Anaël Marit ENS Rennes

THÉORIE DE LA MESURE

- échanger la difficulté contre l'abstraction –

une approche qui s'espère au goût des algébristes

(et pour réconcillier certain.e.s avec les probabilités)

—

Table des matières

Introduction : pour qui et pour quoi ce document ?

Sans conteste, la théorie de la mesure est l'un des gros morceaux du programme de licence. Personnellement (et j'imagine que c'est également le cas pour un certain nombre d'autre étudiantes), j'ai accusé le choc, puis compris qu'en analyse, on n'intègre pour ainsi dire jamais ailleurs que dans $\mathbb{R}^d$ muni de la mesure de Lebesgue, et que toutes ces affreuses constructions abstraites, pleines d'écueils contre lesquels il est facile de s'échouer, pouvaient être reléguées sous le tapis.

En 2026, j'ai préparé l'agrégation externe, et ç'a été l'occasion pour moi de redécouvrir la théorie de la mesure. J'ai notamment pris conscience de deux choses : les mesures abstraites, ça existe en pratique pour peu qu'on veuille bien les voir, et qu'alors les théorèmes élémentaires de théorie de la mesure permettent bien souvent de briser un problème qui paraissait affreux ; et par ailleurs, la théorie de la mesure offre un terrain de jeu dont le niveau d'abstraction savait parler à mon esprit d'algébriste, rendant amusantes plusieurs des leçons d'analyse qui me rebutaient a priori [(i)].

L'objectif de ce document, que j'ai rédigé en grande partie pendant mon année d'agrégation, est de mettre en forme cette approche de la théorie de la mesure que j'ai redécouverte en cette année d'agrégation. **Ce document ne s'adresse pas, ou peu, à des étudiantes qui découvrent la théorie de la mesure.** J'ai considéré que les définitions de bases sont connues ; le but est bien plus de donner du recul sur ces notions que d'en proposer une initiation.

J'ai cherché à m'adresser notamment à mes camarades agrégatives. Quelques petits mots pour celles-ci : l'approche suivie dans ce document va beaucoup plus loin dans l'abstraction que ce qui est attendu par le jury. C'est une approche à double-tranchant : faute d'applications hors du cadre usuel de $\mathbb{R}^d$ et de sa mesure de Lebesgue, ce degré d'abstraction peut-être jugé gratuit et injustifié. En revanche, bien maîtrisé, cela peut être très récompensant tant sur le plan des notes (j'ai par exemple obtenu 19/20 en oral blanc grâce à un plan rempli de théorie de la mesure abstraite) que sur celui de l'amusement : si comme moi vous-êtes plus algébriste qu'analyste, vous apprécierez certainement ce glissement qui peut s'opérer dans certaines leçons via la théorie de la mesure. J'ai cherché à donner des exemples et des applications abstraits qui peuvent se recaser dans des leçons et qui souvent correspondent à ceux que j'utilisais moi-même.

Ce document peut se lire dans le désordre, mais a été pensé pour suivre une évolution progressive et sera certainement plus consistant attaqué dans l'ordre. Quelques mots au sujet du plan :

1. La première partie est consacrée au lemme des classes monotones et à ses applications. Bien qu'intégralement hors-programme à l'agrégation, elle est essentielle à la compréhension de bon nombre de résultats fondamentaux de la théorie de la mesure.
2. La seconde explore des cas d'utilisation concrète des mesures abstraites en analyse et résout plusieurs problèmes courants usuellement attaqués de manière plus directe. Elle parlera probablement aux agrégatives !
3. La troisième partie vise à réexplorer les objets de base des probabilités sous l'angle de la théorie de la mesure. Mon espoir est d'en proposer une approche qui saura plaire, au moins un petit peu, aux algébristes. Que j'y parvienne ou non, cette partie saura certainement nourrir vos plans d'exemples et d'applications concrets de la théorie de la mesure.

(i). Voir par exemple mon plan pour la leçon 239

Dans la suite, j'ai adopté le code suivant :

Le symbole † indiquera des paragraphes plus difficiles, qui débordent plus loin sur le programme de l'agrégation, et dont la lecture n'est pas nécessaire en première approche.

Le symbole ‡ marquera les paragraphes dont la lecture apporte peu à la compréhension générale, et qui sont souvent remplis de considérations techniques parfois laborieuses.

Le symbole ☼ permettra aux agrégatives de repérer des résultats du meilleur effet dans les plans d'agreg.

Prélude : pourquoi tous ces objets horribles ?

Quand on aborde pour la première fois la théorie de la mesure (en L3 la plupart du temps), une question assez naturelle qu'on peut se poser est « *mais pourquoi diable s'embêter autant ?* ». Pour appuyer ces questionnements, commençons par rappeler quelques définitions élémentaires de théorie de la mesure :

Définition I (Tribu, espace mesurable). *Soit X un ensemble. Une* tribu *sur X est un ensemble $\mathfrak{F} \subset \mathcal{P}(X)$ de parties de X satisfaisant les axiomes suivants :*

1. $X \in \mathfrak{F}$
2. $\forall A \in F,\ A^c \in \mathfrak{F}$
3. $\forall (A_n) \in \mathfrak{F}^{\mathbb{N}},\ \bigcup_{n\in\mathbb{N}} A_n \in \mathfrak{F}$

Lorsque $\mathfrak{F}$ est une telle tribu, le couple $(X, \mathfrak{F})$ est appelé espace mesurable.

Les premiers exemples de tribus qui viennent en tête sont probablement la tribu grossière, composée uniquement de l'ensemble vide et de X, et la tribu discrète, qui est $\mathcal{P}(X)$ tout entier. Si le premier n'est, de toute évidence, guère intéressant, le second mérite qu'on s'y intéresse. Si $\mathcal{P}(X)$ convient, à quoi bon s'encombrer d'un tas de définition absconses ? Définitions qui par ailleurs donnent naissances à des objets proprement abjectes à manipuler, comme cette bonne vieille tribu borélienne sur laquelle nous aurons l'occasion de revenir. Eh, c'est en plus un choix qui semble des plus commode ; observons plutôt la définition suivante :

Définition II (Fonction mesurable). *Soient $(X, \mathfrak{F})$ et $(Y, \mathfrak{G})$ deux espaces mesurables. Une application $f : X \to Y$ est dite* mesurable *lorsque :*

$$\forall B \in \mathfrak{G},\ f^{-1}(B) \in \mathfrak{F} \tag{0.1}$$

On le sait, ces fonctions tiennent particulièrement à cœur aux mathématiciennes : il n'y en a toujours que pour elles ! Espaces L^p, distributions, variables aléatoires, tout ça repose essentiellement sur l'idée de fonction mesurable. Or, dans le cas où le domaine de notre fonction est muni de la tribu discrète, il est immédiat que celle-ci est mesurable. Aussi, avec la tribu discrète, plus d'embrouille : toutes les fonctions sont mesurable, et tout va bien dans le meilleur des mondes. N'est-ce pas ?...

Et pourtant, nous savons toutes que l'histoire ne s'arrête pas là. La tribu canonique sur l'espace le plus banal du monde, $\mathbb{R}$, n'est pas la tribu discrète (chose qui n'est pas facile du tout

à prouver, et qui en fait est carrément impossible sans l'axiome du choix). Pourquoi ?

Il y a sans doute beaucoup de bonnes réponses à apporter à cette question, et je n'ai pas la prétention d'avoir percé les raisons profondes qui sont à l'origine de la théorie de la mesure (d'autant que je n'ai, à parler vrai, pas cherché), mais j'ai au moins un bout de réponse à apporter pour faire voir que tout ceci n'est pas fait en vain.

Avant même de chercher à parler de mesures dans son sens le plus abstrait, on peut convenir qu'un des objectifs de tout ça est de fabriquer un objet mathématique qui fournisse une notion satisfaisante de *longueur*, de *surface* et de *volume*. Or il y a un obstacle de nature profondément algébrique qui vient faire obstacle à cette quête : j'ai nommé le terrible et redouté paradoxe de Banach-Tarski.

Le paradoxe de Banach-Tarski

Ce résultat bien connu (et à l'origine de beaucoup de réserves sur l'axiome du choix) peut s'énoncer à peu près de cette manière là :

Définition III (Puzzle-équivalence). *Deux parties A et B de $\mathbb{R}^3$ sont dites* puzzle-équivalentes *lorsque il existe un entier $n \in \mathbb{N}^*$, des partitions $(A_i)_{1\leqslant i\leqslant n}$ et $(B_i)_{1\leqslant i\leqslant n}$ de A et B et $(\varphi_i)_{1\leqslant i\leqslant n}$ des isométries affines de déterminant 1 tels que :*

$$\forall i \in [\![1, n]\!],\ B_i = \varphi_i(A_i) \tag{0.2}$$

En des termes plus prosaïques, cela revient à dire qu'on peut couper A en un nombre fini de morceaux, déplacer ceux-ci à l'aide de transformation *physiquement crédibles*, sans les déformer, les déchirer ou les étirer, et finalement obtenir B. On a alors le troublant phénomène :

Théorème IV (Banach-Tarski). *La boule unité fermée de $\mathbb{R}^3$ est puzzle-équivalente à une union disjointe de deux boules unités.*

Je préfère ne pas rentrer dans les détails de la démonstration, car celle-ci est longue et nous éloignerait de notre propos. Si cela vous intéresse, vous pourrez vous référer à [Wag12]. Notons simplement que cela repose sur d'intéressants arguments de théorie des groupes... et sur l'axiome du choix. Aussi, ceux qui se voient troublés par ce résultat pourront choisir de l'ignorer en rejetant l'axiome, mais s'assiéront du même coup sur les théorèmes de Hahn-Banach, de Steiniz, de Tychonov, le théorème de la base, le lemme de Baire, et d'autres joyeusetés pourtant si plaisantes. Nous ferons quant à nous le choix d'accepter inconditionnellement l'axiome du choix, et ce jusqu'à la fin de ce document.

Quoi qu'il en soit, le paradoxe de Banach-Tarski n'arrange pas nos affaires. En effet, supposons qu'on dispose de λ une mesure de volume sur $\mathcal{P}(\mathbb{R}^3)$. De façon informelle, on voudrait bien entendu qu'une telle mesure soit invariante par isométrie ; déplacer et tourner un objet ne devrait pas changer son volume ! Et pourtant, si (A_n) est une partition de la boule unité, (B_n) une partition de deux boules unités disjointes, et (φ_n) une famille d'isométries qui envoie (A_n)

sur (B_n), fournies par puzzle-équivalence, on aurait alors :

$$\lambda(\bar{B}(0,1)) = \lambda\left(\bigsqcup_{i=1}^{n} A_i\right) \tag{0.3}$$

$$= \sum_{i=1}^{n} \lambda(A_i) \tag{0.4}$$

$$= \sum_{i=1}^{n} \lambda(\varphi_i(A_i)) \tag{0.5}$$

$$= \lambda\left(\bigsqcup_{i=1}^{n} B_i\right) \tag{0.6}$$

$$= 2\lambda(\bar{B}(0,1)) \tag{0.7}$$

et ça, ça n'est pas bon du tout [(ii)]. La moralité de cette histoire, c'est que l'on ne peut pas s'autoriser à mesurer n'importe quelle partie de l'espace ; par suite, il nous faut choisir ce qui sera mesurable et ce qui ne l'est pas : c'est tout le rôle de la tribu. Cette introduction aura, je l'espère, suffit à vous convaincre qu'on a besoin de recourir à des objets sophistiqués pour définir quelque chose d'aussi élémentaire que *le volume* ; en bref, nous avons besoin de tribus.

Remarque : Avant de quitter cette introduction, remarquons quand même que, quelque part, cela rend le paradoxe de Banach-Tarski un peu moins absurde : s'il est possible de découper une pomme en morceaux et de les réassembler pour obtenir la lune, le découpage doit se faire à l'aide d'un couteau magique de matheux capable de tailler avec une précision mathématique des parties non-mesurables de l'espace, parties qui, rappelons-le, sont suffisamment tarabiscotées pour qu'on ne puisse même pas prouver leur existence sans utiliser l'axiome du choix ; on parle donc de morceaux de pomme autrement plus hideux que l'intersection de celle-ci avec $\mathbb{Q}^3$. Ainsi, la pomme peut bien être puzzle-équivalente à la lune, si le réassemblage pratique est plus difficile que de séparer les points rationnels de la pomme des autres, après tout ça ne change pas grand'chose à la vie. ⧫

1 Le lemme des classes monotones

J'imagine que ce terme doit vous parler. Il m'a fallut personnellement attendre le M2 pour enfin comprendre un peu la philosophie qui se cache derrière, et c'est en cessant d'en avoir peur que j'ai pu appréhender certains résultats de façon plus profonde. Il m'a paru essentiel de commencer par là, car il est la racine des beaucoup de choses en théorie de la mesure.

J'aimerai ajouter qu'on peut difficilement le balayer d'un revers de la main en le laissant aux théoriciens de la mesure les plus fondamentaux, car de nombreux résultats cruciaux d'analyse et de probabilités reposent quelque part sur la notion de classes monotones [(iii)] ; ignorer les

(ii). Ce raisonnement informel peut être formalisé pour convaincre la lectrice la plus méticuleuse : on axiomatise une mesure de volume de la même façon que la mesure de Lebesgue, en lui donnant les axiomes usuels d'une mesure ainsi que le fait que la mesure du cube unité doive être 1, et que le volume d'une partie doit être invariant par translation. On peut alors montrer proprement qu'au moins une des parties utilisées dans la puzzle-équivalence des boules ne peut pas être mesurable.

(iii). En analyse, on peut par exemple citer l'unicité de la mesure de Lebesgue ou encore la démonstration du théorème de convergence monotone. En probabilités, les caractérisations les plus communes de la loi d'une variable aléatoire reposent également sur le lemme des classes monotones (par exemple la caractérisation par la fonction de répartition).

classes monotones, c'est se condamner à appliquer sans justification des théorèmes pourtant fondamentaux.

1.1 Philosophie de la méthode

Avant de se lancer dans le formalisme, précisons peut-être un peu le but de tout ça. L'objectif de cette section est de pouvoir répondre à deux questions élémentaires de théorie de la mesure, qu'il est assez immédiat de se poser :

1. Une fonction est-elle mesurable ?
2. Est-ce que deux mesures données sont égales ?

L'apparente simplicité de ces questions cache en réalité un grande difficulté de résolution, et savoir y répondre est crucial pour un bon nombre de résultats usuels et concrets.

En algèbre, il est usuel, lorsqu'on cherche à étudier une certaine structure, de se ramener à un système générateur simple sur lequel travailler, puis d'extrapoler les conclusions sur l'ensemble tout entier : étudier le comportement d'une application linéaire sur une base ; travailler sur des groupes à partir de familles génératrices, etc. L'espace mesuré est, au fond, une structure algébrique comme les autres, quoi qu'un peu plus sauvage puisque sa structure ne repose pas sur des opérations qu'on sait à peu près manipuler. Il semble de prime abord assez naturel de chercher des systèmes générateurs d'une tribu, d'étudier des mesures sur ces systèmes générateurs, puis d'extrapoler à la tribu toute entière.

Hélas, les choses ne sont pas si belles que ça. Nous aurons l'occasion d'y revenir dans l'annexe A, les tribus sont des structures particulièrement difficiles à appréhender, même quand on en connaît des générateurs.

Exemple : La tribu borélienne de $\mathbb{R}$, qu'on note $\mathcal{B}(\mathbb{R})$, est la tribu engendrée par les ouverts de $\mathbb{R}$. Malheureusement, ses éléments n'ont pas le bon goût d'être des unions d'intersections d'ouverts. Ni même des unions d'intersections d'union d'intersections d'ouverts. En fait, ce ne sont pas même pas des unions de... bref. Tout ceci sera précisé dans l'annexe A. ⧫

La complexité des tribus rend certaines tâches a priori démentielles. Prenons un exemple classique, qui semblera familier aux lectrices initiées aux probabilités élémentaires :

Exercice 1.1. Soient X et Y deux variables aléatoires réelles. On suppose que pour toute fonction continue bornée f, on a :

$$\mathbb{E}[f(x)] = \mathbb{E}[f(Y)] \tag{1.1}$$

Montrer que X et Y sont égales en loi.

Cette caractérisation de la loi est essentielle en probabilité, et l'ampleur de la tâche qui nous attend à l'idée de démontrer ce résultat à de quoi décourager. Ce constat navrant montre qu'il va falloir être plus rusée que cela.

1.2 Classes monotones

Définition 1.2 (Classe monotone). *Soit X un ensemble. On appelle* classe monotone*, ou encore* λ-système *de X toute famille $\Lambda \subset \mathcal{P}(X)$ telle que :*

1. $\emptyset \in \Lambda$
2. *Pour toute suite croissante (au sens de l'inclusion) (A_n) d'éléments de Λ, l'union $\bigcup_{n\in\mathbb{N}} A_n$ appartient à Λ.*
3. *Λ est stable par différence propre, c'est-à-dire :*

$$\forall (A,B) \in \Lambda^2,\ A \subset B \implies B \setminus A \in \Lambda \tag{1.2}$$

Donnons quelques exemples fondamentaux de classes monotones, qui vont nous servir à briser des problèmes avec une grande facilité :

Exemple :

- Toute tribu est une classe monotone, mais la réciproque est fausse.
- Supposons X muni d'une tribu $\mathfrak{F}$ et de deux mesures finies μ et ν. Alors l'ensemble :

$$\Lambda := \{A \in \mathfrak{F} \mid \mu(A) = \nu(A)\} \tag{1.3}$$

  est une classe monotone. En effet, $\mu(\emptyset) = \nu(\emptyset) = 0$. Par ailleurs, pour $A \subset B$ deux éléments de Λ, on a :

$$\mu(B \setminus A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B \setminus A) \tag{1.4}$$

  et donc $B \setminus A \in \Lambda$. Enfin, si (A_n) est une suite croissante de Λ, le théorème de continuité croissante des mesures fourni :

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to+\infty} \mu(A_n) = \lim_{n\to+\infty} \nu(A_n) = \nu\left(\bigcup_{n\in\mathbb{N}} A_n\right) \tag{1.5}$$

  et donc Λ est stable par union croissante dénombrable.
  Notons cela-dit que Λ n'est en général pas une tribu. En effet, on peut considérer, par exemple sur $\mathbb{R}$, μ la mesure cardinale qui à une partie $A \subset \mathbb{R}$ associe le cardinal de A (éventuellement infini) et ν la mesure nulle sur $\mathbb{R}$. Clairement, $\mu(\mathbb{R}) \neq \nu(\mathbb{R})$ et donc Λ ne peut pas être une tribu.
- Si $f : (X, \mathfrak{F}) \to (Y, \mathfrak{G})$ est une application (pas nécessairement mesurable. L'ensemble

$$\{B \in \mathfrak{G} \mid f^{-1}(B) \in \mathfrak{F}\} \tag{1.6}$$

  est une sous-classe monotone de $\mathfrak{G}$. Si ces notions sont nouvelles pour vous, je vous invite à le démontrer par vous même. Ceci sera important pour montrer que des applications sont mesurables.

♦

Proposition 1.3. *Soit X un ensemble et $\Pi \subset \mathcal{P}(X)$ un ensemble de parties de X. Alors il existe une unique classe monotone $\lambda(\Pi)$ engendrée par Π, dans le sens où si Λ est une classe monotone contenant Π, on a $\lambda(\Pi) \subset \Lambda$.*

Démonstration. Il suffit de montrer qu'une intersection de classes monotones est encore une classe monotone, ce qui résulte de vérifications élémentaires. Ceci étant fait, il suffit de considérer $\lambda(\Pi)$ l'intersection de toutes les classes monotones contenant Π. □

Remarque : Cette démonstration n'est pas sans rappeler des preuves usuelles d'algèbre qui prouvent l'existence d'un sous-groupe, d'un idéal, d'un espace vectoriel, etc. engendré par une partie. ⧫

Il est temps maintenant de voir en quoi les classes monotones vont nous aider à résoudre des problèmes au niveau des tribus :

Définition 1.4 (π-système). *On appelle π-système de X toute famille $\Pi \subset \mathcal{P}(X)$ telle que :*

1. $X \in \Pi$
2. Π *est stable par intersections finies.*

Exemple : L'ensemble des intervalles de $\mathbb{R}$ est un π-système. Plus simplement, l'ensemble des intervalles de la forme $]-\infty, a]$ avec $a \in \mathbb{R}$ est un π-système. ⧫

Théorème 1.5 (Lemme des classes monotones). *Soit Π un π-système d'un ensemble X. Alors :*

$$\lambda(\Pi) = \sigma(\Pi) \tag{1.7}$$

avec $\sigma(\Pi)$ la tribu engendrée par Π.

Démonstration. Comme $\sigma(\Pi)$ est une tribu, c'est une classe monotone, qui contient Π par définition. Donc $\lambda(\Pi) \subset \sigma(\Pi)$. Pour obtenir l'inclusion réciproque, il suffit de montrer que $\lambda(\Pi)$ est une tribu. Notons que $X \in \Lambda(\Pi)$ par hypothèse.

Stabilité par complémentaire : une classe monotone est stable par différence propre. Ainsi, si $A \in \lambda(\Pi)$, on a :

$$A^c = X \subset A \in \lambda(\Pi) \tag{1.8}$$

puisque $X \in \lambda(\Pi)$ et $A \subset X$.

Stabilité par union dénombrable : on va réduire le problème à montrer que $\lambda(\Pi)$ est stable par intersection finie. En effet, on aurait dans ce cas que $\lambda(\Pi)$ serait stable par union finie en passant au complémentaire, et alors, si (A_n) est une famille dénombrable d'éléments de $\lambda(\Pi)$, on aurait :

$$\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}}^{\uparrow} \underbrace{\bigcup_{k\leqslant n} A_k}_{\in\lambda(\Pi)} \in \lambda(\Pi) \tag{1.9}$$

Montrons donc la stabilité par intersection finie. Soit $E \in \Pi$. On pose :

$$\Lambda_E := \{A \in \lambda(\Pi) \mid A \cap E \in \lambda(\Pi)\} \tag{1.10}$$

Il est clair que Λ_E contient $\emptyset$. Si de plus A et B sont deux éléments de Λ_E tels que $A \subset B$, on a :

$$(B \setminus A) \cap E = (B \cap E) \setminus (A \cap E) \in \lambda(\Pi) \tag{1.11}$$

donc Λ_E est stable par différence propre. Enfin, si (A_n est une suite croissante de Λ_E, on a :

$$\left(\bigcup_{n\in\mathbb{N}} A_n\right) \cap E = \bigcup_{n\in\mathbb{N}}^{\uparrow} \underbrace{(A_n \cap E)}_{\in\lambda(\Pi)} \in \lambda(\Pi) \tag{1.12}$$

Ainsi, Λ_E est une classe monotone. Par ailleurs, comme Π est un π-système, pour $A \in \Pi$, $A \cap E \in \Pi \subset \lambda(\Pi)$. Donc Λ_E contient Π et par suite :

$$\lambda(\Pi) \subset \Lambda_E \tag{1.13}$$

L'inclusion réciproque étant immédiate par construction, on égalité entre ces deux classes monotones. Enfin, si B est un élément quelconque de $\lambda(\Pi)$, la famille :

$$\Lambda_B := \{A \in \lambda(\Pi) \mid A \cap B \in \lambda(\Pi)\} \tag{1.14}$$

est encore une classe monotone, et celle-ci contient Π d'après le raisonnement qui vient d'être mené. Il s'agit donc de $\lambda(\Pi)$, et par suite $\lambda(\Pi)$ est stable par intersection finie : c'est une tribu.

□

1.3 Un peu de sorcellerie

Le niveau d'abstraction du lemme des classes monotones pourrait laisser insensible en masquant complètement sa puissance. Pourtant, il s'agit bel et bien d'un splendide théorème qui va nous permettre d'étudier des tribus en se ramenant en toute allégresse à des parties génératrices bien choisies. Admirez ! nous sommes maintenant en passe de répondre au moins partiellement à nos deux questions phares :

Proposition 1.6. *Soit $f : (X, \mathfrak{F}) \to (Y, \mathfrak{G})$ une application entre espaces mesurables. Soit Π un π-système générateur de $\mathfrak{G}$. Si :*

$$\forall B \in \Pi,\ f^{-1}(B) \in \mathfrak{F} \tag{1.15}$$

alors f est mesurable.

Démonstration. Comme on l'a vu dans l'exemple 1.2, l'ensemble

$$\{B \in \mathfrak{G} \mid f^{-1}(B) \in \mathfrak{F}\} \tag{1.16}$$

est une classe monotone. Par hypothèse, celle-ci contient un π-système générateur de $\mathfrak{G}$: elle contient donc la classe monotone engendrée par Π. Mais celle-ci est égale à $\mathfrak{G}$ par le lemme des classes monotones ! Ainsi, f est mesurable. □

Proposition 1.7. *Soit $(X, \mathfrak{F})$ un espace mesurable muni de deux mesuresfinies μ et ν. On suppose qu'il existe un π-système Π de X qui engendre $\mathfrak{F}$ et tel que :*

$$\forall A \in \Pi,\ \mu(A) = \nu(A) \tag{1.17}$$

Alors $\mu = \nu$.

Démonstration. D'après l'exemple 1.2, la famille

$$\Lambda_n := \{A \in \mathfrak{F} \mid \mu_n(A) = \nu_n(A)\} \tag{1.18}$$

est une classe monotone. Par ailleurs, celle-ci contient Π par hypothèse, et donc $\lambda(\Pi) \subset \Lambda$. Mais par le lemme des classes monotones :

$$\lambda(\Pi) = \sigma(\Pi) = \mathfrak{F} \tag{1.19}$$

Donc $\mathfrak{F} \subset \Lambda$ et $\mu = \nu$. □

Remarque : Ce résultat est l'analogue de faits algébriques usuels, comme par exemple le fait que deux applications linéaires sont égales si, et seulement si, elles coïncident sur une famille génératrice de leur domaine. ⧫

On pourrait être déçu de ce que le résultat ne fonctionne que pour les mesures finies. Toutefois, bien utilisé, celui-ci permet de résoudre des problèmes qui vont au-delà de ce cadre :

Application 1.8. (Unicité de la mesure de Lebesgue) On considère $\mathbb{R}$ muni de sa tribu borélienne $\mathcal{B}(\mathbb{R})$. Il existe au plus une mesure μ définie sur $\mathcal{B}(\mathbb{R})$ telle que :

1. $\mu([0,1]) = 1$
2. $\forall x \in \mathbb{R},\ \forall A \in \mathcal{B}(\mathbb{R}),\ \mu(A + x) = \mu(A)$

Démonstration. Supposons données μ et ν deux telles mesures. Remarquons pour commencer que les singletons ont tous même mesure par invariance par translation, et cette mesure est nécessairement zéro (sinon $[0,1]$ serait de mesure infinie par σ-additivité). Par récurrence immédiate, pour $(n, m \in \mathbb{Z}^2$ tels que $n \leqslant m$, on a :

$$\mu([n,m]) = \nu([n,m]) = m - n \tag{1.20}$$

Par ailleurs, pour $d \in \mathbb{N}^*$, on peut écrire :

$$[0,1] = \{1\} \sqcup \bigsqcup_{k=0}^{d-1} [\frac{k}{d}, \frac{k+1}{d}[\tag{1.21}$$

Donc par σ-additivité et invariance par translation :

$$1 = \mu([0,1]) = d\mu([0, \frac{1}{d}[) = d\mu([0, \frac{1}{d}]) \tag{1.22}$$

En combinant ces deux constats, qui valent également pour ν, on trouve que pour tous rationnels $q \leqslant r$, on a :

$$\mu([q,r]) = \nu([q,r]) = r - q \tag{1.23}$$

La famille $\{\mathbb{R}\} \cup \{[q,r], (q,r) \in \mathbb{Q}^2\}$ est un π-système générateur de $\mathcal{B}(\mathbb{R})$. Hélas, nos mesures μ et ν ne sont pas finies. On va devoir ruser pour ce ramener au cas précédent. Pour $n \in \mathbb{N}^*$, on pose :

$$\forall A \in \mathcal{B}(\mathbb{R}),\ \mu_n(A) := \mu(A \cap [-n,n])\ [\text{resp. } \nu_n(A) := \nu(A \cap [-n,n])] \tag{1.24}$$

Les mesures μ_n et ν_n sont finies. De plus, si I est un intervalle compact à bornes rationnelles, $I \cap [-n,n]$ l'est également et donc par l'analyse précédemment menée, $\mu_n(I) = \nu_n(I)$. Par le lemme des classes monotones, $\mu_n = \nu_n$, ce pour tout n.

Finalement, prenons A un borélien quelconque. Le théorème de continuité croissante des mesures permet d'écrire :

$$\begin{aligned}
\mu(A) &= \mu\left(A \cap \bigcup_{n\in\mathbb{N}^*} [-n,n]\right) && (1.25)\\
&= \lim_{n\to+\infty} \mu_n(A) && (1.26)\\
&= \lim_{n\to+\infty} \nu_n(A) && (1.27)\\
&= \nu(A) && (1.28)
\end{aligned}$$

et donc $\mu = \nu$. □

On peut également résoudre l'exercice 1.1 grâce au lemme des classes monotones. Je vous invite à essayer par vous-même avant de lire la correction ci-dessous ! Rappelons rapidement l'énoncé :

Proposition 1.9. *Soient X et Y deux variables aléatoires réelles. On suppose que pour toute fonction continue bornée f, $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$. Alors X et Y sont égales en loi.*

Démonstration. Notons $\mathbb{P}_X$ et $\mathbb{P}_Y$ les lois de X et Y (qui sont des mesures sur $\mathbb{R}$). Soit a un réel. On va approcher l'indicatrice de $]-\infty, a]$ par une fonction continue bornée en posant :

$$\forall n \in \mathbb{N},\ f_n : x \mapsto \begin{cases} 1 \text{ si } x < a - 1/n \\ 0 \text{ si } x > a \\ 1 - n(x - a + 1/n) \text{ sinon} \end{cases} \tag{1.29}$$

La suite (f_n) est une suite croissante et positive d'approximations trapezoïdales de $\mathbb{1}_{]-\infty,a]}$. Ainsi, par convergence monotone, on a :

$$\begin{aligned}
\mathbb{P}_X(]-\infty,a]) &= \int_{\mathbb{R}} \mathbb{1}_{]-\infty,a]}(x) \mathrm{d}\mathbb{P}_X(x) && (1.30)\\
&= \lim_{n\to+\infty} \int_{\mathbb{R}} f(x) \mathrm{d}\mathbb{P}_X(x) && (1.31)\\
&= \lim_{n\to+\infty} \int_{\mathbb{R}} f(x) \mathrm{d}\mathbb{P}_Y(x) && (1.32)\\
&= \mathbb{P}_Y(]-\infty,a]) && (1.33)
\end{aligned}$$

Par ailleurs, $\mathbb{P}_X(\mathbb{R}) = \mathbb{P}_Y(\mathbb{R}) = 1$ puisqu'il s'agit de mesures de probabilités. Mais la famille $\{\mathbb{R}\} \cup \{]-\infty, a],\ a \in \mathbb{R}\}$ est un π-système générateur de la tribu borélienne. Donc $\mathbb{P}_X = \mathbb{P}_Y$, c'est-à-dire que X et Y sont égales en loi. □

J'aimerai au passage faire remarquer que ce résultat peut se reformuler dans un autre langage :

Proposition 1.10. *On note $\mathscr{P}(\mathbb{R})$ l'ensemble des mesures de probabilités sur $\mathbb{R}$ et C_b^0 l'ensemble des fonctions continues bornées sur $\mathbb{R}$. L'application :*

$$\begin{aligned} T : \mathscr{P}(\mathbb{R}) &\to (C_b^0)' \\ \mathbb{P} &\mapsto \left(T_{\mathbb{P}} : f \mapsto \int_{\mathbb{R}} f(x) \mathrm{d}\mathbb{P}(x) \right) \end{aligned} \tag{1.34}$$

est injective.

En termes plus probabilistes, la forme linéaire décrite par le membre de droite n'est autre que $f \mapsto \mathbb{E}[f(X)]$ lorsque X est une variable aléatoire de loi $\mathbb{P}$. Cette façon de voir les choses est proche de celle qu'on adopte lorsqu'on fait des distributions (on représente une application mesurable, ou une mesure, par une forme linéaire continue sur un espace de fonctions tests) qui est riche en applications. Typiquement, la convergence en loi d'une suite de variables aléatoires revient simplement à la convergence simple de la suite de formes linéaires associées.

2 De l'intérêt de manipuler des mesures abstraites

Pour ma part, mon introduction à la théorie de la mesure s'est à peu près passée comme suit : j'ai appris tout un tas de définitions compliquées (tribu, mesures, espaces σ-finis, etc.), des théorèmes dans un cadre très abstrait, pour ensuite ne plus jamais appliquer tout ceci que dans le cas où notre espace mesuré est $\mathbb{R}$ muni de sa tribu borélienne et de la mesure de Lebesgue, ou bien $\mathbb{N}$ muni de la tribu discrète et de la mesure cardinale. Alors pourquoi s'embêter ? Et puis finalement, en M2, c'est en revenant sur la théorie de la mesure à l'occasion de la préparation à l'agrégation que les choses se sont éclairées : des mesures abstraites, il y en a fréquemment en analyse, pour qui veut bien les voir, et parfois s'autoriser à manipuler des mesures abstraites permet de passer au-travers d'un problème qu'il aurait été extrêmement pénible de résoudre avec des outils plus standards tout en rendant les démonstrations beaucoup plus naturelles. En probabilités, les espaces mesurés abstraits font légion, et savoir jongler entre un point de vue probabiliste et un point de vue théorie de la mesure peut se révéler riche en applicaions (même au niveau de l'agrégation), mais nous reviendront là-dessus en profondeur lors de la section suivante.

Dans cette section, nous aurons l'occasion de voir plusieurs situations communes où l'on peut faire apparaître des mesures plus abstraites que la mesure de Lebesgue, qui trouverons des applications des plus concrètes. Si vous préparez l'agrégation, ces exemples peuvent enrichir avec profits vos leçons tout en justifiant d'énoncer des théorèmes dans des cadres très abstraits en leur fournissant de vraies applications, en plus de faciliter grandement certains problèmes d'écrit.

2.1 Les fonctions sont des mesures

J'ai promis qu'il y a des mesures dans des situations des plus courantes, il est tant de voir *à quel point* ces situations peuvent en effet être banales. Dans toute la suite, on fixe $(X, \mathfrak{F}, \mu)$ un espace mesuré quelconque, et on notera λ_d la mesure de Lebesgue sur $\mathbb{R}^d$.

Proposition 2.1 (Les fonctions sont des mesures). *Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction mesurable à valeurs positives. L'application :*

$$f\mathrm{d}\lambda : \mathcal{B}(\mathbb{R}) \to \mathbb{R}_+ \cup \{+\infty\} \tag{2.1}$$

$$A \mapsto \int_A f(x)\mathrm{d}x \tag{2.2}$$

définit une mesure sur $\mathbb{R}$.

Démonstration. Toutes les vérifications sont élémentaires. La positivité est immédiate de par la positivité de f. De plus, l'intégrale de f sur le vide est évidemment nulle (il s'agit en fait par définition de l'intégrale sur $\mathbb{R}$ de $f\mathbb{1}_\emptyset$ qui est la fonction nulle). Pour la σ-additivité, c'est une conséquence du théorème de de convergence monotone : si $(A_n) \in \mathcal{B}(\mathbb{R})^{\mathbb{N}}$ est une suite de boréliens deux à deux disjoints, alors en notant $A := \bigsqcup A_n$, l'indicatrice de A est la limite simple de la suite croissante $(\mathbb{1}_{\bigsqcup_{k=0}^n A_k})_{n\in\mathbb{N}}$, d'où :

$$f\mathrm{d}\lambda\left(\bigsqcup_{k\in\mathbb{N}} A_k\right) = \int_{\mathbb{R}} \lim_{n\to+\infty} f(x)\mathbb{1}_{\bigsqcup_{k=0}^n A_k}(x)\mathrm{d}x \tag{2.3}$$

$$= \lim_{n\to+\infty} \int_{\bigsqcup_{k=0}^n A_k} f(x)\mathrm{d}x \tag{2.4}$$

$$= \lim_{n\to+\infty} \sum_{k=0}^{n} f\mathrm{d}\lambda(A_k) \tag{2.5}$$

$$= \sum_{k=0}^{+\infty} f\mathrm{d}\lambda(A_k) \tag{2.6}$$

□

Remarque : Cette définition peut être élargie en remplaçant λ par une mesure μ quelconque sur un espace mesuré général $(X, \mathfrak{F})$, et en prenant pour f une fonction à valeurs positives de X dans $\mathbb{R}$. ⧫

†

Remarque : La mesure $f\mathrm{d}\lambda$ est absolument continue par rapport à la mesure de Lebesgue, c'est-à-dire :

$$\forall A \in \mathcal{B}(\mathbb{R}),\ \lambda(A) = 0 \implies f\mathrm{d}\lambda(A) = 0 \tag{2.7}$$

La fonction f est alors la dérivée de Radon-Nikodym $\dfrac{\mathrm{d}[f\mathrm{d}\lambda]}{\mathrm{d}\lambda}$. Réciproquement, toute mesure absolument continue par rapport à Lebesgue s'écrit comme une telle mesure $f\mathrm{d}\lambda$ d'après le théorème de Radon-Nikodym. Voir [BP17], chapitre 10, pour plus de détails. ⧫

Proposition 2.2. *Notons* $\mathrm{Mes}(\mathbb{R})$ *l'ensemble des mesures sur* $\mathbb{R}$*. L'application*

$$L^1_{loc}(\mathbb{R}) \to \mathrm{Mes}(\mathbb{R}) \tag{2.8}$$
$$f \mapsto f\mathrm{d}\lambda \tag{2.9}$$

est injective. Par ailleurs, pour $f : \mathbb{R} \to \mathbb{R}_+$ *une fonction mesurable positive :*

1. $f\mathrm{d}\lambda$ *est finie si, et seulement si,* f *est intégrable.*
2. *Si* f *est localement intégrable,* $f\mathrm{d}\lambda$ *est* σ*-finie.*

Démonstration. Montrons tout d'abord l'injectivité. Soient f et g deux fonctions localement intégrables positives qui ne sont pas égales presque partout. Quitte à permuter f et g, on peut supposer que l'ensemble $E := \{x \in \mathbb{R} \mid f(x) < g(x)\}$ est de mesure non nulle. Il existe nécessairement une entier $n \in \mathbb{N}^*$ tel que la mesure de $A := E \cap [-n, n]$ soit strictement positive. Comme f est localement intégrable, on peut écrire :

$$g\mathrm{d}\lambda(A) - f\mathrm{d}\lambda(A) = \int_A (g - f)(x)\mathrm{d}x > 0 \tag{2.10}$$

et donc $f\mathrm{d}\lambda \neq g\mathrm{d}\lambda$.

Par ailleurs, $f\mathrm{d}\lambda$ est finie si, et seulement si :

$$f\mathrm{d}\lambda(\mathbb{R}) = \int_{\mathbb{R}} f(x)\mathrm{d}x < +\infty \tag{2.11}$$

c'est-à-dire si f est intégrable (puisqu'elle positive). □

Remarque : f peut ne pas être localement intégrable même lorsque $f\mathrm{d}\lambda$ est σ-finie. Par exemple, on peut prendre $f : x \mapsto \dfrac{1}{|x|}$ prolongée en 0 par 0. Cette fonction n'est intégrable sur aucun voisinage de 0, mais en posant, $E_n := \{0\} \cup [-n, -1/n] \cup [1/n, n]$, on obtient un recouvrement dénombrable de $\mathbb{R}$ tel que $f\mathrm{d}\lambda(E_n) < +\infty$ pour tout $n \in \mathbb{N}^*$. ♦

Remarque : Cette façon de voir les choses est particulièrement adaptée aux probabilités : en effet, le fait pour une variable aléatoire réelle X d'admettre une densité f contre la mesure de Lebesgue revient en réalité à l'égalité des mesures :

$$\mathbb{P}_X = f\mathrm{d}\lambda \tag{2.12}$$

où $\mathbb{P}_X$ est la loi de X (dont nous reparlerons dans la définition 3.2). ♦

2.2 Inégalité de Jensen

Il est l'heure maintenant d'introduire le théorème qui peut faire tomber tous les autres si l'on s'y prend correctement : l'inégalité de Jensen. Théorème qui passe relativement inaperçu losqu'on travaille toujours contre la mesure de Lebesgue, mais qui on va le voir se révèle d'une rare puissance correctement utilisé.

Théorème 2.3 (Inégalité de Jensen). *Soit $(X, \mathfrak{F}, \mu)$ un espace probabilisé, soit I un intervalle non-vide de $\mathbb{R}$ et soit $\varphi : I \to \mathbb{R}$ une fonction convexe. Pour toute fonction mesurable $f : X \to I$, on a :*

$$\varphi\left(\int_X f(x)\mathrm{d}\mu(x)\right) \leqslant \int_X \varphi \circ f(x)\mathrm{d}\mu(x) \tag{2.13}$$

Démonstration. Cette démonstration est donnée en exercice dans [BP17]. Par convexité de φ, pour tout $a \in I$, l'application :

$$\begin{aligned} I \setminus \{a\} &\to \mathbb{R} \\ x &\mapsto \frac{\varphi(x) - \varphi(a)}{x - a} \end{aligned} \tag{2.14}$$

est croissante. En particulier, elle admet une limite à droite en a. En d'autres termes, φ est dérivable à droite en tout point. Notons φ'_d sa fonction dérivée à droite. On a de plus :

$$\forall x > a,\ \varphi'_d(a) \leqslant \frac{\varphi(x) - \varphi(a)}{x - a} \tag{2.15}$$

et donc en inversant :

$$\forall x > a,\ \varphi(x) \geqslant \varphi(a) + (x - a)\varphi'_d(a) \tag{$(\star)$}$$

On montre de même que cette inégalité vaut toujours pour $x < a$. Comme elle est trivialement vraie pour $x = a$, elle vaut pour tout x et tout a.

Observons maintenant que $\int_X f(x)\mathrm{d}\mu(x)$ appartient à I. Pour cela, notons m et M l'infimum et le supremum de I (éventuellement infinis), et observons :

$$\forall x \in X,\ m \leqslant f(x) \leqslant M \tag{2.16}$$

avec inégalité stricte à gauche (resp. à droite) si m (resp. M) n'appartient pas à I. Ainsi, par croissance de l'intégrale, et parce que μ est une mesure de masse 1 :

$$m \leqslant \int_X f(x)\mathrm{d}\mu(x) \leqslant M \tag{2.17}$$

avec inégalité stricte à gauche (resp. à droite) lorsque m (resp. M) n'appartient pas à I. Ainsi, l'intégrale de f est bien dans I.

En appliquant l'inégalité $(\star)$ avec $a := \int_X f(x)\mathrm{d}\mu(x)$, on obtient :

$$\forall x \in X,\ \varphi(f(x)) \geqslant \varphi(a) + (f(x) - a)\varphi'_d(a) \tag{2.18}$$

en intégrant en x sur X :

$$\int_X \varphi \circ f(x)\mathrm{d}\mu(x) \geqslant \underbrace{\int_X \varphi(a)\mathrm{d}\mu(x)}_{=\varphi(a)} + \underbrace{\left(\int_X f(x)\mathrm{d}\mu(x) - a\right)}_{=0}\varphi'_d(a) \tag{2.19}$$

et finalement, comme μ est de masse 1 et en remplaçant a par sa valeur, on obtient l'inégalité de Jensen □

La première réaction qu'on peut avoir face à cette inégalité, c'est « génial, mais les mesures de probas, c'est pas tous les jours qu'on en croise. Surtout quand on ne fait pas de probas ». Mais ce serait oublier notre précédent précept : les fonctions sont des mesures ! Observons un rapide petit exemple qui contient toutefois à lui seul tous les ingrédients qui nous serviront dans la suite :

Exemple : Soit $(X, \mathfrak{F}, \mu)$ un espace mesuré, g une fonction positive d'intégrale 1 et f une fonction telle que fg est intégrable. Notons que pour $p \in [1, +\infty[$, la fonction $x \mapsto |x|^p$ est convexe sur $\mathbb{R}$. Par ailleurs, la mesure $g\mathrm{d}\mu$ est une mesure de probabilité sur X. Ainsi, il vient :

$$\left| \int_X f(x)g(x)\mathrm{d}\mu(x) \right|^p \leqslant \int_X |f(x)|^p g(x)\mathrm{d}\mu(x) \tag{2.20}$$

Toute la subtilité de cet exemple réside dans le fait qu'on n'a fait a priori aucune hypothèse sur la mesure μ dont X était muni. L'intégrale qui était initialement à calculer était vu comme l'intégrale de fg contre μ, mais nous avons préféré la voir comme l'intégrale de f contre la *mesure de probabilités* $g\,\mathrm{d}\mu$, obtenant ainsi sans effort une inégalité qu'il aurait pu être extrêmement pénible de montrer autrement. On a troqué la difficulté du calcul initial contre l'abstraction, en remplaçant un objet rassurant (la fonction g) par une mesure de probabilité à densité contre μ. ♦

La section suivante devrait vous convaincre de la beauté de cette inégalité en donnant des cas pratique d'utilisation de l'exemple précédent (qui se recasent merveilleusement bien dans les leçons d'agrégation qui impliquent des fonctions convexes).

2.3 Faire tomber avec élégance les théorèmes usuels

Nous allons à présent redémontrer plusieurs résultats classiques d'intégration à l'aide de l'inégalité de Jensen, sans effort et surtout quasiment sans calcul. Commençons par ce qui est peut-être la première inégalité qu'on apprend en cours de théorie de la mesure :

Théorème 2.4 (Inégalité de Hölder). *Soit $(X, \mathfrak{F}, \mu)$ un espace mesuré. Soit $p \in [1, +\infty]$. Soit q son exposant conjugué, c'est-à-dire l'unique réel tel que $\frac{1}{p} + \frac{1}{q} = 1$. Pour toute fonction $f \in L^p(\mu)$ et toute fonction $g \in L^q(\mu)$, le produit fg est intégrable et on a :*

$$\|fg\|_1 \leqslant \|f\|_p \|g\|_q \tag{2.21}$$

Démonstration. Le cas $p \in \{1, +\infty\}$ est immédiat et résulte de la croissance de l'intégrale. On supposera donc que ni p ni q n'est égal à $+\infty$. On écrit alors :

$$\int_X |f(x)||g(x)|\mathrm{d}\mu(x) = \|f\|_p^p \left(\left(\int_X |f(x)|^{1-p}|g(x)| \frac{|f(x)|^p}{\|f\|_p^p} \mathrm{d}\mu(x) \right)^q \right)^{1/q} \tag{2.22}$$

Mais la fonction $x \mapsto \frac{|f(x)|^p}{\|f\|_p^p}$ est intégrable, positive et d'intégrale 1. Ainsi, la mesure $\frac{|f|^p}{\|f\|_p^p}\mathrm{d}\mu$ est une mesure de probabilités sur X. Comme la fonction $x \mapsto |x|^q$ est convexe sur $\mathbb{R}$, on peut appliquer l'inégalité de Jensen et pousser la puissance q-ième sous l'intégrale :

$$\int_X |f(x)||g(x)|\mathrm{d}\mu(x) \leqslant \|f\|_p^p \left(\int_X |f(x)|^{q(1-p)}|g(x)|^q \frac{|f(x)|^p}{\|f\|_p^p} \mathrm{d}\mu(x) \right)^{1/q} \tag{2.23}$$

Il ne reste qu'à remarquer que $q = \frac{p}{p-1}$ pour conclure, car on a alors :

$$\forall x \in X,\ |f(x)|^{q(1-p)}|f(x)|^p = 1 \tag{2.24}$$

et donc :

$$\int_X |f(x)||g(x)|\mathrm{d}\mu(x) \leqslant \|f\|_p^p \times \|f\|_p^{-p/q} \left(\int_X |g(x)|^q \mathrm{d}\mu(x)\right)^{1/q} \tag{2.25}$$

$$= \|f\|_p \|g\|_q \tag{2.26}$$

□

Elégant n'est-ce pas ? Plus besoin de l'inégalité de Young et autres pénibilités calculatoires, l'inégalité de Hölder découle naturellement de celle de Jensen. Poursuivons dans cette veine en montrant deux autres résultats classiques et plutôt désagréables à montrer sans changer la mesure sur $\mathbb{R}$:

Théorème 2.5 (Convolution L^1/L^p). *Soit $p \in [1, +\infty]$. On note f et g deux fonctions de $\mathbb{R}$ dans lui-même. Si f et L^1 et g est L^p contre la mesure de* Lebesgue [(iv)]*, alors elles sont convolables et leur produit de convolution est de classe L^p. De plus, on a l'inégalité :*

$$\|f * g\|_p \leqslant \|f\|_1 \|f\|_p \tag{2.27}$$

Démonstration. Quitte à ce que ce terme soit infini, on écrit librement l'intégrale positive :

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(y)||g(x-y)|\mathrm{d}y\right)^p \mathrm{d}x = \|f\|_1^p \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |g(x-y)| \frac{|f(y)|}{\|f\|_1}\mathrm{d}y\right)^p \tag{2.28}$$

Vous le voyez maintenant venir : la mesure $\frac{|f|}{\|f\|_1}\mathrm{d}y$ est une mesure de probabilités sur $\mathbb{R}$: on utilise Jensen puis le théorème de Fubini-Tonelli pour intervertir les intégrales en x et en y :

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(y)||g(x-y)|\mathrm{d}y\right)^p \mathrm{d}x \leqslant \|f\|_1^p \int_{\mathbb{R}}\int_{\mathbb{R}} |g(x-y)|^p \frac{|f(y)|}{\|f\|_1}\mathrm{d}x\mathrm{d}y \tag{2.29}$$

$$= \|f\|_1^p \int_{\mathbb{R}} |g(x)|^p \mathrm{d}x \int_{\mathbb{R}} \frac{|f(y)|}{\|f\|_1}\mathrm{d}y \tag{2.30}$$

$$= \|f\|_1^p \|g\|_p^p \tag{2.31}$$

où la dernière étape se fait en utilisant l'invariance par translation de la mesure de Lebesgue et le changement de variable $x \leftarrow x - y$. □

Je suis sûre que vous aussi, vous commencez déjà à devenir accro à ce petit jeu. Voici un dernier cas d'application pour la route :

(iv). Il est dans ce cas précis indispensable d'intégrer contre la mesure de Lebesgue pour profiter de l'invariance par translation de celle-ci lors de la manipulation du produit de convolution.

Théorème 2.6 (Approximation par convolution). *Soit $p \in [1, +\infty[$. On considère $f \in L^p(\lambda_d)$ et $(\Phi_n)_{n\in\mathbb{N}}$ une approximation de l'unité, c'est-à-dire une suite de fonctions mesurables et positives telles que :*

1. $\forall n \in \mathbb{N}, \int_{\mathbb{R}} \Phi_n(x)\mathrm{d}x = 1$
2. $\forall \varepsilon > 0, \int_{\|x\|\geqslant\varepsilon} \Phi_n(x)\mathrm{d}x \xrightarrow[n\to+\infty]{} 0$

On a alors :

$$\lim_{n\to+\infty} \|f * \Phi_n - f\|_p = 0 \tag{2.32}$$

Démonstration. On sait grâce aux théorèmes précédents que f et $f * \Phi_n$ sont dans L^p pour tout n : pas besoin de s'embêter à prendre des pincettes pour écrire les intégrales. On calcule donc directement, pour $n \in \mathbb{N}$ fixé :

$$\|f * \Phi_n - f\|_p^p = \int_{\mathbb{R}} \left| \int_{\mathbb{R}} (f(x-y)\Phi_n(y)\mathrm{d}y) - f(x) \right|^p \mathrm{d}x \tag{2.33}$$

Puisque Φ_n est une fonction positive d'intégrale 1, la mesure $\Phi_n(y)\mathrm{d}y$ est une mesure de probabilités, contre laquelle on a :

$$\forall x \in \mathbb{R}^d,\ f(x) = \int_{\mathbb{R}} f(x)\Phi_n(y)\mathrm{d}y \tag{2.34}$$

On peut donc réécrire notre première égalité en poussant tout le monde sous l'intégrale puis utiliser l'inégalité de Jensen :

$$\|f * \Phi_n - f\|_p^p = \int_{\mathbb{R}} \left| \int_{\mathbb{R}} (f(x-y) - f(x))\Phi_n(y)\mathrm{d}y \right|^p \mathrm{d}x \tag{2.35}$$

$$\leqslant \int_{\mathbb{R}} \int_{\mathbb{R}} |f(x-y) - f(x)|^p \Phi_n(y)\mathrm{d}y\,\mathrm{d}x \tag{2.36}$$

On utilise le théorème de Fubini-Tonelli pour intervertir les intégrales (de termes positifs) et on introduit $\tau_a : g \mapsto (x \mapsto g(x-a))$ les opérateurs de translation sur L^p, pour tout $a \in \mathbb{R}^d$, de sorte à obtenir :

$$\|f * \Phi_n - f\|_p^p \leqslant \int_{\mathbb{R}} \Phi_n(y) \int_{\mathbb{R}} |f(x-y) - f(x)|^p \mathrm{d}x\,\mathrm{d}y \tag{2.37}$$

$$= \int_{\mathbb{R}} \Phi_n(y) \|\tau_y f - f\|_p^p \mathrm{d}y \tag{2.38}$$

Soit $\varepsilon > 0$. Par uniforme continuité des translations dans L^p, il existe $\eta > 0$ tel que :

$$\forall y \in \mathbb{R}^d,\ \|y\| \leqslant \eta \implies \|\tau_y f - f\|_p^p \leqslant \varepsilon \tag{2.39}$$

On découpe alors l'intégrale en morceaux :

$$\int_{\mathbb{R}} \Phi_n(y)\|\tau_y f - f\|_p^p \mathrm{d}y = \int_{\|y\|\leqslant\eta} \Phi_n(y)\|\tau_y f - f\|_p^p \mathrm{d}y + \int_{\|y\|>\eta} \Phi_n(y)\|\tau_y f - f\|_p^p \mathrm{d}y \tag{2.40}$$

$$\leqslant \varepsilon \int_{\|y\|\leqslant\eta} \Phi_n(y)\mathrm{d}y + \int_{\|y\|>\eta} \|\tau_y f - f\|_p^p \Phi_n(y)\mathrm{d}y \tag{2.41}$$

$$\leqslant \varepsilon \underbrace{\int_{\mathbb{R}} \Phi_n(y)\mathrm{d}y}_{=1} + \int_{\mathbb{R}} \|\tau_y f - f\|_p^p \Phi_n(y)\mathrm{d}y \tag{2.42}$$

$$\tag{2.43}$$

Enfin, $\|\tau_y f - f\|_p^p \leqslant 2^p \|f\|_p^p$ car $\|\tau_y f\|_p = \|f\|_p$ (par invariance par translation de la mesure de Lebesgue) et donc :

$$\int_{\|y\|>\eta} \|\tau_y f - f\|_p^p \Phi_n(y)\mathrm{d}y \leqslant 2^p \|f\|_p^p \underbrace{\int_{\|y\|>\eta} \Phi_n(y)\mathrm{d}y}_{\xrightarrow[n\to+\infty]{} 0} \tag{2.44}$$

En recombinant toutes ces inégalités, on obtient :

$$\limsup_{n\to+\infty} \|f * \Phi_n - f\|_p^p \leqslant \varepsilon \tag{2.45}$$

et comme ceci est vrai pour tout $\varepsilon > 0$, on obtient bien que $f * \Phi_n$ converge vers f dans L^p. □

3 Probabilités : où fleurit l'abstraction

C'est ici que ça se corce pour moi... D'expérience, les étudiantes plus enclines à l'algèbre se détournent majoritairement des probabilités avec un dégoût prononcé. Vous faites partie de celles-ci ? De grâce, ne partez pas tout de suite ! Là est mon ambition : vous faire entrevoir que peut-être, tout ça n'est pas si vilain que ce que vous pouvez croire et même... en un sens plus proche de l'algèbre que ne l'est l'analyse ? Quant à savoir si j'y parviendrai, je vous en laisse seule arbitre.

Il me semble (mais ce n'est que mon humble avis d'étudiante de master) que la théorie des probabilités a une nature ambigüe. Parmi toutes les notions que l'on peut étudier, c'est l'une de celle qui laisse le plus facilement croire que tout y est intuitif, voire même d'une certaine manière peu rigoureux. Et c'est là qu'est le piège ! Les objets des probabilités, même les plus élémentaires, ne se comprennent pas sans un degré d'abstraction beaucoup plus élevé que celui nécessaire pour décrire les premiers objets de l'analyse (fonctions continues, dérivables ; suites, limites, etc.). Aussi, il n'est pas vraiment possible de comprendre ce qui se passe « sous le capot » avant la licence 3 et la théorie de la mesure.

Notre objectif ici est de tenter de reprendre les objets les plus élémentaires des probabilités et de les décrire dans un langage qui, peut-être, parlera également aux algébristes. Ce sera également l'occasion de donner un cadre des plus concret pour la manipulation de mesures abstraites, et donc un terrain de chasse des plus commodes pour celles qui préparent actuellement l'agrégation.

3.1 Loi d'une variable aléatoire... un objet universel ? !

Parmi les constructions de base de la théorie des probabilités, un certain nombre peut présenter une apparence abstruse qui a de quoi rebuter. Or certains de ces objets se comprennent en fait fort bien si plutôt que de se demander « qu'est-ce que c'est ? » on cherche plutôt à répondre à « que peut-on faire avec ? ». Commençons par un exemple loin d'être anodin :

Définition 3.1 (Tribu image, mesure image). *Soit $(X, \mathfrak{F}, \mu)$ un espace mesuré, E un ensemble et $f : X \to E$ une application.*

1. *Il existe une unique tribu $f_*(\mathfrak{F})$ sur E maximale pour l'inclusion et telle que $f : (X, \mathfrak{F}) \to (E, f_*(\mathfrak{F}))$ soit mesurable. On l'appelle la* tribu image *de $\mathfrak{F}$ par f.*
2. *Il existe une unique mesure $f_*(\mu)$ sur $(E, f_*(\mathfrak{F}))$ à support dans $f(X)$ telle que f soit une transformation préservant la mesure, dans le sens où :*

$$\forall A \in \mathfrak{F},\ \mu(A) = f_*(\mu)(f(A)) \tag{3.1}$$

On l'appelle la mesure image *de μ par f.*

Démonstration.

1. Si $\mathfrak{G}$ est une tribu sur E telle que $f : (X, \mathfrak{F}) \to (E, \mathfrak{G})$ est mesurable, doit avoir :

$$\forall B \in \mathfrak{G},\ f^{-1}(B) \in \mathfrak{F} \tag{3.2}$$

Il est donc naturel de poser :

$$f_*(\mathfrak{F}) := \{B \in \mathcal{P}(E) \mid f^{-1}(B) \in \mathfrak{F}\} \tag{3.3}$$

On vérifie aisément que c'est une tribu, et il est trivial que celle-ci rend f mesurable. Par ailleurs, si $\mathfrak{G}$ est une tribu sur E qui rend f mesurable, l'analyse menée au-dessus montre que $\mathfrak{G} \subset f_*(\mathfrak{F})$. $f_*(\mathfrak{F})$ est donc maximale pour l'inclusion et nécessairement unique.

2. Soit ν une mesure sur $(E, f_*(\mathfrak{F}))$ à support dans $f(X)$ telle que f préserve la mesure. Comme $f^{-1}(f(X)) = X \in \mathfrak{F}$, $f(X) \in f_*(\mathfrak{F})$ et donc pour tout $B \in f_*(\mathfrak{F})$, on a $B \cap f(X) \in f_*(\mathfrak{F})$. On remarque alors :

$$\forall B \in f_*(\mathfrak{F}),\ \nu(B) = \nu(B \cap f(X)) + \nu(B \setminus f(X)) = \nu(B \cap f(X)) \tag{3.4}$$

Or $f(f^{-1}(B)) = B$ lorsque $B \subset f(X)$. Donc :

$$\forall B \in f_*(\mathfrak{F}),\ \nu(B) = \nu(f(f^{-1}(B \cap f(X)))) = \mu(f^{-1}(B \cap f(X))) = \mu(f^{-1}(B)) \tag{3.5}$$

Il ne reste qu'à vérifier que cette formule définit une mesure pour obtenir d'un seul coup l'existence et l'unicité de la mesure image.

□

Remarque :

1. Si $\mathfrak{G}$ est une tribu donnée sur E, alors $f : (X, \mathfrak{F}) \to (E, \mathfrak{G})$ est mesurable si, et seulement si, $\mathfrak{G} \subset f_*(\mathfrak{F})$ (découle immédiatement de la construction de $f_*(\mathfrak{F})$.

2. Si $f : (X, \mathfrak{F}, \mu) \to (E, \mathfrak{G})$ est une application mesurable, on peut définir une mesure image sur $(E, \mathfrak{G})$ en restreignant la mesure image sur $(E, f_*(\mathfrak{F}))$) à $\mathfrak{G}$.

♦

Les anglophones appellent la mesure image « pushforward measure », terme qui reflète bien le but de cette construction : on dispose d'un espace mesuré $(X, \mathfrak{F}, \mu)$ et l'on aimerait transporter la mesure sur X dans E le long de l'application f. Nous allons illustrer ceci avec un exemple très concret (et commun !) de mesure image :

Définition 3.2 (Loi d'une variable aléatoire). *Soient $(\Omega, \mathfrak{F}, \mathbb{P})$ un espace probabilisé, $(E, \mathfrak{G})$ un espace mesurable et $X : (\Omega, \mathfrak{F}) \to (E, \mathfrak{G})$ une variable aléatoire (c'est-à-dire une application mesurable). On appelle* loi de X *la mesure image $X_*(\mathbb{P})$. C'est une mesure de probabilités sur $(E, \mathfrak{G})$ qu'on note plus fréquemment $\mathbb{P}_X$.*

Voir une loi de probabilité comme une mesure image offre une certaine façon de voir les variables aléatoires : c'est un moyen d'injecter une mesure de probabilité d'un espace Ω, souvent abominablement abstrait, dans un espace qu'on contrôle beaucoup mieux, mais sur lequel il aurait pu être difficile de construire directement la probabilité souhaitée. Pour mieux comprendre ce phénomène, jouons à un petit jeu. Mettons que je m'intéresse à un lancer de pièce équilibrée, et naturellement, je voudrais un espace probabilisé sur lequel définir la probabilité des évènements «pile» et «face». Rien de plus facile, n'est-ce pas ? Je prends l'ensemble $\{0, 1\}$ muni de sa tribu discrète et de la mesure $\frac{1}{2}(\delta_0 + \delta_1)$, et le tour est joué. Maintenant, je veux lancer ma pièce deux fois, et je suppose, cela va de soi, que les lancers sont indépendants. Comment construire mon espace ? Plus compliqué, mais on s'en sort en réfléchissant un peu. Maintenant, je veux lancer ma pièce *une infinité de fois*, de façon indépendante, et comme je suis d'humeur particulièrement tordue je voudrais au passage observer un phénomène numérique N suivant une loi de Poisson de paramètre 1 indépendant de mes lancers de pièce, et je veux manipuler tout cela à la fois, par exemple pour compter combien de fois j'ai obtenu « face » lors de mes N premiers lancers (sans oublier le résultat exact de mes lancers de pièce, cela va sans dire). Une idée de comment construire un espace avec la mesure adéquate ?...

C'est là que la puissance des variables aléatoires entre en jeu. Ce que je veux, c'est un tas de mesures sur $\mathbb{N}$ avec beaucoup de propriétés compliquées, et ça, c'est très difficile à construire explicitement. L'idée des probabilités est plutôt de construire un espace unique, suffisamment gros pour contenir toute l'information dont on a besoin, et de transporter les probabilités dans un espace mieux contrôlé à l'aide d'une batterie de variables aléatoires. Fort heureusement, il existe des théorèmes, sur lesquels nous auront l'occasion de dire quelques mots, affirmant que je peux construire un tel espace abstrait $(\Omega, \mathfrak{F}, \mathbb{P})$, qui hélas n'aura pas le bon goût de se comprendre facilement en termes de nombres entiers. Les variables aléatoires nous serviront à fermer consciencieusement les yeux sur l'aberration chthonienne qu'il aura fallut invoquer pour résoudre nos problèmes. Mais tout cela est pour la section suivante, pour l'heure, notre combat contre les lois de probas n'est pas encore terminé.

Remarque : C'est déjà à ce niveau là qu'on remarque un fossé qui se creuse entre la théorie de la mesure «analytique» et les probabilités. La nature des objets est la même, mais pas la manière dont on s'y intéresse. La loi des variables aléatoires (c'est-à-dire les mesures images) sont au cœur des probabilités lorsqu'on ne s'y intéresse jamais vraiment en analyse, préférant intégrer des fonctions contre des mesures bien établies plutôt que balader des objets abstraits entre espaces mesurables. ♦

On va pour parachever cette section revenir sur des termes de base des probabilités dans ce langage plus abstrait, et utiliser nos nouveaux jouets sophistiqués pour démontrer quelques caractérisations usuelles des lois de proba. Cette partie pourra, j'en suis sûre, particulièrement intéresser l'agrégative.

Définition 3.3 (Densité de probabilité). *Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^d$ de loi $\mathbb{P}_X$. On dit que X admet une densité (sous-entendu contre la mesure de* Lebesgue*) si sa loi est représentée par une fonction f, c'est-à-dire s'il existe une fonction mesurable et positive telle que $\mathbb{P}_X = f\,\mathrm{d}\lambda_d$, avec λ_d la mesure de* Lebesgue *de dimension de d.*

Remarque :

- De la définition de la mesure $f\,\mathrm{d}\lambda_d$, on retrouve immédiatement la définition usuelle (et s'usage pratique) de densité : si A est un borélien de $\mathbb{R}^d$, alors :

$$\mathbb{P}(X \in A) = \mathbb{P}_X(A) = [f\,\mathrm{d}\lambda_d](A) := \int_A f(x)\mathrm{d}x \tag{3.6}$$

- On peut étendre la notion de loi à densité aux lois discrètes. Une variables aléatoire à valeurs entières admet toujours une densité contre la mesure de comptage, donnée par $n \mapsto \mathbb{P}(X = n)$.

⧫

Exemple :

- La loi normale est la mesure sur $\mathbb{R}$ donnée par la densité $x \mapsto \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.
- La loi uniforme sur le pavé unité de $\mathbb{R}^d$ est donnée par la densité $\mathbb{1}_{[0,1]^d}$.
- Une variable aléatoire presque sûrement égale à 0 (c'est-à-dire de loi δ_0) n'a pas de densité. En effet, une loi à densité est diffuse, c'est-à-dire que tous les singletons ont probabilité 0 (c'est une conséquence des propriétés de la mesure de Lebesgue).

⧫

Théorème 3.4. *Si une variable aléatoire admet un densité, celle-ci est unique. Par ailleurs, la densité caractérise la loi.*

Démonstration. C'est une conséquence immédiate de la définition de la mesure $f\,\mathrm{d}\lambda_d$ et de la proposition 2.2. □

Remarque : Il aura fallu le lemme de classe monotones pour démontrer ce résultat, pourtant connu de toutes celles qui ont un jour touché aux probabilités continues ! ⧫

Définition 3.5 (Fonction de répartition). *Soit X une variable aléatoire à valeurs réelles. On appelle fonction de répartition de X la fonctions*

$$F_X : \mathbb{R} \to [0,1] \tag{3.7}$$
$$x \mapsto \mathbb{P}(X \leqslant x) = \mathbb{P}_X(]-\infty, x]) \tag{3.8}$$

Théorème 3.6. *La fonction de répartition caractérise la loi.*

Démonstration. J'en suis sûre, vous savez maintenant comme tout ça va finir.

Soient X et Y deux variables aléatoires de fonction de répartition commune F. On cherche à montrer $\mathbb{P}_X = \mathbb{P}_Y$, c'est-à-dire une égalité entre mesures. Il y a dans notre arsenal un joujou des polus indiqué pour ce faire : le lemme des classes monotones. Remarquons :

$$F_X = F_Y \iff \forall x \in \mathbb{R}, \mathbb{P}_X(]-\infty, x]) = \mathbb{P}_Y(]-\infty, x]) \tag{3.9}$$

Or la famille $(]-\infty, x])_{x\in\mathbb{R}}$ est un π-système générateur de la tribu borélienne. En vertu de la proposition 1.7, $\mathbb{P}_X = \mathbb{P}_Y$. □

Remarque : On l'avait en fait démontré mais sans le dire lors de la démonstration de la proposition 1.9. ⧫

Exercice 3.7. Soient $(U_n)_{n\in\mathbb{N}^*}$ une famille de variables aléatoires uniformes sur $[0,\theta]$, indépendantes [(v)], avec $\theta > 0$. On considère

$$\hat{\theta}_n := \max_{1\leqslant i\leqslant n} U_i \tag{3.10}$$

Montrer que la fonction de répartition de

$$\frac{1}{n}(\hat{\theta}_n - \theta) \tag{3.11}$$

converge simplement vers celle de la loi exponentielle de paramètre $1/\theta$ lorsque n tend vers $+\infty$.

Remarque : Je ne l'ai pas dit pour ne pas heurter les plus réfractaires, mais cet exemple, issu des statistiques, fournit un estimateur de θ dans le modèle uniforme des plus intéressant. Celui-ci est biaisé (c'est-à-dire que son espérance n'est pas égale à θ) et pourtant il converge presque sûrement vers θ à vitesse $1/n$, soit bien plus rapidement que le $1/\sqrt{n}$ d'habitude fourni par le théorème central limite ! Il se recase avec profit dans les différentes leçons de proba à l'agreg, car on peut lui appliquer toute une batterie de résultats au programme pour l'étudier : lemme de Borel-Cantelli pour prouver la convergence P.S., critères de convergence en loi, etc. ⧫

(v). Je sais que l'indépendance fait l'objet de la section suivante, mais cet exercice ne recquiert que la définition standarde d'indépendance. Vous devriez pouvoir vous en tirer sans ce qui suit !

3.2 Indépendance

On arrive maintenant à un point névralgique de la théorie des probabilités, qui n'a pas vraiment son équivalent en théorie de la mesure analytique : l'indépendance. C'est une notion particulièrement traîtreuse, dans la mesure où elle semble très intuitive à manipuler et, hélas, ne l'est pas. Observez plutôt cet exemple un tantinet calculatoire :

Exercice 3.8. Soient $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires de Bernoulli mutuellement indépendantes et de même paramètre $p \in]0,1[$. On souhaite ici modéliser une séquence de lancers de pièce *truquée* : on suppose donc $p \neq 1/2$. On note L_1 la longueur de la première série, c'est-à-dire le nombre le nombre de fois que la pièce est tombée du même côté que lors du premier lancer avant d'avoir changé de côté pour la première fois. On définit L_2 la longueur de la deuxième série de façon analogue. Par exemple, dans cette séquence :

$$\underbrace{\text{Pile Pile}}_{\text{première série}} \quad \underbrace{\text{Face Face Face}}_{\text{deuxième série}} \text{ Pile Face ...} \tag{3.12}$$

on obtient $L_1 = 2$ et $L_2 = 3$. Montrer que L_1 et L_2 ne sont pas indépendantes.

On aurait faussement pu croire que le fait que les lancers de pièce soient indépendants auraient conduit la longueur des séries à l'être également, mais le calcul rigoureux montre qu'il n'en est rien. On peut essayer d'expliquer ce résultat un peu étonnant en disant que la pièce étant truquée, un côté (mettons Face) a plus de chance que l'autre d'être tiré, et donc si la première série est longue, il y a plus de chance qu'il s'agisse d'une série de Face que d'une série de Pile, aussi la deuxième série aura tendance à être plus courte.

Un autre cas très célèbre où l'indépendance s'avère trompeuse est le fameux problème de Monty-Hall (voir par exemple [App23]).

Notre objectif ici n'est pas de faire un cours sur l'indépendance, qu'on trouvera dans tous les ouvrages de référence, mais plutôt d'essayer de voir ce qui, algébriquement, se cache sous la notion d'indépendance, sous l'angle de la théorie de la mesure. Avant toute chose, un peu d'heuristique.

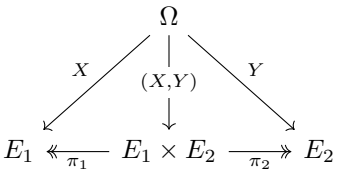
Considérons deux variables aléatoires X et Y définies sur un même espace probabilisé $(\Omega, \mathfrak{F}, \mathbb{P})$ et à valeurs dans des espaces mesurables $(E_1, \mathfrak{G}_1)$ et $(E_2, \mathfrak{G}_2)$. L'indépendance de ces deux variables se comprend mieux comme une propriété du *couple aléatoire* (X, Y), à valeurs dans (E_1, E_2) : sous réserve que cet espace soit muni d'une bonne tribu, l'indépendance de nos variables se traduit comme une propriété de leurs lois :

$$\begin{aligned} \forall (A,B) \in \mathfrak{G}_1 \times \mathfrak{G}_2,\ \mathbb{P}_{(X,Y)}(A\times B) &= \mathbb{P}(X\in A,\ Y\in B) && (3.13)\\ &= \mathbb{P}(X\in A)\mathbb{P}(Y\in B) && (3.14)\\ &= \mathbb{P}_X(A)\mathbb{P}_Y(B) && (3.15) \end{aligned}$$

Si l'on résume tout ceci, notre objectif est de construire une tribu sur l'espae produit $E_1 \times E_2$, de préférence la plus simple possible, qui réponde au cahier des charges suivant :

1. Les évènements de la forme $A \times B$ pour A et B mesurables sont dans cette tribu.
2. Un couple (X, Y) à valeurs dans $E_1 \times E_2$ est mesurable si, et seulement si, X et Y sont mesurable.

On dispose de deux applications de projection naturelles $\pi_i : E_1 \times E_2 \to E_i$, lesquelles fournissent un diagramme commutatif synthétisant la situation :



(3.16)

Une piste pour construire et comprendre cette tribu est d'étudier la mesurabilité des projections.

Proposition 3.9 (Tribu produit). *Soient $(X, \mathfrak{F})$ et $(Y, \mathfrak{G})$ deux espaces mesurables. Il existe une unique tribu sur $X \times Y$ minimale pour l'inclusion qui rende les projections naturelles $\pi_X : X \times Y \to Y$ et $\pi_Y : X \times Y \to Y$ mesurables. On l'appelle la* tribu produit de $\mathfrak{F}$ et $\mathfrak{G}$ *et on la note $\mathfrak{F} \otimes \mathfrak{G}$.*
Il s'agit de la tribu engendrée par les cylindres mesurables, c'est-à-dire :

$$\mathfrak{F} \otimes \mathfrak{G} := \sigma\left(A \times B,\ (A, B) \in \mathfrak{F} \times \mathfrak{G}\right) \tag{3.17}$$

Démonstration. Soit $\mathfrak{H}$ une tribu sur $X \times Y$. La projection π_X est mesurable pour $\mathfrak{H}$ si, et seulement si :

$$\forall A \in \mathfrak{F},\ \pi_X^{-1}(A) = A \times Y \in \mathfrak{H} \tag{3.18}$$

Ainsi, si les deux projections sont mesurables pour $\mathfrak{H}$, on a :

$$\forall (A, B) \in \mathfrak{F} \times \mathfrak{G},\ A \times B = (A \times Y) \cap (X \times B) \in \mathfrak{H} \tag{3.19}$$

et donc :

$$\sigma\left(A \times B,\ (A, B) \in \mathfrak{F} \times \mathfrak{G}\right) \subset \mathfrak{H} \tag{3.20}$$

Comme réciproquement il est clair que $\sigma\left(A \times B,\ (A, B) \in \mathfrak{F} \times \mathfrak{G}\right)$ rend les deux projections mesurables, on obtient le résultat. □

Remarque :

- L'ensemble produit $\mathfrak{F} \times \mathfrak{G}$ n'est pas, en général, une tribu. Par exemple, il est impossible de réaliser l'union des axes des abscisses et des ordonnées de $\mathbb{R}^2$ comme un produit de boréliens de $\mathbb{R}$, bien qu'elle soit dans la tribu $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ comme réunion des cylindres mesurables $\{0\} \times \mathbb{R}$ et $\mathbb{R} \times \{0\}$. En particulier, $\mathfrak{F} \otimes \mathfrak{G} \neq \mathfrak{F} \times \mathfrak{G}$ en général.
- L'ensemble $\mathfrak{F} \times \mathfrak{G}$ est un π-système, par définition générateur de $\mathfrak{F} \otimes \mathfrak{G}$. Ceci ouvrira la voie à l'utilisation du lemme des classes monotones sur la tribu produit. En particulier, l'équation (3.15) qui définit l'indépendance de deux variables aléatoires caractérise entièrement la loi de (X, Y) vu comme variable aléatoire à valeurs dans $(E_1 \times E_2, \mathfrak{G}_1 \otimes \mathfrak{G}_2)$ (on verra plus loin que (X, Y) est bien $\mathfrak{G}_1 \otimes \mathfrak{G}_2$-mesurable).
- Le fait de définir $\mathfrak{F} \otimes \mathfrak{G}$ par propriété universelle a l'avantage de rendre la construction très naturelle et d'en mettre le but en valeur. Toutefois, il est pratique souvent utile de revenir à la description explicite de cette tribu.

♦

Exemple :

- Si X et Y sont des ensembles au plus dénombrables munis de leurs tribus discrètes respectives, on a :
$$\mathcal{P}(X) \otimes \mathcal{P}(Y) = \mathcal{P}(X \times Y) \tag{3.21}$$
- Dans $\mathbb{R}$, on a le comportement fort appréciable suivant :
$$\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^2) \tag{3.22}$$
(voir [BP17]).

♦

Cette tribu répond bien à notre cahier des charges :

Exercice 3.10. Soient $X : (\Omega, \mathfrak{F}) \to (E_1, \mathfrak{G}_1)$ et $Y : (\Omega, \mathfrak{F}) \to (E_2, \mathfrak{G}_2)$ deux applications. Le couple (X, Y) est $\mathfrak{G}_1 \otimes \mathfrak{G}_2$-mesurable si, et seulement si, X et Y sont mesurables.

Munissons notre espace produit d'une mesure adéquate :

Théorème 3.11 (Mesure produit). *Soient $(X, \mathfrak{F}, \mu)$ et $(Y, \mathfrak{G}, \nu)$ deux espaces mesurés σ-finis. Il existe une unique mesure $\mu \otimes \nu$ sur $\mathfrak{F} \otimes \mathfrak{G}$, dite* mesure produit de μ et ν, *telle que :*
$$\forall (A, B) \in \mathfrak{F} \times \mathfrak{G},\ (\mu \otimes \nu)(A \times B) = \mu(A)\nu(B) \tag{3.23}$$
avec la convention $0 \times +\infty = 0$.

Démonstration. Voir [BP17] ou [Che14]. □

Il est essentiel pour la bonne manipulation des mesures produits d'introduire le théorème de Fubini :

Théorème 3.12 (Fubini). *Soient $(X, \mathfrak{F}, \mu)$ et $(Y, \mathfrak{G}, \nu)$ deux espaces mesurés σ-finis. Soit $f : X \times Y \to \mathbb{R}$ une fonction intégrable contre $\mu \otimes \nu$, ou bien à valeurs positives. Les fonctions $x \mapsto \int_Y f(x, y \mathrm{d}\nu(y)$ et $y \mapsto \int_X f(x, y \mathrm{d}\mu(x)$ sont mesurables et :*
$$\int_{X \times Y} f(x, y) \mathrm{d}\mu \otimes \nu(x, y) = \int_X \int_Y f(x, y) \mathrm{d}\nu(y)\, \mathrm{d}\mu(x) = \int_Y \int_X f(x, y) \mathrm{d}\mu(x)\, \mathrm{d}\nu(y) \tag{3.24}$$

Démonstration. Voir [BP17]. □

On est maintenant en mesure de formaliser l'intuition que nous nous sommes faites précédemment sur la notion d'indépendance.

Théorème 3.13 (Caractérisation de l'indépendance). *Soit $(\Omega, \mathfrak{F}, \mathbb{P})$ un espace probabilisé, et soient $(E_1, \mathfrak{G}_1)$ et $(E_2, \mathfrak{G}_2)$ des espaces mesurables. On se donne $X : (\Omega, \mathfrak{F}) \to (E_1, \mathfrak{G}_1)$ et $Y : (\Omega, \mathfrak{F}) \to (E_2, \mathfrak{G}_2)$ deux variables aléatoires. Alors le couple aléatoire (X, Y) est $\mathfrak{G}_1 \otimes \mathfrak{G}_2$-mesurable. De plus, (X, Y) sont indépendantes si, et seulement si :*
$$\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y \tag{3.25}$$

Démonstration. Montrons tout d'abord la $\mathfrak{G}_1 \otimes \mathfrak{G}_2$-mesurabilité. Pour cela, on se donne $(A, B) \in \mathfrak{G}_1 \times \mathfrak{G}_2$. On a alors, en terme d'évènements :

$$((X, Y) \in A \times B) = (X \in A) \cap (Y \in B) \in \mathfrak{F} \tag{3.26}$$

Ainsi, l'ensemble $\{C \in \mathfrak{G}_1 \otimes \mathfrak{G}_2 \mid (X, Y)^{-1}(C) \in \mathfrak{F}\}$ est une classe monotone (vérifications élémentaires) qui contient le π-système générateur formé des cylindres mesurables. Par le lemme des classes monotones 1.5, il s'agit de la tribu produit $\mathfrak{G}_1 \otimes \mathfrak{G}_2$ et donc notre couple aléatoire est bien une variable aléatoire.

Observons alors :

$$X \perp\!\!\!\perp Y \iff \forall (A, B) \in \mathfrak{G}_1 \times \mathfrak{G}_2,\ \mathbb{P}(X \in A,\ Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \tag{3.27}$$

$$\iff \forall (A, B) \in \mathfrak{G}_1 \times \mathfrak{G}_2,\ \mathbb{P}_{(X,Y)}(A \times B) = \mathbb{P}_X(A)\mathbb{P}_Y(B) = \mathbb{P}_X \otimes \mathbb{P}_Y(A \times B) \tag{3.28}$$

On conclut alors par unicité de la mesure produit (qui, comme on peut s'en douter, repose sur un argument de classes monotones). □

Cette caractérisation est plus qu'une façon abstraite de voir l'indépendance : c'est une porte ouverte sur l'utilisatio d'outils de théorie de la mesure qui trouve des application effiaces et pratiques. En voici un exemple :

Application 3.14. Soient X et Y deux variables aléatoires réelles indépendantes définies sur un même espace probabilisé $(\Omega, \mathfrak{F}, \mathbb{P})$. On suppose que X est une variable à densité contre la mesure de Lebesgue, c'est-à-dire qu'il existe une fonction positive d'intégrale 1 $h : \mathbb{R} \to \mathbb{R}$ telle que $\mathbb{P}_X = h(x)\mathrm{d}x$. Alors la variable aléatoire $X + Y$ admet la densité :

$$s \mapsto \int_{\mathbb{R}} h(s - y)\mathrm{d}\mathbb{P}_Y(y) \tag{3.29}$$

Démonstration. On va exploiter la proposition 1.9 et sa reformulation en termes d'intégrales. Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction continue bornée. L'application $\omega \mapsto f(X(\omega) + Y(\omega))$ peut être vue comme la composée des applications :

$$\omega \mapsto (X(\omega), Y(\omega)) \mapsto X(\omega) + Y(\omega) \mapsto f(X(\omega) + Y(\omega)) \tag{3.30}$$

où l'application de somme est mesurable. On peut donc écrire :

$$\mathbb{E}[f(X + Y)] = \int_{\mathbb{R}^2} f(x + y)\mathrm{d}\mathbb{P}_{(X,Y)}(x, y) \tag{3.31}$$

$$= \int_{\mathbb{R}^2} f(x + y)\mathrm{d}[P_X \otimes \mathbb{P}_Y](x, y) \tag{3.32}$$

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} f(x + y)\mathrm{d}\mathbb{P}_X(x)\, \mathrm{d}\mathbb{P}_Y(y) \tag{3.33}$$

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} f(x + y)h(x)\mathrm{d}x\, \mathrm{d}\mathbb{P}_Y(y) \tag{3.34}$$

$$= \int_{\mathbb{R}} f(s) \left(\int_{\mathbb{R}} h(s - y)\mathrm{d}\mathbb{P}_Y(y) \right) \mathrm{d}s \tag{3.35}$$

$$=: \int_{\mathbb{R}} f(s)\mathrm{d}\mathbb{P}_{X+Y}(s) \tag{3.36}$$

en utilisant, dans l'ordre, le théorème de transfert, l'indépendance de X et de Y, le théorème de Fubini et enfin le changement de variables $s \leftarrow x + y$. Ainsi, comme c'est vrai pour toute telle fonction f, on trouve bien :

$$\mathbb{P}_{X+Y} = \left(\int_{\mathbb{R}} h(s-y) \mathrm{d}\mathbb{P}_Y(y) \right) \mathrm{d}s \tag{3.37}$$

□

On va maintenant chercher à étendre nos constructions au cas d'une famille quelconque de variables aléatoiers.

Définition 3.15 (Tribu produit, cas général). *Soit $(X_i, \mathfrak{F}_i)_{i \in I}$ une famille quelconque d'espaces mesurables. Il existe une unique tribu minimale pour l'inclusion sur le produit $\prod_{i \in I} X_i$ qui rende toutes les projections mesurables. On la note $\otimes_{i \in I} \mathfrak{F}_i$*

Proposition 3.16 (Description par générateurs). *La tribu produit $\otimes_{i \in I} \mathfrak{F}_i$ est la tribu engendrée par les cylindres mesurables, c'est-à-dire les parties de la forme $\prod_{i \in I} A_i$, avec $A_i \in \mathfrak{F}_i$ où tous sauf un nombre fini sont égaux) X_i. De plus, pour tout $A \in \otimes_{i \in I} \mathfrak{F}_i$, il existe $J \subset I$ un ensemble au plus dénombrable et $A_J \in \otimes_{i \in J} \mathfrak{F}_i$ tels que :*

$$A = A_J \times \prod_{i \in I \setminus J} X_i \tag{3.38}$$

Démonstration. Le fait que la tribu produit soit engendrée par les cylindres mesurables est immédiat en suivant la démonstration faite dans le cas de deux tribus. Pour la description en extension, on vérifie que les parties de la forme voulue forment une tribu qui contient les cylindres mesurables. L'inclusion réciproque est fastidieuse à obtenir, nous la passons sous silence. □

Théorème 3.17 (Mesure produit, cas général). *Soit I un ensemble d'indices quelconque. On considère $(\Omega_i, \mathfrak{F}_i, \mathbb{P}_i)_{i \in I}$ une famille d'espaces probabilisés indicée par I. Il existe une unique mesure de probabilités définie sur $\otimes_{i \in I} \mathfrak{F}_i$ telle que :*

$$\forall J \in \mathcal{P}_d(I),\ \forall (A_i)_{i \in J} \in \prod_{i \in J} \mathfrak{F}_i,\ \bigotimes_{i \in I} \mathbb{P}_i \left(\prod_{i \in J} A_i \times \prod_{i \in I \setminus J} \Omega_i \right) = \prod_{i \in J} \mathbb{P}_i(A_i) \tag{3.39}$$

avec $\mathcal{P}_d(I)$ l'ensemble des parties au plus dénombrables de I.

Démonstration. Voir [Che14]. □

Théorème 3.18 (Caractérisation de l'indépendance 2). *Soit $(X_i)_{i\in I}$ une famille quelconque de variables aléatoires définies sur un même espace probabilisé, à valeurs respectivement dans un espace mesurable $(E_i, \mathfrak{G}_i)$. Alors l'application $\omega \mapsto (X_i(\omega))_{i\in I}$, à valeurs dans $\prod_{i\in I} E_i$, est $\otimes\mathfrak{G}_i$-mesurable. De plus, la famille (X_i) est mutuellement indépendante si, et seulement si :*

$$\mathbb{P}_{(X_i)_{i\in I}} = \bigotimes_{i\in I} \mathbb{P}_{X_i} \tag{3.40}$$

3.3 Théorèmes d'existence

On va enfin pouvoir énoncer les théorèmes qui affirment que tout les objets que l'on cherche à manipuler en probabilités *existent*, ce qui, comme nous le disions précédemment, est loin d'être évident ! Commeçons par un constat, qui aura l'air d'une trivialité à démontrer mais qu'il est bon d'avoir remarqué.

Lemme 3.19. *Soit μ une mesure de probabilité un espace mesurable $(E, \mathfrak{G})$. Il existe un espace probabilisé $(\Omega, \mathfrak{F}, \mathbb{P})$ et une variable aléatoire $X : (\Omega, \mathfrak{F}) \to (E, \mathfrak{G})$ dont μ est la loi.*

Démonstration. Il suffit en fait de prendre $(\Omega, \mathfrak{F}, \mathbb{P}) = (E, \mathfrak{G}, \mu)$ et X l'indentité sur E. □

Cette petite mise en bouche a le mérite de rassurer. L'objet de notre théorème est l'existence d'une famille quelconques de variables aléatoires indépendantes suivant une loi donnée à l'aide des constructions abstraites précédemment faites.

Théorème 3.20. *Soit $(E, \mathfrak{G})$ un espace mesurable et $(\mu_i)_{i\in I}$ une famille de mesures de probabilités sur $\mathfrak{G}$. Il existe un espace probabilisé $(\Omega, \mathfrak{F}, \mathbb{P})$ et une famille de variables aléatoires $(X_i)_{i\in I}$ toutes définies sur Ω et à valeurs dans E, mutuellement indépendantes, et telles que :*

$$\forall I \in I,\ \mathbb{P}_{X_i} = \mu_i \tag{3.41}$$

Démonstration. On pose :

$$(\Omega, \mathfrak{F}, \mathbb{P}) := \left(E^I,\ \mathfrak{G}^{\otimes I},\ \bigotimes_{i\in I} \mu_i \right) \tag{3.42}$$

et on considère X l'indentité sur Ω. Pour $j \in I$, notons :

$$\begin{aligned} \pi_j : \mathbb{E}^I &\to \mathbb{E} \\ (x_i)_{i\in I} &\mapsto x_j \end{aligned} \tag{3.43}$$

la j-ième projection naturelle. Elle est mesurable par définition de la tribu produit. Ainsi, la famille $(X_i)_{i\in I} := (\pi_i(X))_{i\in I}$ définit une famille de variables aléatoires mutuellement indépendantes suivant les lois respectives μ_i. □

Remarque : Au risque de me répéter, j'aimerais insister sur le fait que l'espace probabilisé sous jacent à un modèle utilisé est donc généralement un espace produit dont la description explicite est peu ragoûtante. Tout le jeu des probabilités est justement d'utiliser des variables aléatoires pour ne jamais travailler que dans un espace unique et facile à manipuler, et surtout ne jamais, au grand jamais, s'intéresser à ce qui se passe réellement dans l'espace de départ.

Cette philosophie rend relativement bien compte de l'utilisation des probabilités qu'on peut faire dans certaines situations physique. Par exemple, un lancer de pièce *est* déterministe : si on connait à la perfection les conditions initiales du lancé, la température de la pièce, le mouvement de l'air, etc. les équation de la dynamique permettront de prédire avec justesse le côté sur lequel la pièce retombera. En pratique, le croisement de toutes ces conditions représente une somme énorme et très variable d'informations dont l'issue dépend de manière très sensible, aussi on préfère au modèle déterministe un modèle probabiliste. Ceci correspond à peu près à juger que l'espace réel des issues (l'univers Ω) est trop compliqué pour que cela vaille la peine de s'y intéresser, aussi le masque-t-on derrière de commodes « variables aléatoire » (qui n'ont donc finalement rien d'aléatoire). ⧫

3.4 Convergence de variables aléatoires... ou pas !

La convergence des variables aléatoires est un thème essentiel et récurrent, aussi bien en pures probabilités qu'en statistiques, et qui occupe une leçon entière à l'agrégation (en plus de fournir un bon nombre d'exemples et d'applications qui se recasent très bien). Comme d'habitude, mon objectif ici n'est pas de faire un cours sur les modes de convergence (vous en trouverez dans tous les bons bouquins de proba) mais plutôt de d'essayer de donner une compréhension intuitive de ces différents modes de convergence, qui ont su rendre confuses un bon nombre de personnes ! Nous consacrerons une belle part de cette section à la convergence en loi, le plus « probabiliste » des modes de convergence, parce qu'il a un comportement radicalement différent des autres (et pour une bonne raison).

Pour notre bonne intelligence, commençons par rappeler quelques définitions :

Définition 3.21 (Convergence P.S., L^p, en probabilités). *Soit (X_n) une suite de variables aléatoires à valeurs dans $\mathbb{R}^d$ et X une variable aléatoire à valeurs dans $\mathbb{R}^d$, toutes définies sur le même espace probabilisé $(\Omega, \mathfrak{F}, \mu)$. On dit que :*

1. (X_n) converge presque sûrement *vers X si :*

$$\mathbb{P}(X_n \to X) = 1 \tag{3.44}$$

2. (X_n) converge dans L^p *vers X si :*

$$\|X_n - X\|_p \to 0 \tag{3.45}$$

3. (X_n) converge en probabilités *vers X si :*

$$\forall \varepsilon > 0,\ \mathbb{P}(\|X_n - X\| \geqslant \varepsilon) \to 0 \tag{3.46}$$

Remarque :

1. La convergence presque sûre et L^p n'ont en fait de probabiliste que leur nom : il s'agit de modes de convergence de fonctions mesurables tout à fait usuel en analyse (convergence

presque partout et dans L^p).

2. L'ensemble
$$\{\omega \in \Omega \mid X_n(\omega) \to X(\omega)\} \tag{3.47}$$
est bien un évènement (i.e. un élément de $\mathfrak{F}$) puisqu'on peut l'exprimer comme :
$$\bigcap_{q\in\mathbb{Q}_+^*} \bigcup_{N\in\mathbb{N}} \bigcap_{n\geqslant N} \{\omega \in \Omega \mid \|X_n(\omega) - X(\omega)\| \leqslant q\} \tag{3.48}$$
qui est une interection dénombrable d'unions dénombrables d'intersections dénombrables de parties mesurables (puisque X_n et X sont des variables aléatoires).

⧫

A simplement regarder les définitions, il peut être relativement difficile de faire la différence entre la convergence presque sûre et la convergence en probabilités, et pourtant cette dernière est strictement plus faible que la première (on en donnera des contre-exemples un peu plus bas). Je pense que la différence entre les deux se comprend assez bien en travaillant sur des variables aléatoires à densité. Avant de poursuivre, je vous invite à essayer de résoudre le petit exercice suivant, qui se révèlera des plus éclairant :

Exercice 3.22. Soient (X_n) des variables aléatoires réelles définies sur le même espace probabilisé $(\Omega, \mathfrak{F}, \mathbb{P})$ et admettant les densités respectives contre la mesure de Lebesgue (f_n). Réexprimer en fonction des densités le fait que (X_n) converge en probabilités vers 0.

Solution. Soit $\varepsilon > 0$. On calcule alors :
$$\mathbb{P}(|X_n| \geqslant \varepsilon) = \mathbb{E}[\mathbb{1}_{|X_n|\geqslant\varepsilon}] \tag{3.49}$$
$$= \int_\Omega \mathbb{1}_{|X_n(\omega)|\geqslant\varepsilon} \,\mathrm{d}\mathbb{P}_{X_n}(\omega) \tag{3.50}$$
$$= \int_{|x|\geqslant\varepsilon} f_n(x)\,\mathrm{d}x \tag{3.51}$$
Par définition, la convergence en probabilités de (X_n) vers 0 est équivalente à ce que ce dernier terme converge vers 0 quelque soit la valeur de ε préalablement fixée. En français, cela signifie que la masse des fonctions (f_n) tend à se concentrer au voisinage de 0. En particulier, il n'est pas nécessaire que *toute* la masse des (f_n) se concentrer au voisinage de 0. □

Remarque : Cette nouvelle expression montre que la convergence en probabilités apparaît elle aussi parfois en analyse, bien que de manière plus cachée. Sauriez-vous trouver où ? Nous avons déjà eu l'occasion d'en observer un occurence au cours de ce document ! [(vi)] ⧫

La différence avec la convergence presque sûre réside ici : avec le même cadre, pour que (X_n) converge *presque sûrement vers 0*, il faudrait que pour tout $\varepsilon > 0$, et pour presque tout $\omega \in \Omega$, tous les $(X_n(\omega))$ soient de valeur absolue plus petite que ε à partir d'un certain rang, ce qui n'arrivera que si la masse des (f_n) se concentre *suffisamment vite* vers 0. A partir de cette idée, on va élaborer un contre-exemple qui je l'espère rendra plus concret ce que j'ai cherché à

(vi). Regardez-bien la définition d'approximation de l'unité dans le théorème 2.6.

exprimer. Pour plus de commodité d'écriture, on prendra des variables aléatoires à densité non pas contre la mesure de Lebesgue mais contre la mesure de comptage sur $\mathbb{N}$ (c'est-à-dire, tout simplement, des variables aléatoires à valeurs entières) ; le principe est exactement le même.

Exemple : Pour $n \in \mathbb{N}^*$, on considère X_n une variable aléatoire à valeurs dans $\{0,1\}$ suivant la loi $(1-1/n)\delta_0 + 1/n\delta_1$. On suppose les (X_n) mutuellement indépendantes. Alors (X_n) converge en probabilités vers 0, mais pas presque sûrement. En effet :

$$\forall 0 < \varepsilon < 1,\ \mathbb{P}(|X_n| \geqslant \varepsilon) = \mathbb{P}(X_n = 0) \tag{3.52}$$
$$= 1 - \frac{1}{n} \tag{3.53}$$
$$\to 0 \tag{3.54}$$

Pourtant, pour $\omega \in \Omega$, $X_n(\omega) \to 0$ si, et seulement si, la suite $(X_n(\omega))$ est ultimement constante égale à 0. On calcule directement, pour $N \in \mathbb{N}^*$:

$$\mathbb{P}\left(\bigcap_{n \geqslant N} (X_n = 0)\right) = \prod_{n \geqslant N} \mathbb{P}(X_n = 0) = \prod_{n \geqslant N} \left(1 - \frac{1}{n}\right) \text{ (vii)} \tag{3.55}$$

On passe au log pour plus de commodité :

$$\log\left(\mathbb{P}\left(\bigcap_{n \geqslant N} (X_n = 0)\right)\right) = \sum_{n \geqslant N} \log\left(1 - \frac{1}{n}\right) \tag{3.56}$$

Comme $\log(1-1/n)$ est équivalent à $-1/n$, qui est le terme d'une série divergente, la somme de droite est égale à $-\infty$ et donc notre produit vaut 0. Ainsi par le théorème de sous-additivité des probabilités :

$$\mathbb{P}(X_n \to 0) = \mathbb{P}\left(\bigcup_{N \in \mathbb{N}^*} \bigcap_{n \geqslant N} (X_n = 0)\right) \leqslant \sum_{N \in \mathbb{N}^*} \mathbb{P}\left(\bigcap_{n \geqslant N} (X_n = 0)\right) = 0 \tag{3.57}$$

♦

Remarque : Bien entendu, cette idée de « converger suffisamment vite vers 0 » est celle qu'expriment les lemmes de Borel-Cantelli. Ici, on a simplement observé à la main le lien avec la convergence des séries $\sum \mathbb{P}(|X_n| \geqslant \varepsilon)$. ♦

On pourra trouver d'autres contre-exemples du même genre dans [FF98]. Ceux-ci sont indispensables à connaître pour l'agrégation !

Cette petite mise en bouche faite, il est temps de passer au parangon de la convergence des variables aléatoires : la convergence en loi.

Définition 3.23 (Convergence en loi). *Soit (X_n) et X des variables aléatoires à valeurs dans $\mathbb{R}^d$, de lois respectives (P_{X_n}) et P_X. On dit que* (X_n) converge en loi *vers X lorsque :*

$$\forall \varphi \in C_b^0(\mathbb{R}^d),\ \mathbb{E}[\varphi(X_n)] \to \mathbb{E}[\varphi(X)] \tag{3.58}$$

(vii). Ce produit est bien défini, donné par le théorèe de continuité décroissante des probabilités.

Ce mode de convergence est nettement plus difficile à appréhender et ne se retrouve pas comme les autres en analyse, et ce pour une bonne raison : **ce n'est pas un mode de convergence de variables aléatoires**. Vous ne me croyez pas ? Résolvez plutôt l'exercice suivant :

Exercice 3.24. Soient (X_n), X et Y des variables aléatoires à valeurs dans $\mathbb{R}^d$. On suppose que (X_n) converge en loi vers X. Donner une condition nécessaire et suffisante sur X et Y pour que (X_n) converge également en loi vers Y.

Trève de suspens : on obtient la convergence simultanée vers X et Y si, et seulement si, X et Y sont égales en loi. Ce constat révélateur suffit à embrasser plusieurs spécificités de la convergence en loi, notamment :

1. La topologie de la convergence en loi n'est pas séparée (une suite peut avoir plusieurs limites) ce qui est relativement déplaisant pour un mode de convergence de variables aléatoires.
2. Pire : (X_n) peut très bien converger en loi vers une variable aléatoire X qui n'est même pas définie sur le même espace probabilisé. Il suffit de regarder en face la définition de loi d'une variable aléatoire pour constater que celle-ci ne fait pas intervenir l'espace sous-jacent [(viii)].

La convergence en loi ne dépendant que de la loi des variables aléatoires manipulées, il serait en fait plus juste de la voir comme un mode de convergence... eh bien des lois de probabilités, comme l'indiquait déjà le nom ! Cet autre façon de voir les choses se formalise très bien à l'aide d'un résultat qu'on a déjà observé. On introduit à nouveau l'opérateur :

$$T : \mathscr{P}(\mathbb{K}^d) \to (C_b^0(\mathbb{K}^d, \mathbb{K}))' \tag{3.59}$$

$$\mathbb{P} \mapsto \left(f \mapsto \int_{\mathbb{K}^d} f(x) \, \mathrm{d}\mathbb{P}(x) \right) \tag{3.60}$$

avec $\mathscr{P}(\mathbb{K}^d)$ l'ensemble des mesures de probabilité boréliennes sur $\mathbb{K}^d$. On a prouvé que cet opérateur est injectif (proposition 1.10). Une réécriture immédiate de la convergence en loi de (X_n) vers X fournit l'équivalence :

$$[X_n \overset{(\mathscr{L})}{\longrightarrow} X] \iff [T(\mathbb{P}_{X_n}) \xrightarrow{\text{c.v.s.}} T(\mathbb{P}_X)] \tag{3.61}$$

qui exprime de manière plus claire le fait qu'on s'intéresse en réalité à une convergence au niveau des mesures de probabilité. A recaser dans la leçon sur les formes linéaires continues !

Remarque :

1. Si vous avez quelques notions de théorie des distributions, ce type de convergence ne vous est certainement pas étranger, et c'est bien normal. On vient en fait de définir la notion de convergence étroite pour certaines classes de mesures (convergence dans le dual des fonctions continues bornées). Mais une mesure est toujours une distribution ! Il se trouve qu'on peut montrer, dans le cas des mesures de probabilité, que la convergence étroite est équivalente à la convergence au sens des distributions, c'est-à-dire qu'on peut remplacer notre classe de fonctions test par C_c^∞. La convergence en loi de (X_n) vers X est donc équivalente à la convergence au sens des distributions des lois des (X_n), vues comme des distributions, vers celle de X.

(viii). Cette petite observation permet de fabriquer très facilement des exemples de suites qui convergent en loi mais pas en probabilités (très utile à l'agrégation !) : on peut par exemple prendre, pour tout $n \in \mathbb{N}$, $X_n : x \mapsto x$ définie sur $[0,1]$, qui suit donc la loi uniforme sur $[0,1]$ et $X : x \mapsto x - 1$ définie de $[1,2]$ dans $[0,1]$. Puisque X suit également la loi uniforme sur $[0,1]$, (X_n) converge en loi vers X, mais il n'y a aucun espoir pour récupérer la convergence en probabilités.

2. Pour celles qui apprécient les notions de topologie abstraite, il s'agit ici d'un cas particulier de convergence faible-$\star$, avec ses théorèmes de compacité qui interviennent par exemple pour une preuve du TCL (on pourra par exemple aller lire le développement (méchamment hors-programme pour l'agrégation) de Thomas Cavalazzi ici : [Cav19]). Si ces notions vous intéressent, vous pourrez trouver de plus amples informations ici [Bré83].

⧫

$\star \qquad \star$

$\star$

A †‡ Construction en extension d'une tribu

On a évoqué plusieurs fois le fait que les tribus sont des objets méchamment sauvages (sans mauvais jeu de mot) et qu'on ne peut pas espérer en faire une simple construction itérative comme unions d'intersections de complémentaires d'une partie génératrice. On va dans cette partie investiguer plus avant ce fait pour mieux comprendre à quel point une tribu est un objet compliqué. Cette étude nous emmènera assez loin dans l'abstraction, et je n'en connais pas de réelles applications. Je pense qu'il vaut mieux considérer tout ceci, au niveau du master, comme des maths récréatives qui constitueront une occasion d'aborder les nombres ordinaux et les constructions transfinies.

Petit disclaimer : cette section porte sur des notions sur lesquelles j'ai lu pour m'amuser, mais que je n'ai jamais proprement travaillées. A prendre avec des pincettes, retourner vers les références en cas de doute.

A.1 Nombres ordinaux

On sent, au fond de nous, qu'il est possible de faire une construction itérative d'une tribu à partir de ses générateurs en prenant des unions d'intersections de complémentaires des éléments du système générateur, puis de recommencer, etc. Malheureusement, on s'apercevra que $\mathbb{N}$, tout infini soit-il, n'est pas assez long pour que le procédé aboutisse.

On va ici introduire la récurrence transfinie, qui permet de faire des constructions et des preuves récursives qui vont « beaucoup plus loin » que $\mathbb{N}$. Pour cela, on aura besoin de la notion de nombres ordinaux. L'idée ici n'est pas de faire une présentation détaillée des ordinaux (ceci étant assez laborieux) mais plutôt d'en faire un débroussaillage intuitif, suffisant pour comprendre la récurrence transfinie, en admettant tous les théorèmes techniques qui justifient que tout est bien posé. Nous renvoyons à [Dug66] pour une présentation détaillée.

Définition 1.1 (Bon ordre). *Soit X un ensemble. On dit qu'un ordre $\preccurlyeq$ sur X est un* bon ordre *si :*

1. $\forall (x,y) \in X^2,\ x \preccurlyeq y \vee y \preccurlyeq x$
2. *Tout partie non vide de X admet un plus petit élément.*

Exemple : L'ordre naturel sur $\mathbb{N}$ est un bon ordre (c'est d'ailleurs pour ça qu'on l'a construit). Étendu à $\mathbb{Z}$, ça n'est plus la cas, bien que ça soit toujours un ordre total. En effet, $\mathbb{Z}$ n'a pas de plus petit élément. ⧫

Les bons ordres sont à la base des raisonnements par récurrence. L'idée des ordinaux va être d'étendre les bons ordres à des ensembles bien plus compliqués que $\mathbb{N}$.

Théorème 1.2 (Zermelo). *Tout ensemble peut être muni d'un bon ordre.*

Démonstration. Cf [Dug66], §II.2.1. □

Pour étudier les bons ordres, on va chercher à établir des « modèles » des ensembles bien ordonnées. L'idée un peu naïve qui se cache derrière est la suivante : supposons qu'on puisse manipuler la classe des ensembles bien ordonnés comme si elle était un ensemble. On quotiente cette classe par la relation d'isomorphisme pour les ensembles bien ordonnés, c'est-à-dire qu'on assimile deux ensembles bien ordonnés s'ils sont en bijection croissante l'un avec l'autre. On fixe alors un système de représentants des classes d'équivalences, qui formeront nos nombres ordinaux, de sorte à ce qu'un ensemble bien ordonné soit par construction isomorphe à un unique nombre ordinal. Le soucis de cette construction est qu'elle déborde du cadre de la théorie des ensembles, ce qui dans certaines situations pose des problèmes logiques. Nous allons plutôt suivre une construction ensembliste des nombres ordinaux.

Définition 1.3 (Nombre ordinal). *Soit α un ensemble. On dit que c'est un* nombre ordinal *si :*

1. $\forall (x,y) \in \alpha,\ [x \in y] \vee [y \in x] \vee [x = y]$
2. $\forall y \in \alpha,\ \forall x \in y,\ x \in \alpha$

Exemple : Si vous vous souvenez de la construction formelle de $\mathbb{N}$, ses éléments sont justement des nombres ordinaux. Observons :

$$0 := \emptyset \quad 1 := \{\emptyset\} \quad 2 := \{\emptyset,\ \{\emptyset\}\} \quad \dots \tag{1.1}$$

Toutefois, il existe d'autres nombres ordinaux, par exemple :

$$\mathbb{N} \cup \{\mathbb{N}\} \tag{1.2}$$

est un nombre ordinal. ⧫

On se contentra dans la suite d'aller à l'essentiel afin de manipuler ces ordinaux, laissant de côté les démonstrations et autres considérations techniques.

Proposition 1.4 (Classe des ordinaux).

1. *Soit α un nombre ordinal. La relation :*

$$\forall (x,y) \in \alpha^2,\ [x \preccurlyeq y] \equiv [x \in y] \tag{1.3}$$

   définit un bon ordre sur α.
2. *Si α est un nombre ordinal, tous les éléments de α sont des nombres ordinaux.*
3. *La classe de tous les nombres ordinaux est une classe bien ordonnée par la relation d'appartenance. Ce n'est pas un ensemble.*
4. *Pour tout nombre ordinal α, la classe de tous les nombres ordinaux inférieurs stricts à α est tout simplement α.*

Théorème 1.5. *Soit $(X, \preccurlyeq)$ un ensemble bien ordonné. Il existe un unique nombre ordinal α tel que $(X, \preccurlyeq)$ soit en bijection croissante avec $(\alpha, \in)$.*

On peut faire des opérations sur les nombres ordinaux qui miment celles sur les entiers naturels. On se contentra de la plus simple d'entre toutes :

Définition 1.6 (Successeur, ordinal limite). *Soit α un nombre ordinal.*

1. *On appelle* successeur de α, *noté $\alpha+1$, l'ensemble $\alpha \cup \{\alpha\}$. C'est un nombre ordinal suppérieur à α.*
2. *On dit qu'un ordinal non-vide α est un ordinal limite s'il n'est le successeur d'aucun nombre ordinal.*

Exemple : $\mathbb{N}$ muni de son ordre naturel est un nombre ordinal limite qu'on note souvent ω. En effet, tout ordinal inférieur strict à $\mathbb{N}$ est par définition un élément de $\mathbb{N}$, donc un entier naturel. Or le successeur de tout entier naturel est encore un entier naturel. Son successeur $\omega+1$ est tout simplement $\mathbb{N} \cup \{\mathbb{N}\}$. ♦

Ne perdons pas de vue notre objectif : faire des raisonnements par récurrence en allant « plus loin » que $\mathbb{N}$. On va exploiter les ordinaux et la notion de successeur. Lorsqu'on travaille usuellement uniquement avec les entiers naturels, le cas des ordinaux limites ne se présente pas, et cette différence va apparaître dans le théorème de récurrence transfinie :

Théorème 1.7 (récurrence transfinie). *Notons $\mathscr{O}$ la classe des ordinaux. Soit $(\mathscr{P}(\alpha))_{\alpha\in\mathscr{O}}$ une famille de propositions indexée par la classe des ordinaux*[ix]*. On suppose :*

1. *$\mathscr{P}(0)$*
2. *$\forall \alpha \in \mathscr{O},\ \mathscr{P}(\alpha) \implies \mathscr{P}(\alpha+1)$*
3. *Pour tout ordinal limite λ :*

$$[\forall \alpha < \lambda, \mathscr{P}(\alpha)] \implies \mathscr{P}(\lambda) \tag{1.4}$$

Alors, $\mathscr{P}(\alpha)$ est vraie pour tout ordinal α.

Démonstration. La démonstration est assez simillaire à la démonstration du théorème usuel de récurrence, révalant qu'on utilise usuellement uniquement la nature d'ordinaux des entiers naturels. Supposons par l'absurde qu'il existe un ordinal α tel que $\mathscr{P}(\alpha)$ est fausse. Puisque la classe des ordinaux est bien ordonnée, il existe un ordinal α_0 minimal tel que $\mathscr{P}(\alpha_0)$ est fausse. Or les hypothèses qu'on a faites montrent :

1. $\alpha_0 > 0$
2. α_0 est un ordinal limite. Dans le cas contraire, il existe un ordinal $\alpha_0 - 1$ tel que $\alpha_0 - 1 + 1 = \alpha_0$. Par construction, $\mathscr{P}(\alpha_0 - 1)$ est vraie, et donc l'hypothèse 2 implique que $\mathscr{P}(\alpha_0)$ est vraie.

Mais alors la minimalité de α_0 contredit l'hypothèse 3. □

(ix). Si le recours à une famille indexée par une classe propre vous gêne, on peut toujours se limiter en fixant un ordinal α^* et en se donnant une famille indexée par l'ensemble des ordinaux inférieurs stricts à α^* (qui n'est autre que α^* lui-même). Dans ce cas la conclusion du théorème est simplement $\mathscr{P}(\alpha)$ pour tout $\alpha < \alpha^*$ (ce théorème englobant déjà le théorème de récurrence usuel en prenant $\alpha^* = \alpha_0$.

Remarque : Il est possible de formuler ce théorème d'une manière beaucoup plus efficace, mais bien moins pratique, à l'aide d'une seule hypothèse. En effet, l'assertion

$$\forall \alpha \in \mathscr{O},\ [\forall \alpha < \alpha, \mathscr{P}(\alpha)] \implies \mathscr{P}(\alpha) \tag{1.5}$$

est équivalente aux trois hypothèses du théorème. En pratique toutefois, on fait souvent les récurrences transfinies en traitant séparément l'initialisation, le passage au successeur et le franchissement des ordinaux limites. ♦

A.2 Construction d'une tribu par récurrence transfinie

Grâce à la théorie des ordinaux, on va pouvoir construire itérativement une tribu à partir d'une famille génératrice. Pour cela, on va suivre la construction dite de la « hiérarchie de Borel ».

Fixons X un ensemble et G une famille de parties de X. Pour construire $\sigma(G)$, on définit la famille transfinie $(A_\alpha)_{\alpha \in \mathscr{O}}$ par les trois règles suivantes :

1. $\Delta_0 := G \cup \{X\}$
2. Pour tout ordinal α, on passe au successeur via :

$$\Delta_{\alpha+1} := \left\{ \left(\bigcup_{n\in\mathbb{N}} A_n \right) \cup \left(\bigcup_{n\in\mathbb{N}} B_n^c \right),\ (A_n), (B_n) \in \Delta_\alpha^{\mathbb{N}} \right\} \tag{1.6}$$

3. On franchit les ordinaux limites de la façon suivante : si λ est un tel ordinal, on pose

$$\Delta_\lambda := \bigcup_{\alpha<\lambda} \Delta_\alpha \tag{1.7}$$

On admet que ces trois règles permettent de définir un objet Δ_α pour tout α. Le résultat est simillaire au théorème de récurrence transfinie et se retrouvera dans [Dug66].

Théorème 1.8. *Notons ω_1 le plus petit ordinal indénombrable. Δ_{ω_1} est une tribu qui contient G.*

Démonstration. On va montrer que Δ_{ω_1} satisfait bien les axiomes d'une tribu. Notons en préambule que la famille (Δ_α) est croissante au sens de l'inclusion et donc par construction, $X \in \Delta_0 \subset \Delta_{\omega_1}$.

Stabilité par complémentaire : Soit $A \in \Delta_{\omega_1}$. Comme ω_1 est un ordinal limite (car tout ordinal inférieur strict à ω_1 est au plus dénombrable, et donc son successeur ne peut pas être ω_1) par construction de Δ_{ω_1}, il existe $\alpha < \omega_1$ un ordinal tel que $A \in \Delta_\alpha$. Or $A^c \in \Delta_{\alpha+1} \subset \Delta_{\omega_1}$.

σ-additivité : Soit $(A_n) \in \Delta_{\omega_1}^{\mathbb{N}}$. Pour tout $n \in \mathbb{N}$, il existe un ordinal $\alpha_n < \omega_1$ tel que $A_n \in \Delta_{\alpha_n}$. On considère α l'union des α_n. C'est un ordinal (quitte à réordonner, on peut le voir comme une union croissante d'ordinaux) dénombrable comme union dénombrable d'ensembles dénombrables. On a donc l'encadrement :

$$\forall n \in \mathbb{N},\ \alpha_n \leqslant \alpha < \omega_1 \tag{1.8}$$

Ainsi, tous les A_n appartiennent à Δ_α et donc par construction, leur union appartient à $\Delta_{\alpha+1} \subset \Delta_{\omega_1}$.

Δ_{ω_1} est donc bien une tribu ! □

A partir de ce théorème, on obtient notamment après des considérations techniques sur lesquelles je ne m'étendrai pas :

Théorème 1.9 (Tribu borélienne). *Avec les mêmes notations, en prenant G l'ensemble des ouverts de $\mathbb{R}$, Δ_{ω_1} est la tribu borélienne de $\mathbb{R}$. De plus, ω_1 est le plus petit ordinal tel que Δ_α est une tribu.*

Bien que n'ayant à ma connaissance pas d'utilité pratique immédiate, ce théorème révèle à quel point les boréliens ne s'expriment pas simplement en terme d'ouverts ! On peut également en déduire ce petit fait des plus divertissants :

Application 1.10. $\mathbb{R}$ est en bijection avec sa tribu borélienne.

Démonstration. Notons G l'ensemble des intervalles de la forme $]-\infty, x]$ pour tout $x \in \mathbb{R}$, qui est évidemment en bijection avec $\mathbb{R}$. La tribu borélienne est la tribu engendrée par G, qui est contenue dans Δ_{ω_1} (avec les notations précédentes). Montrons par récurrence transfinie que Δ_α est en bijection avec $\mathbb{R}$ pour tout ordinal α.

Initialisation : $\Delta_0 = G \cup \{\mathbb{R}\}$: il est en bijection avec $\mathbb{R}$.

Passage au successeur : soit α un ordinal tel que Δ_α est en bijection avec $\mathbb{R}$. Alors :

$$\Delta_{\alpha+1} := \left\{ \left(\bigcup_{n\in\mathbb{N}} A_n \right) \cup \left(\bigcup_{n\in\mathbb{N}} B_n^c \right),\ (A_n), (B_n) \in \Delta_\alpha^{\mathbb{N}} \right\} \tag{1.9}$$

Ainsi, il existe une surjection de $((\Delta_\alpha)^{\mathbb{N}})^2$ dans $\Delta_{\alpha+1}$. Puisque cet ensemble est en bijection avec $\mathbb{R}$ d'après l'hypothèse de récurrence, le théorème de Cantor-Bernstein permet d'affirmer que $\Delta_{\alpha+1}$ est encore en bijection avec $\mathbb{R}$.

Franchissement des ordinaux limites : soit λ un ordinal limite tel que Δ_α est en bijection avec $\mathbb{R}$ pour tout $\alpha < \lambda$. Si $\lambda > \omega_1$, $\Delta_\lambda = \Delta_{\omega_1}$ puisque ce dernier ensemble est une tribu, et donc la suite des Δ_α est stationnaire à partir du rang ω_1. Il est donc immédiat que $\Delta_\lambda = \Delta_{\omega_1}$ est en bijection avec $\mathbb{R}$. Sinon :

$$\Delta_\lambda = \bigcup_{\alpha<\lambda} \Delta_\alpha \tag{1.10}$$

Δ_λ est donc une union sur un ensemble de cardinal au plus celui de $\mathbb{R}$ d'ensembles équipotents à $\mathbb{R}$: il est encore équipotent à $\mathbb{R}$.

Ainsi, par récurrence transfinie, Δ_α est en bijection avec $\mathbb{R}$ pour tout α. C'est en particulier le cas pour $\alpha = \omega_1$. Comme $\mathcal{B}(\mathbb{R})$ est incluse dans Δ_{ω_1} [(x)], le théorème de Cantor-Bernstein permet de conclure. Etonnant non ? □

Remarque : Puisque un ensemble n'est jamais en bijection avec l'ensemble de ses parties, ceci fournit une démonstration alambiquée du fait qu'il existe des parties non mesurables de $\mathbb{R}$. ♦

(x). Ces ensembles sont même égaux, mais on n'a pas besoin de l'égalité pour conclure.

Notations

$\mathbb{N}$	l'ensemble des entiers naturels. **0 est un entier naturel**.
$\mathbb{Z}$	l'ensemble des entiers relatifs.
$\mathbb{Q}$	l'ensemble des nombres rationnels.
$\mathbb{R}$	l'ensemble de snombres réels.
$\mathbb{C}$	l'ensemble des nombres complexes.
$\mathbb{K}$	notation générique qui désigne indifféremment $\mathbb{R}$ ou $\mathbb{C}$ lorsque cela importe peu.
$\mathcal{P}(X)$	l'ensemble des parties d'un ensemble X.
A^c	le complémentaire dans X d'une partie A de X.
$A \sqcup B$	union de deux parties **disjointes** d'un ensemble X.
$\sigma(\Pi)$	la tribu engendrée par un ensemble de parties Π
$\lambda(\Pi)$	la classe monotone engendrée par un ensemble de parties Π
$f_*(\mu)$	mesure image d'une mesure μ par f. Voir définition 3.1
$f\mathrm{d}\mu$	mesure dont la densité contre μ est donnée par f. Voir proposition 2.1
$\mathfrak{F} \otimes \mathfrak{G}$	tribu produit de $\mathfrak{F}$ et $\mathfrak{G}$ (voir proposition 3.9)
$\mu \otimes \nu$	mesure produit de μ et ν (voir théorème 3.11)
$L^p(\mu)$	l'ensemble des fonctions d'un espace mesuré $(X, \mathfrak{F}, \mu)$ dans $\mathbb{K}$, mesurables et de module à la puissance p-ième intégrable, modulo l'égalité presque partout.
$f * g$	le produit de convolution de deux fonctions f et g, lorsqu'il a du sens
$\mathbb{E}$	l'espérance mathématique
$\mathbb{V}$	la variance
$X \perp\!\!\!\perp Y$	on indique ainsi que X et Y sont des variables aléatoires indépendantes.
$\mathbb{P}_X$	lorsque X est une variable aléatoire, on note ainsi la loi de X (voir 3.2)

Références

[App23] Walter Apple. *Probabilités pour les non-probabilistes*. Editions H&K, 2023. isbn : 978-2-35141-410-1.

[BP17] Marc Briane et Gilles Pagès. *Analyse, théorie de l'intégration*. De Boeck Supérieur, 2017. isbn : 978-2-8073-1788-8.

[Bré83] Haïm Brézis. *Analyse fonctionnelle. Théorie et applications*. Masson, 1983. isbn : 2-225-77198-7.

[Cav19] Thomas Cavalazzi. *Théorèmes de Prokhorov et de Lévy*. 2019. url : https://perso.eleves.ens-rennes.fr/people/Thomas.Cavallazzi/assets/pdf/developpements/Prokhorov%20et%20L%C3%A9vy.pdf.

[Che14] Huayi Chen. *Cours M1 Probabilités*. 2014. url : https://webusers.imj-prg.fr/~huayi.chen/Enseignement/Grenoble/2014%5C_2015/M1proba/chap2.pdf.

[Dug66] James Dugundji. *Topologie*. Wm. C. Brown Publishers, 1966. isbn : 0-697-06889-7.

[FF98] Dominique Foata et Aimé Fuchs. *Calcul des probabilités*. Dunod, 1998. isbn : 9782100041046.

[Wag12] Stan Wagon. *The Banach-Tarski paradox*. Cambridge University Press, 2012. isbn : 9780511609596. doi : https://doi.org/10.1017/CBO9780511609596.