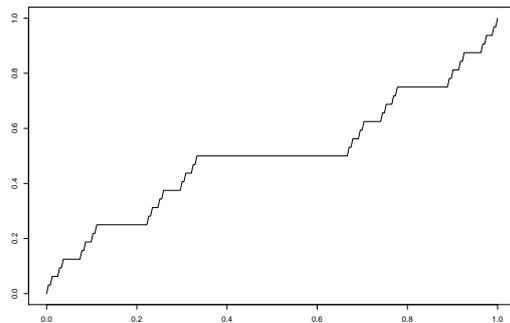
(c) Étape 2.



(d) Étape 3.



(e) Étape 4.



(f) Étape 5.

FIGURE 7.2 – La fonction f_0 est l'identité de $[0, 1]$ dans $[0, 1]$. À l'étape $n + 1$, on subdivise chaque intervalle sur lesquels f_n n'est pas constante en trois sous-intervalles de même longueur. Alors f_{n+1} est constante sur le sous-intervalle central alors qu'elle est affine sur les deux autres sous-intervalles de sorte que f_{n+1} soit continue.

La fonction F est la fonction de répartition d'une *v.a.r.* X continue mais comme on va le montrer tout de suite elle n'est pas à densité. En effet, la fonction f est dérivable au moins sur les plateaux P_n des fonctions f_n . Or,

$$\lambda \left(\bigcup_{n \geq 0} P_n \right) = \sum_{n \geq 0} \frac{1}{3} \left(\frac{2}{3} \right)^n,$$

ce qui montre que f est dérivable presque partout et f' est presque partout nulle.

En fait, il faut revenir au théorème de Radon-Nikodym : la mesure \mathbf{P}_X admet une densité de probabilité par rapport à la mesure de Lebesgue (qui est σ -finie) si $\mathbf{P}_X \prec \lambda$. Il est d'ailleurs clair que la mesure définie par l'exemple 37 n'est pas absolument continue par rapport à la mesure de Lebesgue, on dit qu'elle est singulière. Rappelons la notion d'absolue continuité pour les fonctions.

Définition 7.2.21. Soit $I = [a, b]$ un intervalle. La fonction $F : [a, b] \rightarrow \mathbb{R}$ est dite absolument continue si pour tout réel $\varepsilon > 0$, il existe $\delta > 0$ et $([a_n, b_n])_{n \geq 0}$ des sous-intervalles de I d'intérieurs disjoints tels que

$$\sum_{n \geq 0} (b_n - a_n) \leq \delta \implies \sum_{n \geq 0} |F(a_n) - F(b_n)| \leq \varepsilon.$$

Proposition 7.2.22. Une *v.a.r.* X admet une densité si et seulement si sa fonction de répartition F_X est localement absolument continue.

Remarque 66. En pratique, le plus simple reste donc de vérifier que le candidat naturel, à savoir F'_X , qui existe presque partout, est effectivement une densité.

Caractérisation fonctionnelle d'une loi

La caractérisation suivante est très utile pour, connaissant la loi d'une *v.a.r.* X , déterminer la loi d'une nouvelle variable aléatoire $\varphi(X)$ où φ est une fonction numérique réelle donnée.

Proposition 7.2.23. Deux *v.a.r.* X et Y suivent la même loi si et seulement si pour toute fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ mesurable bornée

$$\mathbf{E}(g(X)) = \mathbf{E}(g(Y)).$$

Démonstration. C'est un corollaire immédiat de 7.1.3. □

Fonctions génératrices

Lorsque X est une variable aléatoire à valeurs dans \mathbb{N} , on peut définir la notion de fonction génératrice

$$\forall z \in \mathbb{C} : G_X(z) = \mathbf{E}(z^X) = \sum_{k \geq 0} \mathbf{P}(X = k) z^k.$$

Le rayon de convergence de cette série entière est plus grand que 1.

Proposition 7.2.24. Si X et Y sont deux variables aléatoires à valeurs dans \mathbb{N} , alors X et Y ont même loi si et seulement si $G_X(z) = G_Y(z)$ pour tout $z \in \mathbb{C}$ tel que $|z| < 1$.

Démonstration. Il suffit de montrer que, pour tout $k \geq 0$, $\mathbf{P}(X = k) = \mathbf{P}(Y = k)$. Ces quantités sont les coefficients d'une série entière et peuvent s'exprimer à l'aide des dérivées successives □

Exercice 29. Soit X une variable aléatoire à valeurs dans \mathbb{N} admettant un moment d'ordre 2. Exprimer les moments d'ordre 1 et d'ordre 2 en fonction des fonctions génératrices.

7.2.3 Exemples de calcul de lois

L'objectif de ce paragraphe est de donner quelques méthodes de calcul de lois de probabilité : on se donne une *v.a.r.* X de loi connue et une fonction $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ — *a priori* arbitraire même si en général on la suppose assez régulière — et on cherche à calculer la loi de $\varphi(X)$.

La méthode de calcul dépend essentiellement de la caractérisation de la loi que l'on choisit, laquelle dépend du problème que l'on traite. On donne ici deux exemples concrets.

À l'aide de la fonction de répartition

Commençons par un exemple très simple : on se donne une variable aléatoire X de loi uniforme sur $[0, 1]$ et on cherche à calculer la loi de X^2 .

La fonction de répartition de X est donnée par F_X définie pour $t \in \mathbb{R}$

$$F_X(t) = \begin{cases} 0 & t < 0 \\ t & t \in [0, 1) \\ 1 & t \geq 1 \end{cases}$$

On cherche à calculer la fonction de répartition de X^2 :

$$F_{X^2}(t) = \mathbf{P}(X^2 \leq t) = \mathbf{1}_{[0, \infty)}(t) \mathbf{P}(X \in [-\sqrt{t}; \sqrt{t}]) = \mathbf{1}_{[0, \infty)}(t) \mathbf{P}(X \leq \sqrt{t}) = \mathbf{1}_{[0, \infty)}(t) F_X(\sqrt{t}).$$

Il arrive parfois que l'on tombe sur une loi remarquable, ce n'est pas le cas ici. Il n'en reste pas moins que la loi a été caractérisée : la fonction F_X est déterminée.

Notons que la fonction de répartition de X^2 est continue, on peut se poser la question de l'existence d'une densité. En dérivant

$$f(t) = F'_{X^2}(t) = \begin{cases} 0 & t < 0 \\ (2\sqrt{t})^{-1} & t \in [0, 1) \\ 0 & t \geq 1 \end{cases}$$

On vérifie facilement que f est une densité.

Rappelons que la fonction de répartition — on la notera φ dans toute la suite — de la loi normale centrée réduite n'est pas explicite, néanmoins il est possible de calculer la densité de X^2 lorsque X suit une $\mathcal{N}(0, 1)$ à l'aide de la méthode impliquant la fonction de répartition. En effet,

$$F_{X^2}(t) = \mathbf{P}(X^2 \leq t) = \mathbf{1}_{[0, \infty)}(t) \mathbf{P}(X \in [-\sqrt{t}; \sqrt{t}]) = \mathbf{1}_{[0, \infty)}(t) \left[\varphi(\sqrt{t}) - \varphi(-\sqrt{t}) \right].$$

Puis on dérive :

$$f(t) = F'_{X^2}(t) = \frac{\mathbf{1}_{[0, \infty)}(t)}{\sqrt{2\pi t}} e^{-t/2}$$

On vérifie que f est une densité. En fait c'est la densité d'une loi connue appelée loi du $\chi^2(1)$. On généralisera ce résultats dans un prochain chapitre ce qui fera apparaître la loi du chi-deux à d degrés de liberté notée $\chi^2(d)$.

À l'aide de la densité

Cette méthode consiste à utiliser le résultat de la proposition 7.2.23 qui caractérise la loi d'une variable aléatoire à l'aide de fonctions tests. Reprenons l'exemple d'une *v.a.r.* X de loi normale centrée réduite dont on cherche à calculer la loi du carré X^2 . La densité de $\mathcal{N}(0, 1)$ est donnée par $f(x) = (2\pi)^{-1/2} e^{-x^2/2}$, $x \in \mathbb{R}$.

Soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction borélienne bornée et calculons en utilisant un argument de parité

$$\mathbf{E}(g(X^2)) = \int_{\mathbb{R}} g(x^2) f(x) dx = 2 \int_0^{\infty} g(x^2) f(x) dx.$$

Par le théorème de changement de variable en posant $y = x^2$, on obtient

$$2 \int_0^{\infty} g(x^2) f(x) dx = 2 \int_0^{\infty} g(y) f(\sqrt{y}) \frac{dy}{2\sqrt{y}} = \int_{\mathbb{R}} \mathbf{1}_{[0, \infty)}(y) g(y) \frac{e^{-y/2}}{\sqrt{2\pi y}} dy.$$

On identifie alors à l'aide de la proposition 7.2.23 la densité de X^2 , à savoir

$$h(y) = \mathbf{1}_{[0, \infty)}(y) \frac{e^{-y/2}}{\sqrt{2\pi y}},$$

soit le résultat montré dans la section précédente

Remarque 67. À la première ligne de calcul, l'intégrale sur \mathbb{R} a été coupée en deux de sorte que la fonction $x \rightarrow x^2$ est un difféomorphisme de $(0, \infty) \rightarrow (0, \infty)$ et $(-\infty, 0) \rightarrow (0, \infty)$ respectivement. Ainsi, il n'est pas utile que la fonction φ considérée soit un difféomorphisme globale.

Notons également que, quoiqu'il arrive, $x \rightarrow x^2$ n'est pas un difféomorphisme en 0, c'est pourquoi il a été retiré du domaine. Cela ne pose pas de problème car $\mathbf{P}(X = 0) = 0$ donc la valeur de l'intégrale ne change pas.

7.2.4 Classification des lois de probabilités sur \mathbb{R}

La preuve de 7.2.18 a montré que la loi de toute *v.a.r.* est un mélange d'une loi discrète et d'une loi continue. L'exemple 37, quant à lui, illustre qu'une loi continue peut être singulière à la mesure de Lebesgue. Cette section vise à donner la décomposition d'une loi en une partie absolument continue (*i.e.* à densité par rapport à Lebesgue), une partie discrète et une partie singulière.

Théorème 7.2.25. Soient μ et ν deux probabilités sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors $\nu = f \cdot \mu + \mu_0$ où $\mu_0 \perp \mu$. De plus, cette décomposition est unique.

Démonstration. C'est un peu la même idée que pour le théorème 6.3 de Radon-Nikodym : on considère la forme linéaire $f \rightarrow \int f d\nu$ continue sur $\mathbf{L}^2(\nu + \mu)$. Nous obtenons donc l'existence d'une fonction $g \in \mathbf{L}^2(\nu + \mu)$ telle que $\int f d\nu = \int fg d(\nu + \mu)$ si bien que

$$\int f(1-g) d\nu = \int fg d\mu.$$

En considérant des fonctions bien choisies, on montre que $g, 1-g \geq 0$, $(\nu + \mu)$ -p.p. si bien que $g(x) \in [0, 1]$ pour tout x quitte à la modifier sur un ensemble de mesure nulle. En posant $B = \{g = 1\}$ on déduit

$$\nu(B) = \int \mathbf{1}_B d\nu = \int g\mathbf{1}_B d(\nu + \mu) = \int \mathbf{1}_B d(\mu + \mu) = \mu(B) + \nu(B),$$

et $\mu(B) = 0$. On pose $\mu_0 = \nu(\cdot \cap B)$ alors μ_0 et μ sont étrangères. On vérifie alors facilement que $\nu(\cdot \cap B^c)$ est à densité $f = \frac{g}{1-g}\mathbf{1}_{B^c}$. \square

Corollaire 7.2.26. Soit μ une mesure σ -finie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors toute mesure de probabilité ν sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ s'écrit de façon unique comme $f \cdot \mu + \mu_0$ où $f \in \mathbf{L}^1_{\mathbb{R}}(\mu)$ est positive et $\mu_0 \perp \mu$.

Démonstration. Il s'agit de généraliser le théorème précédant à une mesure de référence σ -finie. Cela se fait comme dans le théorème 6.3 de Radon-Nikodym. \square

En posant $\mu = \lambda$ la mesure de Lebesgue, l'application de ce corollaire à la loi $\nu = \mathbf{P}_X$ d'une *v.a.r.* implique l'existence d'une fonction f positive λ -intégrable et d'une mesure positive μ_0 étrangère à λ telle que $\mathbf{P}_X = f \cdot \lambda + \mu_0$. En appliquant les arguments de 7.2.18 à μ_0 on montre facilement que toute mesure de probabilité ν sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ se décompose de façon unique en une combinaison convexe de trois lois étrangères : une loi absolument continue par rapport à la mesure de Lebesgue, une loi discrète et une loi singulière.

7.2.5 Simulation de lois

Commençons par un cas simple : soit F une fonction de répartition supposée continue strictement croissante. On considère U une variable aléatoire uniforme sur $[0, 1]$ et on pose $X = F^{-1}(U)$. Alors, pour $t \in \mathbb{R}$

$$\mathbf{P}(X \leq t) = \mathbf{P}(F^{-1}(U) \leq t) = \mathbf{P}(U \leq F(t)) = F(t),$$

car F est croissante. Ainsi, pour générer des nombres aléatoires suivant une loi dont la fonction de répartition est continue strictement croissante, il suffit de savoir générer des nombres suivant une loi uniforme dans $[0, 1]$.

Exemple 38. Soit $X \sim \mathcal{E}(\lambda)$ alors sa fonction de répartition est donnée, pour tout $t \in \mathbb{R}$, par $F(t) = (1 - e^{-\lambda t})\mathbf{1}_{[0, \infty)}(t)$. Pour $p \in (0, 1)$, on obtient $F^{-1}(p) = -\frac{1-p}{\lambda}$ et $Y = F^{-1}(U)$, où U est une loi uniforme sur $[0, 1]$, suit une loi exponentielle de paramètre $\lambda > 0$.

Remarque 68. Si F est une fonction croissante de \mathbb{R} dans \mathbb{R} alors

$$\lim_{x \rightarrow \infty} F(x) = \sup_{x \in \mathbb{R}} F(x) \in \mathbb{R} \cup \{\pm\infty\} \quad \text{et} \quad \lim_{x \rightarrow -\infty} F(x) = \inf_{x \in \mathbb{R}} F(x) \in \mathbb{R} \cup \{\pm\infty\}.$$

On notera ces limites $F(\infty)$ et $F(-\infty)$ respectivement.

Afin de généraliser à une fonction de répartition arbitraire F , il est nécessaire d'introduire la fonction quantile notée H et définie pour tout $p \in [0, 1]$ par

$$H(p) = \inf\{x \in \mathbb{R} : F(x) \geq p\}, \quad \inf \emptyset = \infty \quad \text{et} \quad \inf \mathbb{R} = -\infty.$$

En particulier, $H(0) = -\infty$, $H(1) \in \mathbb{R} \cup \{\infty\}$ et pour tout $p \in (0, 1)$, $H(p) \neq \pm\infty$.

Proposition 7.2.27. *Soit F une fonction de répartition et H la fonction quantile associée. Alors,*

1. *La fonction H est croissante et continue à gauche sur $(0, 1]$. De plus, pour tout $x \in \mathbb{R}$ et $p \in [0, 1]$, $F(x) \geq p$ si et seulement si $x \geq H(p)$.*
2. *Pour tout $p \in [0, 1]$, $F \circ H(p) \geq p$ avec égalité si $H(p) > -\infty$ et F continue en $H(p)$.*
3. *Soit U une variable aléatoire de loi uniforme sur $[0, 1]$. Alors la fonction de répartition de $H(U)$ est égale à F .*
4. *Soit X une variable aléatoire à valeurs dans \mathbb{R} de fonction de répartition F . Si F est continue alors $F(X)$ suit une loi uniforme sur $[0, 1]$.*

Démonstration. Pour tout $p \in [0, 1]$, on note $A_p = \{x \in \mathbb{R} : F(x) \geq p\}$.

1. On commence par démontrer l'équivalence du point 1 : soit $x \in \mathbb{R}$ et $p \in (0, 1]$. Par définition de $H(p)$, si $F(x) \geq p$ alors $x \geq H(p)$. Réciproquement, soient $x \geq H(p)$ et $(x_n)_{n \geq 0}$ une suite de points de A_p qui converge vers $H(p)$. Alors, ou bien $x > H(p)$ et il existe $N \geq 0$ tel que, pour tout $n \geq N$, $H(p) \leq x_n \leq x$ si bien que $p \leq F(x_n) \leq F(x)$ par croissance de F ; ou bien $x = H(p)$ et par continuité à droite de F , $p \leq F(x_n) \rightarrow F(x)$.

Si $p, q \in [0, 1]$ sont tels que $p \leq q$, alors par croissance de F , $A_q \subset A_p$ et $H(p) \leq H(q)$.

Soit $(p_n)_{n \geq 0}$ une suite croissante de points de $[0, 1]$ convergeant vers $p \in (0, 1]$. Par croissance de H , la suite $(H(p_n))_{n \geq 0}$ est croissante et admet une limite $\ell \leq H(p)$. Il s'agit de montrer que $\ell = H(p)$. Supposons au contraire que $\ell < H(p)$. Alors, d'une part $F(\ell) < p$ et d'autre part par l'équivalence ci-dessus $F(\ell) \geq p_n$ pour tout $n \geq 0$. En passant à la limite, on obtient la contradiction voulue.

2. Soit $p \in [0, 1]$ tel que $H(p) \in \mathbb{R}$ alors partant de $H(p) \geq H(p)$, on obtient de l'inégalité précédente $F \circ H(p) \geq p$. Nous avons par ailleurs $H(p) = \infty$ lorsque $p = 1$ et l'inégalité est trivialement satisfaite. De même, si $H(p) = -\infty$ alors $p = 0$ et encore une fois l'inégalité est trivialement satisfaite.

Supposons que $H(p) > -\infty$. Soit $\varepsilon > 0$, puisque $H(p)$ minore A_p , $H(p) - \varepsilon \notin A_p$. Par conséquent, $F(H(p) - \varepsilon) \leq p$. Puisque F est supposée continue en $H(p)$, on obtient $F(H(p)) = \lim_{\varepsilon \rightarrow 0} F(H(p) - \varepsilon) \leq p$.

3. Soit U une variable aléatoire de loi uniforme sur $[0, 1]$. D'après l'équivalence du point 1, pour tout $x \in \mathbb{R}$:

$$\mathbf{P}(H(U) \leq x) = \mathbf{P}(U \leq F(x)) = F(x).$$

4. Les variables aléatoires X et $H(U)$ sont identiquement distribuées. Ainsi, $F(X)$ a même loi que $F(H(U))$, or $F(H(U)) = U$ car F est continue et X est finie presque-sûrement. □

Exemple 39. Considérons le cas le plus simple d'une variable aléatoire X suivant une loi de Bernoulli de paramètre $q \in (0, 1)$. Alors sa fonction de répartition est donnée par

$$F(t) = (1 - q)\mathbf{1}_{[0,1)}(t) + \mathbf{1}_{[1,\infty)}(t) \quad \text{et} \quad H(p) = (-\infty)\mathbf{1}_{p=0} + \mathbf{1}_{(1-q,1]}(p).$$

Si $U \sim \mathcal{U}[0, 1]$, $\mathbf{P}(U = 0) = 0$ si bien que l'on peut considérer $\tilde{H}(U)$ où $\tilde{H}(p) = \mathbf{1}_{(1-q,1]}(p)$. De même on pourra fermer l'intervalle dans l'indicatrice sans changer la loi.

Dans cet exemple très simple, on constate que $H(U)$ est à valeurs dans $\{0, 1\}$ et $\mathbf{P}(H(U) = 1) = \mathbf{P}(U \in (1 - q, 1]) = q$.

7.3 Vecteurs aléatoires

Très souvent il est utile de considérer non pas des variables aléatoires réelles unidimensionnelles mais des vecteurs aléatoires de \mathbb{R}^d . Soit parce que le problème considéré fait naturellement intervenir un vecteur (une position dans l'espace), soit parce que l'on répète d fois une expérience aléatoire. La suite de ce chapitre consiste simplement à adapter les notions au cadre multidimensionnel.

7.3.1 Généralités

Définition 7.3.1. On appelle vecteur aléatoire ou variable aléatoire multivariée toute application mesurable $X : (\Omega, \mathcal{F}, \mathbf{P}) \longrightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Sauf mention contraire, la base de \mathbb{R}^d choisie sera la base canonique. Comme dans le cas déterministe, on peut écrire, dans la base canonique $\{e_i, i = 1, \dots, d\}$,

$$X = \sum_{i=1}^d X_i e_i$$

où la i -ième coordonnée X_i est une *v.a.r.*.

Dans toute la suite, en terme de notation, on choisit la convention vecteur colonne comme c'est l'usage en algèbre linéaire. Cependant, pour des raisons typographiques on écrira souvent $X = (X_1, \dots, X_d)$. Si A est une matrice représentant un morphisme linéaire de \mathbb{R}^d dans \mathbb{R}^q , alors AX est un vecteur de \mathbb{R}^q . L'adjoint d'une matrice (ou d'un vecteur vu comme une matrice) sera noté A^* .

Rappelons que la tribu borélienne de \mathbb{R}^d est engendrée par les pavés de la forme $(a_1, b_1) \times \dots \times (a_d, b_d)$, $a_i, b_i \in \mathbb{R}$, $i = 1, \dots, d$. Autrement dit, $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R})^{\otimes d}$.

7.3.2 Loi d'un vecteur aléatoire, lois marginales

Définition 7.3.2. La loi d'un vecteur aléatoire X est la probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R})^{\otimes d})$, notée \mathbf{P}_X , définie par

$$\mathbf{P}_X(A) = \mathbf{P}(X \in A), \quad A \in \mathcal{B}(\mathbb{R})^{\otimes n}.$$

Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire dans \mathbb{R}^d de loi \mathbf{P}_X . La i ème loi marginale, notée \mathbf{P}_{X_i} est la probabilité image réciproque de \mathbf{P}_X par la projection sur la i ème coordonnée. Plus concrètement, si $A \in \mathcal{B}(\mathbb{R})$,

$$X_i^{-1}(A) = \underbrace{\mathbb{R} \times \dots \times \mathbb{R}}_{i-1} \times A \times \underbrace{\mathbb{R} \times \dots \times \mathbb{R}}_{d-i}$$

et $\mathbf{P}_{X_i}(A) = \mathbf{P}_X(X_i^{-1}(A))$. Ainsi, connaissant la loi d'un vecteur aléatoire X , on peut déterminer la loi de chaque marginale X_i .

Par contre, on ne peut pas, connaissant chaque loi marginale, déterminer la loi du vecteur aléatoire X . La loi \mathbf{P}_X possède intrinsèquement plus d'information que les \mathbf{P}_{X_i} prises toutes ensembles. En fait, il manque de l'information sur la façon dont les marginales dépendent les unes des autres.

7.3.3 Moments

On munit \mathbb{R}^d d'une norme notée $|\cdot|$ ou $|\cdot|_p$, $p \in [1, \infty]$ si l'on veut préciser.

Définition 7.3.3. Un vecteur aléatoire $X \in \mathbb{R}^d$ admet un moment d'ordre $q \geq 1$ si

$$\mathbf{E}(|X|^q) = \int_{\Omega} |X|^q d\mathbf{P} < \infty$$

Définition 7.3.4 (Moyenne, Variance-Covariance). Si un vecteur aléatoire $X \in \mathbb{R}^d$ admet un moment d'ordre 1, l'espérance de X , notée $\mathbf{E}(X)$ est définie par

$$\mathbf{E}(X) = \int_{\Omega} X d\mathbf{P} \in \mathbb{R}^d.$$

Si X admet un moment d'ordre 2, la matrice de Variance-Covariance, ou plus simplement matrice de covariance, est définie par

$$\Sigma(X) = \mathbf{E}[(X - \mathbf{E}(X))(X - \mathbf{E}(X))^*] \in \mathcal{M}_d(\mathbb{R}).$$

Remarque 69. Rappelons que nous utilisons la notation des vecteurs en colonne si bien que $\Sigma(X)$ est bien une matrice de taille $d \times d$.

Remarque 70. Les moments d'ordre supérieurs, en général moins utilisés, ne peuvent s'écrire aussi synthétiquement.

7.3.4 Lois à densité

L'espace mesurable $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est naturellement muni de la mesure de Lebesgue d -dimensionnelle notée λ_d . La définition suivante découle directement de ce fait.

Définition 7.3.5. Un vecteur aléatoire $X = (X_1, \dots, X_d)$ est dit à densité si il existe une fonction $f \in \mathcal{L}^1(\lambda_d)$ positive vérifiant

$$\int_{\mathbb{R}^d} f \, d\lambda_d = 1$$

telle que pour tout borélien $A \in \mathcal{B}(\mathbb{R}^d)$

$$\mathbf{P}_X(A) = \mathbf{P}(X \in A) = \int_{\mathbb{R}^d} \mathbf{1}_A f \, d\lambda_d.$$

Proposition 7.3.6. La loi d'un vecteur aléatoire à densité est caractérisé par sa densité de probabilité : si X et Y sont deux vecteurs aléatoires de \mathbb{R}^d de densités respectives f_X et f_Y alors X et Y ont même loi si et seulement si $f_X = f_Y$ λ_d -p.p..

Démonstration. C'est un corollaire de la proposition 7.1.4. □

Proposition 7.3.7. Soit $X = (X_1, \dots, X_d) \in \mathbb{R}^d$ un vecteur aléatoire à densité f_X . Alors pour tout $i = 1, \dots, d$, la marginale X_i est une v.a.r. à densité. De plus la densité f_{X_i} de X_i est donnée par

$$f_{X_i}(x) = \int_{\mathbb{R}^{d-1}} f_X(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_d) \lambda_{d-1}(dx_1, \dots, dx_{i-1}, dx_{i+1}, \dots, dx_d).$$

Démonstration. Il suffit évidemment de considérer le cas $d = 2$. Soit $A \in \mathcal{B}(\mathbb{R})$, alors le théorème de Fubini implique

$$\mathbf{P}_{X_1}(A) = \mathbf{P}_X(A \times \mathbb{R}) = \int_{\mathbb{R}^2} \mathbf{1}_{A \times \mathbb{R}}(x, y) f_X(x, y) \lambda_2(dx, dy) = \int_{\mathbb{R}} \mathbf{1}_A \underbrace{\int_{\mathbb{R}} f(x, y) \lambda(dy)}_{=f_{X_1}(x)} \lambda(dx).$$

La densité f_{X_2} se calcule de la même façon. □

7.3.5 Fonction de répartition

On peut en dimension supérieure définir une notion de fonction de répartition même si celle-ci n'est que peu utile bien souvent car \mathbb{R}^d , pour $d \geq 2$, n'admet plus d'ordre total naturel.

Définition 7.3.8. La fonction de répartition d'un vecteur $X \in \mathbb{R}^d$ est la fonction $F_X : \mathbb{R}^d \rightarrow [0, 1]$ définie pour $t = (t_1, \dots, t_d) \in \mathbb{R}^d$ par

$$F_X(t_1, \dots, t_d) = \mathbf{P}_X(X_1 \leq t_1, \dots, X_d \leq t_d).$$

Remarque 71. Dans ce cas encore, il y a une distinction entre variable à densité et variable continue. Comme dans le cas unidimensionnel, une variable à densité est continue mais la réciproque est fausse.

Proposition 7.3.9. Deux vecteurs aléatoires $X, Y \in \mathbb{R}^d$ ont même loi si et seulement si $F_X = F_Y$.

Démonstration. C'est une conséquence directe du théorème 2.2.20 car

$$\mathcal{S} = \{(-\infty, a_1] \times \dots \times (-\infty, a_d], \quad (a_1, \dots, a_d) \in \mathbb{R}^d\}$$

est un π -système (non vide et stable par intersection finie). □

7.3.6 Transformation des vecteurs aléatoires à densité

Comme dans le cas des variables aléatoires réelles, se donnant un vecteur aléatoire $X = (X_1, \dots, X_d)$ de loi connue et une fonction $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^p$, on cherche à déterminer la loi de $\varphi(X)$. Pour ce faire, on peut faire usage du théorème de changement de variable comme l'illustre l'exemple suivant.

Exemple 40. Soit (X, Y) un vecteur aléatoire dont la loi est donnée par sa densité

$$h(x, y) = \mathbf{1}_{\mathbb{R}_+ \times \mathbb{R}_+}(x, y) \lambda \mu e^{-\lambda x} e^{-\mu y}.$$

On verra par la suite qu'en fait c'est un couple de variables aléatoires indépendantes de lois exponentielles d'intensités λ et μ respectivement. On cherche à calculer la loi de $(X + Y, X - Y)$. Autrement dit, si $Z = (X, Y)$ et $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ est définie par $\varphi(x, y) = (x + y, x - y)$, on cherche à déterminer la loi de $\varphi(Z)$.

On utilise la caractérisation fonctionnelle de la loi. Pour cela, notons $(u, v) = \varphi(x, y) = (x + y, x - y)$ et donnons nous une fonction $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ mesurable bornée. On calcule,

$$\mathbf{E}[g(\varphi(Z))] = \mathbf{E}(g(X + Y, X - Y)) = \int_{\mathbb{R}^2} g(x + y, x - y) h(x, y) dx dy.$$

On fait le changement de variables $(u, v) = (x + y, x - y)$ qui s'inverse par $(x, y) = ((u + v)/2, (u - v)/2)$. Il est clair que φ est un C^1 -difféomorphisme de \mathbb{R}^2 , c'est en fait un automorphisme linéaire. Le jacobien de φ^{-1} est donné par

$$\det \text{Jac}_{\varphi^{-1}}(u, v) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}$$

Le théorème de changement de variable donne

$$\mathbf{E}[g(\varphi(Z))] = \int_{\mathbb{R}^2} g(u, v) h\left(\frac{u+v}{2}, \frac{u-v}{2}\right) \frac{du dv}{2}.$$

Par la caractérisation avec des fonctions tests, on identifie la densité de $(U, V) = (X + Y, X - Y)$: elle est donnée par

$$h_{(X+Y, X-Y)}(u, v) = h\left(\frac{u+v}{2}, \frac{u-v}{2}\right) = \mathbf{1}_{u+v \geq 0}(u) \mathbf{1}_{u-v \geq 0} \frac{\lambda \mu}{8} e^{-(\lambda+\mu)u/2} e^{-(\lambda-\mu)v/2}$$

que l'on peut simplifier en

$$h_{(X+Y, X-Y)}(u, v) = \mathbf{1}_{u \geq 0} \mathbf{1}_{-u \leq v \leq u} \frac{\lambda \mu}{2} e^{-(\lambda+\mu)u/2} e^{-(\lambda-\mu)v/2}.$$

En exercice complémentaire, on peut donner la densité de $X + Y$, pour ce faire il suffit d'intégrer par rapport à v . Lorsque $\lambda \neq \mu$,

$$\begin{aligned} h_{X+Y}(u) &= \int_{\mathbb{R}} \mathbf{1}_{u \geq 0} \mathbf{1}_{-u \leq v \leq u} \frac{\lambda \mu}{2} e^{-(\lambda+\mu)u/2} e^{-(\lambda-\mu)v/2} dv \\ &= \mathbf{1}_{u \geq 0} \frac{\lambda \mu}{2} e^{-(\lambda+\mu)u/2} \int_{-u}^u e^{-(\lambda-\mu)v/2} dv \\ &= \mathbf{1}_{u \geq 0} \frac{\lambda \mu}{(\lambda - \mu)} e^{-(\lambda+\mu)u/2} \left[e^{(\lambda-\mu)u/2} - e^{-(\lambda-\mu)u/2} \right] = \mathbf{1}_{u \geq 0} \frac{\lambda \mu}{(\lambda - \mu)} \left[e^{-\mu u} - e^{-\lambda u} \right]. \end{aligned}$$

Lorsque $\lambda = \mu$, on obtient :

$$h_{X+Y}(u) = \int_{\mathbb{R}} \mathbf{1}_{u \geq 0} \mathbf{1}_{-u \leq v \leq u} \frac{\lambda \mu}{2} e^{-(\lambda+\mu)u/2} dv = \mathbf{1}_{u \geq 0} \frac{\mu^2}{2} e^{-\mu u} \int_{-u}^u dv = \mathbf{1}_{u \geq 0} \mu^2 u e^{-\mu u}.$$

On finit cette partie sur une proposition dans laquelle est insérée une remarque importante. Si au lieu de déterminer la loi $(X + Y, X - Y)$, on se pose la question de la loi de $X + Y$ seulement, le théorème de changement variables n'est pas applicable directement du fait d'un problème de dimension : $(x, y) \rightarrow x + y$ n'est pas injective. La solution consiste à ajouter une dimension en étudiant, par exemple, la loi de $(X + Y, X)$ et à intégrer par rapport à la seconde variable pour obtenir la loi de $X + Y$. Le choix de l'ajout de variable n'est pas unique mais il faut bien entendu rester dans la simplicité tout en conservant l'injectivité du changement de variable.

Chapitre 8

Indépendance

8.1 Tribus indépendantes

Définition 8.1.1. Une famille $(\mathcal{F}_i)_{i \in I}$ de sous-tribus de \mathcal{F} est dite indépendante si pour tout $J \subset I$ fini et

$$\forall (A_j)_{j \in J} \in \prod_{j \in J} \mathcal{F}_j \implies \mathbf{P} \left(\bigcap_{j \in J} A_j \right) = \prod_{j \in J} \mathbf{P}(A_j).$$

Une famille $(A_i)_{i \in I}$ d'événements est indépendante si la famille des tribus correspondantes $(\sigma(A_i))_{i \in I}$ est indépendante.

À toute fin utile, on rappelle que pour $A \in \mathcal{F}$ on a $\sigma(A) = \{\emptyset, \Omega, A, A^c\}$.

Remarque 72. Il existe une notion plus faible appelée indépendance deux à deux. Des événements $(A_i)_{i \in I}$ sont indépendants deux à deux si pour tout $i \neq j \in I$

$$\mathbf{P}(A_i \cap A_j) = \mathbf{P}(A_i)\mathbf{P}(A_j).$$

La notion d'indépendance de la définition 8.1.1 est parfois appelée indépendance mutuelle. Sauf mention contraire, lorsque nous parlerons d'indépendance sans autre précision, il s'agira toujours de la notion définie en 8.1.1

Exemple 41. On lance deux fois une pièce de monnaie. On considère les événements

$$A = \{\text{“pile au 1er lancé”}\}, \quad B = \{\text{“face au 2e lancé”}\}, \quad C = \{\text{“même tirage au deux lancés”}\}.$$

On vérifie facilement que

- $\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(C) = 1/2$,
 - $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$, $\mathbf{P}(A \cap C) = \mathbf{P}(A)\mathbf{P}(C)$ et $\mathbf{P}(B \cap C) = \mathbf{P}(B)\mathbf{P}(C)$,
- mais que $\mathbf{P}(A \cap B \cap C) = 0$.

La proposition suivante, très utile en pratique, introduit une notion d'indépendance par paquet.

Proposition 8.1.2. Soit $(\mathcal{F}_i)_{i \in I}$ une famille de tribus indépendantes. Soit $(I_k)_{k \in K}$ une partition de I . On note \mathcal{U}_k la tribu engendrée par la famille $(\mathcal{F}_i)_{i \in I_k}$, $k \in K$, autrement dit $\mathcal{U}_k = \sigma(\mathcal{F}_i, i \in I_k)$. Alors, la famille $(\mathcal{U}_k)_{k \in K}$ est indépendante.

Démonstration. On utilise les deux lemmes ci-dessous

Lemme 8.1.3. Soit $(\mathcal{F}_i)_{i \in I}$ une famille de sous-tribus. On suppose que pour tout $i \in I$, la tribu \mathcal{F}_i est engendré par un π -système \mathcal{C}_i contenant Ω . La famille $(\mathcal{F}_i)_{i \in I}$ est indépendante si et seulement si pour tout $J \subset I$ fini

$$\forall (A_j)_{j \in J} \in \prod_{j \in J} \mathcal{C}_j : \quad \mathbf{P} \left(\bigcap_{j \in J} A_j \right) = \prod_{j \in J} \mathbf{P}(A_j). \quad (8.1)$$

Démonstration. Soit $J \subset I$ un sous-ensemble fini d'indices. Si $J = \emptyset$ alors l'égalité (8.1) est trivialement satisfaite. Supposons donc $J \neq \emptyset$. On se donne une énumération de J , c'est à dire $J = \{j_1, \dots, j_k\}$ où $k = \text{card } J$. Pour $r = 0, \dots, k$, on définit la propriété (\mathcal{P}_r) suivante : l'égalité (8.1) est satisfaite pour tout $A_i \in \mathcal{F}_i$, $1 \leq i \leq r$, et tout $A_i \in \mathcal{C}_i$, $r < i \leq k$. On va montrer que la propriété (\mathcal{P}_r) est vraie pour tout $r \in \{0, \dots, k\}$.

La propriété (\mathcal{P}_0) est vraie par hypothèse. Supposons que (\mathcal{P}_{r-1}) est vraie et montrons que (\mathcal{P}_r) est vraie. Pour cela, considérons

$$\mathcal{D} = \left\{ B \in \mathcal{F}_r : \forall (A_i)_{1 \leq i \leq r-1} \in \prod_{i=1}^{r-1} \mathcal{F}_i, (A_i)_{r+1 \leq i \leq k} \in \prod_{i=r+1}^k \mathcal{C}_i, \right. \\ \left. \mathbf{P} \left(\prod_{i=1}^{r-1} A_i \times B \times \prod_{i=r+1}^k A_i \right) = \left(\prod_{i=1}^{r-1} \mathbf{P}(A_i) \right) \mathbf{P}(B) \left(\prod_{i=r+1}^k \mathbf{P}(A_i) \right) \right\}.$$

Montrons que \mathcal{D} est un λ -système. Par hypothèse de récurrence (\mathcal{P}_{r-1}) et puisque $\Omega \in \mathcal{C}_r$, on déduit que $\Omega \in \mathcal{D}$. Soient $B, C \in \mathcal{D}$ tels que $B \subset C$. On note

$$A^- = \bigcap_{i=1}^{r-1} A_i \quad \text{et} \quad A^+ = \bigcap_{i=r+1}^k A_i.$$

Alors,

$$\mathbf{P}(A^- \cap (C \setminus B) \cap A^+) = \mathbf{P}(A^- \cap C \cap A^+) - \mathbf{P}(A^- \cap B \cap A^+).$$

Puis, comme $B, C \in \mathcal{D}$, on obtient

$$\mathbf{P}(A^- \cap (C \setminus B) \cap A^+) = [\mathbf{P}(C) - \mathbf{P}(B)] \prod_{i \neq r} \mathbf{P}(A_i) = \mathbf{P}(C \setminus B) \prod_{i \neq r} \mathbf{P}(A_i),$$

d'où $C \setminus B \in \mathcal{D}$. Soit maintenant $(B_n)_{n \geq 0} \in \mathcal{D}^{\mathbb{N}}$ une suite croissante, en notant $B = \bigcup_{n \geq 0} B_n$, on a

$$\mathbf{P}(A^- \cap B \cap A^+) = \lim_{n \rightarrow \infty} \mathbf{P}(A^- \cap B_n \cap A^+) = \lim_{n \rightarrow \infty} \mathbf{P}(B_n) \prod_{i \neq r} \mathbf{P}(A_i) = \mathbf{P}(B) \prod_{i \neq r} \mathbf{P}(A_i).$$

Ainsi, $B \in \mathcal{D}$. On conclut que \mathcal{D} est un λ -système qui contient le π -système \mathcal{C}_r , donc contient $\sigma(\mathcal{C}_r) = \mathcal{F}_r$. La propriété (\mathcal{P}_r) est donc vraie. Par récurrence, (\mathcal{P}_k) est vraie et ce indépendamment de l'énumération de J choisie. Ceci finit la preuve du lemme. □

et

Lemme 8.1.4. Soit $(\mathcal{F}_i)_{i \in I}$ une famille de sous-tribus. Alors,

$$\mathcal{C} = \left\{ \bigcap_{j \in J} A_j : A_j \in \mathcal{F}_j, J \subset I \text{ fini} \right\} \quad (8.2)$$

est un π -système, contenant Ω , qui engendre la tribu $\sigma(\mathcal{F}_i : i \in I)$.

Démonstration. Puisque $\Omega \in \mathcal{F}_i$ pour tout $i \in I$, on déduit que $\Omega \in \mathcal{C}$ si bien que $\mathcal{C} \neq \emptyset$. Soit $A, B \in \mathcal{C}$, alors, il existe $J_A, J_B \subset I$ des sous-ensembles finis d'indices et des ensembles $C_j \in \mathcal{F}_j$, $j \in J_A$, et des ensembles $D_\ell \in \mathcal{F}_\ell$, $\ell \in J_B$ tels que

$$A = \bigcap_{j \in J_A} C_j \quad \text{et} \quad B = \bigcap_{\ell \in J_B} D_\ell.$$

Par conséquent,

$$A \cap B = \bigcap_{j \in J_A} C_j \cap \bigcap_{\ell \in J_B} D_\ell \in \mathcal{C},$$

quitte à rassembler les ensembles qui sont dans la même sous-tribus. Donc \mathcal{C} est un π -système qui contient Ω . Clairement, $\mathcal{F}_i \subset \mathcal{C}$ pour tout $i \in I$ si bien que $\sigma(\mathcal{F}_i, i \in I) \subset \sigma(\mathcal{C})$. □

Plus précisément, pour tout $k \in K$, \mathcal{U}_k est engendré par le π -système \mathcal{C}_k défini par (8.2) où l'on a remplacé I par I_k . Ces π -systèmes contiennent Ω . Soit $J \subset K$ un sous-ensemble fini, par indépendance de la famille $(\mathcal{F}_i)_{i \in I}$, l'égalité (8.1) est satisfaite. Ceci conclut la preuve de la proposition. \square

8.2 Lemme de Borel-Cantelli

On rappelle les définitions pour $(B_n)_{n \geq 0}$ une suite d'événements mesurables

$$\limsup B_n = \bigcap_{n \geq 0} \bigcup_{k \geq n} B_k \quad \text{et} \quad \liminf B_n = \bigcup_{n \geq 0} \bigcap_{k \geq n} B_n,$$

ainsi que le premier lemme de Borel-Cantelli — *c.f.* la proposition 2.2.7.

Proposition 8.2.1 (Premier lemme de Borel-Cantelli). *Soit $(B_n)_{n \geq 0} \in \mathcal{F}^{\mathbb{N}}$ une suite d'événements. Alors*

$$\sum_{n \geq 0} \mathbf{P}(B_n) < \infty \quad \implies \quad \mathbf{P}(\limsup B_n) = 0.$$

Remarque 73. On remarque si $\sum_{n \geq 0} \mathbf{P}(B_n^c) < \infty$ alors

$$\mathbf{P}(\liminf B_n) = 1, \quad \text{car} \quad (\liminf B_n)^c = \limsup B_n^c.$$

Proposition 8.2.2 (Deuxième lemme de Borel-Cantelli). *Soit $(B_n)_{n \geq 0} \in \mathcal{F}^{\mathbb{N}}$ une suite d'événements indépendants. Alors*

$$\sum_{n \geq 0} \mathbf{P}(B_n) = \infty \quad \implies \quad \mathbf{P}(\limsup B_n) = 1.$$

Remarque 74. Sous les mêmes hypothèses

$$\sum_{n \geq 0} \mathbf{P}(B_n^c) = \infty \quad \implies \quad \mathbf{P}(\liminf B_n) = 0.$$

Ainsi dans le cas d'événements $(B_n)_{n \geq 0}$ indépendants, l'événement $\limsup B_n$ est de probabilité 0 ou 1 et on a le critère suivant

$$\mathbf{P}(\limsup B_n) = 0 \quad \text{si et seulement si} \quad \sum_{n \geq 0} \mathbf{P}(B_n) < \infty.$$

Démonstration. On remarque que $\mathbf{P}(\limsup B_n) = 1 - \mathbf{P}(\liminf B_n^c)$ et on va montrer que $\mathbf{P}(\liminf B_n^c)$ est nulle. Puisque $\liminf B_n^c = \bigcup_{n \geq 0} \bigcap_{k \geq n} B_k^c$, il suffit donc de montrer que, pour tout $n \geq 0$, $\mathbf{P}(\bigcap_{k \geq n} B_k^c)$ est nulle. Fixons $n \geq 0$. Par définition les tribus $\sigma(B_n)$ sont indépendantes et donc les événements B_n^c sont aussi indépendants. Par conséquent

$$\mathbf{P}\left(\bigcap_{k \geq n} B_k^c\right) = \lim_{p \rightarrow \infty} \mathbf{P}\left(\bigcap_{n \leq k \leq p} B_k^c\right) = \lim_{p \rightarrow \infty} \prod_{n \leq k \leq p} \mathbf{P}(B_k^c).$$

D'autre part, notant que $\mathbf{P}(B_n^c) = 1 - \mathbf{P}(B_n)$ et que $1 - x \leq e^{-x}$ pour tout $x \geq 0$, on a

$$\prod_{n \leq k \leq p} \mathbf{P}(B_k^c) = \prod_{n \leq k \leq p} (1 - \mathbf{P}(B_k)) \leq \prod_{n \leq k \leq p} e^{-\mathbf{P}(B_k)} = \exp\left\{-\sum_{n \leq k \leq p} \mathbf{P}(B_k)\right\}.$$

Pour finir, n étant fixé, $\sum_{n \leq k \leq p} \mathbf{P}(B_k)$ tend vers ∞ lorsque $p \rightarrow \infty$. \square

8.3 Variables aléatoires indépendantes

8.3.1 Définition et caractérisation élémentaire

Si $X : (\Omega, \mathcal{F}, \mathbf{P}) \rightarrow (E, \mathcal{E})$ est une variable aléatoire, on note $\sigma(X) \subset \mathcal{F}$ la sous-tribu engendrée par X , c'est à dire la plus petite tribu — au sens de l'inclusion — rendant mesurable l'application X .

Définition 8.3.1. Une famille de variables aléatoires $(X_i)_{i \in I}$, X_i à valeurs dans (E_i, \mathcal{E}_i) , $i \in I$, est indépendante si et seulement si la famille de tribus $(\sigma(X_i))_{i \in I}$ est indépendante.

Proposition 8.3.2. Une famille de variables aléatoires $(X_i)_{i \in I}$ est indépendante si et seulement pour tout sous-ensemble $J \subset I$ fini l'assertion suivante est satisfaite

$$\forall (A_j)_{j \in J} \in \prod_{j \in J} \mathcal{E}_j \implies \mathbf{P} \left(\bigcap_{j \in J} \{X_j \in A_j\} \right) = \prod_{j \in J} \mathbf{P}(X_j \in A_j).$$

Démonstration. Immédiat. □

8.3.2 Constructions de variables aléatoires indépendantes

On peut se poser la question de la construction de variables aléatoires indépendantes (ou de tribus) : étant données deux espaces probabilisés $(\Omega_1, \mathcal{F}_1, \mu_1)$ et $(\Omega_2, \mathcal{F}_2, \mu_2)$, peut-on construire deux variables indépendantes X_1 et X_2 de loi respective μ_1 et μ_2 ?

Pour ce faire, on considère l'espace probabilisé

$$(\Omega, \mathcal{F}, \mathbf{P}) = (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \otimes \mu_2)$$

et on pose X_1 la projection sur Ω_1 et X_2 la projection sur Ω_2 . Concrètement,

$$\forall \omega = (x_1, x_2) \in \Omega_1 \times \Omega_2, \quad X_1(\omega) = x_1, \quad X_2(\omega) = x_2.$$

Les variables aléatoires X_1 et X_2 sont alors des variables aléatoires indépendantes à valeurs dans Ω_1 et Ω_2 respectivement ; X_1 a pour loi μ_1 et X_2 a pour loi μ_2 . En effet, pour tout $B_1 \in \mathcal{F}_1$ et $B_2 \in \mathcal{F}_2$,

$$\mathbf{P}(X_1 \in B_1, X_2 \in B_2) = \mu_1 \otimes \mu_2(B_1 \times B_2) = \mu_1(B_1)\mu_2(B_2).$$

Prenant successivement $B_1 = \Omega_1$ et $B_2 = \Omega_2$, on s'aperçoit que X_1 et X_2 suivent respectivement la loi μ_1 et μ_2 .

Jusqu'ici, on a donc utilisé essentiellement la structure d'espace probabilisé produit. Le passage à une famille quelconque se fait sans trop de problèmes mais repose tout de même sur le théorème 2.2.24 de Carathéodory. Ce dernier théorème donne alors un sens à l'assertion "soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes". Pour cela, nous devons introduire la notion de classe compacte.

Définition 8.3.3 (Classe compacte d'ensembles). Une classe \mathcal{K} de parties d'un ensemble E est dite compacte si, pour toute famille $(K_n)_{n \geq 0} \in \mathcal{K}^{\mathbb{N}}$ telle que $\bigcap_{n \geq 0} K_n = \emptyset$ il existe $N \geq 0$ tel que $\bigcap_{n=0}^N K_n = \emptyset$.

Théorème 8.3.4. Soit, pour tout $i \in I$, $(E_i, \mathcal{F}_i, \mu_i)$ un espace probabilisé. On suppose que, pour tout $i \in I$, il existe une classe compacte $\mathcal{K}_i \subset \mathcal{F}_i$ telle que

$$\forall A \in \mathcal{F}_i, \quad \mu_i(A) = \sup \{ \mu_i(C), C \in \mathcal{K}_i \}.$$

Alors, il existe une unique probabilité μ sur $(\prod_{i \in I} E_i, \otimes_{i \in I} \mathcal{F}_i)$, notée $\otimes_{i \in I} \mu_i$, telle que pour tout $(B_i)_{i \in I} \in \prod_{i \in I} \mathcal{F}_i$ avec $B_i = E_i$ sauf pour un nombre fini de $i \in I$,

$$\left(\otimes_{i \in I} \mu_i \right) \left(\prod_{i \in I} B_i \right) = \prod_{i \in I} \mu_i(B_i).$$

La famille $(X_i)_{i \in I}$ des projections est une famille de variables aléatoires indépendantes telles que, pour tout $i \in I$, X_i soit de loi μ_i .

Démonstration. On introduit l'algèbre de Boole \mathcal{C} constitués des cylindres c'est à dire les ensembles $\prod_{i \in I} B_i$ où $(B_i)_{i \in I} \in \prod_{i \in I} \mathcal{F}_i$ tels que $B_i = E_i$ sauf pour un nombre fini. Par définition, la tribu $\bigotimes_{i \in I} \mathcal{F}_i$ est la plus petite tribu rendant les projections X_i mesurables. On montre facilement que \mathcal{C} engendre cette tribu. Ensuite, pour un cylindre $C \in \mathcal{C}$, alors il existe $J \subset I$ fini tel que

$$C = \prod_{j \in J} C_j \times \prod_{j \in I \setminus J} E_j, \quad \text{avec } C_j \in \mathcal{F}_j, \quad j \in J.$$

Pour ce cylindre C , on pose

$$\mu(C) = \prod_{j \in J} \mu_j(C_j).$$

Il est alors clair que $\mu(\emptyset) = 0$ et que μ est finiment additive. Il reste donc à montrer le dernier point de la définition 2.2.21.

Lemme 8.3.5. *La classe*

$$\mathcal{D} = \left\{ C \times \prod_{j \neq i} E_j, \quad C \in \mathcal{K}_i, \quad i \in I \right\}$$

est compacte. De même, la classe \mathcal{K} formée des intersections dénombrables d'éléments de \mathcal{D} est compacte.

Démonstration. Voir [Nev70, p.78]. □

Soit $(A_n)_{n \geq 0}$ une suite de cylindre telle que $A_{n+1} \subset A_n$ et $\bigcap_{n \geq 0} A_n = \emptyset$. Soit également $\varepsilon > 0$. Pour chaque $n \geq 0$, il existe $J_n \subset I$ fini tel que

$$A_n = \prod_{j \in J_n} A_{n,j} \times \prod_{j \in I \setminus J_n} E_j.$$

Pour chaque $n \geq 0$ et $j \in J_n$, rappelant que J_n est fini, on peut choisir $C_{n,j} \in \mathcal{K}_j$ de sorte que

$$\mu(A_n) \leq \prod_{j \in J_n} \mu_j(C_{n,j}) + \varepsilon. \quad (8.3)$$

On note $C_n = \prod_{j \in J_n} C_{n,j} \times \prod_{j \in I \setminus J_n} E_j$. Observons que la condition de décroissance sur (A_n) implique que $J_n \subset J_{n+1}$ et que pour tout $j \in J_n$, $A_{n+1,j} \subset A_{n,j}$. Ainsi, sans perte de généralité, on peut supposer que $C_{n+1} \subset C_n$.

Maintenant, $\bigcap_{n \geq 0} A_n = \emptyset$ implique $\bigcap_{n \geq 0} C_n = \emptyset$. Par propriété de compacité et décroissance de (C_n) , il existe $N \geq 0$ tel que, pour tout $n \geq N$, $C_n = \emptyset$ et $\mu(C_n) = 0$. Finalement, pour tout $n \geq 0$, $\mu(A_n) \leq \mu(C_n) + \varepsilon = \varepsilon$. Ceci montre que $\lim_{n \rightarrow \infty} \mu(A_n) = 0$. Le théorème est alors une conséquence du théorème 2.2.24. □

Le théorème ci-dessus admet le corollaire suivant sous l'hypothèse polonaise certes plus restrictive mais néanmoins largement suffisante dans la presque totalité des applications.

Corollaire 8.3.6. *Soit, pour tout $i \in I$, $(E_i, \mathcal{F}_i, \mu_i)$ un espace probabilisé. On suppose que, pour tout $i \in I$,*

- E_i est un espace polonais,
- \mathcal{F}_i est la tribu borélienne sur E_i ,
- μ_i est une mesure borélienne.

Alors, il existe une unique probabilité μ sur $(\prod_{i \in I} E_i, \bigotimes_{i \in I} \mathcal{F}_i)$, notée $\bigotimes_{i \in I} \mu_i$, telle que pour tout $(B_i)_{i \in I} \in \prod_{i \in I} \mathcal{F}_i$ avec $B_i = E_i$ sauf pour un nombre fini de $i \in I$,

$$\left(\bigotimes_{i \in I} \mu_i \right) \left(\prod_{i \in I} B_i \right) = \prod_{i \in I} \mu_i(B_i).$$

Démonstration. Il suffit de de poser \mathcal{K}_i l'ensemble des parties compactes de E_i . □

Remarque 75. Dans le chapitre 5, nous avons construit la mesure de produit de façon différente bien que le théorème de Carathéodory nous donne facilement l'existence et l'unicité de la mesure produit. Cependant, de ce théorème nous ne pouvons déduire le théorème de Fubini qui explicite l'intégrale multiple en termes d'intégrales itérées. Dans le cas d'un produit infini, il n'existe plus de telle formulation, en particulier, le calcul de l'intégrale

$$\int_{\prod_{i \in I} E_i} f d \bigotimes_{i \in I} \mu_i$$

n'a rien d'évident.

8.3.3 Caractérisation de l'indépendance de v.a.r.

L'indépendance peut se caractériser à l'aide des trois caractérisations de lois, à savoir, fonctions tests, fonctions de répartition, densités de probabilité. Ces résultats découlent facilement des définitions.

Proposition 8.3.7. Soient $(X_i)_{i \in I}$ une famille de v.a.r et $f_i : E_i \rightarrow \mathbb{R}$, $i \in I$, des fonctions mesurables bornées. Alors la famille $(X_i)_{i \in I}$ est indépendante si et seulement si pour tout $J \subset I$ fini

$$\mathbf{E} \left[\prod_{j \in J} f_j(X_j) \right] = \prod_{j \in J} \mathbf{E}[f_j(X_j)].$$

Proposition 8.3.8. Soit $(X_i)_{i \in I}$ une famille de v.a.r.. Alors $(X_i)_{i \in I}$ est indépendante si et seulement si pour tout $J \subset I$ fini et pour tout $(t_j)_{j \in J} \in \mathbb{R}^J$

$$F_{(X_j)_{j \in J}}((t_j)_{j \in J}) = \mathbf{P} \left(\bigcap_{j \in J} \{X_j \leq t_j\} \right) = \prod_{j \in J} \mathbf{P}(X_j \leq t_j) = \prod_{j \in J} F_{X_j}(t_j).$$

Proposition 8.3.9. Soit $(X_i)_{i \in I}$ une famille de v.a.r.. On suppose que les v.a.r. X_i , $i \in I$, admettent une densité p_i . Alors la famille $(X_i)_{i \in I}$ est indépendante si et seulement si pour tout $j \in J$ fini le vecteur $(X_j)_{j \in J} \in \mathbb{R}^J$ admet pour densité la fonction $p : \mathbb{R}^J \rightarrow \mathbb{R}_+$ telle que

$$p(x) = p((x_j)_{j \in J}) = \prod_{j \in J} p_j(x_j).$$

Exercice 30. Démontrer les quatre propositions ci-dessus, notamment dans le cas d'une famille finie (X_i) de variables aléatoires.

On note $\mathbb{U}^d = \{z \in \mathbb{C}^d : |z_1|^2 + \dots + |z_d|^2 \leq 1\}$. Si $X \in \mathbb{N}^d$ est un vecteur aléatoire à coordonnées entières, alors sa fonction génératrice est définie pour tout $z = (z_1, \dots, z_d) \in \mathbb{U}^d$ par

$$G_X(z) = \mathbf{E}(z_1^{X_1} \dots z_d^{X_d}) = \mathbf{E}(z^X).$$

Proposition 8.3.10. Soit $(X_i)_{i \in I}$ une famille de variables aléatoires à valeurs dans \mathbb{N} . Alors la famille $(X_i)_{i \in I}$ est indépendante si et seulement si pour tout $J \subset I$ fini,

$$\forall (z_j)_{j \in J} \in \mathbb{U}^J, \quad G_{(X_j)_{j \in J}}((z_j)_{j \in J}) = \prod_{j \in J} G_{X_j}(z_j).$$

Dans le chapitre 9, un résultat similaire impliquant les fonctions caractéristiques sera énoncé pour les vecteurs aléatoires de \mathbb{R}^d .

Le corollaire suivant est immédiat mais particulièrement utile en pratique.

Corollaire 8.3.11. Si X_1, \dots, X_n sont des variables aléatoires réelles indépendantes de densités respectives f_{X_i} , alors la densité du vecteur (X_1, \dots, X_n) est donnée par

$$f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n).$$

Remarque 76. Soient X et Y deux *v.a.r.* indépendantes de densités respectives f_X et f_Y . Le corollaire précédant implique que la densité $f_{(X,Y)}$ du couple (X, Y) est $f_{(X,Y)}(x, y) = f_X(x)f_Y(y)$. Comme nous l'avons dit au chapitre précédant, la connaissance de la loi du couple permet de déduire la loi des marginales mais la réciproque est généralement fautive. Si on suppose en outre que les deux variables X et Y du couple (X, Y) sont indépendantes alors on peut déduire des lois marginales la loi du couple.

Corollaire 8.3.12. *Soit (X, Y) un couple de variables aléatoires de densité $h = h_{(X,Y)} = h_X h_Y$ — i.e. X et Y admettent une densité et sont indépendantes. Alors $h_{X+Y}(u) = \int_{\mathbb{R}} h_X(v)h_Y(u-v) dv$ est la densité de $X + Y$. Cette transformation est appelée produit de convolution entre h_X et h_Y et est noté $h_X * h_Y$.*

Démonstration. Soit $g : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ une fonction continue bornée.

$$\mathbf{E}(g(X + Y, X)) = \int_{\mathbb{R}^2} g(x + y, x)h_X(x)h_Y(y) dx dy.$$

On pose $u = x + y$ et $v = x$ donc $x = v$ et $y = u - v$. Le déterminant de la jacobienne vaut en module 1. Après le changement de variable, on obtient

$$\mathbf{E}(g(X + Y, X)) = \int_{\mathbb{R}^2} g(u, v)h_X(v)h_Y(u - v) dudv.$$

On vérifie facilement, après intégration par rapport à v que la densité de $X + Y$ est donnée par

$$h_{(X+Y)}(u) = \int_{\mathbb{R}} h_X(v)h_Y(u - v) dv.$$

□

Remarque 77. Ce corollaire donne une expression explicite de la densité de la somme de deux *v.a.r.*. Pour n *v.a.r.* à densité, il faut itérer ces produits de convolution, soit itérer $n - 1$ intégrales. Ce procédé devient vite infernal, on donnera au chapitre suivant une façon plus facile de calculer la loi d'une somme de *v.a.r.* indépendantes.

Proposition 8.3.13. *Soient X et Y deux *v.a.r.* indépendantes alors $\mathbf{cov}(X, Y) = 0$.*

La réciproque est fautive sauf dans le cas où le couple (X, Y) est gaussien, c'est à dire pour tout $t, s \in \mathbb{R}$ la variable $tX + sY$ est gaussienne (voir le chapitre 10 pour un contre-exemple).

Si $(X_n)_{n \geq 0}$ est suite de variables aléatoires admettant chacune un moment d'ordre 1, alors il vient facilement que

$$\mathbf{E} \left(\sum_{n=0}^N X_n \right) = \sum_{n=0}^N \mathbf{E}(X_n).$$

Sous l'hypothèse d'indépendance, on peut également expliciter la variance.

Proposition 8.3.14. *Soit $(X_n)_{n \geq 0}$ une suite *v.a.r.* indépendantes admettant chacune un moment d'ordre 2. Alors,*

$$\mathbf{V} \left(\sum_{k=0}^n X_k \right) = \sum_{k=0}^n \mathbf{V}(X_k).$$

Démonstration. Calculons

$$\mathbf{V} \left(\sum_{k=0}^n X_k \right) = \mathbf{E} \left[\left(\sum_{k=0}^n X_k \right)^2 \right] - \left(\mathbf{E} \left[\sum_{k=0}^n X_k \right] \right)^2 = \sum_{0 \leq i, j \leq n} \mathbf{E}(X_i X_j) - \sum_{0 \leq i, j \leq n} \mathbf{E}(X_i) \mathbf{E}(X_j).$$

Par indépendance, $\mathbf{E}(X_i X_j) = \mathbf{E}(X_i) \mathbf{E}(X_j)$ sauf lorsque $i = j$. Et donc, après simplification

$$\mathbf{V} \left(\sum_{k=0}^n X_k \right) = \sum_{k=0}^n [\mathbf{E}(X_k^2) - \mathbf{E}(X_k)^2] = \sum_{k=0}^n \mathbf{V}(X_k).$$

□

8.4 Une application du second lemme de Borel-Cantelli

Soit $\mathcal{A} = \{0, 1\}$ un alphabet fini et $(X_n)_{n \geq 1}$ une suite de variables indépendantes. On suppose que pour tout $n \geq 1$, $\mathbf{P}(X_n = 1) = 1 - \mathbf{P}(X_n = 0) = p$ pour $p \in (0, 1)$. Autrement dit les variables aléatoires X_n sont identiquement distribuées de loi commune la loi de Bernoulli de paramètre $p \in (0, 1)$ notée $\mathcal{B}(p)$. On dit que $(X_n)_{n \geq 1}$ est une suite de *v.a.* indépendantes et identiquement distribuées, on note souvent *i.i.d.*.

On définit la variable aléatoire τ à valeurs dans $\overline{\mathbb{N}}^*$ par

$$\tau = \inf\{n \geq 1 : X_n = 1\} \quad \text{et} \quad \inf \emptyset = \infty.$$

Pour tout $n \geq 1$, $\mathbf{P}(\tau = n) = (1 - p)^{n-1}p$. Autrement dit, τ suit une loi géométrique de paramètre p . Cette variable aléatoire τ est finie p.s. :

$$\mathbf{P}(\tau = \infty) = \mathbf{P}\left(\bigcap_{n \geq 1} \{X_n = 0\}\right) = \lim_{q \rightarrow \infty} \mathbf{P}\left(\bigcap_{n=1}^q \{X_n = 0\}\right) = \lim_{q \rightarrow \infty} (1 - p)^q p = 0.$$

Le lemme de Borel-Cantelli implique que

$$\mathbf{P}(\limsup\{X_n = 1\}) = 1 \iff \sum_{n \geq 1} \mathbf{P}(X_n = p) = \infty.$$

Autrement dit, l'événement $\{X_n = 1\}$ arrive infiniment souvent.

En fait, en notant $A_n = \{X_n = 1, X_{n+1} = 1, \dots, X_{n+(k-1)} = 1\}$ pour $k \geq 1$ et $n \geq 1$. Les événements $(A_n)_{n \geq 1}$ ne sont pas indépendants à cause du chevauchement. Par contre les événements $(A_{kn})_{n \geq 1}$ sont indépendants en utilisant la propriété d'indépendance par blocs. De plus $\mathbf{P}(A_{kn}) = p^k$ pour tout $n \geq 1$ et $k \geq 1$ si bien que $\sum_n \mathbf{P}(A_{kn}) = \infty$ pour tout $k \geq 1$. Par conséquent, pour tout $k \geq 1$, l'événement A_{kn} se réalise infiniment souvent. Comme $\{A_{kn} \text{ i.s.}\} \subset \{A_n \text{ i.s.}\}$, il vient que $\mathbf{P}(\limsup A_n) = 1$. Concrètement, cela signifie que dans l'expérience du pile/face (prendre $p = 1/2$ par exemple), on trouve infiniment souvent des sous-suites de k faces consécutives, k étant fixé mais pouvant être arbitrairement grand. Notons $\tau^k(\omega) = \inf\{n \geq 1 : \omega \in A_n\}$ la première occurrence de l'événement A_n . La loi de cette variable aléatoire est plutôt difficile à calculer, car avant cette première occurrence il peut se passer à peu près n'importe quoi sauf, bien entendu, l'occurrence d'un k -bloc de faces. En particulier, il peut y avoir des ℓ -blocs de faces avec $\ell < k$. Si nous étions en mesure de calculer son espérance, nous constaterions que celle-ci croît très vite avec k de sorte qu'il faille attendre très longtemps en moyenne pour voir apparaître ces grands blocs.

Pour terminer, à l'aide de l'inégalité de Bienaymé-Tchebychev, pour tout $\varepsilon > 0$,

$$\mathbf{P}\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - p\right| \geq \varepsilon\right) \leq \frac{\mathbf{V}(X_1)}{\varepsilon^2 n} \rightarrow 0.$$

Cela donne un premier résultat de type loi des grands nombres (on l'appellera loi faible des grands nombres) : la moyenne arithmétique des X_n est avec grande probabilité dans un intervalle de largeur $2\varepsilon > 0$ centrée en $p = \mathbf{E}(X_1)$; cette probabilité est d'autant plus grande que n est grande.

Chapitre 9

Fonctions caractéristiques

La notion de fonction caractéristique provient de l'analyse et plus précisément de la théorie de Fourier. Cette théorie, dont les prémices remontent à un peu plus de deux siècles, est très riches et fait intervenir un grand nombre de concept d'analyse et d'analyse fonctionnelle. Nous nous restreindrons ici aux propriétés utiles dans le cadre des probabilités et des statistiques, lesquelles pour la plupart ne seront pas démontrées.

9.1 Fonction caractéristique d'une *v.a.r.*

Définition 9.1.1. Soit X une *v.a.r.*. La fonction caractéristique de X ou transformée de Fourier de X est la fonction $\varphi_X : \mathbb{R} \rightarrow \mathbb{C}$ définie pour $t \in \mathbb{R}$ par

$$\varphi_X(t) = \mathbf{E}(e^{itX}) = \int_{\mathbb{R}} e^{itx} \mathbf{P}_X(dx). \quad (9.1)$$

Si X est une variable discrète — pour fixer les idées, à valeurs dans \mathbb{N} —, l'équation (9.1) se réécrit

$$\varphi_X(t) = \sum_{n \geq 0} e^{itn} \mathbf{P}(X = n).$$

Remarque 78. La fonction génératrice de X est donnée par

$$G_X(z) = \sum_{n \geq 0} z^n \mathbf{P}(X = n), \quad z \in \mathbb{U}.$$

Ainsi, au moins formellement, la fonction caractéristique de X n'est rien d'autre que la fonction génératrice évaluée en $z = e^{it}$, $t \in \mathbb{R}$.

Si X est une *v.a.r.* de densité f , alors l'équation (9.1) se réécrit

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx} f(x) dx.$$

Eu égard au théorème central limite que l'on démontrera au chapitre 12, il est important de connaître la fonction caractéristique de la loi normale centrée réduite.

Proposition 9.1.2. Soit $X \sim \mathcal{N}(0, 1)$. Alors

$$\varphi_X(t) = e^{-t^2/2}, \quad t \in \mathbb{R}.$$

Démonstration. On souhaite calculer

$$\varphi_X(t) = \int_{\mathbb{R}} e^{-x^2/2} e^{itx} \frac{dx}{\sqrt{2\pi}}.$$

Il est facile de montrer que φ_X est C^1 et la dérivée est donnée par

$$\varphi'_X(t) = \int_{\mathbb{R}} ix e^{-x^2/2} e^{itx} \frac{dx}{\sqrt{2\pi}}.$$

Une intégration par partie (par rapport à x donc) donne

$$\varphi'_X(t) = \int_{\mathbb{R}} ix e^{-x^2/2} e^{itx} \frac{dx}{\sqrt{2\pi}} = \left[-e^{itx} e^{-x^2/2} \right]_{-\infty}^{\infty} - t \int_{\mathbb{R}} e^{itx} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}}.$$

Ainsi, la fonction φ_X satisfait l'équation différentielle linéaire du premier ordre à coefficient non constant $\varphi'_X(t) + t\varphi_X(t) = 0$. Une solution est donnée par

$$\varphi_X(t) = \varphi_X(0) e^{-\int_0^t s ds} = e^{-t^2/2}.$$

La théorie des équations différentielles linéaires implique que c'est l'unique solution. \square

Proposition 9.1.3. *Soit X une v.a.r.. Alors la fonction caractéristique de X est une fonction continue bornée vérifiant, pour $t \in \mathbb{R}$,*

1. $|\varphi_X(t)| \leq 1$;
2. $\varphi_X(-t) = \overline{\varphi_X(t)}$;
3. $\varphi_{aX+b}(t) = e^{itb} \varphi_X(at)$, $a, b \in \mathbb{R}$;
4. Si X est supposée symétrique — X et $-X$ ont même loi — alors $\varphi_X(t) \in \mathbb{R}$.

Démonstration. La continuité de φ_X est une conséquence du théorème de convergence dominée avec une fonction de domination constante égale à 1 et de la continuité pour tout $\omega \in \Omega$ de la fonction $t \rightarrow e^{itX}$.

1. Pour tout $t \in \mathbb{R}$,

$$|\varphi_X(t)| = |\mathbf{E}(e^{itX})| \leq \mathbf{E}|e^{itX}| = 1.$$

2. Pour tout $t \in \mathbb{R}$,

$$\varphi_X(-t) = \mathbf{E}(e^{-itX}) = \overline{\mathbf{E}(e^{itX})}.$$

3. Pour tout $t \in \mathbb{R}$

$$\varphi_{aX+b}(t) = \mathbf{E}(e^{it(aX+b)}) = e^{itb} \mathbf{E}(e^{i(ta)X}) = e^{itb} \varphi_X(at).$$

4. Puisque X et $-X$ ont même loi alors

$$\varphi_X(t) = \mathbf{E}(e^{itX}) = \mathbf{E}(e^{-itX}),$$

et donc par le point 2), $\varphi_X(t) \in \mathbb{R}$. \square

Remarque 79. On vérifie facilement que si $X \sim \mathcal{N}(0, 1)$ alors $Y = \sigma X + \mu$, $\mu \in \mathbb{R}$ et $\sigma \geq 0$, suit une loi normale $\mathcal{N}(\mu, \sigma^2)$. À l'aide de la proposition et de l'expression de la fonction caractéristique d'une loi normale centrée réduite, on obtient

$$\varphi_Y(t) = e^{it\mu} e^{-\sigma^2 t^2/2} = e^{it\mu - \sigma^2 t^2/2}.$$

Comme son nom l'indique, la fonction caractéristique caractérise la loi d'une v.a.r..

Théorème 9.1.4. *Deux v.a.r. X et Y ont même loi si et seulement si leurs fonctions caractéristiques coïncident, i.e. $\varphi_X = \varphi_Y$.*

Ce théorème donne une quatrième méthode pour calculer la loi d'une v.a.r.. Là encore, cela dépendra de la problématique.

Démonstration. Si X et Y ont même loi, il est clair que leurs fonctions caractéristiques coïncident. La réciproque est un peu subtile et provient de la formule d'inversion de Fourier, nous admettrons ce résultat. \square

Il peut être utile parfois de savoir reconnaître qu'une fonction est une fonction caractéristique.

Théorème 9.1.5 (Bochner). *Une fonction $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ est la fonction caractéristique d'une v.a.r. si et seulement si les trois conditions suivantes sont satisfaites :*

1. $|\varphi(t)| \leq 1$ pour $t \in \mathbb{R}$ et $\varphi(0) = 1$;
2. φ est uniformément continue;
3. φ est définie positive :

$$\forall n \in \mathbb{N}^*, \quad \forall (t_1, \dots, t_n) \in \mathbb{R}^n, \quad \forall (z_1, \dots, z_n) \in \mathbb{C}^n : \quad \sum_{k,l} \varphi(t_k - t_l) z_k \bar{z}_l \geq 0$$

Démonstration. Admis. \square

9.2 Fonctions caractéristiques et moments

Comme avec les séries génératrices, il existe une relation entre dérivée de la fonction caractéristique et moment d'une variable aléatoire réelle.

Théorème 9.2.1. *Soit X une v.a.r. admettant un moment d'ordre $p \in \mathbb{N}$. Alors, la fonction caractéristique est de classe C^p . De plus,*

$$\varphi_X^{(k)}(t) = i^k \mathbf{E}(X^k e^{itX}), \quad 0 \leq k \leq p.$$

En particulier,

$$\mathbf{E}(X^k) = \frac{\varphi_X^{(k)}(0)}{i^k}, \quad 0 \leq k \leq p.$$

Démonstration. Il s'agit d'une conséquence du théorème de convergence dominée de Lebesgue et plus spécifiquement du théorème de dérivation sous le signe intégral. \square

Remarque 80. La réciproque est en général fautive : $\varphi_X'(0)$ peut exister sans pour autant que X admette un premier moment (voir l'exemple 42 ci-dessous). Par contre, si φ_X est C^2 alors X admet un second moment (voir [Spi76]).

Exemple 42. Sur \mathbb{Z} on considère $\mu(z) = c/(z^2 \log z)$ pour $z \geq 2$ (et 0 sinon) où $c > 0$ est une constante convenable. Clairement, $\sum_{z \in \mathbb{Z}} z \mu(z) = \infty$. Pourtant, on montre que $\hat{\mu}$ est dérivable en 0.

Par définition, pour tout $t \in \mathbb{R}$,

$$\frac{\hat{\mu}(t) - 1}{t} = \sum_{|k| \geq 2} \frac{e^{itk} - 1}{t} \frac{c}{k^2 \ln(k)} = -2c \sum_{n \geq 2} \frac{1 - \cos(nt)}{tn^2 \ln(n)}.$$

Considérons la série à termes positifs

$$\frac{1}{t} \sum_{n=2}^{\infty} \frac{1 - \cos(nt)}{n^2 \ln(n)}. \quad (9.2)$$

Pour tout $t \in (0, \frac{1}{2})$, nous découpons la série suivant que n est plus grand ou plus petit que $1/t$. En remarquant que $x \rightarrow (\ln(x))^{-1}$ et $x \rightarrow x^{-2}$ sont décroissantes, on obtient d'une part

$$\begin{aligned} \frac{1}{t} \sum_{n \geq 1/t} \frac{1 - \cos(nt)}{n^2 \ln(n)} &\leq -\frac{2}{t \ln(t)} \sum_{n \geq \lfloor \frac{1}{t} \rfloor} \frac{1}{n^2} \leq -\frac{2}{t \ln(t)} \int_{\lfloor \frac{1}{t} \rfloor}^{\infty} \frac{dx}{x^2} \\ &= -\frac{2}{t (\lfloor \frac{1}{t} \rfloor - 1) \ln(t)} \leq -2 \frac{\lfloor \frac{1}{t} \rfloor + 1}{\lfloor \frac{1}{t} \rfloor - 1} \frac{1}{\ln(t)} \rightarrow_{t \rightarrow 0} 0. \end{aligned}$$

D'autre part, en utilisant l'inégalité, $1 - \cos(x) \leq \frac{x^2}{2}$, $x \in \mathbb{R}$, il vient

$$\frac{1}{t} \sum_{2 \leq n < 1/t} \frac{1 - \cos(nt)}{n^2 \ln(n)} \leq t \sum_{2 \leq n < 1/t} \frac{1}{\ln(n)} \leq \frac{t}{\ln(2)} + t \sum_{3 \leq n \leq \lfloor \frac{1}{t} \rfloor} \int_{n-1}^n \frac{dx}{\ln(x)} \leq \frac{t}{\ln(2)} + t \int_0^{\frac{1}{t}} \frac{dx}{\ln(x)}.$$

Il reste donc à montrer que le second terme tend vers 0. Pour cela, il suffit de remarquer

$$\int_2^y \frac{dx}{\ln(x)} \rightarrow_{y \rightarrow \infty} \infty$$

si bien que la règle de l'Hôpital implique

$$\frac{\int_2^y \frac{dx}{\ln(x)}}{y} \sim_{y \rightarrow \infty} \frac{1}{\ln(y)} \rightarrow_{y \rightarrow \infty} 0.$$

9.3 Fonctions caractéristiques de vecteurs aléatoires

Soit X un vecteur aléatoire dans \mathbb{R}^d . La fonction caractéristique de X est la fonction $\varphi_X : \mathbb{R}^d \rightarrow \mathbb{C}$ définie pour tout $t \in \mathbb{R}^d$ par

$$\varphi_X(t) = \mathbf{E}(e^{i\langle t, X \rangle}),$$

où $\langle \cdot, \cdot \rangle$ est le produit scalaire usuel sur \mathbb{R}^d .

Ainsi si X est un vecteur aléatoire de densité $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$, alors

$$\varphi_X(t) = \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} f(x) dx.$$

Proposition 9.3.1. *Soit X un vecteur aléatoire dans \mathbb{R}^q . Alors sa fonction caractéristique φ_X est une fonction continue bornée vérifiant, pour tout $t \in \mathbb{R}^q$,*

1. $|\varphi_X(t)| \leq 1$;
2. $\varphi_X(-t) = \overline{\varphi_X(t)}$;
3. si $A \in \mathcal{M}_{p,q}(\mathbb{R})$ et $B \in \mathbb{R}^p$ alors

$$\varphi_{AX+B}(t) = e^{i\langle t, B \rangle} \varphi_X(A^*t), \quad \forall t \in \mathbb{R}^p,$$

où A^* est l'adjoint de A .

4. si X est supposée symétrique, alors φ_X est à valeurs réelles.

Théorème 9.3.2. *Deux vecteurs aléatoires X et Y ont même loi si et seulement si leurs fonctions caractéristiques coïncident i.e. $\varphi_X = \varphi_Y$.*

Si $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ est C^1 , on note $\nabla\varphi$ le gradient de φ .

Proposition 9.3.3. *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire.*

1. Si X admet un moment d'ordre 1, alors φ_X est C^1 et

$$\mathbf{E}(X) = -i\nabla^* \varphi_X(0).$$

2. Si X admet un moment d'ordre 2, alors φ_X est C^2 et

$$\text{Cov}(X_i, X_j) = -\partial_{ij}^2 \varphi_X(0) + \partial_i \varphi_X(0) \partial_j \varphi_X(0).$$

3. de manière générale, à l'aide de multi-indices, si X admet un moment d'ordre p , alors φ_X est de classe C^p et pour tout $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$ tel que $|\alpha| = \alpha_1 + \dots + \alpha_d \leq p$

$$\mathbf{E}(X^\alpha) = i^{-|\alpha|} \partial_1^{\alpha_1} \dots \partial_d^{\alpha_d} \varphi_X(0),$$

où $X^\alpha = X_1^{\alpha_1} \dots X_d^{\alpha_d}$.

Remarque 81. Il faut encore une fois faire attention au notation. En général, l'opérateur ∇ est représenté en ligne : $\nabla = (\partial_1, \dots, \partial_d)$. Pour être cohérent avec notre convention de vecteur colonne, on doit prendre la transposée ∇^* .

Démonstration. Exercice. □

9.4 Fonctions caractéristiques et indépendance

Théorème 9.4.1. *Deux v.a.r. X et Y sont indépendantes si et seulement si $\varphi_{(X,Y)} = \varphi_X \varphi_Y$.*

Démonstration. Si X et Y sont indépendantes, le résultat suit directement de la proposition 8.3.7. La réciproque fait appel à l'analyse de Fourier et dépasse un peu le cadre de ces notes. □

Exemple 43. Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. i.i.d. et posons $S_n = X_1 + \dots + X_n$. Alors, pour tout $n \geq 1$, et tout $t \in \mathbb{R}$,

$$\varphi_{S_n}(t) = \mathbf{E} \left[e^{it(X_1 + \dots + X_n)} \right] = \mathbf{E} \left[\prod_{i=1}^n e^{itX_i} \right] = \prod_{i=1}^n \mathbf{E} \left[e^{itX_i} \right] = \varphi_{X_1}(t)^n$$

Chapitre 10

Vecteurs gaussiens

10.1 Manipulation des vecteurs gaussiens

Le rôle central de la loi normale en statistique et notamment dans le théorème central limite que l'on montrera au chapitre 12 motive l'étude des vecteurs gaussiens qui sont en quelque sorte l'équivalent en dimension supérieure de la loi normale. On parle d'ailleurs parfois de loi normale sur \mathbb{R}^d .

Rappelons tout d'abord que si $X \sim \mathcal{N}(m, \sigma^2)$ alors

- elle a pour densité la fonction $x \rightarrow \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$ lorsque $\sigma^2 > 0$;
- elle est presque sûrement égale à m lorsque $\sigma^2 = 0$;
- sa fonction caractéristique est donnée par

$$\varphi_X(t) = \exp\left(imt - \frac{\sigma^2 t^2}{2}\right).$$

On se place dans la base canonique de \mathbb{R}^d . Si X est un vecteur aléatoire à valeurs dans \mathbb{R}^d , on considérera toujours X comme un vecteur colonne (même si, encore une fois, X est écrit en ligne pour des raisons typographiques), c'est à dire en notant u^* la transposée de u , $X = (X_1, \dots, X_d)^*$.

Si A est une matrice réelle de taille $q \times d$ et b un vecteur de \mathbb{R}^q , le vecteur aléatoire $Y = AX + b$ appartient à \mathbf{L}^2 dès qu'il en va de même pour X . On vérifie facilement que

$$\mathbf{E}(AX + b) = A\mathbf{E}(X) + b \quad \text{et} \quad \text{Cov}(AX + b) = A\text{Cov}(X)A^*.$$

En particulier, si $u \in \mathbb{R}^d$, la variable réelle u^*X a pour moyenne $u^*\mathbf{E}(X)$ et pour variance $\mathbf{V}(u^*X) = u^*\text{Cov}(X)u$. Par conséquent, $\text{Cov}(X)$ est une matrice réelle symétrique semi-définie positive.

Définition 10.1.1. Soit X un vecteur aléatoire dans \mathbb{R}^d . Le vecteur aléatoire X est dit gaussien si, pour tout $t \in \mathbb{R}^d$, t^*X est *v.a.r.* gaussienne.

Remarque 82. Si X est un vecteur gaussien, chacune de ses coordonnées sont des variables aléatoires gaussiennes.

Exemple 44. Il ne suffit pas que chacune des coordonnées d'un vecteur aléatoire soit gaussienne pour que le vecteur soit gaussien. En effet, si X et ε sont deux *v.a.r.* indépendantes, $X \sim \mathcal{N}(0, 1)$ et $\mathbf{P}(\varepsilon = \pm 1) = 1/2$, alors εX suit une $\mathcal{N}(0, 1)$ mais le couple $(X, \varepsilon X)$ n'est pas un vecteur gaussien. En effet, d'un côté, en utilisant l'indépendance,

$$\mathbf{E}(e^{it\varepsilon X}) = \mathbf{E}(e^{it\varepsilon X} \mathbf{1}_{\varepsilon=1} + e^{it\varepsilon X} \mathbf{1}_{\varepsilon=-1}) = \frac{\mathbf{E}(e^{itX}) + \mathbf{E}(e^{-itX})}{2} = e^{-t^2/2},$$

où l'on reconnaît la fonction caractéristique d'une *v.a.r.* gaussienne. D'un autre côté, toujours en utilisant l'indépendance et en décomposant selon les valeurs que peut prendre ε , on obtient

$$\mathbf{E}(\exp(itX + is\varepsilon X)) = \frac{\varphi_X(t+s) + \varphi_X(t-s)}{2} \neq \varphi_X(t)\varphi_X(s).$$

Proposition 10.1.2. Soient X_1, \dots, X_d des v.a.r. gaussiennes et indépendantes, alors $X = (X_1, \dots, X_d)$ est un vecteur gaussien.

Démonstration. Soit $t \in \mathbb{R}^d$, on vérifie que $t^*X = \sum_{j=1}^d t_j X_j$ suit une loi normale. En effet, par indépendance, en supposant que $X_j \sim \mathcal{N}(m_j, \sigma_j^2)$, on calcule

$$\varphi_X(s) = \mathbf{E}(e^{ist^*X}) = \prod_{j=1}^d \mathbf{E}(e^{ist_j X_j}) = \exp \left(is \sum_{j=1}^d m_j t_j - \frac{s^2}{2} \sum_{j=1}^d \sigma_j^2 t_j^2 \right),$$

où l'on reconnaît la fonction caractéristique d'une v.a.r. gaussienne de moyenne m et de variance σ^2 données par

$$m = \sum_{j=1}^d t_j m_j \quad \text{et} \quad \sigma^2 = \sum_{j=1}^d \sigma_j^2 t_j^2.$$

□

Théorème 10.1.3. Un vecteur aléatoire X est gaussien dans \mathbb{R}^d si et seulement si sa fonction caractéristique est de la forme

$$t \longrightarrow \exp \left(it^*m - \frac{t^*\Gamma t}{2} \right)$$

où $m \in \mathbb{R}^d$ et Γ est matrice $d \times d$ semi-définie positive.

Démonstration. Supposons X gaussien alors t^*X est une v.a.r. gaussienne pour tout $t \in \mathbb{R}^d$. Ainsi,

$$\varphi_X(t) = \varphi_{t^*X}(1) = \exp \left\{ i\mathbf{E}(t^*X) - \frac{\mathbf{V}(t^*X)}{2} \right\} = \exp \left\{ it^*\mathbf{E}(X) - \frac{t^*\mathbf{Cov}(X)t}{2} \right\}.$$

Réciproquement, on suppose que $\varphi_X(t) = \exp \left\{ it^*m - \frac{t^*\Gamma t}{2} \right\}$ alors pour tout $c \in \mathbb{R}$

$$\varphi_{t^*X}(c) = \varphi_X(ct) = \exp \left\{ it^*mc - \frac{t^*\Gamma t}{2} c^2 \right\}.$$

Ceci montre, pour tout $t \in \mathbb{R}^d$, que t^*X est une v.a.r. gaussienne de moyenne t^*m et de variance $t^*\Gamma t$. D'autre part, comme pour tout vecteur aléatoire, nous avons, pour tout $t \in \mathbb{R}^d$, $\mathbf{E}(t^*X) = t^*\mathbf{E}(X)$ et $\mathbf{V}(t^*X) = t^*\mathbf{Cov}(X)t$. Ceci montre que $m = \mathbf{E}(X)$ et $\Gamma = \mathbf{Cov}(X)$. □

Corollaire 10.1.4. Si $X \in \mathbb{R}^d$ est gaussien alors sa loi est complètement caractérisée à l'aide de son espérance et sa matrice de covariance.

Corollaire 10.1.5. Soit $X = (X_1, \dots, X_d) \in \mathbb{R}^d$ un vecteur gaussien. Alors les composantes X_1, \dots, X_d sont des v.a.r. indépendantes si et seulement si la matrice de covariance de X est diagonale.

Proposition 10.1.6. Soient X un vecteur gaussien de \mathbb{R}^d , $b \in \mathbb{R}^q$ et A une matrice réelle de taille $q \times d$. Alors $Y = AX + b$ est un vecteur gaussien de \mathbb{R}^q et de matrice de covariance $A\mathbf{Cov}(X)A^*$.

Démonstration. Soit $t \in \mathbb{R}^q$, alors $t^*Y = t^*(AX) + t^*b = (t^*A)X + t^*b$ est gaussien car $t^*A \in \mathbb{R}^d$ et $t^*b \in \mathbb{R}$ et $X \in \mathbb{R}^d$ est gaussien. Le reste de la proposition est immédiat. □

Proposition 10.1.7. Soient X un vecteur gaussien, A et B deux matrices réelles de tailles respectives $q \times d$ et $r \times d$. Alors AX et BX sont indépendants si et seulement si $A\mathbf{Cov}(X)B^* = 0$.

Remarque 83. Notons que AX et BX sont des vecteurs de \mathbb{R}^q et \mathbb{R}^r respectivement. Ceci reste cohérent avec la définition d'indépendance de deux variables aléatoires puisqu'il n'est pas nécessaire que celles-ci prennent leurs valeurs dans un même espace.

Démonstration. Notons $m = \mathbf{E}(X)$ et $\Gamma = \text{Cov}(X)$. Soit C la matrice réelle de taille $(q+r) \times d$ définie par $C = \begin{pmatrix} A \\ B \end{pmatrix}$ et soit $Y \in \mathbb{R}^{q+r}$ défini par $Y = CX = \begin{pmatrix} AX \\ BX \end{pmatrix}$. D'après la proposition 10.1.6, Y est un vecteur gaussien et on a $\mathbf{E}(Y) = Cm$, $\text{Cov} Y = CTC^*$. Le théorème 10.1.3 implique

$$\forall u \in \mathbb{R}^{q+r}, \quad \varphi_Y(u) = \exp \left\{ iu^*Cm - \frac{1}{2}u^*CTC^*u \right\}.$$

Écrivons $u = \begin{pmatrix} s \\ t \end{pmatrix}$ avec $s \in \mathbb{R}^q$ et $t \in \mathbb{R}^r$. On a $u^*C = s^*A + t^*B$ et donc

$$u^*CTC^*u = s^*A\Gamma A^* + t^*B\Gamma B^*t + s^*A\Gamma B^*t + t^*B\Gamma A^*s = s^*A\Gamma A^*s + t^*B\Gamma B^*t + 2s^*A\Gamma B^*t.$$

Finalement, on calcule

$$\varphi_Y(u) = \exp \left\{ is^*Am - \frac{1}{2}s^*A\Gamma A^*s \right\} \exp \left\{ it^*Bm - \frac{1}{2}t^*B\Gamma B^*t \right\} \exp \{-s^*A\Gamma B^*t\}.$$

Or AX et BX sont naturellement des vecteurs gaussiens et d'après le théorème 10.1.3 l'égalité ci-dessus se réécrit

$$\varphi_Y(u) = \varphi_{AX}(s)\varphi_{BX}(t) \exp\{-s^*A\Gamma B^*t\}.$$

Par conséquent, AX et BX sont indépendants si et seulement si pour tous $s \in \mathbb{R}^q$, $t \in \mathbb{R}^r$, $s^*A\Gamma B^*t = 0$. Autrement dit, si et seulement si $A\Gamma B^* = 0$. \square

Remarque 84. Si X, Y sont deux vecteurs gaussiens indépendants à valeurs dans \mathbb{R}^q et \mathbb{R}^r respectivement, alors $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$ est un vecteur gaussien de \mathbb{R}^{q+r} .

Théorème 10.1.8. Soient $m \in \mathbb{R}^d$ et Γ une matrice $d \times d$ réelle symétrique semi-définie positive.

1. Il existe un vecteur gaussien X à valeurs dans \mathbb{R}^d de loi $\mathcal{N}(m, \Gamma)$ c'est à dire tel que $\mathbf{E}[X] = m$ et $\text{Cov}(X) = \Gamma$.
2. X admet une densité si et seulement si Γ est non dégénérée. Dans ce, la densité de X s'écrit

$$\forall x \in \mathbb{R}^d, \quad p(x) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Gamma}} \exp \left\{ -\frac{(x-m)^*\Gamma^{-1}(x-m)}{2} \right\}.$$

Si X est concentrée sur l'espace affine $m + (\ker \Gamma)^\perp$.

3. Il existe $\alpha > 0$ tel que $\mathbf{E}(e^{\alpha|X|^2}) < \infty$.

Démonstration. Soient Y_1, \dots, Y_d des variables indépendantes de loi $\mathcal{N}(0, 1)$ et notons $Y = (Y_1, \dots, Y_d)^*$. Alors Y est un vecteur gaussien d'espérance nulle et de matrice de covariance $\text{Cov} Y = I$.

La matrice Γ est symétrique semi-définie positive, aussi notant $\Sigma^2 = \text{diag}(\sigma_1^2, \dots, \sigma_d^2)$ la matrice diagonale constituée des valeurs propres (toutes réelles positives) de Γ , il existe une matrice orthogonale A telle que $\Gamma = A\Sigma^2A^*$.

Pour le point 1, on pose $X = m + A\Sigma Y$. Alors X est un vecteur gaussien comme la transformation affine d'un vecteur gaussien. De plus, $\mathbf{E}(X) = A\Sigma\mathbf{E}(Y) + m = m$ et $\text{Cov}(X) = A\Sigma\text{Cov}(Y)\Sigma^*A^* = A\Sigma^2A^* = \Gamma$.

Pour le point 2, supposons Γ inversible si bien que Σ est également inversible ou encore les valeurs propres de Γ sont toutes strictement positives. D'autre part, les composantes de Y étant indépendantes et gaussiennes, la densité de Y est donnée par

$$f_Y(y) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{\|y\|_2^2}{2}}.$$

Soit $g : \mathbb{R}^d \rightarrow \mathbb{R}_+$ mesurable positive, alors

$$\mathbf{E}[f(X)] = \mathbf{E}[f(m + A\Sigma Y)] = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} f(m + A\Sigma y) e^{-\frac{\|y\|_2^2}{2}} dy.$$

Effectuons le changement de variable $x = m + A\Sigma y$, i.e. $y = \Sigma^{-1}A^{-1}(x - m)$. On a alors,

$$\mathbf{E}[f(X)] = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \exp \left\{ -\frac{\|\Sigma^{-1}A^{-1}(x-m)\|_2^2}{2} \right\} |\det \Sigma^{-1}A^{-1}| dx.$$

On remarque pour conclure que, A étant orthogonale, $|\det \Sigma^{-1}A^{-1}| = (\det \Gamma)^{-1/2}$. De plus, $A^* = A^{-1}$ ainsi

$$\|\Sigma^{-1}A^{-1}(x-m)\|_2^2 = (x-m)^*A\Sigma^{-1}\Sigma^{-1}A^*(x-m) = (x-m)^*\Gamma^{-1}(x-m).$$

Par conséquent,

$$\mathbf{E}[f(X)] = \frac{1}{\sqrt{(2\pi)^d \det \Gamma}} \int_{\mathbb{R}^d} f(x) \exp\left\{-\frac{(x-m)^*\Gamma^{-1}(x-m)}{2}\right\} dx,$$

où l'on identifie la densité de X .

Si Γ n'est pas inversible, considérons u_1, \dots, u_r , $1 \leq r \leq d$, une base orthonormale de $\ker \Gamma$. On a donc

$$\{X \in m + (\ker \Gamma)^\perp\} = \bigcap_{j=1}^r \{u_j^*(X-m) = 0\}.$$

Or, pour tout $u \in \ker \Gamma$, on a $\mathbf{V}(u^*(X-m)) = u^* \mathbf{Cov}(X)u = 0$. Par conséquent la *v.a.r.* $u^*(X-m)$ est presque sûrement égale à sa moyenne qui est nulle. D'où $\mathbf{P}(u_i^*(X-m) = 0) = 1$ pour tout $i = 1, \dots, r$ et donc $\mathbf{P}(u_i^*(X-m) = 0, \forall i = 1, \dots, r) = 1$. Dans ce cas la loi de X est supportée par l'espace affine $m + (\ker \Gamma)^\perp$ de dimension strictement plus petite que d . En particulier, X n'admet pas de densité par rapport à λ_d .

Pour le point 3, remarquons que si G est une *v.a.r.* de loi $\mathcal{N}(0, 1)$, alors pour tout $s < 1/2$, en faisant le changement de variable $z = x\sqrt{1-2s}$, on a

$$\beta(s) = \mathbf{E}(e^{sG^2}) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{sx^2} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2(1-2s)}{2}} dx = \frac{1}{\sqrt{1-2s}}.$$

Pour $s \geq 1/2$, $\beta(s) = \infty$. Puisque A est orthogonale, on a

$$\mathbf{E}\left[e^{\alpha\|X-m\|_2^2}\right] = \mathbf{E}\left[e^{\alpha\|\Sigma Y\|_2^2}\right] = \mathbf{E}\left[e^{\alpha\sum_{j=1}^d \sigma_j^2 Y_j^2}\right],$$

et comme les variables Y_i sont *i.i.d.* de loi commune une $\mathcal{N}(0, 1)$ on obtient

$$\mathbf{E}\left[e^{\alpha\|X-m\|_2^2}\right] = \prod_{j=1}^d \mathbf{E}\left[e^{\alpha\sigma_j^2 Y_j^2}\right] = \prod_{j=1}^d \beta(\alpha\sigma_j^2).$$

Cette dernière quantité est donc finie lorsque $\alpha \max_{i \leq d} \sigma_i^2 < 1/2$. Pour finir, remarquons que $\|X\|_2^2 \leq 4\|X-m\|_2^2 + 4\|m\|_2^2$ et donc que

$$\mathbf{E}\left[e^{\alpha\|X\|_2^2}\right] = e^{4\alpha\|m\|_2^2} \mathbf{E}\left[e^{4\alpha\|X-m\|_2^2}\right].$$

D'où l'existence d'un moment exponentiel d'ordre $\alpha > 0$. □

10.2 Loi du χ^2 , moyenne et variance empiriques

Définition 10.2.1. Soit $X = (X_1, \dots, X_d)$ un vecteur gaussien centrée réduite, c'est à dire $m = 0$ et $\Gamma = I_d$. La loi de $|X|^2 = X_1^2 + \dots + X_d^2$ s'appelle la loi du chi-deux à d degrés de libertés; on note $\chi^2(d)$ ou χ_d^2 .

Exemple 45. On rappelle que la loi gamma de paramètres $\alpha > 0$ et $s > 0$, notée $\Gamma_{s,\alpha}$, admet pour densité

$$\gamma_{s,\alpha}(x) = \frac{\alpha^s}{\Gamma(s)} x^{s-1} e^{-\alpha x} \mathbf{1}_{\mathbb{R}_+^*}(x), \quad x \in \mathbb{R}, \quad \text{avec} \quad \Gamma(s) = \int_0^\infty x^{s-1} e^{-x} dx.$$

Si $X \sim \Gamma_{s,\alpha}$ et $Y \sim \Gamma_{t,\alpha}$ sont indépendantes alors la loi de $X+Y$ est une $\Gamma_{s+t,\alpha}$. Pour montrer ce fait, on peut par exemple utiliser le corollaire 8.3.12 :

$$\begin{aligned} \gamma_{s,\alpha} * \gamma_{t,\alpha}(x) &= \int_{\mathbb{R}} \frac{\alpha^s}{\Gamma(s)} y^{s-1} e^{-\alpha y} \mathbf{1}_{\mathbb{R}_+}(y) \frac{\alpha^t}{\Gamma(t)} (x-y)^{t-1} e^{-\alpha(x-y)} \mathbf{1}_{\mathbb{R}_+}(x-y) dy \\ &= \frac{\alpha^{s+t}}{\Gamma(s)\Gamma(t)} e^{-\alpha x} \int_0^x y^{s-1} (x-y)^{t-1} dy = \frac{\alpha^{s+t}}{\Gamma(s+t)} x^{s+t-1} e^{-\alpha x}, \end{aligned}$$

la dernière égalité étant obtenue à l'aide du changement de variable $y = xu$ et en utilisant l'égalité $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ — c.f. par exemple l'exercice 6 de la planche de TD1.

Lemme 10.2.2. *La loi du chi-deux à n degrés de liberté est la loi $\Gamma_{n/2, 1/2}$ donc de densité*

$$x \rightarrow \frac{2^{-n/2}}{\Gamma(n/2)} x^{n/2-1} e^{-x/2} \mathbf{1}_{\mathbb{R}^+}(x).$$

Démonstration. On commence par remarquer que si $G \sim \mathcal{N}(0, 1)$ alors G^2 suit une loi $\Gamma_{1/2, 1/2}$. Donc si X_1, \dots, X_n sont des variables aléatoires indépendantes gaussiennes centrées et réduites, alors la loi de $X_1^2 + \dots + X_n^2$ suit une loi $\Gamma_{n/2, 1/2}$. \square

Théorème 10.2.3 (Théorème de Cochran, version simple). *Soient X un vecteur gaussien de \mathbb{R}^d de loi $\mathcal{N}(0, I)$. L'espace \mathbb{R}^n peut s'écrire comme la somme directe de F et F^\perp . Notons P_F la projection orthogonale sur F . Alors*

1. *les vecteurs aléatoires $P_F X$ et $(I - P_F)X$ sont indépendants de loi respectives $\mathcal{N}(0, P_F)$ et $\mathcal{N}(0, (I - P_F))$;*
2. *les variables aléatoires $\|P_F X\|^2$ et $\|(I - P_F)X\|^2$ sont indépendantes et de lois respectives $\chi^2(q)$ et $\chi^2(d - q)$ où q est la dimension de F ;*

Remarque 85. À toute fin utile, rappelons que $(I - P_F)$ est la projection orthogonale sur F^\perp .

Démonstration. 1. Soit (u_1, \dots, u_q) une base orthonormée de F et (u_{q+1}, \dots, u_d) une base orthonormée de F^\perp . Alors (u_1, \dots, u_d) est une base orthonormée de \mathbb{R}^d . Notons U la matrice de passage de la base standard à la base (u_1, \dots, u_d) . La matrice U est orthogonale, en particulier $U^{-1} = U^*$. Les projections orthogonales sur F et F^\perp s'expriment comme suit dans la base (u_1, \dots, u_d)

$$P_F = UI_q U^* \quad \text{et} \quad (I - P_F) = U(I - I_q)U^*.$$

En effet, si $x = \sum_{i=1}^d x_i u_i$, on calcule

$$P_F x = UI_q U^* x = UI_q \sum_{i=1}^d x_i e_i = U \sum_{i=1}^q x_i e_i = \sum_{i=1}^q x_i u_i.$$

On pose $Y = U^* X$. C'est encore un vecteur gaussien de moyenne $\mathbf{E}(U^* X) = U^* \mathbf{E}(X) = 0$ et de matrice de covariance $\text{Cov}(Y) = U^* I U = I$. Notons que le vecteur Y n'est rien d'autre que le vecteur X exprimée dans la nouvelle base.

On remarque immédiatement que $P_F X = UI_q Y$ et $(I - P_F)X = U(I - I_q)Y$ sont des vecteurs gaussien centrées de covariance respectives $UI_q I_q^* U^* = P_F$ et $U(I - I_q)(I - I_q)^* U^* = (I - P_F)$. De plus, par la proposition 10.1.7, $P_F(I - P_F)^* = UI_q U^* U(I - I_q)^* U^* = 0$ donc $P_F X$ et $(I - P_F)X$ sont indépendants.

2. Pour le deuxième point, en utilisant la nature orthogonale de U , les normes des projections sont données par

$$\|P_F X\|_2^2 = \|UI_q U^* X\|_2^2 = \|UI_q Y\|_2^2 = \|I_q Y\|_2^2 \sim \chi_q^2$$

et

$$\|(I - P_F)X\|_2^2 = \|(I - I_q)Y\|_2^2 \sim \chi_{d-q}^2,$$

car les variables aléatoires Y_1, \dots, Y_d sont indépendantes et de loi gaussiennes centrées réduites. \square

Remarque 86. Ce théorème est un analogue en "loi" du théorème de Pythagore. L'identité $\|x\|_2^2 = \|P_F x\|_2^2 + \|(I - P_F)x\|_2^2$ pour $x \in \mathbb{R}^d$ devient en effet $\|X\|_2^2 \stackrel{d}{=} \|P_F X\|_2^2 + \|(I - P_F)X\|_2^2$.

Le théorème précédent se dérive en de multiples corollaires plus ou moins importants ou utile. À titre d'exemple, nous en donnons deux, le deuxième étant particulièrement intéressant en statistique.

Corollaire 10.2.4 (Théorème de Cochran généralisé). Soit X un vecteur gaussien de \mathbb{R}^d de moyenne $\mu \in \mathbb{R}^d$ et de matrice de covariance $\sigma^2 I$ pour $\sigma^2 > 0$. Soit F_1, \dots, F_k des sous espaces vectoriels, de dimensions respectives d_1, \dots, d_k , deux à deux orthogonaux tels que $\mathbb{R}^d = F_1 \oplus \dots \oplus F_k$. On note P_{F_i} les projecteurs orthogonaux sur F_i . Alors

1. les vecteurs aléatoires $Y_i = \sigma^{-1} P_{F_i}(X - \mu)$, $1 \leq i \leq k$ sont deux à deux indépendants de lois respectives $\mathcal{N}(0, \sigma^{-1} P_{F_i})$;
2. les variables aléatoires réelles $\|Y_i\|^2$, $1 \leq i \leq k$, sont deux à deux indépendantes de lois respectives $\chi^2(d_i)$.

Démonstration. On pose $\tilde{X} = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, I)$ et on procède par induction l'aide du théorème de Cochran simplifié. \square

Corollaire 10.2.5. Soit X un n -échantillon de loi $\mathcal{N}(\mu, \sigma^2)$. On définit la moyenne et la variance empiriques non biaisée de X

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad V_n = \frac{1}{n-1} \sum_{i=1}^n (X_i - M_n)^2.$$

Alors M_n suit une loi $\mathcal{N}(\mu, \sigma^2/n)$, M_n et V_n sont indépendantes et $(n-1)V_n/\sigma^2$ suit la loi du χ_{n-1}^2 .

Démonstration. Soit $Y = (Y_1, \dots, Y_n)^*$ un vecteur gaussien centré réduit. On note

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i \quad \text{et} \quad R^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

La variable $\bar{Y} = \frac{1}{n} \mathbf{1}Y$, c'est l'image de Y par la transformation linéaire $A = \frac{1}{n} \underbrace{(\mathbf{1}, \dots, \mathbf{1})}_n = \frac{1}{n} \mathbf{1}$. Ainsi, \bar{Y}

est une *v.a.r.* gaussienne centrée de variance $1/n$.

On pose $F = \text{vect } \mathbf{1}$. On vérifie que la projection orthogonale sur F notée P_F est définie par $P_F y = \bar{y} \mathbf{1}$ où $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. En effet, $\bar{y} \mathbf{1} \in F$ et

$$\langle y - \bar{y} \mathbf{1}, \mathbf{1} \rangle = \sum_{i=1}^n (y_i - \bar{y}) = 0,$$

si bien que $y - \bar{y} \mathbf{1} \in F^\perp$. Par conséquent $P_F Y = \bar{Y}$ et $Y - \bar{Y} \mathbf{1} = (I - P_F)Y$. En appliquant le théorème de Cochran en remarquant que F^\perp est de dimension $n-1$. Ainsi la variable

$$\|Y - \bar{Y} \mathbf{1}\|^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 = (n-1)R^2 \sim \chi_{n-1}^2$$

et est indépendante de \bar{Y} . Ceci montre le théorème dans le cas centré et réduit.

Si X est un n -échantillon de loi $\mathcal{N}(\mu, \sigma^2)$, alors $Y = \sigma^{-1}(X_1 - \mu, \dots, X_n - \mu)$ est un vecteur gaussien centré réduit et $X = \mu \mathbf{1} + \sigma Y$. Il s'en suit immédiatement que $\bar{X} \mathbf{1} \in F$ et \bar{X} suit une loi gaussienne de moyenne μ et de variance σ^2 . De plus, $X - \bar{X} \mathbf{1} = \sigma(Y - \bar{Y} \mathbf{1}) \in F^\perp$ si bien que

$$(n-1)V_n/\sigma^2 = \frac{1}{\sigma^2} \|X - \bar{X} \mathbf{1}\|^2 = \|Y - \bar{Y} \mathbf{1}\|^2 \sim \chi_{n-1}^2.$$

Enfin, V_n et \bar{X} sont indépendantes car $(I - P_F)X$ et $P_F X$ sont indépendantes et V_n ainsi que \bar{X} sont respectivement $(I - P_F)X$ et $P_F X$ mesurables. \square

Remarque 87. La loi de Student \mathcal{T}_n à n degrés de liberté est la loi de la variable $\frac{X}{\sqrt{Y/n}}$ où $X \sim \mathcal{N}(0, 1)$ est indépendante de $Y \sim \chi_n^2$. Par conséquent, le résultat précédent montre que

$$\sqrt{n} \frac{M_n - \mu}{\sqrt{V_n}} \sim \mathcal{T}_{n-1}.$$

Attention toutefois, nous sommes dans le contexte **gaussien** ! En particulier, moyenne empirique et variance empirique ne sont en général pas indépendantes.

Chapitre 11

Convergences de suites de variables aléatoires

11.1 Convergences trajectorielles

Soit $(\Omega, \mathcal{F}, \mathbf{P})$ un espace probabilisé et on considère $(X_n)_{n \geq 1}$ une suite de variables aléatoires à valeurs dans \mathbb{R}^d . Un certain nombre des résultats suivants s'adaptent sans difficulté majeure aux cas des variables aléatoires à valeurs dans un espace métrique, nous ne considérerons pas ce niveau de détails.

11.1.1 Convergence presque sûre ou presque partout

Définition 11.1.1. On dit que $(X_n)_{n \geq 1}$ converge vers X presque-sûrement si

$$\mathbf{P}(\lim_{n \rightarrow \infty} X_n = X) = \mathbf{P}(\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}) = 1.$$

Proposition 11.1.2. Une suite de v.a.r. $(X_n)_{n \geq 1}$ converge presque sûrement vers X si et seulement si

$$\forall \varepsilon > 0, \quad \mathbf{P}(\limsup \{|X_n - X| > \varepsilon\}) = 0.$$

Remarque 88. Attention, il s'agit de l'ensemble limite supérieure. Autrement dit, pour tout $\varepsilon > 0$, avec probabilité 1, $|X_n - X| \leq \varepsilon$ pour tout $n \geq 1$ sauf pour un nombre fini.

En probabilité on parle de convergence presque-sûre là où en analyse on parle de convergence presque-partout. Bien entendu, dans le dernier cas, la mesure considérée n'est pas nécessairement une probabilité mais les deux notions n'en restent pas moins identiques.

Démonstration. C'est une condition nécessaire. Fixons $\varepsilon > 0$. Si $\omega \in \limsup \{|X_n - X| > \varepsilon\}$, il existe une infinité de n pour lesquels $|X_n(\omega) - X(\omega)| > \varepsilon$ et la suite ne converge pas. Aussi, $\limsup \{|X_n - X| > \varepsilon\} \subset \{\lim_{n \rightarrow \infty} X_n = X\}^c$ et donc

$$\mathbf{P}(\limsup \{|X_n - X| > \varepsilon\}) \leq 1 - \mathbf{P}(\lim_{n \rightarrow \infty} X_n = X) = 0.$$

Pour la réciproque, considérons l'ensemble $N = \bigcup_{k \geq 0} \limsup \{|X_n - X| > 2^{-k}\}$. Il vient facilement que

$$\mathbf{P}(N) \leq \sum_{k \geq 1} \mathbf{P}(\limsup \{|X_n - X| > 2^{-k}\}) = 0.$$

Si $\omega \in N^c$, pour tout $k \geq 1$, $\omega \in \liminf \{|X_n - X| \leq 2^{-k}\}$ et par suite, il existe un entier $N = N(\omega)$ tel que pour tout $n \geq N$, $|X_n(\omega) - X| \leq 2^{-k}$. $X_n(\omega)$ converge donc vers $X(\omega)$ pour tout $\omega \in N^c$ qui est de mesure pleine. (Remarquons tout de même que la convergence n'est pas uniforme). \square

Corollaire 11.1.3. Si, pour tout $\varepsilon > 0$, $\sum_{n \geq 1} \mathbf{P}(|X_n - X| > \varepsilon) < \infty$ alors $(X_n)_{n \geq 1}$ converge presque sûrement vers X .

Démonstration. C'est une conséquence du lemme 8.2.1 de Borel-Cantelli et de la proposition 11.1.2. \square

Exemple 46. Rappelons que si X est une variable aléatoire positive alors

$$\mathbf{E}(X) < \infty \quad \text{ssi} \quad \sum_{n \geq 0} \mathbf{P}(X > n) < \infty.$$

Ce résultat a été vu en exercice lors du cours d'intégration, donnons-en une démonstration alternative.

Tout d'abord, $X = \int_0^\infty \mathbf{1}_{(t, \infty)}(X) dt$ (le vérifier pour les fonctions étagées positives) et en prenant l'espérance le théorème de Fubini implique

$$\mathbf{E}(X) = \mathbf{E} \left(\int_0^\infty \mathbf{1}_{(t, \infty)}(X) dt \right) = \int_0^\infty \mathbf{E}(\mathbf{1}_{(t, \infty)}(X)) dt = \int_0^\infty \mathbf{P}(X > t) dt.$$

D'autre part, comme $t \rightarrow \mathbf{P}(X > t)$ est décroissante,

$$\sum_{n \geq 0} \mathbf{P}(X > n+1) \leq \int_0^\infty \mathbf{P}(X > t) dt = \sum_{n \geq 0} \int_n^{n+1} \mathbf{P}(X > t) dt \leq \sum_{n \geq 0} \mathbf{P}(X > n). \quad (11.1)$$

Si $(X_n)_{n \geq 1}$ est une suite variables aléatoires réelles identiquement distribuées alors X_n/n converge vers 0 *p.s.* dès que $X_1 \in \mathbf{L}^1$. En effet, pour tout $\varepsilon > 0$

$$\sum_{n \geq 1} \mathbf{P}(|X_n| > n\varepsilon) = \sum_{n \geq 1} \mathbf{P}(|X_1| > \varepsilon n) = \sum_{n \geq 1} \mathbf{P}(\varepsilon^{-1}|X_1| > n)$$

qui est finie si $\mathbf{E}(\varepsilon^{-1}|X_1|)$ est finie ou $E|X_1| < \infty$. Si les X_n sont de plus supposées indépendantes alors le second lemme de Borel-Cantelli implique que $\mathbf{P}(\limsup\{|X_n| > n\varepsilon\}) = 0$ si et seulement si $\sum_{n \geq 1} \mathbf{P}(|X_n| > n\varepsilon) < \infty$.

Le lemme suivant est utile lorsque l'on veut montrer qu'une suite converge presque sûrement sans connaître la limite *a priori*.

Lemme 11.1.4. Soit $(\varepsilon_n)_{n \geq 1}$ une suite de réels positifs tels que $\sum_{n \geq 1} \varepsilon_n < \infty$. Supposons que

$$\sum_{n \geq 1} \mathbf{P}(|X_{n+1} - X_n| > \varepsilon_n) < \infty,$$

alors $(X_n)_{n \geq 1}$ converge presque sûrement.

Démonstration. On écrit X_n sous la forme $X_n = X_0 + \sum_{k=0}^{n-1} X_{k+1} - X_k$. Par le lemme de Borel-Cantelli 8.2.1, $\mathbf{P}(\limsup\{|X_{n+1} - X_n| > \varepsilon_n\}) = 0$. Or si $\omega \in \liminf\{|X_{n+1}(\omega) - X_n(\omega)| \leq \varepsilon_n\}$, il existe un entier $N = N(\omega)$ tel que pour tout $k \geq N$, $|X_{k+1}(\omega) - X_k(\omega)| \leq \varepsilon_k$. D'où la convergence de la série. \square

Proposition 11.1.5. Soit $(X_n)_{n \geq 0}$ est une suite de variables aléatoires à valeurs dans \mathbb{R}^d . Alors $(X_n)_{n \geq 0}$ converge vers $X \in \mathbb{R}^d$ presque sûrement si et seulement si pour tout $i \in \{1, \dots, d\}$, la suite de la i ème coordonnées $X_n^{(i)}$ converge presque sûrement vers la i ème coordonnées $X^{(i)}$ de X . Si $f : \mathbb{R}^d \rightarrow \mathbb{R}^p$ est une fonction continue, alors $(f(X_n))_{n \geq 0}$ converge presque sûrement vers $f(X)$.

Démonstration. Élémentaire. \square

11.1.2 Convergence dans \mathbf{L}^p

Définition 11.1.6. Une suite de variables aléatoires $(X_n)_{n \geq 0} \subset \mathbf{L}^p$ converge vers X en moyenne d'ordre p si

$$\lim_{n \rightarrow \infty} \mathbf{E}(|X_n - X|^p) = 0.$$

Autrement dit, c'est la convergence en norme dans \mathbf{L}^p .

Théorème 11.1.7. (Convergence dominée) Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires convergeant vers X *p.s.* et telle qu'il existe une variable aléatoire Y satisfaisant

$$\forall n \geq 0, \quad \|X_n\| \leq Y \quad \mathbf{P} - \text{p.s.}, \quad \text{avec} \quad Y \in \mathbf{L}^p.$$

Alors, $(X_n)_{n \geq 1} \subset \mathbf{L}^p$ et converge vers X dans \mathbf{L}^p .

Remarque 89. Notons qu'une suite X_n bornée convergeant presque-sûrement vers X converge également dans \mathbf{L}^p en choisissant $Y = c$ presque-sûrement pour un $c \geq 0$ convenable.

Proposition 11.1.8. *Soit $(X_n)_{n \geq 0} \subset \mathbf{L}^q$. Si $(X_n)_{n \geq 0}$ converge vers X dans \mathbf{L}^q alors $(X_n)_{n \geq 0}$ converge vers X dans \mathbf{L}^p pour tout $p \leq q$.*

Démonstration. Inégalité de Hölder □

Remarque 90. En fait, comme \mathbf{P} est une probabilité et donc que les constantes sont intégrables, on a que $\mathbf{L}^p \subset \mathbf{L}^q$ pour tout $p \leq q$.

Les cas les plus utiles de la convergence en moyenne sont $p = 1$ et $p = 2$.

Exemple 47. Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires réelles d'espérance μ_n et de variance σ_n^2 . On suppose que

$$\lim_{n \rightarrow \infty} |\mu_n| = \infty \quad \text{et} \quad \lim_{n \rightarrow \infty} \frac{\sigma_n^2}{|\mu_n|} = 0.$$

Alors, X_n/μ_n converge vers 1 dans \mathbf{L}^2 . En effet, par définition :

$$\mathbf{E}(|X_n/\mu_n - 1|^2) = \frac{1}{\mu_n^2} \mathbf{E}|X_n - \mu_n|^2 = \frac{\sigma_n^2}{\mu_n^2} \xrightarrow{n \rightarrow \infty} 0.$$

On remarque que par un calcul très similaire et en appliquant l'inégalité de Cauchy-Schwartz, on obtient la convergence dans \mathbf{L}^1 sous les mêmes hypothèses. Mais ce résultat découle bien entendu également de l'emboîtement des espaces \mathbf{L}^p .

11.1.3 Convergence en probabilité

Définition 11.1.9. Une suite $(X_n)_{n \geq 1}$ converge vers X en probabilité si

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \mathbf{P}(|X_n - X| > \varepsilon) = 0.$$

Remarque 91. Remarquons que le symbole \lim est à l'extérieur de la probabilité contrairement au cas de la convergence presque-sûre. C'est la raison pour laquelle la convergence en probabilité est plus faible (*c.f.* la proposition 11.1.11). Étant plus faible, il est souvent plus facile de montrer une convergence en probabilité qu'une convergence presque-sûre. Cependant, cette convergence apporte moins d'informations.

Proposition 11.1.10. *Soient $(X_n)_{n \geq 0}$ une suite de vecteurs aléatoires et $f : \mathbb{R}^d \rightarrow \mathbb{R}^p$ une fonction continue. Alors*

1. X_n converge vers X en probabilité si et seulement si $X_n^{(i)}$ converge vers $X^{(i)}$ pour tout $i \in \{1, \dots, d\}$;
2. si X_n converge vers X en probabilité alors $f(X_n)$ converge vers $f(X)$.

Démonstration. La preuve du point i) est élémentaire. Pour ii), ce n'est plus aussi immédiat que dans le cadre de la convergence presque-sûre. Puisque f est continue, elle est uniformément continue sur les compacts. Fixons $\varepsilon > 0$ et $a > 0$, alors il existe $\eta = \eta_{a, \varepsilon}$ tel que

$$|x| \leq a \quad \text{et} \quad |x - y| \leq \eta \quad \implies |f(x) - f(y)| \leq \varepsilon.$$

Aussi, $\{|X| \leq a\} \cap \{|X_n - X| \leq \eta\} \subset \{|f(X_n) - f(X)| \leq \varepsilon\}$ et il vient, en passant au complémentaire que

$$\mathbf{P}(|f(X_n) - f(X)| > \varepsilon) \leq \mathbf{P}(\{|X| > a\} \cup \{|X_n - X| > \eta\}) \leq \mathbf{P}(|X| > a) + \mathbf{P}(|X_n - X| > \eta)$$

qui conduit à

$$\limsup_{n \rightarrow \infty} \mathbf{P}(|f(X_n) - f(X)| > \varepsilon) = \mathbf{P}(|X| > a) + \limsup_{n \rightarrow \infty} \mathbf{P}(|X_n - X| > \eta).$$

Nous avons $\varepsilon > 0$, $a > 0$ et $\eta > 0$ et comme X_n converge en probabilité vers X , le second terme à droite est nul. D'un autre côté, lorsque a tend vers ∞ , $\mathbf{P}(|X| > a) \rightarrow 0$ et donc la limite supérieure à gauche de l'inégalité est en fait une limite laquelle est nulle. Ce qui montre le résultat. □

La proposition suivante implique que la convergence en probabilité est la plus faible des convergences trajectorielles.

Proposition 11.1.11. *Si $(X_n)_{n \geq 0}$ converge vers X presque-sûrement ou dans \mathbf{L}^p , $p \geq 1$, alors $(X_n)_{n \geq 0}$ converge vers X en probabilité.*

Démonstration. Supposons tout d'abord que X_n converge *p.s.* vers X alors

$$\mathbf{P}(|X_n - X| > \varepsilon) = \mathbf{E}(\mathbf{1}_{(\varepsilon, \infty)}(|X_n - X|)).$$

Or, $\mathbf{1}_{(\varepsilon, \infty)}(|X_n - X|)$ converge presque sûrement vers 0 et reste bornée donc par le théorème de convergence dominée, on obtient la convergence en probabilité. Si X_n converge vers X dans \mathbf{L}^p alors l'inégalité de Markov donne

$$\mathbf{P}(|X_n - X| > \varepsilon) = \mathbf{P}(|X_n - X|^p > \varepsilon^p) \leq \varepsilon^{-p} \mathbf{E}(|X_n - X|^p) \rightarrow 0.$$

□

Remarque 92. Si X_n converge en probabilité à la fois vers X et vers Y alors $X = Y$ presque sûrement. En effet, soit $\varepsilon > 0$

$$\mathbf{P}(|X - Y| > \varepsilon) \leq \mathbf{P}(|X - X_n| + |X_n - Y| > \varepsilon) \leq \mathbf{P}(|X_n - X| > \varepsilon/2) + \mathbf{P}(|X_n - Y| > \varepsilon/2).$$

Ainsi, sans préjuger de l'éventuelle convergence presque-sûre ou dans \mathbf{L}^p , si X_n converge en probabilité vers X , le bon candidat pour la limite presque-sûre ou dans \mathbf{L}^p est également X .

Proposition 11.1.12. *Soit $(X_n)_{n \geq 0}$ une suite de vecteurs aléatoires dans \mathbf{L}^2 telles que*

$$\lim_{n \rightarrow \infty} \mathbf{E}(X_n) = a \in \mathbb{R}^d \quad \text{et} \quad \lim_{n \rightarrow \infty} \mathbf{V}(X_n) = 0.$$

Alors X_n converge vers a en probabilité.

Démonstration. Soit $\varepsilon > 0$, par hypothèse, on peut trouver $N \geq 0$ tel que pour tout $n \geq N$, $|\mathbf{E}(X_n) - a| < \varepsilon/2$. On remarque alors, par l'inégalité triangulaire, que pour tout $n \geq 0$

$$\{|X_n - a| > \varepsilon\} \subset \{|X_n - \mathbf{E}(X_n)| > \varepsilon/2\}.$$

Appliquant l'inégalité de Bienaymé-Tchebychev

$$\mathbf{P}(|X_n - a| > \varepsilon) \leq \mathbf{P}(|X_n - \mathbf{E}(X_n)| > \varepsilon/2) \leq \frac{2\mathbf{V}(X_n)}{\varepsilon} \rightarrow 0.$$

□

La réciproque de la proposition 11.1.11 est généralement fausse. Toutefois, dans certaines circonstances, la convergence en probabilité implique la convergence presque-sûre et sous des conditions de domination la convergence dans \mathbf{L}^p . C'est l'objet des deux propositions suivantes.

Proposition 11.1.13. *Si une suite de variables aléatoires $(X_n)_{n \geq 0}$ converge en probabilité vers X , alors il existe une sous-suite $(X_{n_r})_{r \geq 0}$ qui converge presque-sûrement.*

Remarque 93. En particulier, toute valeur d'adhérence en probabilité d'une suite $(X_n)_{n \geq 0}$ est valeur d'adhérence presque-sûre.

Démonstration. Pour tout $r \geq 0$, $\lim_{n \rightarrow \infty} \mathbf{P}(|X_n - X| > 2^{-r-1}) = 0$. Il existe donc un entier n_r tel que

$$\forall n \geq n_r, \quad \mathbf{P}(|X_n - X| > 2^{-r-1}) \leq 2^{-r-1}.$$

On peut supposer la suite n_r strictement croissante puisque si n_r convient alors $n_r + 1$ convient également. Ainsi pour tout $r \geq 0$,

$$\mathbf{P}(|X_{n_{r+1}} - X_{n_r}| > 2^{-r}) \leq \mathbf{P}(|X_{n_{r+1}} - X| > 2^{-r-1}) + \mathbf{P}(|X_{n_r} - X| > 2^{-r-1}) \leq 2^{-r}.$$

Alors le lemme 11.1.4 implique la convergence presque-sûre de $(X_{n_r})_{r \geq 0}$. Notons Y la limite. La convergence presque-sûre implique la convergence en probabilité, donc du fait de la remarque 92, $X = Y$ presque-sûrement. □

Proposition 11.1.14. *Si X_n est une suite de variables aléatoires réelles décroissante et convergente vers 0 en probabilité, alors X_n converge presque-sûrement (vers 0).*

Démonstration. Puisque $(X_n)_{n \geq 0}$ converge en probabilité, par la proposition 11.1.13, il existe une sous-suite $n_\ell \uparrow \infty$ d'entiers telle que $(X_{n_\ell})_{\ell \geq 0}$ qui converge presque-sûrement et la limite est nécessairement 0. Pour $k \geq 0$, on introduit $N_k = \sup\{\ell \geq 0 : n_\ell \leq k\}$. Alors, l'hypothèse de décroissance implique

$$X_{n_{N_k-1}} \leq X_k \leq X_{n_{N_k}}.$$

Or, lorsque $k \rightarrow \infty$, $N_k \rightarrow \infty$ puisque $n_\ell \uparrow \infty$ et $X_k \rightarrow 0$ presque-sûrement. \square

11.1.4 Convergence trajectorielle et critère de type Cauchy

Rappelons qu'une suite $(x_n)_{n \geq 0}$ dans $(\mathbb{R}^d, |\cdot|)$ est dite de Cauchy si

$$\forall \varepsilon > 0, \quad \exists N \geq 0 : \quad n, m \geq N \quad \implies \quad |x_n - x_m| \leq \varepsilon.$$

Rappelons également que dans un espace métrique complet les suites de Cauchy sont exactement les suites convergentes. La proposition suivante donne un tel critère de type Cauchy dans le cadre de la convergence trajectorielle.

Proposition 11.1.15. *Soit $(X_n)_{n \geq 0}$ une suite variable aléatoire dans \mathbb{R}^d .*

1. $(X_n)_{n \geq 0}$ converge presque-sûrement si et seulement si

$$\forall \varepsilon > 0 : \quad \lim_{n \rightarrow \infty} \mathbf{P}(\sup_{r \geq 0} |X_{n+r} - X_n| > \varepsilon) = 0 ;$$

2. $(X_n)_{n \geq 0}$ converge en probabilité si et seulement si

$$\forall \varepsilon > 0 : \quad \lim_{n \rightarrow \infty} \sup_{r \geq 0} \mathbf{P}(|X_{n+r} - X_n| > \varepsilon) = 0 ;$$

3. $(X_n)_{n \geq 0}$ converge dans \mathbf{L}^p , $p \in [1, \infty)$ si et seulement si

$$\lim_{n \rightarrow \infty} \sup_{r \geq 0} \mathbf{E}(|X_{n+r} - X_n|^p) = 0.$$

Remarque 94. Le symbole sup est à l'intérieur de la probabilité pour la convergence presque-sûre là où il est à l'extérieur pour la convergence en probabilité. Cela illustre encore une fois le fait que la convergence en probabilité est plus faible que la convergence presque-sûre.

Remarque 95. Pour le dernier point, il s'agit en fait de la complétude des espaces \mathbf{L}^p , $p \in [1, \infty)$. L'espace \mathbf{L}^∞ est également complet, on rappelle que la norme sur \mathbf{L}^∞ est la norme du supremum essentiel :

$$\|X\|_\infty = \inf\{c > 0 : \mathbf{P}(|X| > c) = 0\}.$$

Démonstration. 1. Si $(X_n)_{n \geq 0}$ converge presque-sûrement, elle est donc presque-sûrement de Cauchy, c'est à dire que la variable aléatoire $\sup_{r \geq 0} |X_{n+r} - X_n|$ converge vers 0 presque-sûrement donc en probabilité. Aussi,

$$\forall \varepsilon > 0 : \quad \lim_{n \rightarrow \infty} \mathbf{P}(\sup_{n \geq r} |X_{n+r} - X_n| > \varepsilon) = 0.$$

D'où le résultat. Réciproquement, supposons cette dernière condition satisfaite et introduisons la variable aléatoire V_n définie par

$$V_n = \sup_{p \geq n, q \geq n} |X_p - X_q|.$$

La suite $(V_n)_{n \geq 0}$ est décroissante (pour tout $\omega \in \Omega$ soit dit en passant) et converge vers 0 en probabilité par hypothèse. Ainsi, elle converge vers 0 presque-sûrement par la proposition 11.1.14. Avec probabilité 1, la suite (X_n) est donc de Cauchy, elle converge presque-sûrement car \mathbb{R}^d est complet.

2. Si $(X_n)_{n \geq 0}$ converge vers X en probabilité, il vient, pour tout $r \in \mathbb{N}$, puisque $|X_{n+r} - X_n| \leq |X_{n+r} - X| + |X_n - X|$,

$$\mathbf{P}(|X_{n+r} - X_n| > \varepsilon) \leq \mathbf{P}(|X_{n+r} - X| > \varepsilon/2) + \mathbf{P}(|X_n - X| > \varepsilon/2) \leq 2 \sup_{k \geq n} \mathbf{P}(|X_k - X| > \varepsilon/2).$$

Le majorant, qui ne dépend plus de $r \geq 0$, à droite tend vers la plus grande d'adhérence de $\mathbf{P}(|X_k - X| > \varepsilon/2)$, c'est à dire 0 par hypothèse de convergence en probabilité. Réciproquement, si

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \sup_{r \geq 0} \mathbf{P}(|X_{n+r} - X_n| > \varepsilon) = 0,$$

on peut donc construire une suite strictement croissante d'entiers n_r telle que

$$\sup_{k \geq 0} \mathbf{P}(|X_{n_r+k} - X_{n_r}| > 2^{-r}) \leq 2^{-r}.$$

En particulier, $\mathbf{P}(|X_{n_r+1} - X_{n_r}| > 2^{-r}) \leq 2^{-r}$. D'après le lemme 11.1.4, la suite $(X_{n_r})_{r \geq 0}$ converge presque-sûrement donc en probabilité vers une certaine variable aléatoire X . On a alors

$$\begin{aligned} \mathbf{P}(|X_k - X| > \varepsilon) &\leq \mathbf{P}(|X_{n_k} - X_k| > \varepsilon/2) + \mathbf{P}(|X_{n_k} - X| > \varepsilon/2) \\ &\leq \sup_{r \geq 0} \mathbf{P}(|X_{k+r} - X_k| > \varepsilon/2) + \mathbf{P}(|X_{n_k} - X| > \varepsilon/2), \end{aligned}$$

car $k \leq n_k$. Lorsque $k \rightarrow \infty$, à droite de l'inégalité, le premier terme tend vers 0 par hypothèse et le second par la convergence en probabilité de $(X_{n_k})_{k \geq 0}$.

3. Pour le troisième point, il s'agit de la complétude des espaces \mathbf{L}^p . Il est clair que $(X_n)_{n \geq 0}$ converge en norme \mathbf{L}^p alors $(X_n)_{n \geq 0}$ est de Cauchy dans \mathbf{L}^p . Réciproquement, soit $(X_n)_{n \geq 0}$ une suite de Cauchy dans \mathbf{L}^p , $p \in [1, \infty]$. Afin d'avoir une notation unifiée, on travaille avec les normes plutôt qu'avec les espérances. Soit $(X_n)_{n \geq 0}$ une suite de Cauchy dans \mathbf{L}^p , alors on peut trouver une sous-suite n_k strictement croissante tel que pour tout $k \geq 0$

$$\|X_{n_{k+1}} - X_{n_k}\|_p \leq 2^{-k}.$$

On pose $Y = \sum_{k \geq 0} |X_{n_{k+1}} - X_{n_k}|$, c'est une fonction mesurable positive. De plus, on vérifie facilement, à l'aide du théorème de convergence de Beppo-Lévy et de l'inégalité de Minkowski, que

$$\|Y\|_p \leq \sum_{k \geq 0} \|X_{n_{k+1}} - X_{n_k}\|_p < \infty.$$

Ainsi, la série de variable aléatoire $\sum_{k \geq 0} (X_{n_{k+1}} - X_{n_k})_p$ converge absolument et donc presque-sûrement. La suite de variable aléatoire X_{n_k} converge également presque-sûrement vers un point d'adhérence presque-sûre de X_n , que l'on notera \bar{X} . De plus,

$$\|\bar{X} - X_{n_k}\|_p \leq \sum_{k \geq n} \|X_{n_{k+1}} - X_{n_k}\|_p \leq 2^{-n+1}.$$

Ainsi, \bar{X} est également une valeur d'adhérence de X_n dans \mathbf{L}^p . Finalement, on conclut en remarquant qu'une suite de Cauchy admet au plus une valeur d'adhérence □

Remarque 96. Dans le troisième point, on montre au passage qu'une suite convergente dans \mathbf{L}^p admet une sous-suite convergente presque-sûrement.

11.2 Convergence étroite et convergence en loi

Les différents modes de convergence de la section précédente concernaient les variables aléatoires : ce sont des modes de convergence de suites de fonctions, les variables aléatoires sont vues comme des fonctions. Dans cette partie, on s'intéresse à la convergence des lois de variables aléatoires.

11.2.1 Convergence étroite

Définition 11.2.1. Soit $(\mu_n)_{n \geq 0}$ une suite de probabilités sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. On dit que $(\mu_n)_{n \geq 0}$ converge étroitement vers μ si pour toute fonction continue bornée $f : \mathbb{R}^d \rightarrow \mathbb{R}$

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} f d\mu_n = \int_{\mathbb{R}^d} f d\mu.$$

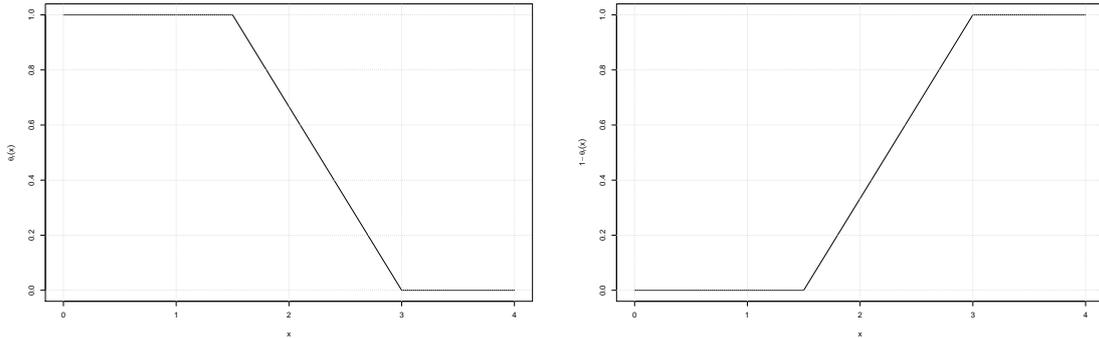
Remarque 97. La continuité des fonctions f est essentielle : si $(\mu_n)_{n \rightarrow \infty}$ converge étroitement vers μ , il est en général faux de dire que $\mu_n(B)$ converge vers $\mu(B)$ pour B un borélien quelconque. Par exemple, $\delta_{1/n}$ converge étroitement vers δ_0 et $\delta_{1/n}(\{0\}) = 0$ pour tout $n \geq 1$ tandis que $\delta_0(\{0\}) = 1$.

Remarque 98. Une suite de probabilité $(\mu_n)_{n \geq 1}$ converge étroitement vers μ , alors μ est une probabilité.

Théorème 11.2.2. Soit $\mathcal{H} \subset \mathcal{C}_b(\mathbb{R}^d)$ tel que $\mathcal{C}_c(\mathbb{R}^d) \subset \overline{\mathcal{H}}^{\|\cdot\|_\infty}$. Une suite $(\mu_n)_{n \geq 0}$ de probabilités converge étroitement vers μ si et seulement si

$$\forall f \in \mathcal{H} : \int_{\mathbb{R}^d} f(x) d\mu_n(x) = \int_{\mathbb{R}^d} f(x) d\mu(x).$$

Démonstration. Montrons tout d'abord le résultat dans le cas où \mathcal{H} est l'espace des fonctions à support compact. On utilise un argument de troncature. Considérons, pour $r > 0$, la fonction $\theta_r : \mathbb{R}_+ \rightarrow [0, 1]$ suivante : $\theta_r(x) = 1$ si $x \in [0, r]$, $\theta_r(x) = 0$ si $x \geq 2r$ et θ_r est affine sur $[r, 2r]$ — *c.f.* Figure 11.1.



(a) Graphe de θ_r .

(b) Graphe de $1 - \theta_r$.

FIGURE 11.1 – La fonction θ_r impliquée dans l'argument de troncature permet d'approcher une fonction continue bornée par une fonction continue à support compact.

Si f est une fonction continue bornée et ν une probabilité sur \mathbb{R}^d , on peut écrire

$$\int_{\mathbb{R}^d} f(x) \nu(dx) = \int_{\mathbb{R}^d} f(x) \theta_r(|x|) \nu(dx) + \int_{\mathbb{R}^d} f(x) [1 - \theta_r(|x|)] \nu(dx),$$

et

$$\left| \int_{\mathbb{R}^d} f(x) [1 - \theta_r(|x|)] \nu(dx) \right| \leq \|f\|_\infty \left(1 - \int_{\mathbb{R}^d} \theta_r(|x|) \nu(dx) \right).$$

Appliquée aux mesures μ_n et μ , on obtient l'inégalité

$$\begin{aligned} \left| \int_{\mathbb{R}^d} f(x) \mu_n(dx) - \int_{\mathbb{R}^d} f(x) \mu(dx) \right| &\leq \left| \int_{\mathbb{R}^d} f(x) \theta_r(|x|) \mu_n(dx) - \int_{\mathbb{R}^d} f(x) \theta_r(|x|) \mu(dx) \right| \\ &\quad + \|f\|_\infty \left(2 - \int_{\mathbb{R}^d} \theta_r(|x|) \mu_n(dx) - \int_{\mathbb{R}^d} \theta_r(|x|) \mu(dx) \right). \end{aligned} \quad (11.2)$$

On remarque que pour tout $r > 0$, les fonctions $x \rightarrow \theta_r(|x|)$ et $x \rightarrow f(x)\theta_r(|x|)$ sont continues à support compact si bien que, par hypothèse,

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} f(x)\theta_r(|x|) \mu_n(dx) = \int_{\mathbb{R}^d} f(x)\theta_r(|x|) \mu(dx)$$

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} \theta_r(|x|) \mu_n(dx) = \int_{\mathbb{R}^d} \theta_r(|x|) \mu(dx). \quad (11.3)$$

Par conséquent, pour tout $r > 0$,

$$\limsup_{n \rightarrow \infty} \left| \int_{\mathbb{R}^d} f(x) \mu_n(dx) - \int_{\mathbb{R}^d} f(x) \mu(dx) \right| \leq 2\|f\|_\infty \left(1 - \int_{\mathbb{R}^d} \theta_r(|x|) \mu(dx) \right).$$

Pour conclure, il suffit d'appliquer le théorème de convergence dominée : $\lim_{r \rightarrow \infty} \theta_r(|x|) = 1$ avec la domination $0 \leq \theta_r(|x|) \leq 1$, la fonction constante égale à 1 étant intégrable puisque μ est une probabilité. Ceci implique que

$$\lim_{r \rightarrow \infty} \int_{\mathbb{R}^d} \theta_r(|x|) \mu(dx) = \mu(\mathbb{R}^d) = 1.$$

On considère désormais le cas d'un $\mathcal{H} \subset C_b(\mathbb{R}^d)$ et tel que $C_c(\mathbb{R}^d) \subset \overline{\mathcal{H}}^{\|\cdot\|_\infty}$. Soit $f \in C_c(\mathbb{R}^d)$, alors pour tout $h \in \mathcal{H}$

$$\left| \int_{\mathbb{R}^d} f(x) \mu_n(dx) - \int_{\mathbb{R}^d} f(x) \mu(dx) \right| \leq \left| \int_{\mathbb{R}^d} h(x) \mu_n(dx) - \int_{\mathbb{R}^d} h(x) \mu(dx) \right| + 2\|f - h\|_\infty.$$

Ainsi, pour toute fonction $h \in \mathcal{H}$,

$$\limsup_{n \rightarrow \infty} \left| \int_{\mathbb{R}^d} f(x) \mu_n(dx) - \int_{\mathbb{R}^d} f(x) \mu(dx) \right| \leq 2\|f - h\|_\infty.$$

Cette inégalité est donc valide pour tout $h \in \mathcal{H}$ si bien que le membre de droite est majoré par $\inf_{h \in \mathcal{H}} \|f - h\|_\infty = d(f, \mathcal{H}) = d(f, \overline{\mathcal{H}}^{\|\cdot\|_\infty}) = 0$ puisque $C_c(\mathbb{R}^d) \subset \overline{\mathcal{H}}^{\|\cdot\|_\infty}$. \square

Théorème 11.2.3 (Portmanteau). *Soient $(\mu_n)_{n \geq 0}$ une suite de probabilités sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et μ une probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Les assertions suivantes sont équivalentes.*

1. La suite μ_n converge étroitement vers μ .
2. Pour tout fermé F , $\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F)$.
3. Pour tout ouvert G , $\mu(G) \leq \liminf_{n \rightarrow \infty} \mu_n(G)$.
4. Pour tout borélien tel que $\mu(\overline{B} \setminus \text{Int } B) = 0$, $\lim_{n \rightarrow \infty} \mu_n(B) = \mu(B)$.
5. Pour toute fonction bornée telle que $\mu(D_f) = 0$, où D_f est l'ensemble des points de discontinuité de f

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} f(x) d\mu_n(x) = \int_{\mathbb{R}^d} f(x) d\mu(x).$$

Démonstration. 1. Supposons que μ_n converge étroitement vers μ et donnons nous un fermé F de \mathbb{R}^d . La fonction $f_k(x) = (1 + d(x, F))^{-k}$ est continue et bornée pour tout $k \geq 1$. De plus, f_k converge en décroissant vers $\mathbf{1}_F$. On a donc pour tout $k \geq 1$

$$\limsup_{n \rightarrow \infty} \mu_n(F) = \limsup_{n \rightarrow \infty} \int_{\mathbb{R}^d} \lim_{k \rightarrow \infty} f_k(x) d\mu_n(x) \leq \limsup_{n \rightarrow \infty} \int_{\mathbb{R}^d} f_k(x) d\mu_n(x) = \int_{\mathbb{R}^d} f_k(x) d\mu(x).$$

Par convergence dominée, $\lim_{k \rightarrow \infty} \int_{\mathbb{R}^d} f_k(x) d\mu(x) = \mu(F)$ et donc $\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F)$.

2. Les points 2 et 3 sont équivalents par passage au complémentaire.
3. Les points 2 et 3 implique le point 4. En effet, $\text{Int } B \subset B \subset \overline{B}$ pour tout borélien B . Par conséquent,

$$\mu(\text{Int } B) \leq \liminf_{n \rightarrow \infty} \mu_n(\text{Int } B) \leq \liminf_{n \rightarrow \infty} \mu_n(B) \leq \limsup_{n \rightarrow \infty} \mu_n(B) \leq \limsup_{n \rightarrow \infty} \mu_n(\overline{B}) \leq \mu(\overline{B}).$$

4. La partie plus délicate consiste à montrer que 4 implique 5. Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction bornée. Par le lemme 7.2.16, l'ensemble $D_f = \{t \in \mathbb{R} : \mu(\{x \in \mathbb{R}^d : f(x) = t\}) > 0\}$ des points de discontinuités de la fonction de répartition de la variable aléatoire réelle f définie sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu)$ est au plus dénombrable. Son complémentaire $D = D_f^c$ est donc dense. Soit $c > 0$ tel que $\|f\|_\infty \leq c$. Soit $\varepsilon > 0$; il existe un nombre fini de points t_1, t_2, \dots, t_r de D tels que $t_1 < -c$, $t_r > c$ et $\max_{1 \leq i \leq r} |t_i - t_{i-1}| \leq \varepsilon$. Considérons la fonction

$$g(x) = \sum_{i=1}^{r-1} t_i \mathbf{1}_{[t_i, t_{i+1})}(f(x)).$$

Alors, pour tout $x \in \mathbb{R}^d$, $|f(x) - g(x)| \leq \max_{i \leq r} |t_i - t_{i-1}| \leq \varepsilon$, de sorte que

$$\left| \int_{\mathbb{R}^d} f(x) d\mu_n(x) - \int_{\mathbb{R}^d} f(x) d\mu(x) \right| \leq \left| \int_{\mathbb{R}^d} g(x) d\mu_n(x) - \int_{\mathbb{R}^d} g(x) d\mu(x) \right| + 2\varepsilon.$$

D'autre part, pour tout $n \geq 0$,

$$\int_{\mathbb{R}^d} g(x) d\mu_n(x) = \sum_{i=1}^{r-1} t_i \mu_n(f \in [t_i, t_{i+1})),$$

et de même pour l'intégrale de g contre μ . Il reste donc à montrer, pour conclure, que pour tout $1 \leq i \leq r-1$, la frontière de $B_i = \{f \in [t_i, t_{i+1})\}$ est de μ -mesure nulle. Remarquons pour cela que $\{x \in D_f^c : t_i < f(x) < t_{i+1}\}$ est l'image réciproque d'un ouvert par une fonction continue ($x \in D_f^c$) qui est contenu dans B_i donc dans $\text{Int } B$. De même, $\overline{B} \subset \{x \in D_f^c : t_i \leq f(x) \leq t_{i+1}\} \cup D_f$. Par conséquent,

$$\overline{B} \setminus \text{Int } B \subset \cup_{i=1}^r \{x \in \mathbb{R}^d : f(x) = t_i\} \cup D_f$$

et comme les t_i sont dans D , le résultat s'en suit.

5. le fait que le point 5 implique le point 1 est immédiat. □

Définition 11.2.4 (Tension). Une famille de probabilités $\mathcal{M} = (\mu_i)_{i \in I}$ sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est dite tendue si pour tout $\varepsilon > 0$ on peut trouver un compact $K \subset \mathbb{R}^d$ tel que, pour tout $i \in I$, $\mu_i(K^c) \leq \varepsilon$.

Exemple 48. Si la famille $\mathcal{M} = \{\mu\}$ ne contient qu'un élément alors \mathcal{M} est trivialement tendue. Il en va de même pour toute famille finie, ou pour toute réunion finie de famille tendue. Si K est un compact non vide de \mathbb{R}^d , on note $\mathcal{M}^1(K)$ l'espace des probabilité sur K . Cette famille de probabilité est trivialement tendue.

Remarque 99. Cette notion se généralise très facilement aux espaces topologiques. Un cas très important que l'on ne verra malheureusement pas dans ce cours est l'espace $\mathcal{C}^0([0, 1])$ qui intervient naturellement lorsqu'on étudie des processus stochastique continu (le mouvement brownien). La tension n'est alors rien d'autre que la compacité faible sur l'espace des probabilités sur $\mathcal{C}^0([0, 1])$.

Théorème 11.2.5 (Prokhorov). *Toute famille de probabilité \mathcal{M} sur \mathbb{R}^d tendue est relativement compact pour la topologie de la convergence étroite. Autrement dit, de toute suite $(\mu_n)_{n \geq 0} \subset \mathcal{M}$, on peut extraire une sous-suite $(\mu_{n_k})_{k \geq 0}$ qui converge étroitement.*

Remarque 100. Le théorème de Prokhorov énonce en fait qu'une famille tendue est compact (pour la topologie de la convergence étroite) si et seulement si elle est fermée. Il se trouve que $\mathcal{M}^1(K)$ est fermée (utiliser le théorème de Portmanteau avec l'ouvert $G = K^c$) et tendue, elle est donc compacte.

Ce théorème peut se montrer en appliquant le théorème de Helly sur les fonctions de répartition. Ce théorème, énoncé et montré ici en dimension 1, se généralise facilement en munissant \mathbb{R}^d d'un ordre partiel *ad-hoc* ce qui permet notamment de définir la notion de continuité à droite.

Théorème 11.2.6 (Helly). *De toute suite de fonction de répartition $(F_n)_{n \geq 0}$, on peut extraire une sous-suite $(F_{n_k})_{k \geq 0}$ telle qu'il existe une fonction F croissante continue à droite avec $F_{n_k}(x) \rightarrow F(x)$ en chaque point de continuité de F .*

Démonstration. À l'aide du procédé d'extraction diagonale, on commence par construire une suite (n_k) croissante telle que $F_{n_k}(x)$ converge en tout point x rationnel. On note $G(x)$ la limite obtenue. C'est une fonction croissante. On définit alors

$$F(x) = \inf\{G(r), r \in \mathbb{Q} \cap (x, \infty)\}.$$

Il est clair que F est croissante. Montrons que F est continue à droite. Soit $x \in \mathbb{R}$ et $\varepsilon > 0$. Par définition de F , il existe $r > x$, $r \in \mathbb{Q}$ tel que $G(r) < F(x) + \varepsilon$. Il vient que

$$\forall y \in [x, r) \quad F(x) \leq F(y) \leq G(r) < F(x) + \varepsilon,$$

ce qui montre la continuité à droite de F . Reste à montrer que F_{n_k} converge en chaque point de continuité de F . Soit un x un tel point et soit $\varepsilon > 0$. Prenons un $y < x$ tel que $F(x) - \varepsilon < F(y)$. Il existe des rationnels r et s vérifiant $y < r < x < s$, $F(y) \leq G(r)$ et $G(s) < F(x) + \varepsilon$. Aussi, on obtient en mettant tout bout à bout

$$F(x) - \varepsilon < G(r) \leq G(s) < F(x) + \varepsilon.$$

De même, pour tout $n \geq 0$, $F_n(r) \leq F_n(x) \leq F_n(s)$, et le long de la sous-suite (n_k) construite précédemment, il vient que

$$\begin{aligned} F(x) - \varepsilon < G(r) &= \lim_{k \rightarrow \infty} F_{n_k}(r) \leq \liminf_{k \rightarrow \infty} F_{n_k}(x) \\ &\leq \limsup_{k \rightarrow \infty} F_{n_k}(x) \leq \lim_{k \rightarrow \infty} F_{n_k}(s) = G(s) < F(x) + \varepsilon. \end{aligned}$$

D'où $\lim_{k \rightarrow \infty} F_{n_k}(x) = F(x)$ ce qui achève la preuve. □

On peut désormais montrer le théorème de Prokhorov

Démonstration. Soit $(\mu_n)_{n \geq 0}$ une suite de probabilité de \mathcal{M} supposée tendue. On note F_n la fonction de répartition de μ_n . D'après le théorème de Helly, on peut trouver une suite strictement croissante n_k et une fonction continue à droite croissante telle que $F_{n_k}(x)$ converge vers $F(x)$ pour tout point de continuité x de F . Une telle fonction F définit une mesure μ telle que pour tout réels $a < b$, $F(b) - F(a) = \mu((a, b])$. Il s'agit de montrer que μ est une mesure de probabilité.

Par construction (limite de fonctions de répartition), F prend ses valeurs dans $[0, 1]$ et donc $\mu(\mathbb{R}) = \lim_{n \rightarrow \infty} \mu((-n, n]) \leq 1$.

Pour l'inégalité inverse, on se donne $\varepsilon > 0$. Par l'hypothèse de tension, il existe K compact de \mathbb{R} tel que $\mu(K) \geq 1 - \varepsilon$ pour tout $\mu \in \mathcal{M}$. Posons $M = \sup\{|x| : x \in K\}$. Comme l'ensemble des points de discontinuité d'une fonction croissante est au plus dénombrable, il existe $a < -M$ et $b > M$ tels que F soit continue en a et en b . Ainsi pour tout $k \geq 1$

$$F_{n_k}(b) - F_{n_k}(a) = \mu_{n_k}((a, b]) \geq \mu_{n_k}(K) \geq 1 - \varepsilon.$$

Comme a, b sont des points de continuité de F , en faisant tendre k vers l'infini, on obtient que $F(b) - F(a) \geq 1 - \varepsilon$ et donc $\mu(\mathbb{R}) \geq \mu((a, b]) \geq 1 - \varepsilon$ pour $\varepsilon > 0$ qui peut être choisi arbitrairement petit. Donc $\mu(\mathbb{R}) = 1$. Comme $F_{n_k}(x)$ converge vers $F(x)$ en chaque point de continuité de F , on conclut que μ_{n_k} converge en étroitement vers μ . □

Remarque 101. Si par ailleurs, il n'y a qu'un seul point limite — ce qui est souvent pas trop difficile à vérifier — alors $(\mu_n)_{n \geq 1}$ converge étroitement vers μ , l'unique point limite.

Proposition 11.2.7. *Soit $(\mu_n)_{n \geq 1}$ une suite de probabilité gaussienne sur \mathbb{R}^d . Pour tout $n \geq 1$, on note respectivement $m_n \in \mathbb{R}^d$ et $\Sigma_n^2 \in \mathcal{M}_d(\mathbb{R})$ la moyenne et la matrice de covariance de μ_n . Alors, la famille $(\mu_n)_{n \geq 1}$ est tendue si et seulement si les familles $(m_n)_{n \geq 0}$ et $(\Sigma_n^2)_{n \geq 0}$ sont bornées dans \mathbb{R}^d et $\mathcal{M}_d(\mathbb{R})$ respectivement.*

Démonstration. Exercice. □

Nous avons vu la notion de fonction caractéristique d'une variable aléatoire. En réalité, ce n'est rien d'autre que la transformée de Fourier d'une mesure de probabilité. On note notamment

$$\hat{\mu} : \mathbb{R}^d \ni t \rightarrow \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} \mu(dx) \in \mathbb{C}.$$

Les preuves des deux théorèmes suivants sont assez longues et ne seront pas présenter ici.

Théorème 11.2.8 (Paul Lévy). *Soit $(\mu_n)_{n \geq 1}$ une suite de probabilités sur \mathbb{R}^d . Si la suite de fonction $(\hat{\mu}_n)_{n \geq 1}$ converge simplement vers une fonction ϕ continue en 0, alors il existe une probabilité μ sur \mathbb{R}^d telle que $\phi = \hat{\mu}$ et $(\mu_n)_{n \geq 0}$ converge étroitement vers μ .*

Corollaire 11.2.9. *Une suite de probabilité $(\mu_n)_{n \geq 0}$ converge étroitement vers μ si et seulement si $(\hat{\mu}_n)_{n \geq 1}$ converge simplement vers $\hat{\mu}$.*

11.2.2 Convergence en loi

Soient $(X_n)_{n \geq 1}$ et X des variables aléatoires dans \mathbb{R}^d .

Définition 11.2.10. La suite $(X_n)_{n \geq 1}$ converge vers X en loi si la suite de probabilités $(\mathbf{P}_{X_n})_{n \geq 1}$ converge étroitement vers \mathbf{P}_X .

De manière équivalente, $(X_n)_{n \geq 1}$ converge en loi vers X si et seulement si pour toute fonction continue bornée

$$\lim_{n \rightarrow \infty} \mathbf{E}(f(X_n)) = \mathbf{E}(f(X)),$$

ou encore que

$$\forall t \in \mathbb{R}^d, \quad \lim_{n \rightarrow \infty} \phi_{X_n}(t) = \mathbf{E}(e^{i\langle t, X_n \rangle}) \rightarrow \phi_X(t) = \mathbf{E}(e^{i\langle t, X \rangle}).$$

Notons que la convergence en loi concerne bien la loi de X_n et ne dit rien en général sur le comportement de $X_n(\omega)$, $\omega \in \Omega$. Du reste, il n'est pas nécessaire que tous les X_n partagent le même espace de probabilité.

Proposition 11.2.11. *Soient $(X_n)_{n \geq 1}$ et X des v.a.r.. La suite $(X_n)_{n \geq 1}$ converge en loi vers X si et seulement si, pour tout $t \in \mathbb{R}$ où F_X est continue, $\lim_{n \rightarrow \infty} F_{X_n}(t) = F_X(t)$.*

Remarque 102. En fait l'existence d'une fonction F telle que la condition soit vérifiée implique l'existence d'une variable aléatoire X telle que X_n converge en loi vers X .

Démonstration. C'est une conséquence de la remarque précédente et du théorème de Portmanteau. \square

Exemple 49. Soit $(U_n)_{n \geq 0}$ une suite v.a. i.i.d. de loi uniforme sur $[0, \theta]$, $\theta > 0$. On pose pour $n \geq 1$, $X_n = \max\{U_i : 1 \leq i \leq n\}$. D'une part, on montre que X_n converge en probabilité vers θ . En effet, soit $\varepsilon > 0$,

$$\mathbf{P}(|X_n - \theta| > \varepsilon) = \mathbf{P}(X_n < \theta - \varepsilon) = \left(\frac{\theta - \varepsilon}{\theta}\right)^n \rightarrow 0,$$

lorsque $n \rightarrow \infty$.

Autrement dit, X_n est un estimateur consistant¹ de $\theta > 0$. On peut s'intéresser à la vitesse de convergence de cet estimateur. Pour cela, on doit établir une convergence en loi. Soit $Z_n = n(\theta - X_n)$ pour tout $n \geq 0$. On calcule la fonction de répartition de Z_n . Soit $t \in \mathbb{R}$

$$\begin{aligned} F_{Z_n}(t) &= \mathbf{P}(Z_n \leq t) = \mathbf{P}(X_n \geq \theta - t/n) \\ &= 1 - \mathbf{P}(X_n < \theta - t/n) \\ &= 1 - \mathbf{1}_{[0, n\theta]}(t) \left(1 - \frac{t}{n\theta}\right)^n. \end{aligned}$$

Lorsque $n \rightarrow \infty$, $F_{Z_n}(t)$ converge simplement vers $F(t) = 1 - \mathbf{1}_{[0, \infty]}(t)e^{-t/\theta}$ (sauf peut-être en 0). On reconnaît la fonction de répartition d'une loi exponentielle de paramètre $1/\theta$.

1. Ces notions seront abordées de manière plus approfondis dans le cours de Statistiques inférentielles. Un estimateur n'est ni plus ni moins qu'une fonction mesurable d'un échantillon, noté ici (U_1, \dots, U_n) . Il est dit faiblement consistant car il converge en probabilité vers θ . Il est même fortement consistant, c'est à dire que la convergence est presque sûre, car (X_n) est croissante.

Proposition 11.2.12. Si $(X_n)_{n \geq 1}$ converge en loi vers X et si $g : \mathbb{R}^d \rightarrow \mathbb{R}^q$ est continue alors $(g(X_n))_{n \geq 1}$ converge en loi vers $g(X)$.

Démonstration. Trivial. □

Remarque 103. En particulier, si $((X_n, Y_n))_{n \geq 0}$ converge en loi vers (X, Y) alors $(X_n)_{n \geq 0}$ converge en loi vers X et $(Y_n)_{n \geq 0}$ converge en loi vers Y ; de même que $(X_n + Y_n)_{n \geq 0}$ et $(X_n Y_n)_{n \geq 0}$ converge en loi vers $X + Y$ et XY respectivement.

La proposition suivante montre que la convergence en loi est le mode de convergence le plus faible parmi ceux évoqués jusqu'ici.

Proposition 11.2.13. Si $(X_n)_{n \geq 1}$ converge en probabilité vers X alors $(X_n)_{n \geq 1}$ converge en loi vers X .

Démonstration. Signalons que toutes les variables aléatoires sont définies sur le même espace de probabilité et montrons que ϕ_{X_n} converge simplement vers ϕ_X . Soit $t \in \mathbb{R}^d$ alors

$$|\phi_{X_n}(t) - \phi_X(t)| \leq \mathbf{E}|e^{i\langle t, X_n \rangle} - e^{i\langle t, X \rangle}| \leq \mathbf{E}(\min(2, |t||X_n - X|)),$$

de sorte que, pour tout $\varepsilon > 0$ en écrivant $1 = \mathbf{1}_{[0, \varepsilon]}(|X_n - X|) + \mathbf{1}_{(\varepsilon, \infty)}(|X_n - X|)$,

$$|\phi_{X_n}(t) - \phi_X(t)| \leq \varepsilon |t| \mathbf{P}(|X_n - X| \leq \varepsilon) + 2\mathbf{P}(|X_n - X| > \varepsilon) \leq \varepsilon |t| + 2\mathbf{P}(|X_n - X| > \varepsilon).$$

Par conséquent, pour tout $\varepsilon > 0$, $\limsup_{n \rightarrow \infty} |\phi_{X_n}(t) - \phi_X(t)| \leq \varepsilon |t|$ d'où le résultat. □

Il y a en réalité une réciproque partielle à ce résultat.

Proposition 11.2.14. On suppose les variables $(X_n)_{n \geq 1}$ définies sur le même espace probabilisé. Soit $c \in \mathbb{R}^d$. Si $(X_n)_{n \geq 1}$ converge vers c en loi, alors la convergence a aussi lieu en probabilité.

Remarque 104. Il est ici nécessaire de supposer les X_n sur un même espace de probabilité de sorte que la convergence en probabilité fasse sens.

Démonstration. On se ramène au cas réel en considérant les composantes de X_n . Soit $\varepsilon > 0$.

$$\mathbf{P}(|X_n - c| > \varepsilon) = \mathbf{P}(X_n < c - \varepsilon) + \mathbf{P}(X_n > c + \varepsilon) = F_{X_n}(c - \varepsilon) + 1 - F_{X_n}(c + \varepsilon).$$

Puisque X_n converge en loi vers c , d'après la proposition 11.2.11, pour tout $t \neq c$, $F_{X_n}(t)$ converge vers $\mathbf{1}_{\mathbb{R}_+}(t - c)$. Ceci montre que $\mathbf{P}(|X_n - c| > \varepsilon)$ tend vers 0 pour tout $\varepsilon > 0$. □

La convergence en loi des marginales ne permet pas en général de conclure à la convergence en loi du vecteur. Encore ici, il existe une réponse partielle.

Lemme 11.2.15 (Lemme de Slutsky). Soit $((X_n, Y_n))_{n \geq 0}$ une suite de vecteurs aléatoire définis sur un même espace de probabilité. Si $(X_n)_{n \geq 1}$ converge en loi vers X et que $(Y_n)_{n \geq 1}$ converge en probabilité vers c , alors $((X_n, Y_n))_{n \geq 1}$ converge en loi vers (X, c) .

Exemple 50. En anticipant légèrement sur le chapitre 12, considérons une suite $(X_n)_{n \geq 1}$ de variables aléatoires réelles *i.i.d.* admettant un moment d'ordre 2 et notons

$$M_n = \frac{1}{n} \sum_{k=1}^n X_k \quad \text{et} \quad V_n = \frac{1}{n-1} \sum_{k=1}^n (X_k - M_n)^2 = \frac{1}{n-1} \sum_{k=1}^n X_k^2 - \frac{n}{n-1} M_n^2.$$

À la fin du chapitre 10, le corollaire 10.2.5 établit que si (X_1, \dots, X_n) est un **vecteur gaussien**, alors M_n et V_n sont indépendantes de loi respectives $\mathcal{N}(0, 1/n)$ et $\chi^2(n-1)$ si bien que

$$\sqrt{n} \frac{M_n - \mathbf{E}(X_1)}{\sqrt{V_n}} \sim \mathcal{T}_{n-1}.$$

Sous l'hypothèse plus faible où l'échantillon n'est plus gaussien, nous avons que $\sqrt{n} \frac{M_n - \mathbf{E}(X_1)}{\sqrt{V_n}}$ **converge en loi** vers une $\mathcal{N}(0, 1)$. Nous avons en effet par la loi des grands nombre M_n converge en probabilité vers $\mathbf{E}(X_1)$, de même que V_n converge presque-sûrement vers $\mathbf{V}(X_1)$. Nous pouvons donc écrire

$$\sqrt{n} \frac{M_n - \mathbf{E}(X_1)}{\sqrt{V_n}} = \sqrt{n} \frac{M_n - \mathbf{E}(X_1)}{\sqrt{\mathbf{V}(X_1)}} \sqrt{\frac{\mathbf{V}(X_1)}{V_n}} = f \left(\sqrt{n} \frac{M_n - \mathbf{E}(X_1)}{\sqrt{\mathbf{V}(X_1)}}, \frac{\mathbf{V}(X_1)}{V_n} \right)$$

où $f(x, y) = x\sqrt{y}$. Par le théorème central limite que $\sqrt{n} \frac{M_n - \mathbf{E}(X_1)}{\sqrt{\mathbf{V}(X_1)}}$ converge en loi vers une $\mathcal{N}(0, 1)$ et le rapport $\frac{\mathbf{V}(X_1)}{V_n}$ converge vers 1 presque-sûrement donc en probabilité. Le lemme de Slutsky permet de déduire la convergence en loi du couple, on conclut en remarquant que f est continue.

Ce résultat permet d'écrire un intervalle de confiance pour $\mathbf{E}(X_1)$ lorsque la variance n'est pas connue, ce qui est pratiquement toujours le cas. La contrainte étant que le résultat est asymptotique donc l'échantillon doit être suffisamment grand.

Pour établir un tel intervalle de confiance, on se fixe *a priori* un niveau confiance $\alpha \in [0, 1]$. En général, on peut prendre $\alpha = 0.95$. Puis on cherche $t \geq 0$ de sorte que pour n assez grand :

$$\begin{aligned} \alpha = \mathbf{P} \left(-t \leq \sqrt{n} \frac{M_n - \mathbf{E}(X_1)}{\sqrt{V_n}} \leq t \right) &= \mathbf{P} \left(M_n - t\sqrt{\frac{V_n}{n}} \leq \mathbf{E}(X_1) \leq M_n + t\sqrt{\frac{V_n}{n}} \right) \\ &\approx \frac{1}{\sqrt{2\pi}} \int_{-t}^t \exp(-x^2/2) dx. \end{aligned}$$

On trouve alors $t \approx 1.96$. Aussi, avec une probabilité supérieur à $\alpha = 0.95$, la valeur de la vraie moyenne $\mathbf{E}(X_1)$ se trouve dans l'intervalle de confiance $\text{IC}_{0.95} = [M_n - 1.96\sqrt{V_n/n}; M_n + 1.96\sqrt{V_n/n}]$.

Remarquons enfin que lorsque n est grand, la loi de Student à $n - 1$ degrés de libertés se rapproche d'une gaussienne. Ainsi, lorsque la variance est inconnue, ou bien nous pouvons faire une hypothèse gaussienne et on obtient un intervalle de confiance à l'aide de la loi de Student, ou bien l'échantillon est suffisamment grand et cette fois-ci l'intervalle de confiance est obtenu à l'aide de la loi normale.

Exemple 51. Outre l'exemple traité ci-dessus, le lemme de Slutsky intervient régulièrement en statistiques par exemple dans la Δ -méthode.

Pour fixer les idées, considérons une suite $(X_n)_{n \geq 1}$ de variables aléatoires *i.i.d.* toute de loi exponentielle $\mathcal{E}(\lambda)$, $\lambda > 0$. La loi (faible) des grands nombres donne que

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{n \rightarrow \infty} \mathbf{E}(X_1) = 1/\lambda,$$

la convergence ayant lieu en probabilité. Ainsi, nous disposons non pas d'un estimateur de λ mais de $1/\lambda$. Notons $\theta = 1/\lambda$ et posons $f(\theta) = 1/\theta = \lambda$. Alors c'est un exercice de montrer que $f(\bar{X}_n) = 1/\bar{X}_n \rightarrow \lambda$ en probabilité. On cherche alors à établir un théorème central limite afin d'établir un intervalle de confiance par exemple. Pour cela, on calcule un développement de Taylor à l'ordre 2 de f en θ :

$$f(\bar{X}_n) = f(\theta) + f'(\theta)(\bar{X}_n - \theta) + \frac{1}{2}f''(\theta)(\bar{X}_n - \theta)^2 + o((\bar{X}_n - \theta)^2).$$

Ainsi,

$$\sqrt{n}(f(\bar{X}_n) - f(\theta)) = f'(\theta)\sqrt{n}(\bar{X}_n - \theta) + \frac{1}{2}f''(\theta)\sqrt{n}(\bar{X}_n - \theta)^2 + o(\sqrt{n}(\bar{X}_n - \theta)^2). \quad (11.4)$$

Une première application du lemme de Slutsky, du théorème central limite et de la loi des grands nombres implique $\sqrt{n}(\bar{X}_n - \theta)^2$ converge vers 0 en probabilité. En effet, on écrit

$$\sqrt{n}(\bar{X}_n - \theta)^2 = \underbrace{\sqrt{n}(\bar{X}_n - \theta)}_{(I)} \underbrace{(\bar{X}_n - \theta)}_{(II)}.$$

Alors le facteur (I) converge en loi vers $\mathcal{N}(0, \mathbf{V}(X_1))$ par le théorème central limite ; le facteur (II) quant à lui converge vers 0 en probabilité par la loi faible des grands nombres. Ainsi, le produit converge vers 0 en probabilité par le lemme de Slutsky.

Ceci implique en particulier que les deux derniers termes de l'équation (11.4) convergent vers 0 en probabilité. Le premier terme de (11.4) lui converge en loi vers $\mathcal{N}(0, [f'(\theta)]^2 \mathbf{V}(X_1))$. A l'aide du lemme de Slutsky, on obtient que

$$\sqrt{n}(f(\bar{X}_n) - f(\theta)) = \sqrt{n} \left(\frac{1}{\bar{X}_n} - \lambda \right) \text{ converge en loi vers } \mathcal{N}(0, [f'(\theta)]^2 \mathbf{V}(X_1)).$$

La loi asymptotique dépend encore du paramètre λ et par conséquent il faudrait encore arranger un peu les choses à la manière de l'exemple précédent. Ceci devrait convaincre du caractère "boîte à outils" du lemme de Slutsky.

Terminons par remarquer que les hypothèses sur f sont relativement faibles : f doit être C^2 au voisinage de θ et $f'(\theta) \neq 0$. Par ailleurs, on établira au chapitre 12 un théorème central limite multivarié. Cette méthodologie s'étend alors très facilement au cas multivarié.

Démonstration. Notons que la fonction caractéristique de (X, c) est $\phi_{(X,c)}(s, t) = \phi_X(s) e^{i(t,c)}$. On a alors

$$\begin{aligned} \left| \phi_{(X_n, Y_n)}(s, t) - \phi_X(s) e^{i(t,c)} \right| &= \left| \mathbf{E}(e^{i\langle s, X_n \rangle} (e^{i\langle t, Y_n \rangle} - e^{i\langle t, c \rangle}) + e^{i\langle t, c \rangle} (\phi_{X_n}(s) - \phi_X(s))) \right| \\ &\leq \mathbf{E}|e^{i\langle t, Y_n \rangle} - e^{i\langle t, c \rangle}| + |\phi_{X_n}(s) - \phi_X(s)|. \end{aligned}$$

Comme dans la preuve de la proposition précédentes, on a pour tout $\varepsilon > 0$

$$\left| \phi_{(X_n, Y_n)}(s, t) - \phi_X(s) e^{i(t,c)} \right| \leq \varepsilon |t| + 2\mathbf{P}(|Y_n - c| > \varepsilon) + |\phi_{X_n}(s) - \phi_X(s)|.$$

Et il suffit de prendre la limite supérieure pour conclure. \square

Corollaire 11.2.16. *Soit $(X_n)_{n \geq 0}$ et $(Y_n)_{n \geq 0}$ deux suites de variables aléatoires à valeurs dans \mathbb{R}^d définies sur le même espace probabilisé. On suppose que $(X_n)_{n \geq 0}$ converge en loi vers X et que $|X_n - Y_n|$ converge vers 0 en probabilité. Alors $(Y_n)_{n \geq 0}$ converge en loi vers X .*

Démonstration. C'est un corollaire immédiat du lemme de Slutsky en posant $Y_n = X_n - (X_n - Y_n)$ et l'application continue $g(x, d) = x + d$. \square

On termine cette partie par un analogue du lemme de Fatou qui permet de donner une condition intégrabilité de la limite en loi.

Proposition 11.2.17. *Si $(X_n)_{n \geq 0}$ convergent en loi vers X , alors*

$$\mathbf{E}|X| \leq \liminf_{n \rightarrow \infty} \mathbf{E}|X_n|.$$

Démonstration. Puisque X_n converge en loi vers X , $|X_n|$ converge en loi vers $|X|$ par continuité de $|\cdot|$. Soit $k \geq 1$ un entier,

$$\mathbf{E}(|X| \wedge k) = \lim_{n \rightarrow \infty} \mathbf{E}(|X_n| \wedge k) = \liminf_{n \rightarrow \infty} \mathbf{E}(|X_n| \wedge k) \leq \liminf_{n \rightarrow \infty} \mathbf{E}|X_n|.$$

On conclut par convergence monotone. \square

Remarque 105. On peut bien entendu on peut remplacer la valeur absolue par n'importe quelle fonction continue positive.

11.3 Loi du 0-1 de Kolmogorov et séries aléatoires

On s'intéresse dans cette partie à la convergence (dans \mathbb{R}) des séries aléatoires réelles indépendantes (non nécessairement de même loi). Ces résultats seront utiles pour démontrer la loi des grands nombres dite forte. La convergence a alors lieu presque-sûrement et non plus en probabilité comme pour la loi faible. On commence par énoncer la loi du 0-1 de Kolmogorov. Pour ce faire, il est nécessaire d'introduire la notion de tribu asymptotique.

Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires. On note $\mathcal{A}_n = \sigma(X_n, X_{n+1}, \dots)$ la tribu engendrée par les variables X_m pour tout $m \geq n$. La tribu asymptotique, notée \mathcal{A}_∞ , est définie comme l'intersection des tribus \mathcal{A}_n : $\mathcal{A}_\infty = \bigcap_{n \geq 0} \mathcal{A}_n$. Intuitivement, un événement $A \in \mathcal{A}_\infty$ si il dépend du comportement asymptotique de $(X_n)_{n \geq 0}$, ou encore si l'occurrence de A ne dépend pas de la valeur prise par un nombre fini de X_n . Par exemple, si $S_n = X_1 + \dots + X_n$,

1. $\{\lim_{n \rightarrow \infty} S_n \text{ existe}\}$ est un événement asymptotique : la modification de la valeur de X_k pour un nombre fini de k ne modifie pas la nature convergente ou divergente de la série ;
2. $\{\limsup_{n \rightarrow \infty} S_n \geq 0\}$ n'est pas un événement asymptotique car il dépend de toute les variables X_n .
3. soit $(B_n)_{n \geq 0}$ une suite événements, i.e. $B_n \in \mathcal{F}$ pour tout $n \geq 0$. Alors les événements

$$\limsup\{X_n \in B_n\} \quad \text{et} \quad \liminf\{X_n \in B_n\}$$

sont des événements asymptotiques.

Théorème 11.3.1 (Loi du 0-1 de Kolmogorov). *Soient $(X_n)_{n \geq 0}$ une suite variables aléatoires indépendantes et $A \in \mathcal{A}_\infty$ un événement asymptotique. Alors $\mathbf{P}(A) \in \{0, 1\}$.*

Démonstration. On va montrer que A est indépendant de lui-même, i.e. $\mathbf{P}(A \cap A) = \mathbf{P}(A)\mathbf{P}(A)$, d'où l'on déduit $\mathbf{P}(A) \in \{0, 1\}$.

Pour cela, notons que $\mathcal{A}_\infty \subset \sigma(X_{k+n}, k \geq 0)$ pour tout $n \geq 0$ et que $\cup_{k \geq 0} \sigma(X_0, \dots, X_k)$ et \mathcal{A}_∞ sont des π -systèmes contenant Ω qui engendrent respectivement $\sigma(X_n, n \geq 0)$ et \mathcal{A}_∞ . Par le lemme 8.1.3, pour vérifier l'indépendance de ces deux tribus, il suffit de vérifier l'indépendance sur les π -systèmes les engendrant. Or, il est clair par la proposition 8.1.2 que pour tout $n \geq 0$, la tribu $\sigma(X_0, \dots, X_n)$ et la tribu $\sigma(X_{n+k+1}, k \geq 0)$ sont indépendantes.

Ainsi, les tribus $\sigma(X_n, n \geq 0)$ et \mathcal{A}_∞ sont indépendantes. Or, si $A \in \mathcal{A}_\infty$ alors $A \in \sigma(X_n, n \geq 0)$ et par conséquent l'événement A appartient à deux tribus indépendantes, il est indépendant de lui-même. \square

Le lemme suivant fait partie du folklore probabiliste et donne une information sur les fluctuations de somme de variables indépendantes.

Lemme 11.3.2 (Inégalité de Lévy-Ottoviani). *Soient ξ_1, \dots, ξ_p des variables aléatoires indépendantes. On note, pour $r = 1, \dots, p$, $Z_r = \sum_{1 \leq i \leq r} \xi_i$. Alors pour tout $\eta > 0$ et $\delta \geq 0$,*

$$\inf_{1 \leq r < p} \mathbf{P}(|Z_p - Z_r| \leq \delta) \times \mathbf{P}(\sup_{1 \leq r \leq p} |Z_r| > \eta + \delta) \leq \mathbf{P}(|Z_p| > \eta).$$

Démonstration. Notons $\tau = \inf\{i = 1, \dots, p : |Z_i| > \eta + \delta\}$ avec la convention $\inf \emptyset = \infty$. On cherche donc à majorer la probabilité de l'événement $\{\sup_{1 \leq r \leq p} |Z_r| > \eta + \delta\} = \{\tau \leq p\}$ par celle de l'événement $\{|Z_p| > \eta\}$. On remarque que $\{\tau = 1\} = \{|Z_1| > \eta + \delta\}$ et, pour $1 < r \leq p$,

$$\{\tau = r\} = \{|Z_1| \leq \eta + \delta\} \cap \dots \cap \{|Z_{r-1}| \leq \eta + \delta\} \cap \{|Z_r| > \eta + \delta\}.$$

Ceci implique que $\{\tau = p\} \subset \{|Z_p| > \eta + \delta\} \subset \{|Z_p| > \eta\}$ et, pour $r = 1, \dots, p-1$,

$$\{\tau = r\} \cap \{|Z_p - Z_r| \leq \delta\} \subset \{|Z_r| > \eta + \delta\} \cap \{|Z_p - Z_r| \leq \delta\} \subset \{|Z_p| > \eta\},$$

car $|Z_r| \leq |Z_p| + |Z_r - Z_p|$. Il s'en suit que

$$\mathbf{P}(|Z_p| > \eta) \geq \mathbf{P}(\tau = p) + \sum_{r=1}^{p-1} \mathbf{P}(\tau = r, |Z_p - Z_r| \leq \delta).$$

Les événements $\{\tau = r\}$ et $\{|Z_p - Z_r| \leq \delta\}$ sont indépendants car $\{\tau = r\}$ est dans la tribu $\sigma(\xi_1, \dots, \xi_r)$ et $\{|Z_p - Z_r| \leq \delta\}$ est dans la tribu $\sigma(\xi_{r+1}, \dots, \xi_p)$. D'où, l'inégalité

$$\mathbf{P}(|Z_p| > \eta) \geq \mathbf{P}(\tau = p) + \sum_{r=1}^{p-1} \mathbf{P}(\tau = r) \mathbf{P}(|Z_p - Z_r| \leq \delta) \geq \alpha \sum_{r=1}^p \mathbf{P}(\tau = r),$$

où $\alpha = \inf_{1 \leq r < p} \mathbf{P}(|Z_p - Z_r| \leq \delta)$. Il suffit alors de remarquer que $\mathbf{P}(\tau \leq p) = \sum_{r=1}^p \mathbf{P}(\tau = r)$ pour conclure. \square

Théorème 11.3.3 (Paul Lévy). *Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. indépendantes. Pour $n \geq 1$, on note $S_n = X_1 + \dots + X_n$. Les assertions suivantes sont équivalentes*

1. $(S_n)_{n \geq 1}$ converge presque sûrement vers une variable aléatoire réelle ;
2. $(S_n)_{n \geq 1}$ converge en probabilité vers une variable aléatoire réelle ;
3. $(S_n)_{n \geq 1}$ converge en loi vers une variable aléatoire réelle.

Démonstration. Montrons tout d'abord que si $(S_n)_{n \geq 0}$ converge en probabilité alors elle converge presque sûrement. Pour ce faire, on utilise le critère de type Cauchy du théorème 11.1.15. Soit $\varepsilon > 0$ alors, par monotonie,

$$\mathbf{P}(\sup_{r \geq 0} |S_{n+r} - S_n| > \varepsilon) = \lim_{p \rightarrow \infty} \mathbf{P}(\sup_{0 \leq r \leq p} |S_{n+r} - S_n| > \varepsilon).$$

On cherche à appliquer l'inégalité de Lévy-Ottoviani. On remarque que

$$\forall 1 \leq r \leq p, \quad S_{n+r} - S_n = \sum_{j=n+1}^{n+r} X_j = \sum_{i=1}^r X_{n+i} = \sum_{i=1}^r \xi_i,$$

avec $\xi = X_{i+n}$. Avec ces notations, $S_{n+r} - S_n$ joue le rôle de Z_r de l'inégalité de Lévy-Ottoviani qu'on peut appliquer au couple $(\eta, \delta) = (\varepsilon/2, \varepsilon/2)$. On obtient alors

$$\inf_{1 \leq r < p} \mathbf{P}(|Z_p - Z_r| \leq \varepsilon/2) \times \mathbf{P}(\sup_{1 \leq r \leq p} |Z_r| > \varepsilon) \leq \mathbf{P}(|Z_p| > \varepsilon/2).$$

Puisque $Z_r = S_{n+r} - S_n$, cette inégalité se réécrit

$$\inf_{1 \leq r < p} \mathbf{P}(|S_{n+p} - S_{n+r}| \leq \varepsilon/2) \times \mathbf{P}(\sup_{1 \leq r \leq p} |S_{n+r} - S_n| > \varepsilon) \leq \mathbf{P}(|S_{n+p} - S_n| > \varepsilon/2).$$

On pose $\beta_n = \sup\{\mathbf{P}(|S_{q+n} - S_{p+n}| > \varepsilon/2) : p, q \geq 0\}$. On a, pour tout $p \geq 1$,

$$\mathbf{P}(|S_{n+p} - S_n| > \varepsilon/2) \leq \beta_n, \quad \text{et} \quad \inf_{1 \leq r < p} \mathbf{P}(|S_{n+p} - S_{n+r}| \leq \varepsilon/2) \geq 1 - \beta_n,$$

d'où $(1 - \beta_n)\mathbf{P}(\sup_{1 \leq r \leq p} |S_{n+r} - S_n| > \varepsilon) \leq \beta_n$. D'autre part, comme

$$\mathbf{P}(|S_{n+p} - S_{n+q}| > \varepsilon/2) \leq \mathbf{P}(|S_{n+p} - S_n| > \varepsilon/4) + \mathbf{P}(|S_{n+q} - S_n| > \varepsilon/4),$$

$\beta_n \leq 2 \sup_{r \geq 0} \mathbf{P}(|S_{n+r} - S_n| > \varepsilon/4)$. Pour conclure, puisque $(S_n)_{n \geq 0}$ converge en probabilité, le critère de type Cauchy pour la convergence en probabilité de la proposition 11.1.15 implique que $\lim_{n \rightarrow \infty} \beta_n = 0$. Ainsi, pour tout $n \geq 1$ suffisamment grand de sorte que $1 - \beta_n > 0$, on a pour tout $p \geq 1$,

$$\mathbf{P}(\sup_{1 \leq r \leq p} |S_{n+r} - S_n| > \varepsilon) \leq \frac{\beta_n}{1 - \beta_n}$$

et donc

$$\mathbf{P}(\sup_{r \geq 0} |S_{n+r} - S_n| > \varepsilon) = \sup_{p \geq 1} \mathbf{P}(\sup_{1 \leq r \leq p} |S_{n+r} - S_n| > \varepsilon) \leq \frac{\beta_n}{1 - \beta_n}.$$

Ceci montre que $(S_n)_{n \geq 1}$ converge presque sûrement.

Il reste à montrer que la convergence en loi implique la convergence en probabilité. Supposons au contraire que $(S_n)_{n \geq 1}$ ne converge pas en probabilité. Encore une fois, la proposition 11.1.15 implique qu'il existe $\varepsilon > 0$ et $\alpha > 0$ tel que

$$\forall n \geq 1, \quad \exists (p_n, q_n) \in \mathbb{N}^2, \quad n \leq p_n < q_n, \quad \mathbf{P}(|S_{q_n} - S_{p_n}| > \varepsilon) > \alpha. \quad (11.5)$$

Posons $Z_n = S_{q_n} - S_{p_n}$ et montrons que $(Z_n)_{n \geq 1}$ converge en loi vers 0. Puisque S_{p_n} est indépendante de Z_n , on a en écrivant $S_{q_n} = S_{p_n} + Z_n$,

$$\forall t \in \mathbb{R}, \quad \phi_{S_{q_n}}(t) = \phi_{S_{p_n}}(t)\phi_{Z_n}(t).$$

Puisque $(S_n)_{n \geq 1}$ converge en loi vers S_∞ , ϕ_{S_n} converge simplement vers la fonction caractéristique ϕ de S_∞ . La fonction ϕ est continue sur \mathbb{R} et $\phi(0) = 1$. Il existe donc une constante $c > 0$ tel que $|t| \leq c$ implique $|\phi(t)| > 0$. Comme $n \leq p_n < q_n$, les suites $\phi_{S_{p_n}}$ et $\phi_{S_{q_n}}$ convergent simplement vers ϕ lorsque $n \rightarrow \infty$ si bien que pour $|t| \leq c$ implique $\lim_{n \rightarrow \infty} \phi_{Z_n}(t) = 1$.

Puisque, pour tout $x \in \mathbb{R}$, $1 - \cos(2x) \leq 4(1 - \cos(x))$, on a,

$$\forall t \in \mathbb{R}, \quad 0 \leq 1 - \operatorname{Re} \phi_{Z_n}(2t) \leq 4[1 - \operatorname{Re} \phi_{Z_n}(t)].$$

Par conséquent, pour tout $t \in \mathbb{R}$, il existe $n \geq 0$ tel que $2^{-n}|t| \leq c$ et donc $\lim_{n \rightarrow \infty} \operatorname{Re} \phi_{Z_n}(t) = 1$. Finalement, puisque $|\phi_{Z_n}(t)| \leq 1$, on déduit que $\lim_{n \rightarrow \infty} \phi_{Z_n}(t) = 1$. Ceci implique Z_n converge en loi vers δ_0 . La convergence en loi vers une v.a.r constante presque sûrement implique la convergence probabilité vers cette constante. Contradiction avec (11.5). \square

La proposition suivante est une application du théorème de Lévy ci-dessus. Elle sera par ailleurs utile dans la démonstration de la loi forte des grands nombres de Kolmogorov au chapitre suivant.

Proposition 11.3.4 (Séries centrées). *Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. indépendantes. On suppose que, pour tout $n \geq 1$, $X_n \in \mathbf{L}^2$ et $\mathbf{E}(X_n) = 0$. Alors $\sum_{n \geq 1} \mathbf{E}[X_n^2] < \infty$ implique que $(S_n)_{n \geq 1}$ converge presque sûrement et dans \mathbf{L}^2 vers une variable aléatoire réelle.*

Démonstration. Puis les variables aléatoires X_n sont centrées, $\mathbf{E}(X_n^2) = \mathbf{V}(X_n)$. Puis, pour tout $n, r \in \mathbb{N}^*$, par indépendance des X_n ,

$$\mathbf{E}[|S_{n+r} - S_n|^2] = \mathbf{V}\left(\sum_{i=n+1}^{n+r} X_i\right) = \sum_{i=n+1}^{n+r} \mathbf{V}(X_i) = \sum_{i=n+1}^{n+r} \mathbf{E}(X_i^2) \leq \sum_{i>n} \mathbf{E}(X_i^2),$$

qui est le reste d'une série convergente. La suite $(S_n)_{n \geq 1}$ est donc de Cauchy dans \mathbf{L}^2 , elle converge dans \mathbf{L}^2 vers S_∞ . La convergence en moyenne quadratique implique la convergence en probabilité et le théorème de Paul Lévy implique la convergence presque sûre vers une variable aléatoire S'_∞ . Enfin, $S_\infty = S'_\infty$ puisque $(S_n)_{n \geq 1}$ converge en probabilité vers S_∞ et presque sûrement, donc en probabilité, vers S'_∞ . \square

On termine cette partie par l'énoncé du théorème des trois séries de Kolmogorov. Celui-ci est une conséquence de l'inégalité maximale de Kolmogorov suivante et du théorème des deux séries.

Proposition 11.3.5 (Inégalité maximale de Kolmogorov). *Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. indépendantes, de carré intégrable et centrées. On note, pour tout $k \geq 1$, $S_k = X_1 + \dots + X_k$. Alors, pour tout $a > 0$ et tout $n \geq 1$,*

$$\mathbf{P}(\sup\{S_k : k = 1, \dots, n\} \geq a) \leq \frac{\mathbf{V}(S_n)}{a^2 + \mathbf{V}(S_n)}.$$

De plus,

$$\mathbf{P}\left(\sup_{1 \leq k \leq n} |S_k| \geq a\right) \leq \frac{\mathbf{E}[S_n^2]}{a^2}.$$

Démonstration. Soient $a > 0$ et $\tau = \inf_{k \geq 1} \{S_k \geq t\}$ alors les ensembles $A_k = \{\tau = k\}$, $k = 1, \dots, n$, sont deux à deux disjoints et

$$A = \bigcup_{k=1}^n A_k = \left\{ \sup\{S_k : k = 1, \dots, n\} \geq t \right\}.$$

Soit $c \geq 0$. La variable aléatoire $(S_k + c)\mathbf{1}_{A_k}$ est $\sigma(X_1, \dots, X_k)$ -mesurable et $S_n - S_k$ est $\sigma(X_{k+1}, \dots, X_n)$ -mesurable. Ces deux variables sont donc indépendantes et

$$\mathbf{E}[(S_k + c)\mathbf{1}_{A_k}(S_n - S_k)] = \mathbf{E}[(S_k + c)\mathbf{1}_{A_k}] \mathbf{E}[S_n - S_k] = 0.$$

Puisque les A_k sont disjoints, $\sum_{k=1}^n \mathbf{1}_{A_k} = \mathbf{1}_A \leq 1$. On obtient donc

$$\begin{aligned}
\mathbf{V}(S_n) + c^2 &= \mathbf{E}(S_n^2) + 2c \underbrace{\mathbf{E}(S_n)}_{=0} + c^2 = \mathbf{E}[(S_n + c)^2] \\
&\geq \mathbf{E} \left[\sum_{k=1}^n (S_n + c)^2 \mathbf{1}_{A_k} \right] = \sum_{k=1}^n \mathbf{E}[(S_n + c)^2 \mathbf{1}_{A_k}] = \sum_{k=1}^n \mathbf{E}[(S_k + c + S_n - S_k)^2 \mathbf{1}_{A_k}] \\
&= \sum_{k=1}^n \mathbf{E}[(S_k + c)^2 + 2(S_k + c)(S_n - S_k) + (S_n - S_k)^2] \mathbf{1}_{A_k} \\
&= \sum_{k=1}^n \mathbf{E}[(S_k + c)^2 \mathbf{1}_{A_k}] + \sum_{k=1}^n \mathbf{E}[(S_n - S_k)^2 \mathbf{1}_{A_k}] \\
&\geq \sum_{k=1}^n \mathbf{E}[(S_k + c)^2 \mathbf{1}_{A_k}]. \quad (11.6)
\end{aligned}$$

Comme $c \geq 0$ et par définition de τ , nous avons $(S_k + c)^2 \mathbf{1}_{A_k} \geq (t + c)^2 \mathbf{1}_{A_k}$. Ainsi, le calcul précédent donne

$$\mathbf{V}(S_n) + c^2 \geq \sum_{k=1}^n \mathbf{E}[(t + c)^2 \mathbf{1}_{A_k}] = (t + c)^2 \mathbf{P}(A).$$

Pour obtenir la première inégalité maximale, il suffit de poser $c = \mathbf{V}(S_n)/t \geq 0$.

Pour l'autre inégalité, on pose $\bar{\tau} = \inf_{k=1, \dots, n} \{|S_k| \geq t\}$, $\bar{A}_k = \{\bar{\tau} = k\}$ ainsi que $\bar{A} = \{\bar{\tau} \leq n\}$. On ne peut plus faire aboutir le calcul ci-dessus pour $c > 0$ mais il est encore valide pour $c = 0$. Dans ce cas, $S_k^2 \mathbf{1}_{\bar{A}_k} \geq t^2 \mathbf{1}_{\bar{A}_k}$. Le même calcul donne $\mathbf{P}(\bar{A}) \leq \mathbf{V}(S_n)/t^2$. \square

Exercice 31. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes dans \mathbf{L}^2 telles que $\mathbf{E}(X_n) = 0$ pour tout $n \geq 1$ et $V = \sup_{n \geq 1} \mathbf{V}(X_n) < \infty$ (famille bornée dans \mathbf{L}^2). Alors pour tout $\varepsilon > 0$,

$$\limsup_{n \rightarrow \infty} \frac{|S_n|}{n^{1/2}(\ln n)^{1/2+\varepsilon}} = 0, \quad \text{p.s..}$$

Théorème 11.3.6 (des deux séries de Kolmogorov). *Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. indépendantes de carré intégrable telles que $\sum_{n \geq 1} \mathbf{E}(X_n)$ et $\sum_{n \geq 1} \mathbf{V}(X_n)$ convergent dans \mathbb{R} . Alors $\sum_{n \geq 1} X_n$ converge presque sûrement vers une v.a.r..*

Démonstration. Sans perte de généralité, on peut supposer $\mathbf{E}(X_n) = 0$, car en recentrant on ne change pas la variance. On pose $S_n = \sum_{k=1}^n X_k$ et on va montrer que

$$\mathbf{P} \left(\limsup_{n \rightarrow \infty} S_n - \liminf_{n \rightarrow \infty} S_n = 0 \right) = 1.$$

Soit $m \geq 0$, on a

$$\limsup_{n \rightarrow \infty} S_n - \liminf_{n \rightarrow \infty} S_n = \limsup_{n \rightarrow \infty} (S_n - S_m) - \liminf_{n \rightarrow \infty} (S_n - S_m) \leq 2 \sup_{k \geq 1} \left| \sum_{i=1}^k X_{m+i} \right|.$$

Ainsi, pour tout $m \geq 1$ et tout $\varepsilon > 0$,

$$\begin{aligned}
\mathbf{P} \left(\limsup_{n \rightarrow \infty} (S_n - S_m) - \liminf_{n \rightarrow \infty} (S_n - S_m) \geq \varepsilon \right) &\leq \mathbf{P} \left(2 \sup_{k \geq 1} \left| \sum_{i=1}^k X_{m+i} \right| \geq \varepsilon \right) \\
&\leq \mathbf{P} \left(\max_{k=1, \dots, p} \left| \sum_{i=1}^k X_{m+i} \right| \geq \varepsilon/2 \right) \\
&\leq \limsup_{p \rightarrow \infty} 4\varepsilon^{-2} \sum_{i=m+1}^{m+p} \mathbf{V}(X_i) \leq 4\varepsilon^{-2} \sum_{i>m} \mathbf{V}(X_i), \quad (11.7)
\end{aligned}$$

par l'inégalité maximale de Kolmogorov. En faisant tendre $m \rightarrow \infty$, on obtient que $\limsup S_n = \liminf S_n$ donc $(S_n)_{n \geq 0}$ converge presque sûrement. \square

Théorème 11.3.7 (des trois séries de Kolmogorov). *Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. indépendantes. La série aléatoire $\sum_{n \geq 1} X_n$ converge presque sûrement dans \mathbb{R} si et seulement si pour un certain $A > 0$ les trois séries suivantes convergent :*

$$\sum_{n \geq 1} \mathbf{P}(|X_n| \geq A), \quad \sum_{n \geq 1} \mathbf{E}[X_n \mathbf{1}_{|X_n| \leq A}], \quad \text{et} \quad \sum_{n \geq 1} \mathbf{V}(X_n \mathbf{1}_{|X_n| \leq A}).$$

Dans ce théorème, il s'agit bien d'une équivalence. Néanmoins, la preuve de la nécessité de ces conditions est un peu technique, nous nous contenterons de montrer qu'elles sont suffisantes. Remarquons toutefois que si $\sum_{n=1}^{\infty} X_n$ converge presque sûrement alors la première condition est satisfaite, car sinon, par le deuxième lemme de Borel-Cantelli, on aurait que le terme général de la série ne tend pas vers 0.

Démonstration. Soit $A > 0$ tel que les trois conditions soient vérifiées. On pose $Y_n = X_n \mathbf{1}_{|X_n| \leq A}$. La convergence de la première série et le premier lemme de Borel-Cantelli implique que $|X_n| \leq A$ sauf pour un nombre (aléatoire) fini. Ainsi, pour tout n suffisamment grand, $X_n = Y_n$ presque sûrement et donc $\sum_{n=1}^{\infty} X_n$ converge si et seulement si $\sum_{n=1}^{\infty} Y_n$ converge.

Par le théorème des deux séries de Kolmogorov, les deux dernières conditions impliquent la convergence de la série $\sum_{n=1}^{\infty} Y_n$. □

Chapitre 12

Loi des grands nombres et Théorème Central Limite

Ce chapitre traite des deux principaux théorèmes en théorie des probabilités. À eux deux, ils justifient le bon choix de l'axiomatique de Kolmogorov pour appréhender les phénomènes aléatoires.

12.1 Loi des grands nombres

On commence par la loi faible des grands nombres dans le contexte \mathbf{L}^2 .

Théorème 12.1.1 (Loi Faible des Grands Nombres dans \mathbf{L}^2). *Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. i.i.d. admettant un moment d'ordre 2. Alors, la convergence suivante a lieu dans \mathbf{L}^2 et en probabilité*

$$\frac{1}{n} \sum_{k=1}^n X_k \longrightarrow \mathbf{E}(X_1).$$

Démonstration. On montre d'abord la convergence dans \mathbf{L}^2 en utilisant le caractère *i.i.d.* :

$$\mathbf{E} \left[\left(\frac{1}{n} \sum_{k=1}^n X_k - \mathbf{E}(X_1) \right)^2 \right] = \frac{1}{n^2} \sum_{k=1}^n \mathbf{E}[(X_k - \mathbf{E}(X_1))^2] = \frac{\mathbf{V}(X_1)}{n} \rightarrow 0.$$

Comme la convergence dans \mathbf{L}^2 implique la convergence en probabilité en utilisant l'inégalité de Bienaymé-Tchebychev, le théorème est montré. \square

L'hypothèse \mathbf{L}^2 semble un peu forte et surtout ne semble pas naturelle puisque dans la convergence établie la variance n'apparaît pas, d'où le corollaire.

Corollaire 12.1.2 (Loi faible des grands nombres dans \mathbf{L}^1). *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires i.i.d. à valeurs dans \mathbb{R}^d et admettant un moment d'ordre 1. Alors la convergence suivante a lieu en probabilité et dans \mathbf{L}^1 .*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k = \mathbf{E}(X_1).$$

Démonstration. Sans perte de généralité, on peut supposer les variables X_k positives. En effet, nous avons

$$\mathbf{E}|\overline{X_n} - \mathbf{E}(X_1)| \leq \mathbf{E}|\overline{X_n^+} - \mathbf{E}(X_1^+)| + \mathbf{E}|\overline{X_n^-} - \mathbf{E}(X_1^-)|.$$

Ainsi, si on montre la convergence, dans \mathbf{L}^1 , de la moyenne empirique des parties positives et négatives des variables X_k , on obtiendra le résultat du corollaire. Soit $M > 0$, et considérons les variables $X_k \wedge M$.

Il est facile de voir que

$$\begin{aligned} \mathbf{E} \left| \frac{1}{n} \sum_{k=1}^n X_k - \mathbf{E}(X_1) \right| &\leq \mathbf{E} \left| \frac{1}{n} \sum_{k=1}^n (X_k - X_k \wedge M) \right| \\ &\quad + \mathbf{E} \left| \frac{1}{n} \sum_{k=1}^n X_k \wedge M - \mathbf{E}(X_1 \wedge M) \right| \\ &\quad + |\mathbf{E}(X_1 \wedge M) - \mathbf{E}(X_1)| \\ &\leq 2\mathbf{E}|X_1 \wedge M - X_1| \\ &\quad + \mathbf{E} \left| \frac{1}{n} \sum_{k=1}^n X_k \wedge M - \mathbf{E}(X_1 \wedge M) \right|. \end{aligned}$$

Prenant la limite supérieure en $n \rightarrow \infty$, on a, par la loi faible des grands nombres dans \mathbf{L}^2 du théorème 12.1.1, pour tout $M > 0$

$$\limsup_{n \rightarrow \infty} \mathbf{E} \left| \frac{1}{n} \sum_{k=1}^n X_k - \mathbf{E}(X_1) \right| \leq 2\mathbf{E}|X_1 \wedge M - X_1|.$$

On remarque ensuite que $|X_1 \wedge M - X_1|$ tend presque-sûrement vers 0 lors $M \rightarrow \infty$. De plus, pour tout $M \geq 0$,

$$|X_1 \wedge M - X_1| = \mathbf{1}_{X_1 > M} |M - X_1| \leq M \mathbf{1}_{X_1 > M} + X_1 \leq 2X_1,$$

et le théorème de convergence dominée implique que

$$\limsup_{n \rightarrow \infty} \mathbf{E} \left| \frac{1}{n} \sum_{k=1}^n X_k - \mathbf{E}(X_1) \right| = 0.$$

La convergence dans \mathbf{L}^1 impliquant la convergence en probabilité, cela termine la preuve du corollaire. \square

Théorème 12.1.3 (Loi Forte des Grands Nombres dans le cadre \mathbf{L}^2). *Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. i.i.d. admettant un moment d'ordre 2. Alors,*

$$\frac{1}{n} \sum_{k=1}^n X_k \longrightarrow \mathbf{E}(X_1), \quad \mathbf{P} - p.s..$$

Démonstration. Supposons dans un premier temps que pour tout $n \geq 1$, $X_n \geq 0$. On introduit la notation

$$M_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

Pour montrer la convergence presque sûre de $(M_n)_{n \geq 1}$ vers $\mathbf{E}(X_1)$, on établit d'abord la convergence presque sûre de $(M_{n^2})_{n \geq 1}$ vers $\mathbf{E}(X_1)$, puis nous passerons à la suite toute entière. Ce faisant, par un calcul très similaire à celui de la preuve du théorème 12.1.1, on a pour tout $n \geq 1$, $\mathbf{E}(M_n) = \mathbf{E}(X_1)$ et $\mathbf{V}(M_n) = n^{-2}\mathbf{V}(X_1)$. Soit $\varepsilon > 0$, on utilise à nouveau l'inégalité de Bienaymé-Tchebychev pour estimer la probabilité

$$\mathbf{P}(|M_{n^2} - \mathbf{E}(X_1)| > \varepsilon) \leq \frac{\mathbf{E}(|M_{n^2} - \mathbf{E}(X_1)|^2)}{\varepsilon^2} = \frac{\mathbf{V}(X_1)}{n^2 \varepsilon^2}. \quad (12.1)$$

En sommant sur $n \geq 1$ de part et d'autre de l'inégalité (12.1), le corollaire 11.1.3 montre que M_{n^2} converge presque sûrement vers $\mathbf{E}(X_1)$.

Montrons désormais que la suite $(M_n)_{n \geq 0}$ converge presque-sûrement vers $\mathbf{E}(X_1)$, c'est ici qu'on se sert de la positivité des incréments. Pour tout $n > 1$, $[\sqrt{n}] \leq \sqrt{n} \leq [\sqrt{n} + 1]$ et donc, en notant $q_n = [\sqrt{n}]$, $q_n^2 \leq n \leq (q_n + 1)^2$. Comme les variables sont positives, on obtient les inégalités

$$S_{q_n^2} \leq S_n \leq S_{(q_n+1)^2} \quad \text{et} \quad n^{-1} S_{q_n^2} \leq M_n \leq n^{-1} S_{(q_n+1)^2}.$$

Par conséquent,

$$n^{-1} q_n^2 M_{q_n^2} \leq M_n \leq n^{-1} (q_n + 1)^2 M_{(q_n+1)^2}.$$

Rappelant que q_n/\sqrt{n} tend vers 1 et que M_{n^2} converge vers $\mathbf{E}(X_1)$, on obtient la convergence voulue pour $(M_n)_{n \geq 1}$.

Pour le cas général, il suffit d'écrire $X_k = X_k^+ - X_k^-$ et de vérifier que X_k^+ et X_k^- vérifie les conditions du résultat que l'on vient de montrer. □

Remarque 106. Remarquons que dans la preuve, l'hypothèse d'indépendance intervient pour montrer la convergence en probabilité et la convergence presque-sûre est obtenue par monotonie. Pour cette dernière, nous aurions pu invoquer la proposition 11.1.14.

De la même façon que pour la loi faible, on peut affaiblir l'hypothèse \mathbf{L}^2 et considérer X_1 dans \mathbf{L}^1 .

Théorème 12.1.4 (Loi forte des grands nombres de Kolmogorov). *Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. i.i.d. ; on note pour tout $n \geq 1$, $M_n = n^{-1}(X_1 + \dots + X_n)$.*

1. Si X_1 est intégrable, $(M_n)_{n \geq 1}$ converge presque sûrement et dans \mathbf{L}^1 vers $\mathbf{E}(X_1)$.
2. Si X_1 n'est pas intégrable, au moins un des deux événements $\{\limsup M_n = \infty\}$ ou $\{\liminf M_n = -\infty\}$ est de probabilité 1.

La démonstration de ce théorème nécessite trois lemmes usuels d'analyse que le lecteur assidu n'aura pas manqué de démontrer dans la première planche de TD de théorie de la mesure.

Lemme 12.1.5 (Lemme de Stolz-Cesàro). *Soient $(b_n)_{n \geq 1}$ une suite croissante de réels strictement positifs telle que $\lim_{n \rightarrow \infty} b_n = \infty$ et $(x_n)_{n \geq 0}$ une suite de réels convergeant vers $x \in \mathbb{R}$. Alors, en posant $b_0 = 0$,*

$$\lim_{n \rightarrow \infty} \frac{1}{b_n} \sum_{i=1}^n (b_i - b_{i-1}) x_i = x.$$

Démonstration. Exercice. □

Lemme 12.1.6 (Lemme de Kronecker). *Soient $(b_n)_{n \geq 1}$ une suite croissante de réels strictement positifs et $(x_n)_{n \geq 1}$ une suite de réels. Si la série $\sum_{n \geq 1} b_n^{-1} x_n$ est convergente (dans \mathbb{R}) alors*

$$\lim_{n \rightarrow \infty} b_n^{-1} \sum_{i=1}^n x_i = 0.$$

Démonstration. Exercice. □

Enfin, les estimées suivantes sont des conséquences standards des comparaisons séries/intégrales.

Lemme 12.1.7. *Pour tout $\alpha > 1$ et $k \geq 1$, $\sum_{n \geq k+1} n^{-\alpha} \leq k^{1-\alpha}/(\alpha - 1)$.*

Démonstration. Exercice. □

Preuve de la LGN de Kolmogorov. On suppose dans un premier temps que X_1 admet un moment d'ordre 1 et que $\mathbf{E}(X_1) = 0$. Introduisons quelques notations : pour tout $n \geq 1$,

$$\widehat{X}_n = X_n \mathbf{1}_{|X_n| < n}, \quad \widehat{M}_n = \frac{1}{n} \sum_{i=1}^n \widehat{X}_i, \quad \widetilde{X}_n = \widehat{X}_n - \mathbf{E}(\widehat{X}_n), \quad \text{et} \quad \widetilde{M}_n = \frac{1}{n} \sum_{i=1}^n \widetilde{X}_i.$$

Notons que les suites $(\widehat{X}_n)_{n \geq 1}$ et $(\widetilde{X}_n)_{n \geq 1}$ sont constituées de variables aléatoires *i.i.d.*

Pour montrer la convergence presque sûre de $(M_n)_{n \geq 0}$, nous allons procéder en deux temps :

1. tout d'abord, nous établirons que

$$M_n \xrightarrow{\text{p.s.}} 0 \iff \widehat{M}_n \xrightarrow{\text{p.s.}} 0 \iff \widetilde{M}_n \xrightarrow{\text{p.s.}} 0;$$

2. puis nous montrerons que $(\widetilde{M}_n)_{n \geq 1}$ converge vers 0 presque sûrement.

Considérons la première équivalence du point 1 en montrant que $M_n - \widehat{M}_n$ converge presque sûrement vers 0. On a, pour tout $n \geq 1$, $M_n - \widehat{M}_n = n^{-1} \sum_{i=1}^n X_i \mathbf{1}_{|X_i| > i}$. Or, les variables aléatoires $(X_n)_{n \geq 1}$ étant *i.i.d.*, il vient que

$$\sum_{n \geq 1} \mathbf{P}(|X_n| \geq n) = \sum_{n \geq 1} \mathbf{P}(|X_1| \geq n) \leq 1 + \mathbf{E}|X_1| < \infty.$$

Ainsi, d'après le premier lemme de Borel-Cantelli, $\mathbf{P}(\limsup\{|X_n| \geq n\}) = 0$. Ainsi, il existe $N = \limsup\{|X_n| \geq n\}$ négligeable tel que, pour tout $\omega \notin N$, il existe un entier $n_\omega \geq 1$ tel que $n \geq n_\omega$ implique $|X_n(\omega)| < n$ implique $X_n(\omega) = \widehat{X}_n(\omega)$. D'où, pour tout $\omega \notin N$,

$$\forall n \geq n_\omega, \quad M_n(\omega) - \widehat{M}_n(\omega) = \frac{1}{n} \sum_{i=1}^{n_\omega} X_i \mathbf{1}_{|X_i| \geq i}.$$

Il s'agit d'une somme finie renormalisée par n ce qui établit la première équivalence.

Pour la seconde équivalence, en utilisant la même démarche, on obtient

$$\forall n \geq 1, \quad \widehat{M}_n - \widetilde{M}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{E}[|X_1| \mathbf{1}_{|X_1| < i}],$$

car les variables aléatoires X_i sont identiquement distribuées. Puisque X_1 est intégrable, X_1 est finie presque sûrement et donc $X_1 \mathbf{1}_{|X_1| < i}$ converge presque sûrement vers X_1 . De plus, $|X_1 \mathbf{1}_{|X_1| < i}| \leq |X_1|$ pour tout $i \geq 1$ et comme X_1 est intégrable, le théorème de convergence dominée implique que $\lim_{i \rightarrow \infty} \mathbf{E}(X_1 \mathbf{1}_{|X_1| < i}) = \mathbf{E}(X_1) = 0$. Le lemme de Cesàro implique que $\widehat{M}_n - \widetilde{M}_n$ tend vers 0.

Montrons désormais le second point : $(\widetilde{M}_n)_{n \geq 1}$ converge presque sûrement vers 0. Nous appliquons le lemme de Kronecker : pour montrer que $n^{-1} \sum_{i=1}^n \widetilde{X}_i$ converge vers 0 presque sûrement, il suffit de montrer que la série $\sum_{i=1}^n i^{-1} \widetilde{X}_i$ converge presque sûrement dans \mathbb{R} . Or les variables aléatoires $(n^{-1} \widetilde{X}_n)_{n \geq 1}$ sont indépendantes et, pour tout $n \in \mathbb{N}^*$, $|n^{-1} \widetilde{X}_n| \leq 2$ donc de carré intégrable et finalement $\mathbf{E}[n^{-1} \widetilde{X}_n] = 0$. D'après le résultat sur les séries centrées de la proposition 11.3.4, il suffit de vérifier que $\sum_{n \geq 1} n^{-2} \mathbf{E}(\widetilde{X}_n^2) < \infty$ pour obtenir la convergence presque sûre de $\sum_{n \geq 1} n^{-1} \widetilde{X}_n$ dans \mathbb{R} . On calcule

$$\sum_{n \geq 1} n^{-2} \mathbf{E}(\widetilde{X}_n^2) = \sum_{n \geq 1} n^{-2} \mathbf{E} \left[\left(\widehat{X}_n - \mathbf{E}[\widehat{X}_n] \right)^2 \right] = \sum_{n \geq 1} n^{-2} \mathbf{V}(\widehat{X}_n) \leq \sum_{n \geq 1} n^{-2} \mathbf{E}[\widehat{X}_n^2].$$

Or, les variables $(X_n)_{n \geq 1}$ sont *i.i.d.* si bien que

$$\forall n \geq 1, \quad \mathbf{E}(\widehat{X}_n^2) = \mathbf{E}(X_n^2 \mathbf{1}_{|X_n| < n}) = \mathbf{E}(X_1^2 \mathbf{1}_{|X_1| < n}).$$

Par convergence monotone, on trouve finalement que

$$\sum_{n \geq 1} n^{-2} \mathbf{E}(\widetilde{X}_n^2) \leq \sum_{n \geq 1} n^{-2} \mathbf{E}(X_1^2 \mathbf{1}_{|X_1| < n}) = \mathbf{E} \left[\sum_{n \geq 1} n^{-2} X_1^2 \mathbf{1}_{|X_1| < n} \right].$$

Or, pour tout $x \geq 0$, on obtient par le lemme de comparaison séries/intégrales

$$\sum_{n \geq 1} n^{-2} x^2 \mathbf{1}_{x < n} = x^2 \sum_{n \geq \lfloor x \rfloor + 1} n^{-2} = \frac{x^2}{(\lfloor x \rfloor + 1)^2} + x^2 \sum_{n \geq \lfloor x \rfloor + 2} n^{-2} \leq \frac{x^2}{(\lfloor x \rfloor + 1)^2} + \frac{x^2}{\lfloor x \rfloor + 1} \leq 2x.$$

Par conséquent, $\sum_{n \geq 1} n^{-2} \mathbf{E}(\widetilde{X}_n^2) \leq 2\mathbf{E}|X_1| < \infty$. On a donc convergence presque sûre de $(\widetilde{M}_n)_{n \geq 0}$ vers 0 et donc de $(M_n)_{n \geq 1}$ vers 0.

Il s'agit de montrer que $(M_n)_{n \geq 1}$ converge également dans \mathbf{L}^1 . Soit $k \in \mathbb{N}^*$ et écrivons $|M_n| = \min(|M_n|, k) + (|M_n| - k)^+$. La fonction $x \rightarrow (x - k)^+$ est convexe et croissante si bien que

$$(|M_n| - k)^+ \leq \frac{1}{n} \sum_{i=1}^n (|X_i| - k)^+.$$

Puisque les variables sont identiquement distribuées, il vient que

$$\mathbf{E}|M_n| \leq \mathbf{E}[\min(|M_n|, k)] + \frac{1}{n} \sum_{i=1}^n \mathbf{E}[(|X_i| - k)^+] = \mathbf{E}[\min(|M_n|, k)] + \mathbf{E}[(|X_1| - k)^+].$$

On remarque alors que $\min(|M_n|, k)$ converge presque sûrement vers 0 quand $n \rightarrow \infty$. De même, ces variables aléatoires sont uniformément bornée en n par k et le théorème de convergence dominée montre que le premier terme à droite de l'inégalité tend vers 0 lorsque $n \rightarrow \infty$ si bien que

$$\forall k \in \mathbb{N}^*, \quad \limsup_{n \rightarrow \infty} \mathbf{E}|M_n| \leq \mathbf{E}[(|X_1| - k)^+].$$

Comme $|X_1|$ est intégrable, $(|X_1| - k)^+$ converge presque sûrement vers 0 lorsque $k \rightarrow \infty$. De plus, $(|X_1| - k)^+ \leq |X_1|$ et par convergence dominée

$$\limsup_{n \rightarrow \infty} \mathbf{E}|M_n| \leq \lim_{k \rightarrow \infty} \mathbf{E}[(|X_1| - k)^+] = 0,$$

d'où la convergence \mathbf{L}^1 . Ceci termine la démonstration dans le cas X_1 intégrable d'espérance nulle.

Supposons désormais X_1 intégrable mais $m = \mathbf{E}(X_1) \neq 0$. Observons que, en notant $\bar{X}_n = X_n - m$ pour tout $n \geq 1$, on a

$$M_n - m = \frac{1}{n} \sum_{i=1}^n (X_i - m) = \frac{1}{n} \sum_{i=1}^n \bar{X}_i = \bar{M}_n.$$

Or les variables aléatoires $(\bar{X}_n)_{n \geq 1}$ sont *i.i.d.*, \bar{X}_1 est intégrable et $\mathbf{E}\bar{X}_1 = 0$. Ainsi, \bar{M}_n converge presque sûrement et dans \mathbf{L}^1 vers 0. D'où l'on déduit que M_n converge presque sûrement et dans \mathbf{L}^1 vers m .

On considère désormais le cas où X_1 n'est pas intégrable. Les variables aléatoires $\liminf M_n$ et $\limsup M_n$ sont des variables asymptotiques de la suite de variables indépendantes $(X_n)_{n \geq 1}$. D'après la loi du 0-1 de Kolmogorov 11.3.1, les événements $\{\liminf M_n = -\infty\}$ et $\{\limsup M_n = \infty\}$ sont de probabilité 0 ou 1. En fait, il existe c^* et $c_* \in \mathbb{R} \cup \{\pm\infty\}$ tels que, presque sûrement, $\liminf M_n = c_*$ et $\limsup M_n = c^*$.

Supposons que les deux événements $\{\limsup M_n = \infty\}$ et $\{\liminf M_n = -\infty\}$ soient négligeables, alors $-\infty < c_* \leq c^* < \infty$. Il vient alors que $\frac{X_n}{n} = M_n - \frac{n-1}{n} M_{n-1}$ si bien que

$$\limsup \frac{X_n}{n} \leq c^* - c_* \quad \text{et} \quad \liminf \frac{X_n}{n} \geq c_* - c^*.$$

Soit $c > c^* - c_*$. Comme $\limsup\{X_n \geq cn\} \subset \{\limsup \frac{X_n}{n} \geq c\}$, on déduit que $\mathbf{P}(\limsup\{X_n \geq cn\}) = 0$. Puis, les variables $(X_n)_{n \geq 1}$ étant *i.i.d.*, le deuxième lemme de Borel-Cantelli implique

$$\sum_{n \geq 1} \mathbf{P}(X_1^+ \geq cn) = \sum_{n \geq 1} \mathbf{P}(X_1 \geq cn) = \sum_{n \geq 1} \mathbf{P}(X_n \geq cn) < \infty.$$

Ainsi, X_1^+ est intégrable. De même, $\limsup\{X_n \leq -cn\} \subset \{\liminf \frac{X_n}{n} \leq -c\}$, et en utilisant un argument similaire, on obtient que X^- est intégrable. C'est une contradiction. \square

Remarque 107. Si $(X_n)_{n \geq 1}$ est une suite de variables aléatoires positives *i.i.d.* avec $\mathbf{E}(X_1) = \infty$, alors, presque sûrement, $\lim_{n \rightarrow \infty} M_n = \infty$.

En effet, pour tout $k \in \mathbb{N}^*$, $\liminf \frac{1}{n} \sum_{i=1}^n X_i \geq \liminf \frac{1}{n} \sum_{i=1}^n \min(X_i; k)$. D'après la loi forte des grands nombres, il existe N_k négligeable tel que

$$\forall \omega \notin N_k, \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \min(X_i; k) = \mathbf{E}[\min(X_1; k)].$$

Posons $N = \cup_{k \geq 1} N_k$. On vérifie que N est négligeable et

$$\forall \omega \notin N, \quad \forall k \geq 1, \quad \liminf \frac{1}{n} \sum_{i=1}^n X_i(\omega) \geq \liminf \frac{1}{n} \sum_{i=1}^n \min(X_i; k) = \mathbf{E}[\min(X_1; k)].$$

Par convergence monotone, l'espérance à droite tend vers ∞ lorsque $k \rightarrow \infty$.

En fait, on peut même supprimer l'hypothèse de positivité et supposer que la partie positive X_1^+ ou la partie négative X_1^- dans \mathbf{L}^1 . La moyenne empirique converge alors vers $-\infty$ ou ∞ respectivement.

Dans le cas X_1^+ et X_1^- non intégrable, on peut également dire quelque chose sur le comportement de la moyenne empirique, mais cette fois-ci il faut comparer les queues de distribution de la partie positive et négative. On renvoie à [Kes70] et [Eri73] pour ces considérations.

12.2 Théorème Central Limite

Théorème 12.2.1. Soit $(X_n)_{n \geq 0}$ une suite de v.a.r. i.i.d. avec X_1 de carré intégrable ; on note $m = \mathbf{E}(X_1)$ et $\sigma^2 = \mathbf{V}(X_1)$. Considérons, pour tout $n \geq 1$,

$$T_n = \sqrt{n} \left(\frac{S_n}{n} - m \right), \quad \text{où } S_n = X_1 + \cdots + X_n.$$

Alors la suite $(T_n)_{n \geq 1}$ converge en loi, lorsque $n \rightarrow \infty$, vers une v.a.r. de loi $\mathcal{N}(0, \sigma^2)$.

Démonstration. On calcule la fonction caractéristique de T_n et on simplifie en utilisant le caractère *i.i.d.* :

$$\phi_{T_n}(t) = \mathbf{E}(e^{itT_n}) = \phi_{X_1 - m}(t/\sqrt{n})^n. \quad (12.2)$$

Or, $\phi(0) = 1$, $\phi'(0) = 0$, car $X_1 - m$ est centrée, et $\phi''(0) = -\sigma^2$. Donc le développement de Taylor à l'ordre 2 de ϕ_{T_n} donne

$$\phi_{T_n}(t) = \left(1 - \frac{t^2 \sigma^2}{2n} + \frac{t^2}{n} \varepsilon(t/\sqrt{n}) \right)^n.$$

Lemme 12.2.2. Soit $(z_n)_{n \geq 0}$ est une suite de nombre complexe telle que $\lim_{n \rightarrow \infty} n z_n = z$ alors $\lim_{n \rightarrow \infty} (1 + z_n)^n = e^z$.

Comme $n \left(-\frac{t^2 \sigma^2}{2n} + \frac{t^2}{n} \varepsilon(t/\sqrt{n}) \right) \rightarrow -\frac{t^2 \sigma^2}{2}$ on obtient que

$$\forall t \in \mathbb{R}, \quad \lim_{n \rightarrow \infty} \phi_{T_n}(t) = e^{-\frac{t^2 \sigma^2}{2}}.$$

On reconnaît ici la fonction caractéristique d'une $\mathcal{N}(0, \sigma^2)$. □

12.3 TCL multivarié

Le TCL univarié se généralise facilement à la dimension supérieure. C'est en fait un corollaire.

Théorème 12.3.1. Soit $(X_n)_{n \geq 1}$ une suite de vecteurs aléatoires dans \mathbb{R}^d , i.i.d. avec $X_1 \in \mathbf{L}^2$. On note $m = \mathbf{E}(X_1)$ et Γ la matrice de covariance de X_1 . Alors la suite de vecteurs aléatoires $(T_n)_{n \geq 1}$ définis pour tout $n \geq 1$ par

$$T_n = \sqrt{n} \left(\frac{X_1 + \cdots + X_n}{n} - m \right)$$

converge en loi vers un vecteur gaussien de loi $\mathcal{N}(0, \Gamma)$.

Démonstration. En utilisant les fonctions caractéristiques, on a pour tout $t \in \mathbb{R}^d$

$$\phi_{T_n}(t) = \mathbf{E} \left[e^{it^* T_n} \right] = \phi_{t^* X_1}(1)$$

et

$$t^* T_n = \sqrt{n} \left(\frac{t^* X_1 + \cdots + t^* X_n}{n} - t^* m \right).$$

La suite de v.a.r. $(t^* X_n)_{n \geq 1}$ est *i.i.d.* et de carré intégrable avec $\mathbf{E}(t^* X_1) = t^* m$ et $\mathbf{V}(t^* X_1) = t^* \Gamma t$. D'après le TCL univarié, lorsque $n \rightarrow \infty$, $t^* T_n$ converge en loi vers une variable réelle de loi $\mathcal{N}(0, t^* \Gamma t)$. Par conséquent,

$$\lim_{n \rightarrow \infty} \phi_{T_n}(t) = \lim_{n \rightarrow \infty} \phi_{t^* T_n}(1) = \exp \left(-\frac{t^* \Gamma t}{2} \right).$$

D'où le résultat. □

12.4 Applications de la loi des grands nombres

Théorème 12.4.1 (fondamental de la Statistique). *Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. indépendantes et identiquement distribuées suivant la loi μ sur \mathbb{R}^d . Pour tout $n \geq 1$, on note*

$$\mu_n^\omega = \frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)}, \quad \omega \in \Omega,$$

la mesure empirique. Alors, presque sûrement, $(\mu_n)_{n \geq 1}$ converge étroitement vers μ .

Remarque 108. Précisons sa signification : pour tout borélien $B \in \mathbb{R}^d$ et tout $\omega \in \Omega$,

$$\mu_n^\omega(B) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)}(B) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_B(X_i(\omega)).$$

De façon plus générale, si $f : \mathbb{R}^d \rightarrow \mathbb{R}$ est borélienne bornée ou borélienne positive, nous pouvons écrire

$$\int_{\mathbb{R}^d} f(x) \mu_n^\omega(dx) = \frac{1}{n} \sum_{i=1}^n f(X_i(\omega)).$$

Le théorème précédent affirme qu'il existe un ensemble négligeable N tel que pour tout $\omega \notin N$, $(\mu_n^\omega)_{n \geq 1}$ converge vers μ étroitement, *i.e.*

$$\forall \omega \in N^c, \quad \forall f \in C_b(\mathbb{R}^d), \quad \int_{\mathbb{R}^d} f(x) \mu_n^\omega(dx) = \frac{1}{n} \sum_{i=1}^n f(X_i(\omega)) \rightarrow \int_{\mathbb{R}^d} f(x) \mu(dx).$$

Démonstration. L'espace $C_c(\mathbb{R}^d)$ des fonctions continues à support compact est séparable. C'est une conséquence du théorème de Stone-Weierstrass établissant la densité des fonctions polynomiales pour $\|\cdot\|_\infty$ dans l'espace des fonctions continues sur un compact. Il existe ainsi une famille dénombrable $\mathcal{H} = (h_r)_{r \in \mathbb{N}} \subset C_c$ dense dans C_c pour $\|\cdot\|_\infty$.

Soit $r \in \mathbb{N}$, nous avons

$$\int_{\mathbb{R}^d} h_r(x) \mu_n^\omega(dx) = \frac{1}{n} \sum_{i=1}^n h_r(X_i(\omega)).$$

Les variables aléatoires réelles $(h_r(X_n))_{n \geq 1}$ sont indépendantes, identiquement distribués, bornées et donc de carré intégrable. Par la loi forte des grands nombres dans le cadre \mathbf{L}^2 , il existe N_r négligeable tel que si $\omega \notin N_r$,

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} h_r(x) \mu_n^\omega(dx) = \mathbf{E}[h_r(X_1)] = \int_{\mathbb{R}^d} h_r(x) \mu(dx).$$

Comme la famille \mathcal{H} est dénombrable, on peut même définir universellement un tel ensemble négligeable. En effet, notons $N = \cup_{r \in \mathbb{N}} N_r$, alors $\mathbf{P}(N) \leq \sum_{r \geq 0} \mathbf{P}(N_r) = 0$. De plus, si $\omega \notin N$, on a

$$\forall r \in \mathbb{N}, \quad \lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} h_r(x) \mu_n^\omega(dx) = \int_{\mathbb{R}^d} h_r(x) \mu(dx).$$

D'après le théorème 11.2.2, il vient que, pour tout $\omega \in N^c$,

$$\forall f \in C_b, \quad \lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} f(x) \mu_n^\omega(dx) = \int_{\mathbb{R}^d} f(x) \mu(dx).$$

Ceci termine la preuve du théorème. □

Le théorème fondamentale de la Statistique établit donc que la mesure empirique a tendance à converger (étroitement) vers la loi théorique sauf peut-être pour certains n -échantillons exceptionnels. Dans le contexte des *v.a.r.*, ce théorème peut se traduire en terme de fonction de répartition.

Théorème 12.4.2 (Glivenko-Cantelli). Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. indépendantes et identiquement distribuées. On note F la fonction de répartition de X_1 et, pour tout $n \geq 1$, F_n la fonction de répartition empirique, i.e.

$$\forall \omega \in \Omega, \quad \forall t \in \mathbb{R}, \quad F_n^\omega(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{]-\infty, t]}(X_i(\omega)).$$

Alors, presque sûrement, $(F_n)_{n \geq 1}$ converge vers F uniformément sur \mathbb{R} .

Remarque 109. Ce théorème signifie qu'il existe N négligeable tel que

$$\forall \omega \in N^c, \quad \lim_{n \rightarrow \infty} \sup_{t \in \mathbb{R}} |F_n^\omega(t) - F(t)| = 0.$$

Remarque 110. La fonction (aléatoire) de répartition empirique est en fait la fonction de répartition de la mesure (aléatoire) empirique : pour tout $\omega \in \Omega$, F_n^ω est la fonction de répartition de μ_n^ω .

Démonstration. Le fait que $(F_n^\omega(t))_{n \geq 1}$ converge presque sûrement pour tout $t \in \mathbb{R}$ fixé est simplement une conséquence de la loi forte des grands nombres. En effet, les variables aléatoires $(\mathbf{1}_{]-\infty, t]}(X_n))_{n \geq 1}$ sont indépendantes, identiquement distribuées, de loi commune la loi de Bernoulli de paramètre $\mathbf{P}(X_1 \leq t)$, et bornées donc de carré intégrable. Ceci établit donc l'existence pour tout $t \in \mathbb{R}$ d'un ensemble N_t négligeable tel que dès que $\omega \notin N_t$, $\lim_{n \rightarrow \infty} F_n^\omega(t) = F(t)$. De la même manière, il existe pour chaque $t \in \mathbb{R}$ un ensemble N_t^- négligeable tel que

$$\mathbf{P}(X_1 < t) = F(t^-) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbf{1}_{]-\infty, t[}(X_i(\omega)) = \lim_{n \rightarrow \infty} F_n^\omega(t^-).$$

La suite de la preuve consiste en deux choses : d'abord il s'agit de montrer que la convergence est uniforme ; d'autre part, il faut construire un ensemble N négligeable universel (indépendant de $t \in \mathbb{R}$). Pour ce faire, considérons deux fonctions de répartitions F et F_n ainsi qu'une subdivision finie $\tau = (t_i)_{i=1, \dots, p}$ avec $t_1 \leq t_2 \leq \dots \leq t_p$, $p \in \mathbb{N}^*$. On note

$$\delta_F(\tau) = \max \left\{ 1 - F(t_p), (F(t_p^-) - F(t_{p-1}))^+, \dots, (F(t_2^-) - F(t_1))^+, F(t_1^-) \right\},$$

et

$$R_n(\tau) = \max_{i=1, \dots, p} \left\{ (F(t_i) - F_n(t_i))^+, (F_n(t_i^-) - F(t_i^-))^+ \right\}.$$

On va montrer que $\|F - F_n\|_\infty = \sup_{t \in \mathbb{R}} |F(t) - F_n(t)| \leq \delta_F(\tau) + R_n(\tau)$. En effet, les fonctions F et F_n étant croissantes et positives,

1. si $t < t_1$, $F(t) - F_n(t) \leq F(t_1^-) \leq \delta_F(\tau)$ et

$$F_n(t) - F(t) \leq F_n(t_1^-) \leq F_n(t_1^-) - F(t_1^-) + F(t_1^-) \leq R_n(\tau) + \delta_F(t);$$

2. si $t \in [t_{i-1}, t_i]$, $i = 2, \dots, p$, alors d'une part

$$F(t) - F_n(t) \leq F(t_i^-) - F_n(t_{i-1}) \leq F(t_i^-) - F(t_{i-1}^-) + F(t_{i-1}^-) - F_n(t_{i-1}) \leq \delta_F(\tau) + R_n(\tau),$$

et, d'autre part,

$$F_n(t) - F(t) \leq F_n(t_i^-) - F(t_{i-1}) \leq F_n(t_i^-) - F(t_i^-) + F(t_i^-) - F(t_{i-1}) \leq R_n(\tau) + \delta_F(\tau);$$

3. enfin, si $t \geq t_p$, $F_n(t) - F(t) \leq 1 - F(t_p) \leq \delta_F(\tau)$ et

$$F(t) - F_n(t) \leq 1 - F_n(t_p) \leq 1 - F(t_p) + F(t_p) - F_n(t_p) \leq \delta_F(\tau) + R_n(\tau).$$

Notons, pour tout $x \in]0, 1[$,

$$C(x) = \inf\{u \in \mathbb{R} : F(u) \geq x\}.$$

Puisque $\lim_{t \rightarrow \infty} F(t) = 1$, l'ensemble $A_x = \{u \in \mathbb{R} : F(u) \geq x\}$ est non vide. Comme $\lim_{t \rightarrow -\infty} F(t) = 0$, A_x est minoré. Ceci montre l'existence de $C(x)$ pour tout $x \in (0, 1)$. Puisque F est croissante, la

fonction C est elle-même croissante. De plus, A_x est une demi-droite et comme F est continue à droite, $F(C(x)) \geq x$ et donc $C(x) \in A_x$. Autrement dit, $A_x = [C(x), \infty)$ d'où

$$C(x) \leq t \iff x \leq F(t).$$

En particulier, $F(C(x)^-) \leq x$ puisque pour $s < C(x)$, $F(s) < x$.

On considère l'ensemble N défini par

$$N = \bigcup_{q \in \mathbb{Q} \cap (0,1)} \left(N_{C(q)} \cup N_{C(q)}^- \right).$$

Il est immédiat que $\mathbf{P}(N) = 0$. D'autre part si $\omega \notin N^{\mathbb{G}}$,

$$\forall q \in \mathbb{Q} \cap (0,1), \quad F_n^\omega(C(q)) \rightarrow F(C(q)), \quad \text{et} \quad F_n^\omega(C(q)^-) \rightarrow F(C(q)^-).$$

On va montrer que la convergence a lieu uniformément en $t \in \mathbb{R}$ pour tout $\omega \in N^{\mathbb{G}}$. Soit donc $\omega \in N^{\mathbb{G}}$ fixé et $p \in \mathbb{N}^*$. Pour tout $i = 1, \dots, p$, on pose $t_i = C(i(p+1)^{-1})$. On remarque que $F(t_i^-) \leq i/(p+1)$ et $F(t_i) \geq i/(p+1)$ pour tout $i = 1, \dots, p$. Par conséquent, $\delta_F(\tau) \leq 1/(p+1)$ et par l'inégalité établie plus haut, $\|F_n^\omega - F\|_\infty \leq 1/(p+1) + R_n(\tau)$.

Par définition de $R_n(\tau)$, pour tout $\omega \in N^{\mathbb{G}}$ fixé et tout $p \in \mathbb{N}^*$ fixé, $R_n(\tau) \rightarrow 0$ si $n \rightarrow \infty$ si bien que

$$\limsup_{n \rightarrow \infty} \|F - F_n^\omega\|_\infty \leq 1/(p+1) + \limsup_{n \rightarrow \infty} R_n(\tau) = 1/(p+1).$$

Le membre de gauche de l'inégalité ne dépend plus de p et l'inégalité est valable pour tout $p \geq 1$ donc la limite supérieure à gauche est nulle. Ceci achève la preuve de ce résultat. □

Ce théorème peut être affiné en précisant la vitesse de convergence des fonctions de répartition empirique vers la fonction de répartition théorique. Ce dernier résultat est à la base du test de Kolmogorov-Smirnov — on pourra pour cela se référer à [Bil68, Théorème 13.5, p.105].

Chapitre 13

Espérance conditionnelle

Dans ce chapitre, sauf mention contraire, on considère un espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$ et X à valeurs dans \mathbb{R}^d une variable aléatoire définie sur $(\Omega, \mathcal{F}, \mathbf{P})$.

13.1 Conditionnement par un événement

Définition 13.1.1 (Probabilité conditionnelle). Soit $B \in \mathcal{F}$. La probabilité conditionnelle sachant B est une fonction d'ensemble, notée $\mathbf{P}(\cdot|B)$, de la tribu \mathcal{F} dans $[0, 1]$ définie par :

$$\forall A \in \mathcal{F} : \quad \mathbf{P}(A|B) = \begin{cases} \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} & \text{si } \mathbf{P}(B) > 0, \\ 0 & \text{sinon.} \end{cases}$$

Remarque 111. La spécification de la probabilité conditionnelle lorsque $\mathbf{P}(B) = 0$ est arbitraire est sans importance particulière.

Proposition 13.1.2. Soit $B \in \mathcal{F}$ tel que $\mathbf{P}(B) > 0$. La fonction d'ensemble $\mathbf{P}(\cdot|B) : \mathcal{F} \rightarrow [0, 1]$ est une probabilité sur \mathcal{F} .

Démonstration. On vérifie facilement que $\mathbf{P}(\Omega|B) = 1$. De plus, si $(A_n)_{n \geq 0}$ est une famille dénombrable d'éléments de \mathcal{F} deux à deux disjoints, alors il en va de même de la famille $(A_n \cap B)_{n \geq 0}$. Ainsi,

$$\mathbf{P}\left(\bigcup_{n \geq 0} A_n \middle| B\right) = \frac{\mathbf{P}\left(\left(\bigcup_{n \geq 0} A_n\right) \cap B\right)}{\mathbf{P}(B)} = \frac{\mathbf{P}\left(\bigcup_{n \geq 0} (A_n \cap B)\right)}{\mathbf{P}(B)} = \sum_{n \geq 0} \frac{\mathbf{P}(A_n \cap B)}{\mathbf{P}(B)} = \sum_{n \geq 0} \mathbf{P}(A_n|B).$$

□

Remarque 112. La probabilité conditionnelle par rapport à un événement B est parfois notée \mathbf{P}_B . Cette notation a l'avantage de mettre en exergue le fait que la probabilité conditionnelle est une probabilité, cependant elle est peu pratique.

Proposition 13.1.3. Soit $B \in \mathcal{F}$ tel que $\mathbf{P}(B) > 0$. Pour tout événement $A \in \mathcal{F}$ indépendant de B on a $\mathbf{P}(A|B) = \mathbf{P}(A)$. De manière plus générale, si $A, B \in \mathcal{F}$ sont tels que $\mathbf{P}(A)\mathbf{P}(B) > 0$ alors A et B sont indépendants si et seulement si $\mathbf{P}(A|B) = \mathbf{P}(A)$ si et seulement si $\mathbf{P}(B|A) = \mathbf{P}(B)$.

Démonstration. Le premier point est immédiat puisque par hypothèse $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$. Pour le deuxième point, il suffit de remarquer que $\mathbf{P}(A \cap B) = \mathbf{P}(A|B)\mathbf{P}(B) = \mathbf{P}(A)\mathbf{P}(B)$. □

Proposition 13.1.4 (Formule des probabilités totales). Soit $(B_n)_{n \geq 0} \in \mathcal{F}^{\mathbb{N}}$ une partition (modulo 0) de Ω . Alors pour tout $A \in \mathcal{F}$:

$$\mathbf{P}(A) = \sum_{n \geq 0} \mathbf{P}(A|B_n)\mathbf{P}(B_n).$$

Démonstration. Puisque $(B_n)_{n \geq 0}$ est une partition, les ensembles $A \cap B_n$, $n \geq 0$, sont deux à deux disjoints et $A = A \cap (\cup_{n \geq 0} B_n)$, ainsi

$$\mathbf{P}(A) = \sum_{n \geq 0} \mathbf{P}(A \cap B_n) = \sum_{n \geq 0} \mathbf{P}(A|B_n)\mathbf{P}(B_n).$$

□

Proposition 13.1.5 (Formule de Bayes). *Soit $(B_n)_{n \geq 0}$ une famille d'événements de \mathcal{F} formant une partition (modulo 0) de Ω . Alors pour tout $A \in \mathcal{F}$ tel que $\mathbf{P}(A) > 0$ et tout $n \geq 0$:*

$$\mathbf{P}(B_n|A) = \frac{\mathbf{P}(A|B_n)\mathbf{P}(B_n)}{\sum_{k \geq 0} \mathbf{P}(A|B_k)\mathbf{P}(B_k)}.$$

Démonstration. Par la formule des probabilités totales :

$$\mathbf{P}(B_n|A) = \frac{\mathbf{P}(B_n \cap A)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A|B_n)\mathbf{P}(B_n)}{\sum_{k \geq 0} \mathbf{P}(A|B_k)\mathbf{P}(B_k)}.$$

□

Comme la probabilité conditionnelle $\mathbf{P}(\cdot|B)$ est en particulier une probabilité, on peut calculer la moyenne d'une variable aléatoire intégrable par rapport à cette nouvelle probabilité. Cette espérance est appelée espérance conditionnelle.

Définition 13.1.6. Soient X est une v.a. \mathbf{P} -intégrable à valeurs dans \mathbb{R}^d et $B \in \mathcal{F}$ un événement tel que $\mathbf{P}(B) > 0$. L'espérance conditionnelle de X sachant B , notée $\mathbf{E}(X|B)$, est l'espérance de X par rapport à la probabilité $\mathbf{P}(\cdot|B)$. Ainsi, par définition,

$$\mathbf{E}(X|B) = \int_{\Omega} X(\omega)\mathbf{P}(d\omega|B).$$

Proposition 13.1.7. *Soient X une v.a. \mathbf{P} -intégrable à valeurs dans \mathbb{R}^d et $B \in \mathcal{F}$. Alors,*

$$\mathbf{E}(X|B) = \begin{cases} \frac{\mathbf{E}(X\mathbf{1}_B)}{\mathbf{P}(B)}, & \text{si } \mathbf{P}(B) > 0 \\ 0, & \text{sinon.} \end{cases}$$

De plus, si $A \in \mathcal{F}$, alors $\mathbf{P}(A|B) = \mathbf{E}(\mathbf{1}_A|B)$.

Démonstration. Là encore, lorsque $\mathbf{P}(B) = 0$, on définit la valeur de l'espérance conditionnelle de manière arbitraire. Si $\mathbf{P}(B) > 0$, l'égalité $\mathbf{P}(A|B) = \mathbf{E}(\mathbf{1}_A|B)$ provient de la définition. Puis, on commence par vérifier l'égalité pour des fonctions en escaliers positives.

$$\mathbf{E}\left(\sum_{i \in I} \alpha_i \mathbf{1}_{A_i}\right) = \sum_{i \in I} \alpha_i \mathbf{P}(A_i|B) = \frac{1}{\mathbf{P}(B)} \mathbf{E}\left(\sum_{i \in I} \alpha_i \mathbf{1}_{A_i} \mathbf{1}_B\right).$$

Pour des variables aléatoires X intégrables, on procède par approximation, puis, si $X \in \mathbb{R}^d$, on raisonne composantes par composantes. □

Soit $(B_i)_{i \in I}$ une partition dénombrable (modulo 0) de Ω formée d'ensembles \mathcal{F} -mesurables et posons $\mathcal{G} = \sigma(B_i, i \in I)$. Pour une variable aléatoire X supposée \mathbf{P} -intégrable, on définit la variable aléatoire à valeurs dans \mathbb{R}^d et définie sur $(\Omega, \mathcal{F}, \mathbf{P})$ par l'égalité

$$\mathbf{E}(X|\mathcal{G})(\omega) = \sum_{i \in I} \mathbf{E}(X|B_i)\mathbf{1}_{B_i}(\omega).$$

Dans cette expression, si $\mathbf{P}(B_i) = 0$, alors on pose arbitrairement $\mathbf{E}(X|B_i) = 0$. *In fine*, cette égalité est définie presque-sûrement.

Proposition 13.1.8. *La variable aléatoire $\mathbf{E}(X|\mathcal{G})$ est \mathcal{G} -mesurable. De plus, si X est \mathbf{P} -intégrable, alors il en va de même de $\mathbf{E}(X|\mathcal{G})$. De plus, pour toute variable aléatoire \mathcal{G} -mesurable bornée Z , on a*

$$\int_{\Omega} Z\mathbf{E}(X|\mathcal{G}) d\mathbf{P} = \int_{\Omega} ZX d\mathbf{P} \iff \mathbf{E}(Z\mathbf{E}(X|\mathcal{G})) = \mathbf{E}(ZX).$$

Démonstration. L'application notée $\mathbf{E}(X|\mathcal{G})$ est (limite d') une fonction étagée sur des ensembles \mathcal{F} -mesurables et même \mathcal{G} -mesurables, c'est donc une variable aléatoire \mathcal{G} -mesurable. L'égalité des espérances est triviale si Z est \mathcal{G} -étagée bornée. \square

On a ainsi défini une espérance conditionnellement à une sous-tribu engendrée par une partition qui est consistante avec la définition de probabilité conditionnelle. Peut-on faire de même avec une sous-tribu arbitraire? La réponse est oui.

13.2 Espérance conditionnelle

On se sert des propriétés de l'espérance conditionnelle sachant une sous-tribu engendré par une partition pour proposer une définition de l'espérance conditionnelle en général, *i.e.*, pour des sous-tribus arbitraires. Puis, on vérifie que cette définition est consistante dans le sens qu'une telle espérance conditionnelle existe effectivement et qu'elle est caractérisée par les conditions de la définition (unicité en un certain sens).

On ne montrera pas dans ce cours le résultat de dualité entre partitions mesurables et σ -algèbres, mais il est important d'avoir conscience qu'il s'agit peu ou prou de la même notion.

Définition 13.2.1 (Espérance conditionnelle). Soient $\mathcal{G} \subset \mathcal{F}$ une sous-tribu et X une variable aléatoire \mathbf{P} -intégrable. Une variable aléatoire Y est appelée espérance conditionnelle de X sachant \mathcal{G} , et on note $Y = \mathbf{E}(X|\mathcal{G})$ si

1. Y est \mathcal{G} -mesurable;
2. pour toute variable aléatoire \mathcal{G} -mesurable bornée Z , $\mathbf{E}(YZ) = \mathbf{E}(XZ)$.

Si $B \in \mathcal{F}$, alors $\mathbf{P}(B|\mathcal{G}) = \mathbf{E}(\mathbf{1}_B|\mathcal{G})$ est appelée probabilité conditionnelle de B sachant \mathcal{G} .

La proposition suivante donne une caractérisation équivalente de l'espérance conditionnelle. Elle s'avère parfois plus commode.

Proposition 13.2.2. *Soient $\mathcal{G} \subset \mathcal{F}$ une sous-tribu et X une variable aléatoire \mathbf{P} -intégrable. Alors une variable aléatoire \mathcal{G} -mesurable Y est l'espérance conditionnelle de X sachant \mathcal{G} si et seulement si, pour tout $A \in \mathcal{G}$, $\mathbf{E}(\mathbf{1}_A Y) = \mathbf{E}(\mathbf{1}_A X)$.*

Démonstration. Dans la définition de l'espérance conditionnelle, en posant $Z = \mathbf{1}_A$, l'égalité $\mathbf{E}(\mathbf{1}_A Y) = \mathbf{E}(\mathbf{1}_A X)$ est immédiate.

Réciproquement, si pour tout $A \in \mathcal{G}$, $\mathbf{E}(\mathbf{1}_A Y) = \mathbf{E}(\mathbf{1}_A X)$, alors il en va de même pour toute variable aléatoire \mathcal{G} -mesurable Z étagée positive, c'est à dire $Z = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$ avec $A_i \in \mathcal{G}$ et $\alpha_i \geq 0$, $i = 1, \dots, n$. Par le théorème de convergence monotone, cette égalité reste valide pour les variables aléatoires \mathcal{G} -mesurables positives. On conclut pour toute variable aléatoire \mathcal{G} -mesurable Z en décomposant $Z = Z^+ - Z^-$, les parties positives et négatives étant trivialement \mathcal{G} -mesurables. \square

Théorème 13.2.3. *Si $\mathcal{G} \subset \mathcal{F}$ est une sous-tribu et X une variable aléatoire \mathbf{P} -intégrable, alors $\mathbf{E}(X|\mathcal{G})$ existe et est unique.*

Démonstration. On commence par montrer l'unicité. Soit Y et Y' satisfaisant la condition de l'espérance conditionnelle, *i.e.* pour toute variable aléatoire Z \mathcal{G} -mesurable bornée $\mathbf{E}(XZ) = \mathbf{E}(YZ) = \mathbf{E}(Y'Z)$. On pose $Z = \mathbf{1}_{Y > Y'}$, qui est \mathcal{G} -mesurable et bornée. On obtient que $0 = \mathbf{E}[(Y - Y')\mathbf{1}_{Y > Y'}]$, d'où $Y = Y'$ presque-sûrement en intervertissant les rôles de Y et Y' .

Pour l'existence, quitte à raisonner composantes par composantes, on peut supposer X à valeurs réelles. On décompose alors $X = X^+ - X^-$ en partie positive et négative. Traitons le cas de la partie positive et définissons la mesure positive Q par

$$Q(A) = \mathbf{E}(X^+ \mathbf{1}_A), \quad A \in \mathcal{G}.$$

Alors, Q est absolument continue par rapport à \mathbf{P} et le théorème de Radon-Nikodym implique qu'il existe une densité \mathcal{G} -mesurable Y^+ tel que $Q(A) = \mathbf{E}(Y^+ \mathbf{1}_A) = \mathbf{E}(X^+ \mathbf{1}_A)$. On a de même une densité \mathcal{G} -mesurable Y^- tel que $\mathbf{E}(Y^- \mathbf{1}_A) = \mathbf{E}(X^- \mathbf{1}_A)$. Ainsi, il existe une variable aléatoire \mathcal{G} -mesurable $Y = Y^+ - Y^-$ tel que pour tout $A \in \mathcal{G}$, $\mathbf{E}(Y \mathbf{1}_A) = \mathbf{E}(X \mathbf{1}_A)$. \square

Notons que si $\mathcal{G} \subset \mathcal{F}$ est une sous-tribu, alors la probabilité conditionnelle sachant \mathcal{G} est définie comme $\mathbf{P}(B|\mathcal{G}) = \mathbf{E}(\mathbf{1}_B|\mathcal{G})$ pour tout $B \in \mathcal{F}$. Si X est une variable aléatoire de \mathbb{R}^d , on appelle loi conditionnelle de X sachant \mathcal{G} la probabilité sur \mathbb{R}^d qui à chaque borélien A assigne la probabilité $\mathbf{P}(A|\mathcal{G})$. Notez que c'est une variable aléatoire.

Définition 13.2.4. Si X, Y sont deux variables aléatoires telles que $X \in \mathbf{L}^1$, on définit l'espérance conditionnelle de X sachant Y , notée $\mathbf{E}(X|Y)$, comme l'espérance conditionnelle de X sachant la tribu engendrée par Y , *i.e.* $\mathbf{E}(X|Y) = \mathbf{E}(X|\sigma(Y))$.

13.3 Propriétés de l'espérance conditionnelle

Théorème 13.3.1. Soient $\mathcal{G} \subset \mathcal{H} \subset \mathcal{F}$ des sous-tribus et X, Y des variables aléatoires \mathbf{P} -intégrables et Z une variable aléatoire. Alors,

1. $\mathbf{E}[\mathbf{E}(X|\mathcal{G})] = \mathbf{E}[X]$ (formule des probabilités totales);
2. pour tout $\lambda \in \mathbb{R}$, $\mathbf{E}(\lambda X + Y|\mathcal{G}) = \lambda \mathbf{E}(X|\mathcal{G}) + \mathbf{E}(Y|\mathcal{G})$ p.s. (linéarité);
3. si $Y \leq X$ presque-sûrement, alors $\mathbf{E}(Y|\mathcal{G}) \leq \mathbf{E}(X|\mathcal{G})$ p.s. (monotonie);
4. si $\mathbf{E}|XY| < \infty$ et Y est \mathcal{G} -mesurable alors

$$\mathbf{E}(XY|\mathcal{G}) = Y\mathbf{E}(X|\mathcal{G}) \quad \text{p.s.} \quad \text{et} \quad \mathbf{E}(Y|\mathcal{G}) = \mathbf{E}(Y|Y) = Y \quad \text{p.s.};$$

5. $\mathbf{E}[\mathbf{E}(X|\mathcal{G})|\mathcal{H}] = \mathbf{E}[\mathbf{E}(X|\mathcal{H})|\mathcal{G}] = \mathbf{E}(X|\mathcal{G})$ p.s. (conditionnements emboîtés);
6. $|\mathbf{E}(X|\mathcal{G})| \leq \mathbf{E}(|X|\mathcal{G})$ p.s. (inégalité triangulaire);
7. si $\sigma(X)$ et \mathcal{G} sont deux tribus indépendantes, alors $\mathbf{E}(X|\mathcal{G}) = \mathbf{E}(X)$ p.s.;
8. Si pour tout $A \in \mathcal{G}$, $\mathbf{P}(A) \in \{0, 1\}$, alors $\mathbf{E}(X|\mathcal{G}) = \mathbf{E}(X)$ p.s..

Démonstration. 1. C'est la caractérisation de l'espérance conditionnelle appliquée à la variable aléatoire bornée $Z = \mathbf{1}_\Omega$.

2. Il est clair que $\lambda \mathbf{E}(X|\mathcal{G}) + \mathbf{E}(Y|\mathcal{G})$ est \mathcal{G} -mesurable. Soit Z une application \mathcal{G} -mesurable bornée. Alors, par linéarité de l'espérance et définition de l'espérance conditionnelle

$$\begin{aligned} \mathbf{E}(Z(\lambda \mathbf{E}(X|\mathcal{G}) + \mathbf{E}(Y|\mathcal{G}))) &= \lambda \mathbf{E}(Z\mathbf{E}(X|\mathcal{G})) + \mathbf{E}(Z\mathbf{E}(Y|\mathcal{G})) \\ &= \lambda \mathbf{E}(ZX) + \mathbf{E}(ZY) \\ &= \mathbf{E}(Z(\lambda X + Y)). \end{aligned}$$

3. Soit $Z = \mathbf{1}_{\mathbf{E}(X|\mathcal{G}) < \mathbf{E}(Y|\mathcal{G})}$. Clairement, Z est \mathcal{G} -mesurable bornée. En particulier, par définition de l'espérance conditionnelle, et puisque $X \geq Y$ presque-sûrement

$$0 \geq \mathbf{E}(Z(\mathbf{E}(X|\mathcal{G}) - \mathbf{E}(Y|\mathcal{G}))) = \mathbf{E}(Z(X - Y)) \geq 0.$$

Ainsi, $Z = 0$ presque-sûrement.

4. Soit $A \in \mathcal{G}$, et supposons d'abord que $Y = \mathbf{1}_B$ pour un $B \in \mathcal{G}$. Alors,

$$\mathbf{E}(\mathbf{1}_A \mathbf{E}(\mathbf{1}_B X|\mathcal{G})) = \mathbf{E}(\mathbf{1}_A \mathbf{1}_B X) = \mathbf{E}(\mathbf{1}_A \mathbf{1}_B \mathbf{E}(X|\mathcal{G}))$$

car $\mathbf{1}_A \mathbf{1}_B$ est \mathcal{G} -mesurable borné. Par linéarité de l'espérance (classique), cette relation est toujours satisfaite si Y est étagée positive, puis par convergence dominée en utilisant l'hypothèse $\mathbf{E}|XY| < \infty$, c'est encore vrai lorsque Y est \mathcal{G} -mesurable. Puisque $A \in \mathcal{G}$ est arbitraire, presque-sûrement $\mathbf{E}(XY|\mathcal{G}) = Y\mathbf{E}(X|\mathcal{G})$.

Pour la deuxième égalité, on utilise la première égalité avec $X = \mathbf{1}_\Omega$ et on remarque que $\mathbf{E}(\mathbf{1}_\Omega|\mathcal{G}) = 1$ presque-sûrement ($\mathbf{1}_\Omega$ est trivialement \mathcal{G} -mesurable et la variable aléatoire constante égale à 1 vérifie l'égalité de la définition de l'espérance conditionnelle).

5. Soit Z une variable aléatoire \mathcal{G} -mesurable bornée. Alors Z est également \mathcal{H} -mesurable bornée, ainsi, par les points 1) et 4),

$$\mathbf{E}(Z\mathbf{E}(\mathbf{E}[X|\mathcal{H}]|\mathcal{G})) = \mathbf{E}(\mathbf{E}(\mathbf{E}[ZX|\mathcal{H}]|\mathcal{G})) = \mathbf{E}(ZX)$$

Par définition de l'espérance conditionnelle, on obtient

$$\mathbf{E}(\mathbf{E}[X|\mathcal{H}]|\mathcal{G}) = \mathbf{E}(X|\mathcal{G}).$$

Pour l'autre égalité, on utilise le point 4) et le fait qu'une variable aléatoire \mathcal{G} -mesurable est \mathcal{H} -mesurable si $\mathcal{G} \subset \mathcal{H}$.

6. C'est une conséquence du point 1) et 2) en posant $X = X^+ - X^-$.
 7. Soit Z une variable aléatoire \mathcal{G} -mesurable bornée, alors $\sigma(X)$ et \mathcal{G} étant indépendantes, Z est une variable aléatoire indépendante de X . Ainsi,

$$\mathbf{E}(Z\mathbf{E}(X|\mathcal{G})) = \mathbf{E}(ZX) = \mathbf{E}(Z)\mathbf{E}(X) = \mathbf{E}(Z\mathbf{E}(X)).$$

Et il vient que $\mathbf{E}(X|\mathcal{G}) = \mathbf{E}(X)$ presque-sûrement.

8. Si \mathcal{G} est trivial, c'est à dire tout $A \in \mathcal{G}$ vérifie $\mathbf{P}(A) \in \{0, 1\}$, alors une variable aléatoire \mathcal{G} -mesurable bornée est presque-sûrement constante.

$$\mathbf{E}(Z\mathbf{E}(X|\mathcal{G})) = \mathbf{E}(ZX) = Z\mathbf{E}(X) = \mathbf{E}(Z\mathbf{E}(X)),$$

car Z et $Z\mathbf{E}(X)$ sont constantes presque-sûrement. □

Les théorèmes de convergences de type Beppo-Lévy, Fatou et convergence dominée de Lebesgue se généralisent facilement aux espérances conditionnelles.

Théorème 13.3.2 (Beppo-Lévy, Fatou, convergence dominée de Lebesgue). *Soit $\mathcal{G} \subset \mathcal{F}$ une sous-tribu de $(\Omega, \mathcal{F}, \mathbf{P})$*

1. Convergence monotone conditionnelle : *Soit $(X_n)_{n \geq 0}$ une suite croissante de variables aléatoires à valeurs réelles positives et \mathbf{P} -intégrables. Alors $\lim_{n \rightarrow \infty} X_n$ existe dans $[0, \infty]$ et*

$$\lim_{n \rightarrow \infty} \mathbf{E}(X_n|\mathcal{G}) = \mathbf{E}(X|\mathcal{G}), \quad p.s..$$

2. Lemme de Fatou conditionnel : *Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires positives et \mathbf{P} -intégrables. Alors*

$$\mathbf{E}(\liminf_{n \rightarrow \infty} X_n|\mathcal{G}) \leq \liminf_{n \rightarrow \infty} \mathbf{E}(X_n|\mathcal{G}).$$

3. *Soient Y une variable positive \mathbf{P} -intégrable et $(X_n)_{n \geq 0}$ une suite de variables aléatoires telles que $|X_n| \leq Y$ pour tout $n \geq 0$ et X_n converge vers X presque-sûrement. Alors*

$$\lim_{n \rightarrow \infty} \mathbf{E}(X_n|\mathcal{G}) = \mathbf{E}(X|\mathcal{G})$$

presque-sûrement et dans \mathbf{L}^1 .

Remarque 113. La version conditionnelle des lemmes de convergence monotone et de Fatou suppose que les variables X_n sont \mathbf{P} -intégrables, ceci pour assurer l'existence de l'espérance conditionnelle.

Démonstration. 1. On considère, pour tout $n \geq 0$, $Y_n = \mathbf{E}(X_n|\mathcal{G})$. La suite $(Y_n)_{n \geq 0}$ est monotone croissante de variable aléatoires positives par monotonie de l'espérance conditionnelle. Ainsi, la suite $(Y_n)_{n \geq 0}$ converge presque sûrement vers une variable aléatoire que l'on note Y . Il s'agit donc de montrer que $Y = \mathbf{E}(X|\mathcal{G})$. Pour ce faire, considérons $A \in \mathcal{G}$, alors par le théorème de convergence monotone classique, en notant $X = \lim_{n \rightarrow \infty} X_n$

$$\mathbf{E}(X\mathbf{1}_A) = \lim_{n \rightarrow \infty} \mathbf{E}(X_n\mathbf{1}_A) = \lim_{n \rightarrow \infty} \mathbf{E}(Y_n\mathbf{1}_A) = \mathbf{E}(Y\mathbf{1}_A).$$

2. On applique le point 1) à la suite $Y_n = \inf_{k \geq n} X_k$ et la monotonie de l'espérance conditionnelle.

3. On définit $W_n = \sup_{k \geq n} |X_k - X|$. Alors $0 \leq W_n \leq 2Y$ et W_n converge vers 0 presque-sûrement. Donc $\mathbf{E}(W_n)$ converge vers 0. Par l'inégalité triangulaire

$$\mathbf{E}|\mathbf{E}(X_n|\mathcal{G}) - \mathbf{E}(X|\mathcal{G})| \leq \mathbf{E}(\mathbf{E}(|X_n - X||\mathcal{G})) = \mathbf{E}(|X_n - X|) = \mathbf{E}(W_n),$$

si bien $\mathbf{E}(X_n|\mathcal{G})$ converge vers $\mathbf{E}(X|\mathcal{G})$ dans \mathbf{L}^1 . Comme $(W_n)_{n \geq 0}$ est décroissante, la monotonie de l'espérance conditionnelle implique que $(\mathbf{E}(W_n|\mathcal{G}))_{n \geq 0}$ est également décroissante et converge presque-sûrement vers une variable aléatoire $W \geq 0$. Alors par le lemme de Fatou

$$0 \leq \mathbf{E}(W) \leq \liminf_{n \rightarrow \infty} \mathbf{E}\mathbf{E}(W_n|\mathcal{G}) = \lim_{n \rightarrow \infty} \mathbf{E}(W_n) = 0.$$

Par conséquent, $W = 0$ presque-sûrement et $\mathbf{E}(W_n|\mathcal{G})$ converge vers 0 presque-sûrement. Mais,

$$|\mathbf{E}(X_n|\mathcal{G}) - \mathbf{E}(X|\mathcal{G})| \leq \mathbf{E}(W_n|\mathcal{G}).$$

□

13.4 Inégalité de Jensen et de Markov conditionnelles

Proposition 13.4.1. Soit $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction convexe et X une v.a.r. telle que X et $\varphi(X) \in \mathbf{L}^1$, alors $\varphi(\mathbf{E}(X|\mathcal{G})) \leq \mathbf{E}(\varphi(X)|\mathcal{G})$.

Remarque 114. Pour ne pas se tromper dans le sens de l'inégalité, penser à la fonction valeur absolue.

Démonstration. La version conditionnelle de l'inégalité de Jensen se montre de la même façon que la version classique.

Il est connu que pour tout $x_0 \in \mathbb{R}$ il existe $a, b \in \mathbb{R}$ (qui dépendent de x_0 et non nécessairement uniques) tels que

$$\forall x \in \mathbb{R}, \quad \varphi(x) \geq ax + b \quad \text{et} \quad \varphi(x_0) = ax_0 + b.$$

Choisissons $x_0 = \mathbf{E}(X|\mathcal{G})$ et $x = X$ puis après passage aux espérances conditionnelles, il vient par linéarité de l'espérance conditionnelle

$$\varphi(\mathbf{E}(X|\mathcal{G})) = a\mathbf{E}(X|\mathcal{G}) + b = \mathbf{E}(aX + b|\mathcal{G}) \leq \mathbf{E}(\varphi(aX + b)|\mathcal{G}).$$

□

Proposition 13.4.2 (Inégalité de Markov conditionnelle). Soit $\mathcal{G} \subset \mathcal{F}$ une sous-tribu et X une variable aléatoire réelle positive. Alors,

$$\forall \lambda > 0 : \quad \mathbf{P}(X > \lambda|\mathcal{G}) \leq \frac{\mathbf{E}(X|\mathcal{G})}{\lambda}.$$

Remarque 115. Bien entendu, comme dans le cas de l'espérance classique, on peut aussi montrer une inégalité de Bienaymé-Tchebychev conditionnelle.

Démonstration. Encore une fois, La preuve est très identique à celle donnée dans le cas l'espérance classique. En fait, $X > \lambda \mathbf{1}_{X > \lambda}$ presque-sûrement, puis en utilisant la monotonie de l'espérance conditionnelle, on obtient $\mathbf{E}(X|\mathcal{G}) \geq \mathbf{E}(\lambda \mathbf{1}_{X > \lambda}|\mathcal{G}) = \lambda \mathbf{P}(X > \lambda|\mathcal{G})$. □

13.5 Conditionnement des vecteurs gaussiens

Proposition 13.5.1. Soit (Y, X_1, \dots, X_d) un vecteur gaussien de \mathbb{R}^{d+1} tel que $X = (X_1, \dots, X_d)$ possède une matrice de variance inversible Σ . Notons $a = \Sigma^{-1}(\text{cov}(Y, X_1), \dots, \text{cov}(Y, X_d))$, alors

$$\mathbf{E}(Y|X_1, \dots, X_d) = \mathbf{E}(Y) + a^*(X - \mathbf{E}X).$$

En particulier, une variable (ou vecteur) gaussienne conditionnée par rapport à un vecteur gaussien est encore gaussienne.

Démonstration. On suppose que (Y, X_1, \dots, X_d) est centré et on note $\hat{Y} = a^*X$. On vérifie facilement que $\text{cov}(Y - \hat{Y}, X_i) = \mathbf{E}[(Y - \hat{Y})X_i] = 0$ pour tout $i = 1, \dots, d$. Ainsi, puisque le vecteur $(X_1, \dots, X_d, Y - \hat{Y})$ est gaussien, que $(Y - \hat{Y})$ est indépendant de (X_1, \dots, X_d) , on obtient que

$$\mathbf{E}(Y|X_1, \dots, X_d) = \mathbf{E}(Y - \hat{Y}|X_1, \dots, X_d) + \hat{Y} = \mathbf{E}(Y - \hat{Y}) + \hat{Y} = \hat{Y}.$$

□

13.6 Point de vue hilbertien des espérances conditionnelles

L'espérance conditionnelle lorsque X est de carré intégrable s'interprète géométriquement dans le cadre de la théorie des espaces de Hilbert. L'application de la théorie des espaces de Hilbert aux espérances conditionnelles est illustrée par le théorème suivant.

Théorème 13.6.1. *Soient X une variable aléatoire de carré \mathbf{P} -intégrable et $\mathcal{G} \subset \mathcal{F}$ une sous-tribu. L'espace des variables aléatoires \mathcal{G} -mesurables de carré \mathbf{P} -intégrable est un s.e.v. fermé noté $F_{\mathcal{G}}$, et l'espérance conditionnelle de X sachant \mathcal{G} est la projection orthogonale de X sur $F_{\mathcal{G}}$.*

Démonstration. On doit montrer que pour toute variable aléatoire \mathcal{G} -mesurable Y telle que $\mathbf{E}|Y|^2 < \infty$, on a

$$\mathbf{E}((X - Y)^2) \geq \mathbf{E}((X - \mathbf{E}(X|\mathcal{G}))^2),$$

avec égalité si $Y = \mathbf{E}(X|\mathcal{G})$.

On vérifie tout d'abord que $\mathbf{E}[\mathbf{E}(X|\mathcal{G})^2] < \infty$ par l'inégalité de Jensen. Soit Y une variable aléatoire \mathcal{G} -mesurable telle que $\mathbf{E}(Y^2) < \infty$, alors par Cauchy-Schwarz, XY est intégrable. Alors, d'une part $\mathbf{E}(XY) = \mathbf{E}(Y\mathbf{E}(X|\mathcal{G}))$ et d'autre part,

$$\mathbf{E}(X\mathbf{E}(X|\mathcal{G})) = \mathbf{E}[\mathbf{E}(X\mathbf{E}\{X|\mathcal{G}\}|\mathcal{G})] = \mathbf{E}(\mathbf{E}\{X|\mathcal{G}\}^2).$$

En utilisant ces deux égalités, le calcul suivant termine la preuve

$$\begin{aligned} \mathbf{E}[(X - Y)^2] - \mathbf{E}[(X - \mathbf{E}(X|\mathcal{G}))^2] &= \mathbf{E}[X^2 - 2XY + Y^2 - X^2 + 2X\mathbf{E}(X|\mathcal{G}) - \mathbf{E}(X|\mathcal{G})^2] \\ &= \mathbf{E}[Y^2 - 2Y\mathbf{E}(X|\mathcal{G}) + \mathbf{E}(X|\mathcal{G})^2] \\ &= \mathbf{E}[(Y - \mathbf{E}(X|\mathcal{G}))^2] \geq 0. \end{aligned}$$

□

Remarque 116. Dans la littérature il existe deux points de vues équivalents pour construire l'espérance conditionnelle : ou bien, on utilise le théorème de Radon-Nikodym, ou bien on utilise le théorème de projection dans un Hilbert qui permet de définir l'espérance conditionnelle pour des variables aléatoires de carré intégrable. La seconde méthode s'étend facilement aux variables aléatoires intégrables. En réalité, les deux méthodes sont strictement équivalentes puisque le théorème de Radon-Nikodym découle du théorème de projection. Néanmoins, les deux points de vues restent intéressants pour eux-mêmes, l'un est très centré sur la théorie de la mesure alors que l'autre est plus géométrique.

13.7 Lois conditionnelles régulières

13.7.1 Densité conditionnelle

Lorsque la variable aléatoire par rapport à laquelle on conditionne, que l'on note Y ici, est à valeurs discrètes, disons dans \mathbb{N} pour simplifier, il est facile de définir la notion de loi conditionnelle de X sachant Y : c'est la famille de lois $\{\mathbf{P}_{X|Y=y}, y \in \mathbb{N}\}$ telle que si $\mathbf{P}(Y = y) > 0$,

$$\mathbf{P}_{X|Y=y}(B) = \mathbf{P}(X \in B|Y = y), \quad B \in \mathcal{B}(\mathbb{R}),$$

et dans le cas contraire, $\mathbf{P}_{X|Y=y}$ est une probabilité quelconque.

Malheureusement, lorsque la variable aléatoire Y n'est plus à valeurs discrètes, ce procédé ne s'étend pas aussi simplement : si par exemple Y est de loi continue, $\mathbf{P}(Y = y) = 0$ pour tout $y \in \mathbb{R}$. L'objet

de cette section est de contourner ce problème et plus spécifiquement d'énoncer le théorème 13.7.3 dont la démonstration (partielle) sera donnée à la fin de cette section. Ce théorème permet de généraliser la formule de Bayes aux lois à densités. On commence par un lemme technique qui peut être également utile dans d'autres contextes.

Lemme 13.7.1. *Soit X une variable aléatoire à valeurs dans \mathbb{R}^d et soit Z une variable aléatoire $\sigma(X)$ -mesurable à valeurs dans \mathbb{R}^p . Alors il existe une application $h : \mathbb{R}^d \rightarrow \mathbb{R}^p$ mesurable telle que $Z = h(X)$. De plus h est définie \mathbf{P}_X -p.s..*

Rappelons que $\sigma(X)$ est la plus petite tribu sur Ω rendant X mesurable. En particulier, puisque X est \mathcal{F} -mesurable (c'est une variable aléatoire), nous avons $\sigma(X) \subset \mathcal{F}$.

Démonstration. Si $Z = h(X)$ alors Z est clairement $\sigma(X)$ -mesurable puisque, pour tout borélien $A \in \mathcal{B}(\mathbb{R}^p)$, $h^{-1}A \in \mathcal{B}(\mathbb{R}^d)$ et donc, puisque X est $\sigma(X)$ -mesurable par définition, il vient que $X^{-1}h^{-1}(A) \in \sigma(X)$.

Pour la réciproque, il suffit de montrer que tout $A \in \sigma(X)$ s'écrit $X^{-1}B$ pour $B \in \mathcal{B}(\mathbb{R}^d)$. Pour ce faire, notons

$$\mathcal{A} = \{A \in \sigma(X) : \exists B \in \mathcal{B}(\mathbb{R}^d), X^{-1}B = A\}.$$

Clairement, $\mathcal{A} \subset \sigma(X)$. De plus, $\emptyset = X^{-1}\emptyset \in \mathcal{A}$ alors que \mathcal{A} est stable par passage au complémentaire et réunion dénombrable, ce sont les propriétés de l'image réciproque. Ainsi, \mathcal{A} est une tribu contenu dans $\sigma(X)$. Cependant, il est facile de voir que X est \mathcal{A} -mesurable, donc $\mathcal{A} = \sigma(X)$.

Ce résultat montre que si Z est une application $\sigma(X)$ -mesurable étagée positive, alors il existe $\alpha_1, \dots, \alpha_n \in \mathbb{R}_+$ et des boréliens B_1, \dots, B_n de $\mathcal{B}(\mathbb{R}^d)$ tels que

$$Z = \sum_{i=1}^n \alpha_i \mathbf{1}_{X^{-1}B_i} = \sum_{i=1}^n \alpha_i \mathbf{1}_{B_i} \circ X.$$

Il suffit alors de poser $h = \sum_{i=1}^n \alpha_i \mathbf{1}_{B_i}$ qui est une application borélienne étagée positive de \mathbb{R}^d dans \mathbb{R}_+ . L'extension du cas Z étagée positive à Z à valeurs dans \mathbb{R}^p se fait comme d'habitude : Z positive par approximation, Z réelle en décomposant en partie positive et partie négative, Z à valeurs dans \mathbb{R}^p en raisonnant composantes par composantes. □

Ce lemme permet de donner un sens à la notation $\mathbf{E}(X|Y = y)$, d'où la définition suivante.

Définition 13.7.2. Soit X à valeurs dans \mathbb{R}^p et \mathbf{P} -intégrable et soit Y une variable aléatoire à valeurs dans \mathbb{R}^q . On définit l'espérance conditionnelle de X sachant $Y = y$ par $\mathbf{E}(X|Y = y) = \varphi(y)$ où φ est une fonction mesurable satisfaisant $\varphi(Y) = \mathbf{E}(X|Y)$. De même, pour $A \in \mathcal{F}$, on définit $\mathbf{P}(A|Y = y) = \mathbf{E}(\mathbf{1}_A|Y = y)$.

Théorème 13.7.3. *Supposons que la loi jointe de (X, Y) admette une densité f par rapport à la mesure de Lebesgue λ_{p+q} . Alors, Y admet une densité f_Y et la loi conditionnelle régulière de X sachant $Y = y$ admet une densité pour \mathbf{P}_Y -p.t. $y \in \mathbb{R}^q$, notée $f_{X|Y=y}$, définie pour $x \in \mathbb{R}^p$ par*

$$f_{X|Y=y}(x) = \frac{f(x, y)}{f_Y(y)} \mathbf{1}_{\{f_Y > 0\}}.$$

De plus, pour toute fonction réelle φ mesurable telle que $\varphi(X) \in \mathbf{L}^1$ et pour \mathbf{P}_Y -p.t. $y \in \mathbb{R}^q$,

$$\mathbf{E}(\varphi(X)|Y = y) = \int_{\mathbb{R}^p} \varphi(x) f_{X|Y=y}(x) dx.$$

Remarque 117. Le \mathbf{P}_Y -p.p. provient du fait que $f_Y(y) = \int_{\mathbb{R}^p} f(x, y) \lambda_p(dx)$ est seulement définie \mathbf{P}_Y -p.p.. Dans les faits, elle est souvent définie partout — typiquement si $y \rightarrow f(x, y)$ est continue pour presque tout $x \in \mathbb{R}^p$, mais pas seulement. Dans ce cas, l'expression de $f_{X|Y=y}$ est toujours valide.

Exemple 52. Soit $(X, Y) \in \mathbb{R}^2$ de loi jointe $f_{(X,Y)}(x, y) = \mathbf{1}_{\mathbb{R}_+ \times \mathbb{R}_+}(x)ye^{-yx}e^{-y}$. Alors, Y admet pour densité

$$f_Y(y) = \mathbf{1}_{\mathbb{R}_+}(y)e^{-y} \int_0^\infty ye^{-yx} dx = \mathbf{1}_{\mathbb{R}_+}(y)e^{-y}.$$

La densité conditionnelle $f_{X|Y=y}$ est donnée pour presque tout $y \in \mathbb{R}_+$ par

$$f_{X|Y=y}(x) = \mathbf{1}_{\mathbb{R}_+}(x)ye^{-yx}.$$

Autrement dit, Y suit une loi exponentielle de paramètre 1 et la loi de X sachant Y est une loi exponentielle de paramètre (aléatoire) Y , on note $\mathcal{L}(X|Y) = \mathcal{E}(Y)$.

Corollaire 13.7.4. *Supposons X et Y indépendantes et de densités respectives, par rapport à λ_p et λ_q , f_X et f_Y . Alors la loi conditionnelle régulière de X sachant Y admet f_X pour densité.*

13.7.2 Noyau de transition et loi conditionnelle régulière

Dans toute la suite, X et Y sont des variables aléatoires définies sur un même espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$ et à valeurs dans \mathbb{R}^p et \mathbb{R}^q respectivement. Tous ces résultats restent valables si X et Y sont à valeurs dans un espace polonais, le lemme 13.7.1 étant même vrai si elles sont à valeurs dans un espace mesurable.

Définition 13.7.5. Une application $K : \mathbb{R}^q \times \mathcal{B}(\mathbb{R}^p) \rightarrow [0, 1]$ est appelée noyau de transition si les deux conditions suivantes sont satisfaites :

1. pour tout $x \in \mathbb{R}^q$, $K(x, \cdot)$ est une probabilité sur $\mathcal{B}(\mathbb{R}^p)$;
2. pour tout $A \in \mathcal{B}(\mathbb{R}^p)$, l'application $x \rightarrow K(x, A)$ est $\mathcal{B}(\mathbb{R}^q)$ -mesurable.

Si K est un noyau de transition, alors on peut faire agir une probabilité μ à gauche et une fonction mesurable bornée f à droite de telle sorte que

$$\langle \mu K, f \rangle = \langle \mu, Kf \rangle = \int_{\mathbb{R}^q \times \mathbb{R}^p} \mu(dx)K(x, dy)f(y).$$

Plus directement, on définit la probabilité μK pour tout $A \in \mathcal{B}(\mathbb{R}^p)$ par

$$\mu K(A) = \int_{\mathbb{R}^q} \mu(dx)K(x, A),$$

et on définit la fonction Kf pour tout $x \in \mathbb{R}^q$

$$Kf(x) = \int_{\mathbb{R}^p} K(x, dy)f(y).$$

Définition 13.7.6. Soient X et Y des variables aléatoires à valeurs dans \mathbb{R}^p et \mathbb{R}^q respectivement. Un noyau de transition $K : \mathbb{R}^q \times \mathcal{B}(\mathbb{R}^p) \rightarrow [0, 1]$ est appelée loi conditionnelle régulière de X sachant Y si pour presque tout $y \in \mathbb{R}^q$

$$\mathbf{P}(X \in B|Y = y) = K(y, B) \quad \text{ou} \quad \mathbf{E}(\varphi(X)|Y = y) = K\varphi(y).$$

où $B \in \mathcal{B}(\mathbb{R}^p)$ et $\varphi : \mathbb{R}^p \rightarrow \mathbb{R}$ borélienne bornée.

Au vu du lemme de factorisation 13.7.1, on voit tout de suite que pour chaque fonction φ borélienne bornée on peut associer une fonction h_φ telle que $\mathbf{E}(\varphi(X)|Y = y) = h_\varphi(y)$. Cela définit un opérateur linéaire (dans un sens généralisé), c'est à dire un sous-espace linéaire de $m\mathcal{M}_b(\mathbb{R}^p) \times m\mathcal{M}_b(\mathbb{R}^q)$, où $m\mathcal{M}_b(\mathbb{R}^d)$ représente l'espace des fonctions mesurables (partout) bornées sur \mathbb{R}^d . Cependant, la fonction h_φ n'est définie que \mathbf{P}_Y -presque partout et l'ensemble négligeable N correspondant dépend *a priori* de φ et X . Mais l'espace des fonctions φ boréliennes bornées n'est en général pas dénombrable, ainsi on ne peut pas trouver un ensemble négligeable N universel, c'est à dire valable pour chaque fonction φ . Néanmoins, on vérifie facilement que si φ est \mathbf{P}_X -intégrable alors h_φ est \mathbf{P}_Y -intégrable. En effet :

$$\mathbf{E}|h_\varphi(Y)| = \mathbf{E}|\mathbf{E}(\varphi(X)|Y)| \leq \mathbf{E}|\varphi(X)|.$$

Ainsi, on peut restreindre l'opérateur linéaire à l'espace $\mathbf{L}_{\mathbf{P}_X}^1(\mathbb{R}^q) \times \mathbf{L}_{\mathbf{P}_Y}^1(\mathbb{R}^q)$. Or il se trouve que \mathbf{L}^1 est séparable, c'est à dire contient un sous-ensemble dénombrable dense (c'est le cas dès que la tribu est engendré par une famille dénombrable de parties). Ainsi, pour chacune des fonctions dans cette partie dense, il sort un certain ensemble négligeable, leur réunion dénombrable est toujours négligeable et convient à toutes les fonctions de la partie dense. Finalement, à l'aide d'un argument de continuité, on peut choisir l'ensemble négligeable de façon universelle. C'est ce que raconte le théorème suivant (admis), il implique l'existence de la loi conditionnelle de X sachant Y .

Théorème 13.7.7. *Soient X et Y des variables aléatoires à valeurs dans \mathbb{R}^p et \mathbb{R}^d respectivement. Alors, il existe un noyau de transition $K : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^p) \rightarrow [0, 1]$ tel que $\mathbf{P}_X = \mathbf{P}_Y K$.*

Ce théorème permet de donner un sens à l'écriture $\mathbf{P}_{X|Y=y}$ puisqu'en l'espèce

$$\mathbf{P}_{X|Y=y}(A) = \mathbf{P}(X \in A | Y = y) = K(y, A), \quad A \in \mathcal{B}(\mathbb{R}^p).$$

De même, pour une fonction $\varphi : \mathbb{R}^p \rightarrow \mathbb{R}$, on aura

$$\mathbf{E}(\varphi(X) | Y = y) = \int_{\mathbb{R}^p} K(y, dx) \varphi(x).$$

On peut désormais montrer le théorème 13.7.3.

Preuve du théorème 13.7.3. Calculons pour $h \times g : \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}$ borélienne bornée

$$\begin{aligned} \int_{\mathbb{R}^{p+q}} h \times g(x, y) f(x, y) \, dx dy &= \mathbf{E}(h \times g(X, Y)) \\ &= \mathbf{E}(g(Y) \mathbf{E}(h(X) | Y)) \\ &= \mathbf{E}(g(Y) K h(Y)) \\ &= \int_{\mathbb{R}^q} g(y) f_Y(y) \int_{\mathbb{R}^p} h(x) K(y, dx) \, dy \\ &= \int_{\mathbb{R}^{p+q}} h \times g(x, y) f_Y(y) K(y, dx) \, dy, \end{aligned}$$

où la dernière égalité provient du théorème de Fubini. Ainsi, presque-partout, $K(y, \cdot)$ est absolument continue par rapport à λ_p , on note $f_{X|Y=y}$ sa densité. Alors $f_{(X,Y)}(x, y) = f_Y(y) f_{X|Y=y}(x)$ presque-partout. \square

Lois usuelles

13.8 Lois discrètes

Nom de la loi	Support	$\mathbf{P}(X = k)$	$\mathbf{E}(X)$	$\mathbf{V}(X)$
Uniforme	$\{1, \dots, n\}$	$\frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$
Bernoulli, $\mathcal{B}(p)$	$\{0, 1\}$	$\{1-p, p\}$	p	$p(1-p)$
Binomiale, $\mathcal{B}(n, p)$	$\{0, \dots, n\}$	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
Hypergéométrique	$\{0, \dots, n\}$	$\frac{\binom{N_1}{k} \binom{N_2}{n-k}}{\binom{N_1+N_2}{n}}$	$\frac{nN_1}{N_1+N_2}$	$\frac{nN_1N_2(N_1+N_2-n)}{(N_1+N_2)^2(N_1+N_2-1)}$
Géométrique, $\mathcal{G}(p)$	$\mathbf{N} \setminus \{0\} = \mathbf{N}^*$	$p(1-p)^{k-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson, $\mathcal{P}(a)$	\mathbf{N}	$e^{-a} \frac{a^k}{k!}$	a	a

13.9 Lois continues

Nom de la loi	Support	Densité $f(x)$	$\mathbb{E}(X)$	$\mathbb{V}(X)$
Uniforme, $\mathcal{U}[a, b]$	$[a, b]$	$\frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Gaussienne, $\mathcal{N}(\mu, \sigma^2)$	\mathbb{R}	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Exponentielle, $\mathcal{E}(\lambda)$	\mathbb{R}_+	$\lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma, $\Gamma(a, \theta)$	\mathbb{R}_+^*	$\frac{\theta^a}{\Gamma(a)} e^{-\theta x} x^{a-1} \mathbf{1}_{\mathbb{R}_+^*}(x)$	$\frac{a}{\theta}$	$\frac{a}{\theta^2}$
Chi-deux, $\chi^2(d)$	\mathbb{R}_+	$\frac{1}{2^{d/2}\Gamma(d/2)} x^{d/2-1} e^{-x/2} \mathbf{1}_{\mathbb{R}_+}(x)$	d	$2d$
Cauchy (centrée), $\mathcal{C}(a)$	\mathbb{R}	$\frac{a}{\pi(a^2+x^2)}$	NON DÉFINIE	NON DÉFINIE

Fonction Gamma : $\Gamma(z) := \int_0^\infty t^{z-1} e^{-t} dt, \quad \text{Re } z > 0.$

Bibliographie

- [Bil68] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, Inc., New York-London-Sydney, 1968.
- [BP04] M. Briane and G. Pagès. *Théorie de l'intégration : Cours et exercices, licence & master de mathématiques*. Vuibert, 2004.
- [Car67] Henri Cartan. *Calcul différentiel*. Hermann, Paris, 1967.
- [Eri73] K. Bruce Erickson. The strong law of large numbers when the mean is undefined. *Trans. Amer. Math. Soc.*, 185 :371–381 (1974), 1973.
- [Kes70] Harry Kesten. The limit points of a normalized random walk. *Ann. Math. Statist.*, 41 :1173–1205, 1970.
- [Kin73] J. F. C. Kingman. Subadditive ergodic theory. *Ann. Probability*, 1 :883–909, 1973. With discussion by D. L. Burkholder, Daryl Daley, H. Kesten, P. Ney, Frank Spitzer and J. M. Hammersley, and a reply by the author.
- [Nev70] Jacques Neveu. *Bases mathématiques du calcul des probabilités*. Masson et Cie, Éditeurs, Paris,, 1970. Préface de R. Fortet, Deuxième édition, revue et corrigée.
- [Rud87] Walter Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.
- [Spi76] Frank Spitzer. *Principles of random walks*. Springer-Verlag, New York, second edition, 1976. Graduate Texts in Mathematics, Vol. 34.