Année universitaire 2010-2011

UNIVERSITÉ HENRI POINCARÉ

Olivier GARET

Intégration et Probabilités

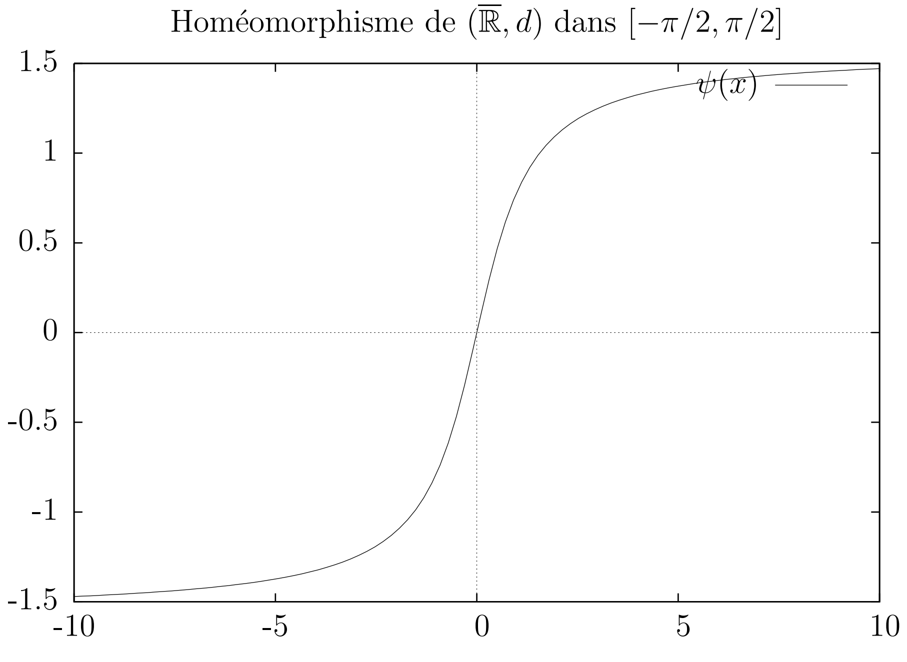
Table des matières

Chapitre 1

Compléments d'analyse

1.1 La droite réelle achevée

On ajoute deux points à $\mathbb{R}$ que l'on note $-\infty$ et $+\infty$. On définit ainsi la droite réelle achevée $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty; +\infty\}$. Notons $\psi(x) = \operatorname{atan} x$ pour x réel $\psi(+\infty) = \pi/2$, $\psi(-\infty) = -\pi/2$.



Pour x, y dans $\overline{\mathbb{R}}$, on note $d(x, y) = |\psi(x) - \psi(y)|$. Il n'est pas très difficile de vérifier que pour tous x, y, z dans $\overline{\mathbb{R}}$, on a

- $d(x, z) \leq d(x, y) + d(y, z)$
- $d(x, y) = 0 \iff x = y$

– $d(x,y) = d(y,x)$.

On dit alors que $(\overline{\mathbb{R}}, d)$ est un espace métrique.

Si $(x_n)_{n\geq 1}$ est une suite à valeurs dans $\overline{\mathbb{R}}$, on dit que (x_n) converge vers x si $d(x, x_n)$ tend vers 0.

On peut remarquer que ψ réalise un homéomorphisme croissant de $(\overline{\mathbb{R}}, d)$ dans $[-\pi/2, \pi/2]$ (l'homéomorphisme réciproque est bien sûr un prolongement de la fonction tangente).

Corollaire 1. *De toute suite (a_n) à valeurs dans $(\overline{\mathbb{R}}, d)$, on peut extraire une sous-suite convergente.*

Démonstration. $\psi(a_n)$ est à valeurs dans l'intervalle compact $[-\pi/2, \pi/2]$, donc, il existe une suite $\varphi(n)$ d'entiers strictement croissante et $y \in [-\pi/2, \pi/2]$ tel que $\psi(a_{\varphi(n)})$ tend vers y. Par continuité de ψ^{-1}, $(a_{\varphi(n)})$ tend vers $\psi^{-1}(y)$. □

Pour une suite $(x_n)_{n\geq 1}$ à valeurs dans $\mathbb{R}$, on peut vérifier que $(x_n)_{n\geq 1}$ converge vers a dans $\overline{\mathbb{R}}$ si et seulement si elle converge vers a dans $\mathbb{R}$, que $(x_n)_{n\geq 1}$ converge vers $+\infty$ dans $\overline{\mathbb{R}}$ si et seulement si elle tend vers $+\infty$ lorsque n tend vers l'infini, que $(x_n)_{n\geq 1}$ converge vers $-\infty$ dans $\overline{\mathbb{R}}$ si et seulement si elle tend vers $-\infty$ lorsque n tend vers l'infini.

On peut prolonger la relation d'ordre $\leq$ sur $\overline{\mathbb{R}}$, en disant que sont vraies les relations "$-\infty \leq a$" "$a \leq +\infty$" pour tout $a \in \mathbb{R}$ ainsi que "$-\infty \leq +\infty$".

On peut alors énoncé le théorème suivant

Théorème 1. *Toute suite monotone de $\overline{\mathbb{R}}$ converge.*

Démonstration. On va le prouver pour une suite croissante. Si la suite est constante égale à $-\infty$, elle converge. Sinon, à partir d'un certain rang, elle est à valeurs dans $]-\infty, +\infty]$, donc on peut se ramener au cas où elle est à valeurs $]-\infty, +\infty]$. Maintenant, si elle contient $+\infty$, elle est constante à partir d'un certain rang, donc elle converge. On s'est donc finalement ramené au cas où la suite est à valeurs réelles : si elle est croissante, majorée, elle converge dans $\mathbb{R}$, si elle est croissante non majorée, elle converge vers $+\infty$. □

1.2 Limite supérieure

La limite supérieure d'une suite (a_n) à valeurs dans $\overline{\mathbb{R}}$ est

$$\overline{\lim_{n\to+\infty}}\, a_n = \lim_{n\to+\infty} \sup_{k\geq n} a_k.$$

Cette limite existe bien car la suite (v_n) définie par $v_n = \sup_{k\geq n} a_k$ est décroissante. De même, la limite inférieure d'une suite (a_n) à valeurs dans $\overline{\mathbb{R}}$ est

$$\underline{\lim}_{n\to+\infty} a_n = \lim_{n\to+\infty} \inf_{k\geq n} a_k.$$

Cette limite existe bien car la suite (w_n) définie par $w_n = \inf_{k\geq n} a_k$ est croissante.

Lemme 1. *Soit* $(x_n)_{n\geq 1}$ *une suite à valeurs dans* $\overline{\mathbb{R}}$, f *une fonction croissante continue de* $(\overline{\mathbb{R}}, d)$ *dans* $(\overline{\mathbb{R}}, d)$. *Alors,* $\sup\{f(x_i); i \geq 1\} = f(\sup\{x_i; i \geq n\})$.

Démonstration. La suite $\max\{x_i; 1 \leq i \leq k\}$ converge vers $\sup\{x_i; i \geq 1\}$ lorsque k tend vers l'infini. Donc par continuité $f(\max\{x_i; 1 \leq i \leq k\})$ converge vers $f(\sup\{x_i; i \geq 1\})$. Or $f(\max\{x_i; 1 \leq i \leq k\}) = \max\{f(x_i); 1 \leq i \leq k\}$, qui elle-même converge vers $\sup\{f(x_i); i \geq 1\}$ lorsque k tend vers l'infini. Finalement $\sup\{f(x_i); i \geq 1\} = f(\sup\{x_i; i \geq n\})$. □

Théorème 2. $\overline{\lim}_{n\to+\infty} a_n$ *est la plus grande valeur d'adhérence de* a_n *dans* $\overline{\mathbb{R}}$.

Démonstration. Posons $\bar{l} = \overline{\lim}_{n\to+\infty} a_n$, et, comme précédemment $v_n = \sup_{k\geq n} a_k$. Commençons par montrer que toute valeur d'adhérence a de (a_n) vérifie $a \leq \bar{l}$. Soit $a = \lim_{n\to+\infty} a_{\varphi(n)}$ une valeur d'adhérence. Si $a = -\infty$ ou $\bar{l} = +\infty$, il n'y a rien a montrer. Sinon, prenons $\varepsilon > 0$. Il existe N tel que $n \geq N$ entraîne $v_n \leq \bar{l} + \varepsilon$, et donc $a_k \leq \bar{l} + \varepsilon$ pour $k \geq N$. Comme $\varphi(n)$ tend vers l'infini, il existe M tel que $n \geq M$ entraîne $\varphi(n) \geq N$. Finalement on a $a_{\varphi(n)} \leq \bar{l} + \varepsilon$ pour $n \geq M$, d'où $a \leq \bar{l} + \varepsilon$. Comme ε est quelconque, on a $a \leq \bar{l}$. Reste à montrer que $\bar{l}$ est valeur d'adhérence.

On pose $\varphi(1) = 1$, puis

$$\varphi(k+1) = \inf\left\{n > \varphi(k); \psi(v_{\varphi(k)}) \geq \psi(a_n) \geq \psi(v_{\varphi(k)}) - 1/k\right\}$$

$\varphi(k+1)$ est bien défini, car $\sup_{k\geq\varphi(n)} \psi(a_k) = \psi(\sup_{k\geq\varphi(n)} a_k)$, d'après le lemme ($\psi$ est un homéomorphisme, donc est continu). Pour $k \geq 1$, on a

$$\psi(v_{\varphi(k)}) \geq \psi(a_{\varphi(k+1)}) \geq \psi(v_{\varphi(k)}) - 1/k,$$

ce qui montre que $\psi(a_{\varphi(k)})$ tend vers $\psi(\bar{l})$, et donc, comme ψ^{-1} est continue, que $a_{\varphi(k)}$ tend vers $\bar{l}$. □

De même

Théorème 3. $\underline{\lim}_{n\to+\infty} a_n$ *est la plus petite valeur d'adhérence de* a_n *dans* $\overline{\mathbb{R}}$.

Théorème 4.

$$\overline{\lim_{n\to+\infty}}\ a_n = \sup\{x \in \overline{\mathbb{R}}; \{n \geq 1; a_n \geq x\} \text{ est infini.}\}$$

Démonstration. Supposons que x est tel que $\{n \geq 1; a_n \geq x\}$ est infini. On peut donc en extraire une suite $\varphi(n)$ strictement croissante d'entiers telle que $a_{\varphi(n)}$ converge vers $z \in \overline{\mathbb{R}}$ et $a_{\varphi(n)} \geq x$ pour tout n : comme $\overline{\lim}_{n\to+\infty} a_n$ est plus grand que toutes les valeurs d'adhérence, on a donc

$$\overline{\lim_{n\to+\infty}}\ a_n \geq z \geq x$$

En prenant le sup sur tous les x tels que $\{n \geq 1; a_n \geq x\}$ est infini, on obtient

$$\overline{\lim_{n\to+\infty}}\ a_n \geq \sup\{x \in \mathbb{R}; \{n \geq 1; a_n \geq x\} \text{ est infini.}\}$$

Maintenant raisonnons par l'absurde et supposons que

$$L = \overline{\lim_{n\to+\infty}}\ a_n > S = \sup\{x \in \mathbb{R}; \{n \geq 1; a_n \geq x\} \text{ est infini.}\}$$

Soit $\varepsilon > 0$ tel que $L > S + \varepsilon$. Comme L est la plus grande valeur d'adhérence de a_n, L est la limite d'une suite extraite $a_{\varphi(n)}$. Pour n assez grand, on a $a_{\varphi(n)} > S + \varepsilon$, ce qui entraîne que l'ensemble des n tels que a_n dépasse $S + \varepsilon$ est infini, ce qui contredit la définition de S. □

De même

Théorème 5.

$$\underline{\lim}_{n\to+\infty}\ a_n = \inf\{x \in \overline{\mathbb{R}}; \{n \geq 1; a_n \leq x\} \text{ est infini.}\}$$

Théorème 6. *Une suite* (a_n) *à valeurs dans* $\overline{\mathbb{R}}$ *converge si et seulement si* $\underline{\lim}_{n\to+\infty} a_n = \overline{\lim}_{n\to+\infty} a_n$, *qui est alors la limite.*

Démonstration. Si $\underline{\lim}_{n\to+\infty} a_n = \overline{\lim}_{n\to+\infty} a_n = +\infty$, alors pour tout $x \in \mathbb{R}$, $\{n \geq 1; a_n \leq x\}$ est fini, ce qui montre que a_n tend vers $+\infty$. De même, si $\underline{\lim}_{n\to+\infty} a_n = \overline{\lim}_{n\to+\infty} a_n = -\infty$, alors pour tout $x \in \mathbb{R}$, $\{n \geq 1; a_n \geq x\}$ est fini, ce qui montre que a_n tend vers $-\infty$. Passons au cas où $\underline{\lim}_{n\to+\infty} a_n = \overline{\lim}_{n\to+\infty} a_n = l \in \mathbb{R}$. Soit $\varepsilon > 0$. Comme

$$\overline{\lim}_{n\to+\infty} a_n = \sup\{x \in \overline{\mathbb{R}}; \{n \geq 1; a_n \geq x\} \text{ est infini.}\} = l \in \mathbb{R},$$

l'ensemble des n tels que $a_n \geq l + \varepsilon$ est fini. De même, comme

$$\underline{\lim}_{n\to+\infty} a_n = \inf\{x \in \overline{\mathbb{R}}; \{n \geq 1; a_n \leq x\} \text{ est infini.}\} = l \in \mathbb{R},$$

l'ensemble des n tels que $a_n \leq l - \varepsilon$ est fini. Finalement, l'ensemble des n tels que $|a_n - l| \geq \varepsilon$ est fini . Ainsi, pour tout $\varepsilon > 0$, à partir d'un certain rang, $|a_n - l| < \varepsilon$ ce qui montrer que a_n tend vers l.

Réciproquement, si a_n converge vers $l \in \overline{\mathbb{R}}$, $\underline{\lim}_{n\to+\infty} a_n$ et $\overline{\lim}_{n\to+\infty} a_n$ sont égales à l puisque ce sont des valeurs d'adhérence de (a_n). □

Théorème 7. *Soit* (u_n), u'_n *deux suites avec* $u_n \leq u'_n$ *pour tout* n. *On a* $\overline{\lim}_{n\to+\infty} u_n \leq \overline{\lim}_{n\to+\infty} u'_n$ *et* $\underline{\lim}_{n\to+\infty} u_n \leq \underline{\lim}_{n\to+\infty} u'_n$

Démonstration. Pour tout n, $\sup_{k\geq n} u_k \leq \sup_{k\geq n} u'_k$, d'où la première inégalité en faisant tendre n vers $+\infty$. Pour tout n, $\inf_{k\geq n} u_k \leq \inf_{k\geq n} u'_k$, d'où la deuxième inégalité en faisant tendre n vers $+\infty$. □

Corollaire 2. *Soit* $l \in \mathbb{R}$. *On suppose que pour tout* $\varepsilon > 0$,

$$\underline{\lim}_{n\to+\infty} u_n \geq l - \varepsilon$$

et

$$\overline{\lim}_{n\to+\infty} u_n \leq l + \varepsilon.$$

Alors, u_n *converge vers* l.

Démonstration. En faisant tendre $\varepsilon > 0$ dans les deux inégalités, on obtient

$$\underline{\lim}_{n\to+\infty} \ u_n \geq l$$

$$\overline{\lim}_{n\to+\infty} \ u_n \leq l.$$

Finalement

$$l \leq \underline{\lim}_{n\to+\infty} \ u_n \leq \overline{\lim}_{n\to+\infty} \ u_n \leq l,$$

comme les termes extrêmes sont égaux, ceci entraîne que tous les termes sont égaux. □

1.3 Exercices sur les limites supérieures et inférieures

1. $\underline{\lim}_{n\to+\infty} \ (-1)^n\big(1+\frac{1}{n}\big) \qquad \overline{\lim}_{n\to+\infty} \ (-1)^n\big(1+\frac{1}{n}\big)$

   $$\underline{\lim}_{n\to+\infty} \ \cos n \qquad \overline{\lim}_{n\to+\infty} \ \cos n \qquad \underline{\lim}_{n\to+\infty} \ \frac{1}{1+\cos n} \qquad \overline{\lim}_{n\to+\infty} \ \frac{1}{1+\cos n}$$

2. Soit $(a_i,\, i \in I)$ une famille non vide d'éléments de $\overline{\mathbb{R}}$.
   (a) Démontrer que $\inf(a_i,\, i\in I) = -\sup(-a_i,\, i \in I)$.
   (b) Soit α un élément de $\mathbb{R}$. Démontrer que

   $$\inf(\alpha + a_i,\, i\in I) = \alpha + \inf(a_i,\, i\in I)$$

   et en déduire que

   $$\sup(\alpha + a_i,\, i\in I) = \alpha + \sup(a_i,\, i\in I).$$

3. Démontrer que $\underline{\lim}_{n\to+\infty} \ a_n = -\overline{\lim}_{n\to+\infty} \ (-a_n)$, pour toute suite $(a_n)_{n\geq 1}$ d'éléments de $\overline{\mathbb{R}}$.
4. Soit $(a_n)_{n\geq 1}$ et $(b_n)_{n\geq 1}$ deux suites dans $\mathbb{R}$.
   (a) Démontrer que

   $$\underline{\lim}_{n\to+\infty} \ (a_n+b_n) \geq \underline{\lim}_{n\to+\infty} \ a_n + \underline{\lim}_{n\to+\infty} \ b_n.$$

et

$$\overline{\lim_{n\to+\infty}} (a_n+b_n) \le \overline{\lim_{n\to+\infty}} a_n + \overline{\lim_{n\to+\infty}} b_n.$$

Montrer que les inégalités peuvent être strictes.

(b) On suppose que $(a_n)_{n\geq 1}$ converge dans $\mathbb{R}$. Démontrer que

$$\underline{\lim_{n\to+\infty}} (a_n+b_n) = \lim_{n\to+\infty} a_n + \underline{\lim_{n\to+\infty}} b_n; \quad \underline{\lim_{n\to+\infty}} (a_n-b_n) = \lim_{n\to+\infty} a_n - \overline{\lim_{n\to+\infty}} b_n$$

$$\overline{\lim_{n\to+\infty}} (a_n+b_n) = \lim_{n\to+\infty} a_n + \overline{\lim_{n\to+\infty}} b_n; \quad \overline{\lim_{n\to+\infty}} (a_n-b_n) = \lim_{n\to+\infty} a_n - \underline{\lim_{n\to+\infty}} b_n$$

5. *Suites sous-additives (lemme de Fekete)*
Soit $(u_n)_{n\geq 1}$ une suite vérifiant

$$\forall n,p \geq 1 \quad u_{n+p} \leq u_n + u_p.$$

Le but de l'exercice est de montrer que $\frac{u_n}{n}$ converge vers $\inf\limits_{n\geq 1} \frac{u_n}{n}$.

(a) Soit k un entier naturel non nul fixé, r un entier entre 0 et $k-1$.
Montrer que

$$\frac{u_{kn+r}}{nk+r} \leq \frac{nu_k}{nk+r} + \frac{u_r}{nk+r}.$$

En déduire $\overline{\lim\limits_{n\to+\infty}} \frac{u_{kn+r}}{nk+r} \leq \frac{u_k}{k}$.

(b) Montrer que $\overline{\lim\limits_{n\to+\infty}} \frac{u_n}{n} \leq \inf_{k\geq 1} \frac{u_k}{k}$.

(c) Conclure.

(d) Application 1. Pour $A \in M_n(\mathbb{R})$, on pose $|||A||| = \sup\limits_{y\in\mathbb{R}^n\backslash\{0\}} \frac{\|Ay\|_\infty}{\|y\|_\infty}$.
Montrer que la suite $|||A^n|||^{1/n}$ converge vers un réel positif.

(e) Application 2. Soit E une partie finie de $\mathbb{R}^d$. On note A_n l'ensemble des suites $(u_1,\dots u_n)$ qui vérifient
– $u_1 \in E$
– $u_{i+1} - u_i \in E$ pour tout $i \in \{1,\dots,n-1\}$
– $i \mapsto u_i$ est injective
Montrer que la suite $|A_n|^{1/n}$ converge vers un réel positif.

6. Soit $(a_n)_{n\geq 1}$ et $(b_n)_{n\geq 1}$ deux suites de réels telles que, pour tout $n \geq 1$, $a_n > 0$, $b_n > 0$ et $\lim\limits_{n\to+\infty} (a_n)^n = a > 0$, $\lim\limits_{n\to+\infty} (b_n)^n = b > 0$. Soient $p,q > 0$ avec $p+q=1$. Déterminer $\lim\limits_{n\to+\infty} (pa_n+qb_n)^n$.

7. Soit $\alpha > 1$. Montrer qu'il existe une unique suite $(u_k)_{k\geq 1}$ vérifiant pour tout N
$$1 = \sum_{i=1}^{N} \left(\sum_{k=i}^{N} u_k \right)^{-\alpha}.$$
Montrer que $\varliminf_{n\to+\infty} u_n \leq \zeta(\alpha)^{1/\alpha} \leq \varlimsup_{n\to+\infty} u_n$. Montrer que $(u_n)_{n\geq 1}$ converge vers une limite que l'on déterminera (cette dernière question est plutôt un défi, l'auteur de ces lignes ne connaît pas la réponse).

Chapitre 2

Un peu de théorie de la mesure

La théorie des probabilités décrit les événements comme des sous-ensembles d'un ensemble Ω représentant tous les résultats possibles *a priori* – même s'il peut s'avérer ensuite que certains n'arrivent jamais. Remarquons bien qu'il n'est pas possible de modéliser un phénomène aléatoire quelconque si l'on ne connaît pas les résultats possibles *a priori*.

Soit donc Ω un ensemble. Pour tout $A \subset \Omega$, on note A^c le complémentaire de A dans Ω :

$$A^c = \{x \in \Omega; x \notin A\}.$$

2.1 Tribus

2.1.1 Axiomes de base

On dit qu'une partie $\mathcal{F} \subset \mathcal{P}(\Omega)$ est une tribu si elle vérifie les propriétés suivantes :

1. $\varnothing \in \mathcal{F}$.
2. $\forall A \in \mathcal{F} \quad A^c \in \mathcal{F}$.
3. Pour toute suite $(A_i)_{i \in \mathbb{N}}$ d'éléments de $\mathcal{F}$, $\bigcup\limits_{i=1}^{+\infty} A_i \in \mathcal{F}$.

2.1.2 Propriétés

Les propositions suivantes sont alors des conséquences relativement faciles des axiomes de base :

- $\Omega \in \mathcal{A}$.
- Pour toute suite $(A_i)_{i \in \mathbb{N}}$ d'éléments de $\mathcal{A}$, $\bigcap\limits_{i=1}^{+\infty} A_i \in \mathcal{A}$.

– Pour toute suite $(A_i)_{1\leq i\leq n}$ d'éléments de $\mathcal{A}$, $\overset{n}{\underset{i=1}{\cup}} A_i \in \mathcal{A}$.
– Pour toute suite $(A_i)_{1\leq i\leq n}$ d'éléments de $\mathcal{A}$, $\overset{n}{\underset{i=1}{\cap}} A_i \in \mathcal{A}$.

Une fois que Ω et $\mathcal{A}$ sont fixés, on appelle événement tout élément de $\mathcal{A}$.

Exercice :

Montrer qu'une partie $\mathcal{A} \subset \mathcal{P}(\Omega)$ est une tribu si elle vérifie les propriétés suivantes :

1. $\Omega \in \mathcal{A}$.
2. $\forall (A,B) \in \mathcal{A}\times\mathcal{A} \quad (A \subset B) \Longrightarrow (B\backslash A \in \mathcal{A})$.
3. $\forall (A,B) \in \mathcal{A}^2 \quad A\cup B \in \mathcal{A}$.
4. Pour toute suite $(A_i)_{i\in\mathbb{N}}$ d'éléments de $\mathcal{A}$ deux à deux disjoints, $\overset{+\infty}{\underset{i=1}{\cup}} A_i \in \mathcal{A}$.

Étant donné un espace mesuré $(\Omega, \mathcal{F})$, on dira qu'une propriété $\mathcal{P}(x)$ est vraie μ-presque partout ou encore pour μ-presque tout x si il existe $A \in \mathcal{F}$ tel que

$$\forall x \in \Omega\backslash A; \mathcal{P}(x)$$

et $\mu(A) = 0$.

Pour A et B dans $\mathcal{F}$, on dira parfois que A et B sont égaux μ presque partout pour signifier que $\mu(A\Delta B) = 0$.

2.1.3 Sous-tribus

Si $\mathcal{A}$ est une tribu et que la partie $\mathcal{B} \subset \mathcal{A}$ est une tribu, alors on dit que $\mathcal{B}$ est une sous-tribu de $\mathcal{A}$.[1]

2.1.4 Opérations sur les tribus

Intersection de tribus

Soit Ω un ensemble et T un ensemble de tribus sur Ω. T est supposé non vide.[2] Il peut être fini ou infini, voire même infini non dénombrable. Alors $\mathcal{A} = \underset{A\in T}{\cap} \mathcal{A}$ est une tribu.

Démonstration. Il suffit de vérifier les 3 axiomes de base des tribus.

[1] Une erreur classique à ne pas commettre : si $\mathcal{B}$ est une sous-tribu de $\mathcal{A}$, que $B \subset A$ avec $A \in \mathcal{A}$, alors rien ne permet d'affirmer que $B \in \mathcal{B}$ ni que $B \in \mathcal{A}$.

[2] T est donc un ensemble d'ensembles d'ensembles, lol.

- $\forall \mathcal{A} \in T \quad \varnothing \in \mathcal{A}$. Donc $\varnothing \in \bigcap\limits_{\mathcal{A} \in T} \mathcal{A} = \mathcal{A}$.
- Soit $A \in \mathcal{A}$. On doit montrer que $A^c \in \mathcal{A}$. Soit $\mathcal{A} \in T$. Comme les $A \in \mathcal{A}$ et que $\mathcal{A}$ est une tribu, $A^c \in \mathcal{A}$. Comme ceci est vrai pour tout $\mathcal{A} \in T$, on a $A^c \in \bigcap\limits_{\mathcal{A} \in T} \mathcal{A} = \mathcal{A}$.
- Soit $(A_i)_{i \in I}$ une famille dénombrable d'éléments de $\mathcal{A}$. On doit montrer que $\bigcup\limits_{i \in I} A_i \in \mathcal{A}$. Soit $\mathcal{A} \in T$. Comme les A_i sont dans $\mathcal{A}$ et que $\mathcal{A}$ est une tribu, $\bigcup\limits_{i \in I} A_i \in \mathcal{A}$. Comme ceci est vrai pour tout $\mathcal{A} \in T$, on a $\bigcup\limits_{i \in I} A_i \in \bigcap\limits_{\mathcal{A} \in T} \mathcal{A} = \mathcal{A}$.

□

Tribu engendrée par une famille de tribus

Soit $(\mathcal{A}_i)_{i \in I}$ une famille de tribus sur Ω. L'ensemble des tribus contenant des tribus contenant toutes les $\mathcal{A}_i$ est non vide, puisque $\mathcal{P}(\Omega)$ est une telle tribu. D'après le résultat énoncé ci-dessus, l'intersection de toutes ces tribus est une tribu. Par construction, cette tribu est la plus petite tribu contenant toutes les $\mathcal{A}_i$. On la note

$$\bigwedge_{i \in I} \mathcal{A}_i.$$

Tribu engendrée par une famille d'ensembles

Soit $(A_i)_{i \in I}$ une famille de parties de Ω.

Pour tout i, la plus petite tribu contenant A_i est la tribu $\mathcal{A}_i = (\varnothing, A_i, A_i^c, \Omega)$. Ainsi, la plus petite tribu contenant les ensembles A_i est

$$\bigwedge_{i \in I} \mathcal{A}_i.$$

On note cette tribu $\sigma(A_i; i \in I)$.

2.1.5 Tribu borélienne, fonctions mesurables

Soit $(A, \mathcal{A})$ et $(B, \mathcal{B})$ deux espaces mesurés. On dit qu'une application f de A dans B est mesurable de $(A, \mathcal{A})$ dans $(B, \mathcal{B})$ si quelque soit $X \in \mathcal{B}$, son image réciproque $f^{-1}(X)$ est dans $\mathcal{A}$.

Commençons par une remarque simple : si f de A dans B est $(A, \mathcal{A}) - (B, \mathcal{B})$ mesurable et que g de B dans C est $(B, \mathcal{B}) - (C, \mathcal{C})$ mesurable, alors $g \circ f$ est $(A, \mathcal{A}) - (C, \mathcal{C})$ mesurable.

Théorème 8 (Théorème fondamental de la mesurabilité)**.** *Soit f une application quelconque d'un ensemble Ω dans un ensemble Ω'. Alors*
- *Pour toute tribu $\mathcal{T}$ sur Ω', $f^{-1}(\mathcal{T})$ est une tribu sur Ω.*
- *Pour tout $\mathcal{A} \in \mathcal{P}(\mathcal{P}(\Omega'))$, $\sigma(f^{-1}(\mathcal{A})) = f^{-1}(\sigma(\mathcal{A}))$*

Démonstration.
- Vérifions que $f^{-1}(\mathcal{T})$ vérifie les axiomes des tribus
  - $\varnothing \in f^{-1}(\mathcal{T})$ car $\varnothing = f^{-1}(\varnothing)$ et $\varnothing \in \mathcal{T}$
  - Soit $A \in f^{-1}(\mathcal{T})$: il existe $B \in \mathcal{T}$ avec $A = f^{-1}(B)$. $A^c = (f^{-1}(B))^c = f^{-1}(B^c)$; or $B^c \in \mathcal{T}$ donc $A^c \in f^{-1}(\mathcal{T})$
  - Soient $(A_i)_{i\geq 1}$ des éléments de $f^{-1}(\mathcal{T})$: pour tout i, il existe $B_i \in \mathcal{T}$ avec $A_i = f^{-1}(B_i)$. $\cup_i A_i = \cup_i(f^{-1}(B_i)) = f^{-1}(\cup_i B_i)$; or $\cup_i B_i \in \mathcal{T}$ donc $\cup_i A_i \in f^{-1}(\mathcal{T})$
- $\mathcal{A} \subset \sigma(\mathcal{A})$, donc $f^{-1}(\mathcal{A}) \subset f^{-1}(\sigma(\mathcal{A}))$, puis

  $$\sigma(f^{-1}(\mathcal{A})) \subset \sigma(f^{-1}(\sigma(\mathcal{A}))) = f^{-1}(\sigma(\mathcal{A})),$$

  où l'égalité provient de la première partie du théorème. Il reste à montrer que $f^{-1}(\sigma(\mathcal{A})) \subset \sigma(f^{-1}(\mathcal{A}))$. Notons

  $$\mathcal{C} = \{X \in \sigma(\mathcal{A}); f^{-1}(X) \in \sigma(f^{-1}(\mathcal{A}))\}.$$

  Il n'est pas difficile de démontrer que $\mathcal{C}$ est une tribu (laissé en exercice). Mais $\mathcal{C}$ contient $\mathcal{A}$, donc $\mathcal{C}$ est égal à $\sigma(\mathcal{A})$ tout entier, ce qui montre l'inclusion voulue.

□

Si il n'y a pas d'ambiguité sur la tribu $\mathcal{B}$ de l'espace d'arrivée, on note $\sigma(f)$ la tribu $f^{-1}(\mathcal{B})$; c'est la plus petite tribu $\mathcal{A}$ sur A telle que f soit une application mesurable de $(A, \mathcal{A})$ dans $(B, \mathcal{B})$. On dit que c'est la tribu engendrée par l'application f.

Corollaire 3. *Soit $(A, \mathcal{A})$ et $(B, \mathcal{B})$ deux espaces mesurés. On suppose que $\mathcal{B} = \sigma(\mathcal{C})$. Une application f de A dans B est mesurable de $(A, \mathcal{A})$ dans $(B, \mathcal{B})$ si quel que soit $X \in \mathcal{C}$, son image réciproque $f^{-1}(X)$ est dans $\mathcal{A}$.*

Si A est un ensemble muni d'une topologie, on appelle tribu borélienne de A et l'on note $\mathcal{B}(A)$ la tribu engendrée par les ouverts de A.

Lorsque l'ensemble d'arrivée d'une fonction est $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on parle fréquemment d'application mesurable sans préciser l'espace d'arrivée.

Théorème 9. *La tribu borélienne de $\mathbb{R}^d$ est également la tribu engendrée par les pavés ouverts de $\mathbb{R}^d$ dont les côtés ont des extrémités rationnelles ; les ensembles de la forme $\prod_{i=1}^{d}]a_i, b_i[$, avec $a_i < b_i$ et a_i, b_i dans $\mathbb{Q}$.*

Démonstration. Soit $\mathcal{T}$ la tribu engendrée par ces pavés : $\mathcal{T} \subset \mathcal{B}(\mathbb{R}^d)$ car ces pavés sont eux-mêmes des ouverts de $\mathbb{R}^d$. Pour obtenir l'inclusion réciproque, il suffit de montrer que chaque ouvert O de $\mathbb{R}^d$ est dans $\mathcal{T}$. Soit donc O un ouvert de $\mathbb{R}^d$. Soit $x \in \mathbb{R}^d$: il existe $\varepsilon > 0$ tel que $x+]-\varepsilon,+\varepsilon[^d \subset O$. Comme $\mathbb{Q}$ est dense dans $\mathbb{R}$, on peut trouver des rationnels $a_i(x)$ et $b_i(x)$ avec $x_i - \varepsilon < a_i(x) < x_i < b_i(x) < x_i + \varepsilon$. Posons alors $U(x) = \prod_{i=1}^d]a_i(x), b_i(x)[$. On a

$$O = \bigcup_{x \in O} U(x)$$

On peut définir une relation d'équivalence sur O par $u \sim v$ si et seulement si $U(u) = U(v)$. Evidemment, l'application U passe au quotient, et l'on peut écrire

$$O = \bigcup_{x \in O \backslash \sim} U(x)$$

Mais $O \backslash \sim$ est au plus dénombrable car U est à valeur dans $\mathbb{Q}^{2d}$ qui est dénombrable. Ainsi, O est réunion dénombrable d'éléments de $\mathcal{T}$, donc O est dans $\mathcal{T}$. □

On peut en déduire aisément que la tribu borélienne de $\mathbb{R}$ est engendrée par les ensembles de la forme $]-\infty, a[$, où a décrit $\mathbb{R}$. Ce résultat pourra éventuellement être traité en exercice.

Corollaire 4. *Soit $(A, \mathcal{A})$ un espace mesuré, f une application de A dans $\mathbb{R}$. Si pour tout $a \in \mathbb{R}$, l'ensemble $f^{-1}(]-\infty, a[)$ est dans $\mathcal{A}$, alors f est mesurable de $(A, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*

Corollaire 5. *Soit A et B deux espaces topologiques. Toute application continue de $(A, \mathcal{B}(A))$ dans $(B, \mathcal{B}(B))$ est mesurable de $(A, \mathcal{B}(A))$ dans $(B, \mathcal{B}(B))$.*

On note couramment $\mathcal{V}(A, \mathcal{A})$ l'ensemble des applications mesurables de $(A, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. De même, on note $\overline{\mathcal{V}}(A, \mathcal{A})$ l'ensemble des applications mesurables de $(A, \mathcal{A})$ dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ et $\overline{\mathcal{V}}_+(A, \mathcal{A})$ l'ensemble des applications mesurables de $(A, \mathcal{A})$ dans $(\overline{\mathbb{R}}_+, \mathcal{B}(\overline{\mathbb{R}}_+))$

Tribu produit

Soit $(\Omega, \mathcal{A})$ et $(\Omega', \mathcal{A}')$ deux espaces mesurés. On appelle tribu produit sur $\Omega \times \Omega'$ la tribu engendrée par les ensembles $A \times B \in \mathcal{A} \times \mathcal{B}$. On note $\mathcal{A} \otimes \mathcal{B}$ cette tribu.

Commençons par une remarque simple : si π_1 est l'application de $\Omega \times \Omega'$ dans Ω qui à $(x, y) \in \Omega \times \Omega'$ associe $\pi_1(x, y) = x$, alors π_1 (la projection sur la première coordonnée) est une application $(\Omega \times \Omega', \mathcal{A} \otimes \mathcal{B}) - (\Omega, \mathcal{A})$

mesurable. En effet, si $A \in \mathcal{A}$, $\pi_1^{-1}(A) = A \times \Omega' \in \mathcal{A} \times \mathcal{B} \subset \mathcal{A} \otimes \mathcal{B}$. De même, si π_2 est l'application de $\Omega \times \Omega'$ dans Ω' qui à $(x, y) \in \Omega \times \Omega'$ associe $\pi_2(x, y) = y$, alors π_2 (la projection sur la deuxième coordonnée) est une application $(\Omega \times \Omega', \mathcal{A} \otimes \mathcal{B}) - (\Omega', \mathcal{B})$ mesurable.

Théorème 10. *On suppose que* $\mathcal{A} = \sigma((A_i)_{i\in I})$ *et* $\mathcal{B} = \sigma((B_i)_{j\in J})$. *Alors* $\mathcal{A} \otimes \mathcal{B} = \sigma((A_i \times B_j)_{(i,j)\in I\times J})$.

Démonstration. Notons $\mathcal{O}$ la tribu engendrée par les $(A_i \times B_j)_{(i,j)\in I\times J}$. Pour $A \subset \Omega'$, on note $\mathcal{C}_A = \{B \in \mathcal{B} : A \times B \in \mathcal{O}\}$. On montre facilement que pour tout A, $\mathcal{C}_A$ est une sous-tribu de $\mathcal{B}$ sur Ω'. (le faire !) Supposons maintenant qu'il existe i avec $A = A_i$: on voit alors que $(B_j)_{i\in J} \subset \mathcal{C}_A$. Alors, $\sigma((B_j)_{j\in J}) \subset \sigma(\mathcal{C}_A)$, soit $\mathcal{B} \subset \mathcal{C}_A$, d'où $\mathcal{B} = \mathcal{C}_A$. Maintenant, notons $\mathcal{D} = \{A \in \mathcal{A} : \mathcal{C}_A = \mathcal{B}\}$. On peut montrer que $\mathcal{D}$ est une sous tribu de $\mathcal{A}$. (le faire !) Mais $\mathcal{D}$ contient les A_i qui engendrent $\mathcal{A}$, donc $\mathcal{D} = \mathcal{A}$, ce qui signifie que pour tout $(A, B) \in \mathcal{A} \times \mathcal{B}$, on a $A \times B \in \mathcal{O}$. En considérant les tribus enendrées, on a $\mathcal{A} \otimes \mathcal{B} \subset \mathcal{O}$. L'inclusion réciproque est évidente. □

Théorème 11. *Soient* f *une application de* C *dans* Ω, g *une application de* C *dans* Ω'.*On définit une application* F *de* C *dans* $\Omega \times \Omega'$ *par* $F(x) = (f(x), g(x))$. *L'application* F *est* $(C, \mathcal{C}) - (\Omega \times \Omega', \mathcal{A} \otimes \mathcal{B})$ *mesurable si et seulement si* f *est* $(C, \mathcal{C}) - (\Omega, \mathcal{A})$ *mesurable et* g $(C, \mathcal{C}) - (\Omega', \mathcal{B})$ *mesurable.*

Démonstration. La condition est nécessaire car $f = \pi_1 \circ F$ et $g = \pi_2 \circ F$: ainsi lorsque F est $(C, \mathcal{C}) - (\Omega \times \Omega', \mathcal{A} \otimes \mathcal{B})$ mesurable, comme π_2 est $(\Omega \times \Omega', \mathcal{A} \otimes \mathcal{B}) - (\Omega, \mathcal{A})$ mesurable, f est mesurable comme composée d'applications mesurables. Pour les mêmes raisons, g est mesurable. Supposons maintenant que f est $(C, \mathcal{C}) - (\Omega, \mathcal{A})$ mesurable et g $(C, \mathcal{C}) - (\Omega', \mathcal{B})$ mesurable et intéressons-nous à F. Soit $A \times B \in \mathcal{A} \times \mathcal{B}$: $F^{-1}(A \times B) = f^{-1}(A) \cap g^{-1}(B)$. Comme f est g sont mesurables, $f^{-1}(A)$ et $g^{-1}(B)$ sont dans $\mathcal{C}$, donc leur intersection aussi. Ainsi pour tout $A \times B \in \mathcal{A} \times \mathcal{B}$, $F^{-1}(A \times B) \in \mathcal{C}$. Mais les ensembles $A \times B \in \mathcal{A} \times \mathcal{B}$ engendrent $\mathcal{A} \otimes \mathcal{B}$, donc F est bien $(C, \mathcal{C}) - (\Omega \times \Omega', \mathcal{A} \otimes \mathcal{B})$ mesurable. □

Théorème 12. *Soit* $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2), (\Omega_3, \mathcal{F}_3)$ *trois espaces mesurés. L'application* $\Psi : ((\Omega_1 \times \Omega_2) \times \Omega_3) \to \Omega_1 \times (\Omega_2 \times \Omega_3)$ *qui à* $((x, y), z)$ *associe* $(x, (y, z)$ *est bi-mesurable de* $((\Omega_1 \times \Omega_2) \times \Omega_3), (\mathcal{F}_1 \otimes \mathcal{F}_2) \otimes \mathcal{F}_3)$ *vers* $(\Omega_1 \times (\Omega_2 \times \Omega_3), \mathcal{F}_1 \otimes (\mathcal{F}_2 \otimes \mathcal{F}_3))$. *Ainsi les deux tribus* $(\mathcal{F}_1 \otimes \mathcal{F}_2) \otimes \mathcal{F}_3$ *et* $\mathcal{F}_1 \otimes (\mathcal{F}_2 \otimes \mathcal{F}_3)$ *peuvent s'identifier et on notera simplement* $\mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \mathcal{F}_3$ *cette tribu sur* $\Omega_1 \times \Omega_2 \times \Omega_3$.

Démonstration. En utilisant le théorème 10, on voit que les ensembles $(A_1 \times A_2) \times A_3$ et $A_1 \times (A_2 \times A_3)$ engendrent respectivement les deux tribus considérées. Le corollaire 3 permet alors de conclure. □

L'extension au produit d'un nombre quelconque d'espaces mesurés se fait alors aisément par récurrence.

Théorème 13. *Pour tout entier $d \geq 2$, on a*

$$\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R})^{\otimes d}$$

Démonstration. Il suffit de montrer que pour tout $d \geq 1$, $\mathcal{B}(\mathbb{R}^{d+1}) = \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R})$, puis de conclure par récurrence. Or, d'après le théorème 9 la tribu $\mathcal{B}(\mathbb{R}^d)$ est la tribu engendrée par les ensembles A de la forme $\prod_{i=1}^{d}]a_i, b_i[$, avec $a_i < b_i$ et a_i, b_i dans $\mathbb{Q}$, tandis que $\mathcal{B}(\mathbb{R})$ est la tribu engendrée par les ensembles B de la forme $]a_{d+1}, b_{d+1}[$, avec $a_{d+1} < b_{d+1}$ et a_{d+1}, b_{d+1} dans $\mathbb{Q}$. D'après le théorème 10, les produits $A \times B$ engendrent la tribu $\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R})$; mais ces ensembles sont exactement les ensembles de la forme $\prod_{i=1}^{d+1}]a_i, b_i[$, avec $a_i < b_i$ et a_i, b_i dans $\mathbb{Q}$, qui, toujours d'après le théorème 9 engendrent la tribu $\mathcal{B}(\mathbb{R}^{d+1})$. □

Théorème 14. *Soit f, g deux applications mesurables de $(C, \mathcal{C})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et G une application mesurable de $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors H définie par $H(x) = G(f(x), g(x))$ est mesurable de $(C, \mathcal{C})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*
En particulier les choix $G(x, y) = x + y$ et $G(x, y) = xy$ nous disent que $f + g$ et fg sont mesurables de $(C, \mathcal{C})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Démonstration. Avec les notations du Théorème 11, $H = G \circ F$. Pour le cas particulier, notons que comme H est une application continue de $\mathbb{R}^2$ dans $\mathbb{R}$, c'est une application $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)) - (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable, ou de manière équivalente $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})) - (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable les applications continues sont mesurables par rapport aux tribus boréliennes associées aux topologies correspondantes. □

2.2 Mesures

2.2.1 Algèbres

On dit qu'une partie $\mathcal{A} \subset \mathcal{P}(\Omega)$ est une algèbre si elle vérifie les propriétés suivantes :

1. $\varnothing \in \mathcal{A}$.
2. $\forall A \in \mathcal{A} \quad A^c \in \mathcal{A}$.
3. Pour tous A et B dans $\mathcal{A}$, $A \cup B \in \mathcal{A}$

Remarque : il n'est pas difficile de démontrer qu'une algèbre est stable par union finie ou intersection finie.

On voit tout de suite que la différence avec la définition d'une tribu est que la stabilité par réunion dénombrable n'est pas requise. En fait, les tribus sont parfois appelés σ-algèbres, la lettre σ étant traditionnellement attachée aux propriétés liées à des familles dénombrables.

Remarque : en anglais

– algèbre se dit "field", plus rarement algebra

– tribu (σ-algèbre) se dit "σ-field".

2.2.2 Espace mesuré

Soit $\mathcal{A}$ une algèbre. On appelle mesure sur $(\Omega, \mathcal{A})$ toute application

$$\mu : \ \mathcal{A} \ \rightarrow [0, +\infty]$$

vérifiant les propriétés suivantes :

1. $\mu(\varnothing) = 0$.
2. Pour toute suite $(A_i)_{i\in\mathbb{N}}$ d'éléments de $\mathcal{A}$ deux à deux disjoints et telle que $\overset{+\infty}{\underset{i=1}{\cup}} A_i \in \mathcal{A}$, alors ,
$\mu(\overset{+\infty}{\underset{i=1}{\cup}} A_i) = \sum\limits_{i=1}^{+\infty} \mu(A_i)$.

Dans le cas où $\mathcal{A}$ est une tribu, le triplet $(\Omega, \mathcal{A}, \mu)$ est appelé espace mesuré.

Si $\mu(\Omega) < +\infty$, on dit μ est une mesure finie. Si il existe une suite A_n d'éléments de $\mathcal{A}$ avec $\mu(A_n) < +\infty$ pour tout n et que $\Omega = \overset{+\infty}{\underset{i=1}{\cup}} A_i$, on dit que μ est σ-finie.

Les propositions suivantes sont alors des conséquences relativement faciles des définitions :

1. Pour toute suite $(A_i)_{1\leq i\leq n}$ d'éléments de $\mathcal{A}$ deux à deux disjoints,
$\mu(\overset{n}{\underset{i=1}{\cup}} A_i) = \sum\limits_{i=1}^{n} \mu(A_i)$.
2. $\forall A, B \in \mathcal{A} \quad (A \cap B = \varnothing) \Longrightarrow (\mu(A \cup B) = \mu(A) + \mu(B))$
3. $\forall (A, E) \in \mathcal{A}^2$ avec $A \subset E$ et $\mu(A) < +\infty$ on a $\quad \mu(E\backslash A) = \mu(E) - \mu(A)$
4. $\forall A, B \in \mathcal{A} \quad \mu(A \cap B) < +\infty \Longrightarrow \mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$
5. $\forall A, B \in \mathcal{A} \quad \mu(A \cup B) \leq \mu(A) + \mu(B)$
6. $\forall A, B \in \mathcal{A} \quad (A \subset B) \Longrightarrow (\mu(A) \leq \mu(B))$

7. $\forall A, B \in \mathcal{A} \quad \mu(A \cap B) \leq \min(\mu(A), \mu(B))$
8. $\forall A, B \in \mathcal{A} \quad \mu(A \cup B) \geq \max(\mu(A), \mu(B))$
9. Pour toute suite $(A_i)_{i \geq 1}$ d'éléments de $\mathcal{A}$ telle que $\overset{+\infty}{\underset{i=1}{\cup}} A_i \in \mathcal{A}$,

   $\mu(\overset{+\infty}{\underset{i=1}{\cup}} A_i) \leq \sum_{i=1}^{+\infty} \mu(A_i)$.
10. Si $(A_i)_{i \in \mathbb{N}}$ est une suite croissante d'événements de $\mathcal{A}$ (c'est à dire que $\forall n \in \mathbb{N} \quad A_n \subset A_{n+1})$) telle que $A = \overset{+\infty}{\underset{i=1}{\cup}} A_i \in \mathcal{A}$, alors la suite $(\mu(A_n))_{n \in \mathbb{N}}$ est monotone, croissante, et converge vers $\mu(A)$.
11. Si $(A_i)_{i \in \mathbb{N}}$ est une suite décroissante d'événements (c'est à dire que $\forall n \in \mathbb{N} \quad A_{n+1} \subset A_n))$, avec $\mu(A_1) < +\infty$ et que $A = \overset{+\infty}{\underset{i=1}{\cap}} A_i \in \mathcal{A}$, alors la suite $(\mu(A_n))_{n \in \mathbb{N}}$ est monotone, décroissante, et converge vers $\mu(A)$.

Démonstration. 1. Il suffit de poser $A_i = \varnothing$ pour $i \geq n+1$ et d'appliquer l'axiome 2.

2. Il suffit d'appliquer la propriété 1 avec $n = 2$, $A_1 = A$ et $A_2 = B$.
3. Il suffit d'appliquer la propriété 2 avec $B = E \backslash A$: A et B sont disjoints donc $\mu(A) + \mu(A^c) = \mu(A \cup A^c) = \mu(E)$.
4. Les ensembles $A \backslash B$, $B \backslash A$ et $A \cap B$ sont disjoints et leur réunion est $A \cup B$, donc l'après la propriété 1, on a

   $$\begin{aligned} \mu(A \cup B) &= \mu(A \backslash B) + \mu(B \backslash A) + \mu(A \cap B) \\ \mu(A \cup B) &= (\mu(A \backslash B) + \mu(A \cap B) + (\mu(B \backslash A) + \mu(A \cap B)) - \mu(A \cap B) \\ &= \mu(A) + \mu(B) - \mu(A \cap B) \end{aligned}$$

   car $A \backslash B$ $A \cap B$ sont disjoints, de réunion A, tandis que $B \backslash A$ et $A \cap B$ sont disjoints, de réunion B.
5. Il suffit d'appliquer la relation 4 en remarquant que $\mu(A \cap B) \geq 0$
6. Si $A \subset B$, on a B est la réunion disjointe de A et de $B \backslash A$. Donc $\mu(B) = \mu(A) + \mu(B \backslash A) \geq \mu(A)$.
7. $(A \cap B) \subset A$, donc d'après la propriété 6 $\mu(A \cap B) \leq \mu(A)$. De même $\mu(A \cap B) \leq \mu(B)$. Finalement $\mu(A \cap B) \leq \min(\mu(A), \mu(B))$.
8. $A \subset (A \cup B)$, donc d'après la propriété 6 $\mu(A) \leq \mu(A \cup B)$. De même $\mu(B) \leq \mu(A \cup B)$. Finalement $\max(\mu(A), \mu(B)) \leq \mu(A \cup B)$.

9. Posons $B_1 = A_1$ et, pour tout $n \geq 2$ $B_n = A_n \backslash (\bigcup_{i=1}^{n-1} B_i)$. Par construction, les $(B_k)_{k\geq 1}$ sont deux à deux disjoints. De plus, on peut montrer par récurrence sur n que
$$\forall n \geq 1 \quad \bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} B_i.$$
On en déduit
$$\bigcup_{i=1}^{+\infty} A_i = \bigcup_{i=1}^{+\infty} B_i.$$
Donc
$$\begin{aligned} \mu(\bigcup_{i=1}^{+\infty} A_i) &= \mu(\bigcup_{i=1}^{+\infty} B_i) \\ &= \sum_{i=1}^{+\infty} \mu(B_i) \\ &\leq \sum_{i=1}^{+\infty} \mu(A_i). \end{aligned}$$

10. Comme $A_n \subset A_{n+1}$, on a $\mu(A_n) \leq \mu(A_{n+1})$, donc la suite est croissante. Comme on a pour tout n : $A_n \subset A$, la suite $(\mu(A_n))_{n\geq 1}$ est majorée par $\mu(A)$. Posons $B_1 = A_1$ et, pour tout $n \geq 2$ $B_n = A_n \backslash A_{n-1}$. On a :
$$\forall n \geq 1 \quad A_n = \bigcup_{i=1}^{n} B_i$$
et
$$A = \bigcup_{i=1}^{+\infty} A_i = \bigcup_{i=1}^{+\infty} B_i$$
Ainsi,
$$\begin{aligned} \mu(A) &= \mu(\bigcup_{i=1}^{+\infty} B_i) \\ &= \sum_{i=1}^{+\infty} \mu(B_i) \\ &= \lim_{n\to+\infty} \sum_{i=1}^{n} \mu(B_i) \\ &= \lim_{n\to+\infty} \mu(\bigcup_{i=1}^{n} B_i) \\ &= \lim_{n\to+\infty} \mu(A_n) \end{aligned}$$

11. On applique le résultat précédent à la suite croissante $(A'_n)_{n\geq 1}$ définie par $A'_n = A_1 \backslash A_n$.

□

2.2.3 Masse de Dirac

Soit $(\Omega, \mathcal{F})$ un espace et une tribu. Soit $x \in \Omega$. On appelle mesure de Dirac en x et on note δ_x la mesure définie par

$$\forall A \in \mathcal{F} \quad \delta_x(A) = 1\!\!1_A(x).$$

Vérifions brièvement que δ_x est une mesure : il est évident que δ_x est à valeurs dans $[0, +\infty]$. Maintenant, soit $(A_n)_{n\geq 1}$ une famille d'éléments de $\mathcal{F}$ deux à deux disjoints. Si x n'est dans aucun des A_i, il n'est pas dans leur réunion, et donc on a

$$\delta_x(\cup_{i\geq 1} A_i) = 0 = \sum_{i\geq 1} \delta_x(A_i).$$

Si x est dans un des A_i, il est dans un unique A_i, puisque les A_i sont deux à deux disjoints ; ainsi

$$\sum_{i\geq 1} \delta_x(A_i) = 1 = \delta_x(\cup_{i\geq 1} A_i).$$

2.2.4 Mesure de comptage

Soit $(\Omega, \mathcal{F})$ un espace et une tribu. On appelle mesure de comptage sur Ω la mesure C définie par

$$\forall A \in \mathcal{F} \quad C(A) = |A|,$$

où $|A|$ est le cardinal de A (le nombre d'éléments de A si A est fini, $+\infty$ sinon). Vérifions brièvement que C est une mesure : il est évident que C est à valeurs dans $[0, +\infty]$. Maintenant, soit $(A_n)_{n\geq 1}$ une famille d'éléments de $\mathcal{F}$ deux à deux disjoints. Si un des A_i est infini, il est évident que $\sum_{i\geq 1} C(A_i) = +\infty = C(\cup_{i\geq 1} A_i)$. De même si il y a une infinité de A_i non vides $\cup_{i\geq 1} A_i$ est infini et la somme $\sum_{i\geq 1} C(A_i)$ a une infinité de termes qui dépasse 1 donc encore une fois $\sum_{i\geq 1} C(A_i) = +\infty = C(\cup_{i\geq 1} A_i)$. Reste le cas où aucun des A_i n'est infini et où seul un nombre fini est non-vide : c'est donc une réunion finie d'ensemble finis et alors la formule recherchée est bien connue.

2.2.5 Opération simples

La somme de deux mesures est une mesure ; en multipliant une mesure par une constante positive, on a encore une mesure.

La preuve est simple et est laissée en exercice.

2.2.6 Extension d'une mesure – mesure de Lebesgue

On va présenter maintenant un théorème abstrait qui sera peu employé dans ce cours, mais est important pour fonder les bases de la théorie de l'intégration de Lebesgue.

Théorème 15 (Théorème de prolongement de Hahn)**.** *Étant donnée une algèbre $\mathcal{F}$ de parties d'un ensemble Ω et une mesure μ sur $\mathcal{F}$, la fonction d'ensemble $\bar{\mu}$ définie sur la tribu $\sigma(\mathcal{F})$ de parties de Ω engendrée par $\mathcal{F}$ par*

$$\bar{\mu}(A) = \inf(\{ \sum_{n=1}^{\infty} \mu(A_n); (A_n, n \geq 1) \subseteq \mathcal{F} \text{ et } A \subseteq \bigcup_{n=1}^{\infty} A_n\})$$

est une mesure sur $\sigma(\mathcal{F})$ qui prolonge μ. Ce prolongement est unique si μ est une mesure est σ-finie.

La preuve de ce théorème est basée sur le concept de mesure extérieure, développé notamment par Carathéodory.

Ce théorème difficile est admis.

Le théorème suivant ne sera pas utilisé dans ce cours, mais mérite d'être mentionné car il est très pratique dans certains problèmes théoriques que le lecteur pourra rencontrer dans le futur.

Théorème 16. *Si μ est une mesure finie sur $\mathcal{F}$ et $\mathcal{A}$ une algèbre engendrant $\mathcal{F}$, alors pour tout $A \in \mathcal{F}$ et tout $\varepsilon > 0$, il existe $A' \in \mathcal{A}$ tel que $\mu(A\Delta A') \leq \varepsilon$.*

Démonstration. Il suffit de montrer que l'ensemble $\mathcal{T}$ des $A \in \mathcal{F}$ tels que pour tout $\varepsilon > 0$, il existe $A' \in \mathcal{A}$ tel que $\mu(A\Delta A') \leq \varepsilon$, est une tribu. Comme $A\Delta A' = A^c\Delta A'^c$, la stabilité par passage au complémentaire est évidente. Soit $(A_n)_{n\geq 1}$ une suite d'éléments de $\mathcal{T}$. On pose $A = \cup_{n\geq 1} A_n$ et on se donne $\varepsilon > 0$. Pour tout n, soit $A'_n \in \mathcal{A}$ tel que $\mu(A_n\Delta A'_n) \leq \frac{\varepsilon}{2^{n+1}}$. Soit n tel que $\mu(A\backslash \bigcup_{k=1}^{n} A_k) \leq \varepsilon/2$. On a

$$\mu(A\Delta \bigcup_{k=1}^{n} A'_k) \leq \mu(A\Delta \bigcup_{k=1}^{n} A_k)+\mu(\bigcup_{k=1}^{n} A'_k\Delta \bigcup_{k=1}^{n} A_k) \leq \varepsilon/2+\sum_{k=1}^{n} \frac{\varepsilon}{2^{k+1}} \leq \varepsilon.$$

□

Passons au théorème d'existence de la mesure de Lebesgue.

Théorème 17. *Il existe une unique mesure λ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que quels que soient les réels a et b avec $a < b$, on ait*
- $\lambda(]-\infty, a]) = \lambda(]-\infty, a[) = \lambda([b, +\infty[) = \lambda(]b, +\infty[) = +\infty$
- $\lambda(]a, b]) = \lambda(]a, b[) = \lambda([a, b]) = \lambda([a, b[) = b - a.$
- $\lambda(\{a\}) = 0.$

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

Démonstration. Idée de preuve : on définit λ pour les réunions dénombrables d'intervalles (bornés ou pas), puis on applique le théorème de prolongement de Hahn. □

2.2.7 Mesure image

Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré, f une application de Ω dans Ω'. On pose

$$\tilde{\sigma}(\mathcal{F}, f) = \{A \in \mathcal{P}(\Omega'); f^{-1}(A) \in \mathcal{F}\}.$$

On appelle mesure image de μ par f et on note μ_f la mesure définie sur $\tilde{\sigma}(\mathcal{F}, f)$ par

$$\mu_f(A) = \mu(f^{-1}(A)).$$

Si f est une application qui est mesurable comme application de $(\Omega, \mathcal{F})$ dans $(\Omega', \mathcal{G})$, μ_f est évidemment définie sur $\mathcal{G}$, puisque $\mathcal{G}$ est une sous-tribu de $\tilde{\sigma}(\mathcal{F}, f)$.

2.3 Convergence et mesurabilité

2.3.1 Tribu borélienne de $\overline{\mathbb{R}}$

Rappelons brièvement quelques notions de base de la topologie de $\overline{\mathbb{R}}$. On a $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$. On définit φ sur $[-\pi/2, \pi/2]$, par $\varphi(x) = \tan x$ si $x \in]-\pi/2, \pi/2[$,

On définit une métrique sur $\overline{\mathbb{R}}$ par $d(x, y) = |\varphi^{-1}(x) - \varphi^{-1}(y)|$. Une boule ouverte pour d n'est rien d'autre que l'image par φ d'une boule ouverte de $[-\pi/2, \pi/2]$, ainsi la tribu borélienne sur $\overline{\mathbb{R}}$ n'est autre que la tribu image de la tribu borélienne de $[-\pi/2, \pi/2]$ par l'application φ. En particulier, il s'ensuit que la tribu borélienne de $\overline{\mathbb{R}}$ est engendrée par les ensembles de la forme $]x, +\infty]$. D'autre part, les boréliens de $\mathbb{R}$ ainsi que les singletons $\{+\infty\}$ et $\{-\infty\}$ sont dans la tribu borélienne de $\overline{\mathbb{R}}$.

2.3.2 Importance de la séparabilité de $\mathbb{R}$

$\mathbb{R}$ est un espace séparable, c'est à dire qu'il possède (au moins) une partie dénombrable dense : en effet, on sait bien que l'ensemble $\mathbb{Q}$ des rationnels est dense dans $\mathbb{R}$ Cette propriété est très souvent utilisée en théorie de la mesure, par exemple dans le résultat suivant, qui met en oeuvre une technique classique à connaître.

Théorème 18. *Soit f, g deux applications mesurables de $(\Omega, \mathcal{F})$ dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$. Alors*

$$\{f > g\} = \{\omega \in \Omega; f(\omega) > g(\omega)\} \in \mathcal{F}.$$

Démonstration.

$$\begin{aligned} \{f > g\} &= \bigcup_{q \in \mathbb{Q}} \{f > q > g\} \\ &= \bigcup_{q \in \mathbb{Q}} \{f > q\} \cap \{q > g\} \\ &= \bigcup_{q \in \mathbb{Q}} f^{-1}(]q, +\infty]) \cap g^{-1}([-\infty, q[) \end{aligned}$$

Comme f est mesurable de $(\Omega, \mathcal{F})$ dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ et que $]q, +\infty] \in \mathcal{B}(\overline{\mathbb{R}})$, $f^{-1}(]q, +\infty]) \in \mathcal{F}$. De même $g^{-1}([-\infty, q]) \in \mathcal{F}$, leur intersection est encore dans $\mathcal{F}$, et une union dénombrable d'éléments de $\mathcal{F}$ est dans $\mathcal{F}$, ce qui donne le résultat voulu. □

2.3.3 Convergence et mesurabilité

Théorème 19. *Soit $(f_n)_{n \geq 1}$ une suite d'applications mesurables de $(\Omega, \mathcal{F})$ dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$. Alors les applications suivantes et les événements suivants sont mesurables :*

1. $\sup_{n \geq 1} f_n$
2. $\inf_{n \geq 1} f_n$
3. $\overline{\lim}_{n \to +\infty} f_n$
4. $\underline{\lim}_{n \to +\infty} f_n$
5. $\{f_n$ *converge vers* $+\infty\}$
6. $\{f_n$ *converge vers* $-\infty\}$

7. *$\{f_n$ converge dans $\overline{\mathbb{R}}\}$*
8. *$\{f_n$ converge dans $\mathbb{R}\}$*

Démonstration. 1. Posons $f = \sup_{n\geq 1} f_n$. On a

$$f^{-1}(]x,+\infty]) = \bigcup_{n\geq 1} f_n^{-1}(]x,+\infty]),$$

où, en adoptant le formalisme probabiliste :

$$\{f > x\} = \bigcup_{n\geq 1} \{f_n > x\}.$$

2. On peut simplement remarquer que $\inf_{n\geq 1} f_n = -\sup_{n\geq 1}(-f_n)$, et appliquer le point précédent, sachant que l'opposé d'une fonction mesurable est mesurable(par exemple car $-f = (x \mapsto -x) \circ f$.

3. $\overline{\lim}_{n\to+\infty} f_n = \inf_{n\geq 1} g_n$, avec $g_n = \sup_{k\geq n} f_k$. La mesurabilité des (g_n) provient du point 1 ; on applique alors le point 2.

4. Preuve analogue, ou $\underline{\lim}_{n\to+\infty} f_n = -\overline{\lim}_{n\to+\infty}(-f_n)$

5. $\{f_n$ converge vers $+\infty\}$ est l'image réciproque de $\{+\infty\}$ par l'application mesurable $\underline{\lim}_{n\to+\infty} f_n$.

6. $\{f_n$ converge vers $-\infty\}$ est l'image réciproque de $\{-\infty\}$ par l'application mesurable $\overline{\lim}_{n\to+\infty} f_n$.

7. On a vu dans les points 3. et 4. que les fonctions $\overline{\lim}_{n\to+\infty} f_n$ et $\underline{\lim}_{n\to+\infty} f_n$ étaient $(\Omega,\mathcal{F}) - (\overline{\mathbb{R}},\mathcal{B}(\overline{\mathbb{R}}))$ mesurables.
   Or $\{f_n$ converge dans $\overline{\mathbb{R}}\}$ est le complémentaire de $\{\underline{\lim}_{n\to+\infty} f_n < \overline{\lim}_{n\to+\infty} f_n\}$ qui est dans $(\Omega,\mathcal{F})$ d'après le paragraphe sur la séparabilité.

8. C'est une conséquence immédiate des trois points précédents.

□

2.4 Exercices de théorie de la mesure

1. (a) Soient f et g deux applications quelconques de Ω dans $\mathbb{R}$ qui vérifient
$$\forall x \in \mathbb{R} \quad \{f < x\} = \{g < x\}.$$
Montrer que $f = g$ sur Ω. Que peut-on dire si la condition ci-dessus n'est vérifiée que pour tout x rationnel ?

   (b) On suppose maintenant que f et g sont mesurables de l'espace mesuré $(\Omega, \mathcal{F})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Montrer que si pour tout x réel $\{f < x\} = \{g < x\}$ μ-presque partout, alors $f = g$ μ-presque partout.

2. On pose
$$\mu_n = \sum_{k=1}^{2n} \frac{1}{k} \delta_k.$$
Dites brièvement pourquoi μ_n est une mesure et calculer $\mu_5([0, \pi])$. Étudier le comportement asymptotique de $\mu_n([n, +\infty[)$.

3. Soit a un réel et $\tau_a : \mathbb{R} \to \mathbb{R}$ la translation définie par $\tau_a(x) = x + a$. Montrer que la famille $\mathcal{A}_a = \{A \in \mathcal{P}(\mathbb{R}); \tau_a(A) = A\}$ des parties invariantes par τ_a est une tribu sur $\mathbb{R}$.
Plus généralement, si f est une application de $\mathbb{R}$ dans $\mathbb{R}$, donner une condition suffisante sur f pour que la famille $\mathcal{A} = \{A \in \mathcal{P}(\mathbb{R}); f(A) = A\}$ soit une tribu.

4. Soient $\mathcal{A}$ et $\mathcal{B}$ deux tribus sur un ensemble Ω. Montrer que la tribu engendrée par $\mathcal{A}$ et $\mathcal{B}$ coïncide avec la tribu engendrée par les ensembles de la forme $A \cap B$, où (A, B) décrit $\mathcal{A} \times \mathcal{B}$.

5. On rappelle que la tribu borélienne de $\mathbb{R}$ est la tribu engendrée par les ensembles de la forme $]a, b[; (a, b)^2 \in \mathbb{Q}$. Montrer que la tribu borélienne est également la tribu engendrée par les familles
   - $\mathcal{C} = \{]-\infty, a[; a \in \mathbb{Q}\}$.
   - $\mathcal{D} = \{]-\infty, a]; a \in \mathbb{Q}\}$.
   - $\mathcal{E} = \{[a, b[; (a, b)^2 \in \mathbb{Q}\}$.
   - $\mathcal{F} = \{[a, b]; (a, b)^2 \in \mathbb{Q}\}$.

6. Pour n entier strictement positif, on note $A_n = n\mathbb{N}^*$. Notons $\mathcal{P}$ l'ensemble des nombres premiers positifs et $\mathcal{T}$ la sous-tribu de $(\mathbb{N}^*, \mathcal{B}(\mathbb{N}^*))$ engendrée par les $(A_p)_{p \in \mathcal{P}}$.

   (a) Montrer que l'ensemble C des entiers qui sont premiers avec 2000 est $\mathcal{T}$-mesurable.

   (b) Montrer que l'ensemble $B = \{2^k; k \in \mathbb{N}^*\}$ des puissances de deux est $\mathcal{T}$-mesurable.

7. Soit (E, d) un espace métrique. Montrer que la tribu borélienne $\mathcal{B}(E)$ engendrée par les ouverts de E est aussi la plus petite tribu rendant mesurables toutes les applications continues de (E, d) dans $\mathbb{R}$ (muni de la tribu borélienne et de la topologie usuelle).

8. *Lemme de Doob*
   Soit f et g deux applications mesurables de $(\Omega, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Montrer que g est $\sigma(f)$-mesurable si et seulement si il existe une application mesurable u de $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ dans lui-même telle que $g = u \circ f$.

9. *Support d'une mesure sur* $\mathbb{R}^d$
   (a) Soit μ une mesure sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. On appelle support de μ l'ensemble des $x \in \mathbb{R}^d$ tels que tout ouvert contenant x est de mesure positive. Montrer que le support de μ est fermé.
   (b) Soit O un ouvert de $\mathbb{R}^d$. Montrer qu'il existe un ensemble dénombrable D, des familles $(x_n)_{n\in D}$, $(r_n)_{n\in D}$ avec $x_n \in \mathbb{Q}^d$ et $r_n \in \mathbb{Q}_*^+$ et $O = \underset{n\in D}{\cup} B(x_n, r_n)$.
   (c) Soit μ une mesure sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Montrer qu'il existe un ouvert O maximal (pour l'inclusion) de mesure nulle, puis que $\mathbb{R}^d \backslash O$ est le support de μ.

   On notera que la définition donnée à la première question est encore valide dans tout espace topologique muni de sa tribu borélienne. La caractérisation donnée dans la dernière question est encore vraie si l'espace est à base dénombrable (c'est le cas, comme ici, des espaces métriques séparables (avec une partie dénombrable dense)) ou si μ est une mesure finie.

Chapitre 3

Espace probabilisé

Voyons maintenant la définition d'une probabilité sur $(\Omega, \mathcal{F})$.

3.1 Espace probabilisé

On appelle
– probabilité
– ou mesure de probabilité
– ou loi
sur $(\Omega, \mathcal{F})$ toute application

$$\mathbb{P} : \ \mathcal{F} \ \rightarrow [0,1]$$

vérifiant les propriétés suivantes :

1. $\mathbb{P}(\varnothing) = 0$, $\mathbb{P}(\Omega) = 1$.
2. Pour toute suite $(A_i)_{i \in \mathbb{N}}$ d'éléments de $\mathcal{F}$ deux à deux disjoints, $\mathbb{P}(\bigcup_{i=1}^{+\infty} A_i) = \sum_{i=1}^{+\infty} \mathbb{P}(A_i)$.

Alors, le triplet $(\Omega, \mathcal{F}, \mathbb{P})$ est appelé espace probabilisé.

On remarque qu'un espace probabilisé est très exactement un espace mesuré associé à une mesure positive de masse totale 1.

Remarque sur le vocabulaire : l'image $\mathbb{P}(A)$ d'un événemement A par l'application $\mathbb{P}$ est appelée probabilité de cet événement. Ainsi le mot “probabilité” peut-il désigner à la fois une application et la valeur de cette application en un point. Le contexte doit permettre de lever toute ambiguité.

Les propositions suivantes sont alors des conséquences relativement faciles des définitions :

1. Pour toute suite $(A_i)_{1\leq i\leq n}$ d'éléments de $\mathcal{F}$ deux à deux disjoints, $\mathbb{P}(\overset{n}{\underset{i=1}{\cup}} A_i) = \sum\limits_{i=1}^{n} \mathbb{P}(A_i)$.
2. $\forall A, B \in \mathcal{F} \quad (A \cap B = \varnothing) \Longrightarrow (\mathbb{P}(A\cup B) = \mathbb{P}(A) + \mathbb{P}(B))$
3. $\forall A \in \mathcal{F} \quad \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
4. $\forall A, B \in \mathcal{F} \quad \mathbb{P}(A\cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A\cap B)$
5. $\forall A, B \in \mathcal{F} \quad \mathbb{P}(A\cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$
6. $\forall A, B \in \mathcal{F} \quad (A \subset B) \Longrightarrow (\mathbb{P}(A) \leq \mathbb{P}(B))$
7. $\forall A, B \in \mathcal{F} \quad \mathbb{P}(A\cap B) \leq \min(\mathbb{P}(A), \mathbb{P}(B))$
8. $\forall A, B \in \mathcal{F} \quad \mathbb{P}(A\cup B) \geq \max(\mathbb{P}(A), \mathbb{P}(B))$
9. Pour toute suite $(A_i)_{i\geq 1}$ d'éléments de $\mathcal{F}$, $\mathbb{P}(\overset{+\infty}{\underset{i=1}{\cup}} A_i) \leq \sum\limits_{i=1}^{+\infty} \mathbb{P}(A_i)$.
10. Si $(A_i)_{i\in\mathbb{N}}$ est une suite croissante d'événements (c'est à dire que $\forall n \in \mathbb{N} \quad A_n \subset A_{n+1})$) et que l'on pose $A = \overset{+\infty}{\underset{i=1}{\cup}} A_i$, alors la suite $(\mathbb{P}(A_n))_{n\in\mathbb{N}}$ est monotone, croissante, et converge vers $\mathbb{P}(A)$.
11. Si $(A_i)_{i\in\mathbb{N}}$ est une suite décroissante d'événements (c'est à dire que $\forall n \in \mathbb{N} \quad A_{n+1} \subset A_n)$) et que l'on pose $A = \overset{+\infty}{\underset{i=1}{\cap}} A_i$, alors la suite $(\mathbb{P}(A_n))_{n\in\mathbb{N}}$ est monotone, décroissante, et converge vers $\mathbb{P}(A)$.

Démonstration. Il suffit de particuliser aux cas d'une mesure de masse 1 les propriétés des mesures démontrées au chapitre précédent. □

3.2 Partitions et probabilités

Le théorème très simple qui suit est très fréquemment utilisé. Il traduit le fait que pour calculer une probabilité, il faut parfois diviser les cas.

Théorème 20. *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace probabilisé. Soit* I *un ensemble d'index fini ou dénombrable et* $(\Omega_i)_{i\in I}$ *une partition de* Ω. *Alors on a*

$$\forall A \in \mathcal{F} \quad \mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(A\cap\Omega_i).$$

Démonstration. Comme la famille $(\Omega_i)_{i\in I}$ une partition de Ω, la famille $(A\cap\Omega_i)_{i\in I}$ est une partition de A. A est donc réunion disjointe des $(A\cap\Omega_i)_{i\in I}$, donc $\mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(A\cap\Omega_i)$. □

3.3 Probabilité conditionnelle

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et B un événement observable de probabilité non nulle. On appelle probabilité conditionnelle sachant B l'application

$$\begin{array}{rcl} \mathbb{P}(.|B) : \mathcal{F} & \rightarrow & \mathbb{R} \\ A & \mapsto & \mathbb{P}(A|B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \end{array}$$

$\mathbb{P}(A|B)$ se lit "Probabilité de A sachant B".
On a évidemment

$$\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A|B). \tag{3.1}$$

Remarque : L'application "probabilité conditionnelle" est une probabilité. Elle vérifie donc toutes les propriétés énoncées précédemment.

3.3.1 Conditionnements en chaîne

Si A, B sont deux événements observables avec $A \subset B$ et $\mathbb{P}(B) \neq 0$, la formule (3.1) devient

$$\mathbb{P}(A) = \mathbb{P}(B)\mathbb{P}(A|B). \tag{3.2}$$

On a la généralisation suivante :

Théorème 21. *Soient $n \geq 2$ et $E_1, \dots, E_n$ des événements observables vérifiant*

$$E_n \subset E_{n-1} \subset \cdots \subset E_1$$

et $\mathbb{P}(E_{n-1}) > 0$. Alors on a

$$\mathbb{P}(E_n) = \mathbb{P}(E_n|E_{n-1})\mathbb{P}(E_{n-1}|E_{n-2}) \dots \mathbb{P}(E_2|E_1)\mathbb{P}(E_1)$$

Démonstration. La formule se montre par récurrence sur n. Pour $n = 2$, c'est une conséquence immédiate de (3.2). Pour $n > 2$, on applique d'abord la formule pour $n = 2$ aux événements E_n et E_{n-1} :

$$\mathbb{P}(E_n) = \mathbb{P}(E_n|E_{n-1})\mathbb{P}(E_{n-1}),$$

puis on applique la propriété de récurrence au rang $n - 1$. □

Exemple : (d'après André Franquin) Chez les papous, il y a les papous à poux et les papous pas à poux. La probabilité pour qu'un papou ait des poux vaut 0.1. De plus, chez les papous, il y a les papous papas et les papous pas papas. La probabilité pour qu'un papou à poux soit papa vaut 0.6. Or, chez les poux, il y a les poux papas et les poux pas papas : la probabilité pour qu'un papou à poux possède au moins un pou papa est de 0.8.

Question : on tire au hasard un papou. Quelle est la probabilité pour que ce soit un papa papou à poux papa ? Réponse : $0.8 \times 0.6 \times 0.1 = 0.048$.

Ce théorème est parfois énoncé sous la forme plus compliquée – mais équivalente – suivante.

Théorème 22. *Soient $n \geq 2$ et $A_1, \dots, A_n$ des événements observables avec $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$. Alors*

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = (\prod_{k=1}^{n-1} \mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_{k+1} | A_1 \cap A_2 \cap \dots \cap A_k))\mathbb{P}(A_1)$$

Démonstration. Il suffit de poser, pour $1 \leq i \leq n$, $E_i = \bigcap_{k=1}^{i} A_k$ et d'appliquer le théorème précédent. $\square$

3.3.2 Conditionnement par tous les cas possibles

Ceci est l'expression en termes de probabilités conditionnelles du principe de partition.

Théorème 23. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Soit I un ensemble d'index fini ou dénombrable et $(\Omega_i)_{i \in I}$ une partition de Ω. Alors on a*

$$\forall A \in \mathcal{F} \quad \mathbb{P}(A) = \sum_{i \in J} \mathbb{P}(A|\Omega_i)\mathbb{P}(\Omega_i),$$

où $J = \{i \in I ; \mathbb{P}(\Omega_i) > 0\}$.

Démonstration. D'après le théorème 20, on a

$$\begin{aligned} \mathbb{P}(A) &= \sum_{i \in I} \mathbb{P}(A \cap \Omega_i) \\ &= \sum_{i \in J} \mathbb{P}(A \cap \Omega_i) + \sum_{i \in I \setminus J} \mathbb{P}(A \cap \Omega_i) \\ &= \sum_{i \in J} \mathbb{P}(A \cap \Omega_i) \\ &= \sum_{i \in J} \mathbb{P}(A|\Omega_i)\mathbb{P}(\Omega_i) \end{aligned}$$

□

3.3.3 Formule de Bayes

Théorème 24. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Soit I un ensemble d'index fini ou dénombrable et $(\Omega_i)_{i\in I}$ une partition de Ω telle que pour tout $i \in I$, $\mathbb{P}(\Omega_i)$ soit non nul. Soit A un élément de probabilité non nulle.*

Alors on a, pour tout $j \in I$, la formule

$$\mathbb{P}(\Omega_j|A) = \frac{\mathbb{P}(A|\Omega_j)\mathbb{P}(\Omega_j)}{\sum_{i\in I} \mathbb{P}(A|\Omega_i)\mathbb{P}(\Omega_i)}.$$

Démonstration.

$$\begin{aligned}\mathbb{P}(\Omega_j|A) &= \frac{\mathbb{P}(A\cap\Omega_j)}{\mathbb{P}(A)}\\ &= \frac{\mathbb{P}(A|\Omega_j)\mathbb{P}(\Omega_j)}{\mathbb{P}(A)}\end{aligned}$$

et on applique le théorème précédent.

□

Exemple :

– 60% des étudiants qui vont en T.D. obtiennent l'examen.
– 10% des étudiants qui ne vont pas en T.D. obtiennent l'examen.
– 70% des étudiants vont en T.D.

Quelle proportion des lauréats a séché les cours ? On note A l'événement "être assidu en cours". On a $\mathbb{P}(A) = 0.7$, et donc $\mathbb{P}(A^c) = 0.3$. On note L l'événement "obtenir l'examen" : on a $\mathbb{P}(L|A^c) = 0.1$ et $\mathbb{P}(L|A) = 0.6$. On a alors

$$\mathbb{P}(A^c|L) = \frac{\mathbb{P}(L|A^c)\mathbb{P}(A^c)}{\mathbb{P}(L|A^c)\mathbb{P}(A^c) + \mathbb{P}(L|A)\mathbb{P}(A)} = \frac{0.1\times 0.3}{0.1\times 0.3 + 0.6\times 0.7} = \frac{3}{45} = \frac{1}{15}.$$

3.4 Indépendance

3.4.1 Événements indépendants

On dit que deux événements observables A et B sont indépendants si on a

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

Soit $(A_i)_{i\in G}$ une partie d'éléments de $\mathcal{F}$ indexés par un ensemble G. On dit que les événements constituant la famille $(A_i)_{i\in G}$ sont globalement indépendants si l'on a pour tout ensemble fini $I \subset G$:

$$\mathbb{P}(\bigcap_{i\in I} A_i) = \prod_{i\in I} \mathbb{P}(A_i).$$

3.4.2 Tribus indépendantes

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé ; $\mathcal{A}$ et $\mathcal{B}$ deux sous-tribus de $\mathcal{F}$. On dit que les tribus $\mathcal{A}$ et $\mathcal{B}$ sont indépendantes sous P si

$$\forall A \in \mathcal{A} \quad \forall B \in \mathcal{B} \quad \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Plus généralement, si $(\mathcal{A}_i)_{i\in I}$ est une famille de sous-tribus de $\mathcal{F}$, on dit que cette famille est indépendante sous P si pour tout ensemble fini $J \subset I$, on a

$$\forall (A_i)_{i\in J} \in \prod_{i\in J} \mathcal{A}_i \quad \mathbb{P}(\bigcap_{i\in J} A_i) = \prod_{i\in J} \mathbb{P}(A_i).$$

Remarque : Si I est fini et que $(\mathcal{A}_i)_{i\in I}$ est une famille de sous-tribus de $\mathcal{F}$, cette famille est indépendante sous P si et seulement si on a

$$\forall (A_i)_{i\in I} \in \prod_{i\in I} \mathcal{A}_i \quad \mathbb{P}(\bigcap_{i\in I} A_i) = \prod_{i\in I} \mathbb{P}(A_i).$$

Il suffit en effet de poser $A_i = \Omega$ pour $i \in I \backslash J$ pour exprimer une intersection indexée par J en une intersection indexée par I.

Exercice : Soient $A, B \in \mathcal{F}$. Montrer que A est indépendant de B si et seulement si la tribu $\sigma(A)$ est indépendante de la tribu $\sigma(B)$.

Remarque utile : Si les tribus $\mathcal{A}$ et $\mathcal{B}$ sont indépendantes sous P, que $\mathcal{A}'$ est une sous-tribu de $\mathcal{A}$ et $\mathcal{B}'$ est une sous-tribu de $\mathcal{B}$, alors les tribus $\mathcal{A}'$ et $\mathcal{B}$' sont indépendantes sous P.

3.4.3 Indépendance et tribus engendrées

Définition On dit qu'une famille $\mathcal{C}$ de parties de Ω est un π-système si

$$\forall (A, B) \in \mathcal{C} \times \mathcal{C} \quad A \cap B \in \mathcal{C}.$$

On donne maintenant un résultat général de théorie de la mesure très utile. Sa preuve, basée sur le théorème $\lambda - \pi$ de Dynkin, est admise ici.[1]

[1] Le lecteur intéressé pourra se référer à la dernière section de ce chapitre ainsi qu'à la section 3.3 de l'ouvrage de Patrick Billingsley : *Probability and measure*, précisément aux théorèmes 3.2 et 3.3.

Proposition 1 (Critère d'identitification d'une probabilité)**.** *Soit P et Q deux probabilités sur l'espace mesuré $(\Omega, \mathcal{F})$. On suppose qu'il existe un π-système $\mathcal{C}$ qui engendre $\mathcal{F}$ ($\sigma(\mathcal{C}) = \mathcal{F}$) et sur lequels P et Q coïncident, c'est à dire que*

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

Alors $P = Q$.

Théorème 25. *Soit $\mathcal{C}$ et $\mathcal{D}$ deux familles de parties mesurables de $(\Omega, \mathcal{F})$. On suppose que $\mathcal{C}$ et $\mathcal{D}$ sont des π-systèmes et que pour tout $(A, B) \in \mathcal{C} \times \mathcal{D}$, on a $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Alors, les tribus $\mathcal{A} = \sigma(\mathcal{C})$ et $\mathcal{B} = \sigma(\mathcal{D})$ sont indépendantes.*

Démonstration. Pour $A \in \mathcal{A}$, on pose $\mathcal{T}_A = \{B \in \mathcal{B}, \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\}$.

Regardons d'abord le cas où $A \in \mathcal{C}$. Si $\mathbb{P}(A) = 0$, alors A est indépendant de tout, donc $\mathcal{T}_A = \mathcal{B}$. Si $\mathbb{P}(A) \neq 0$, on peut définir sur $\mathcal{B}$ la probabilité conditionnelle P_A par

$$\forall B \in \mathcal{B} \quad P_A(B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

Les probabilités P et P_A coïncident sur $\mathcal{D}$. Comme $\mathcal{D}$ est un π-système qui engendre $\mathcal{B}$, P et P_A coïncident sur $\mathcal{B}$. On en déduit que lorsque $A \in \mathcal{C}$, on a $\mathcal{T}_A = \mathcal{B}$.

On a donc montré que si $\mathcal{C}$ et $\mathcal{D}$ sont des π-systèmes, alors

$$\begin{aligned} & \forall (A, B) \in \mathcal{C} \times \mathcal{D} \quad \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \\ \Longrightarrow \; & \forall (A, B) \in \mathcal{C} \times \sigma(\mathcal{D}) \quad \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \end{aligned}$$

Mais, $\mathcal{B} = \sigma(\mathcal{D})$ est lui-même un π-système. Le résultat que l'on vient de démontrer s'applique cette fois avec $(\mathcal{B}, \mathcal{C})$ à la place de $(\mathcal{C}, \mathcal{D})$, et on obtient que

$$\begin{aligned} & \forall (A, B) \in \mathcal{C} \times \mathcal{D} \quad \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \\ \Longrightarrow \; & \forall (A, B) \in \sigma(\mathcal{C}) \times \sigma(\mathcal{D}) \quad \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \end{aligned}$$

ce qui était notre objectif

□

Théorème 26. *Les deux propriétés suivantes sont équivalentes :*

- *Les tribus $(\mathcal{A}_i)_{i \in I}$ sont indépendantes*
- *Pour tout $j \in I$, la tribu $\mathcal{A}_j$ est indépendante de la tribu $\bigwedge_{i \in I \backslash \{j\}} \mathcal{A}_i$.*

Démonstration. – Preuve de $1 \Longrightarrow 2$: Soit $j \in I$. On considère le π-système $\mathcal{C}$ défini par

$$\mathcal{C} = \bigcup_{F \subseteq I \backslash \{j\}} \{ \bigcap_{x \in F} A_x ; \forall x \in F \quad A_x \in \mathcal{A}_x \}.$$

Il est facile de voir que $\mathcal{C}$ est un π-système qui engendre $\bigwedge_{i \in I \backslash \{j\}} \mathcal{A}_i$ et que $\forall (A, B) \in \mathcal{A}_j \times \mathcal{C} \quad \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Le théorème 25 permet de conclure.

– Preuve de $2 \Longrightarrow 1$:
On montre par récurrence sur n la proposition $\mathcal{P}_n$

$$\mathcal{P}_n : \ |I| = n \Longrightarrow \forall \prod_{x \in I} A_x \in \prod_{x \in I} \mathcal{A}_x \quad \mathbb{P}(\bigcap_{x \in I} A_x) = \prod_{x \in I} \mathbb{P}(A_x).$$

Il est clair que $\mathcal{P}_0$ et $\mathcal{P}_1$ sont vraies. Montrons $\mathcal{P}_n \Longrightarrow \mathcal{P}_{n+1}$. Soit I un ensemble de cardinal $n+1$. On peut écrire $I = J \cup \{x_0\}$ avec $|J| = n$. Soit $E_1 = \prod_{x \in I} A_x \in \prod_{x \in I} \mathcal{A}_x$. On a $E_1 = A_{x_0} \cap E_2$, avec $E_1 = \bigcap_{x \in J} A_x$. Comme $A_{x_0} \in \mathcal{A}_{x_0}$ et $E_2 \in \bigwedge_{i \in I \backslash \{x_0\}} \mathcal{A}_i$, l'hypothèse 2 implique $\mathbb{P}(E_1) = \mathbb{P}(A_{x_0})\mathbb{P}(E_2)$. Mais d'après $\mathcal{P}_n$, on a

$$\mathbb{P}(E_2) = \mathbb{P}(\bigcap_{x \in J} A_x) = \prod_{x \in J} \mathbb{P}(A_x),$$

d'où

$$\mathbb{P}(E_1) = \mathbb{P}(A_{x_0})\mathbb{P}(E_2) = \prod_{x \in I} \mathbb{P}(A_x),$$

ce qui achève la preuve. □

3.5 Théorème $\lambda - \pi$ de Dynkin (*)

Cette section peut être omise en première lecture.

On dit que $\mathcal{L} \subset \mathcal{P}(\Omega)$ est un λ-système si on a
- $\Omega \in \mathcal{L}$
- $\forall A, B \in \mathcal{L} \quad A \subset B \Longrightarrow B \backslash A \in \mathcal{L}$
- Pour toute suite croissante $(A_n)_{n \geq 1}$ d'éléments de $\mathcal{L}$, $\bigcup_{n \geq 1} A_n \in \mathcal{L}$

On peut déjà remarquer que si $\mathcal{A} \subset \mathcal{P}(\Omega)$ est à la fois un λ-système et un π-système, alors $\mathcal{A}$ est une tribu. En effet, soit $(A_n)_{n\geq 1}$ une suite d'éléments de $\mathcal{A}$. Si on pose $A'_n = \overset{n}{\underset{k=1}{\cup}} A_k = \Omega \backslash (\overset{n}{\underset{k=1}{\cap}} A_k^c)$, on a $A'_n \in \mathcal{A}$ (on utilise la stabilité par intersection finie et par complémentation). Donc, comme $\mathcal{A} \subset \mathcal{P}(\Omega)$ est un λ-système $A = \underset{n\geq 1}{\cup} A_n = \underset{n\geq 1}{\cup} A'_n \in \mathcal{A}$, d'après le troisième axiome d'un λ-système.

Théorème 27 (Théorème $\lambda - \pi$)**.** *Si $\mathcal{P}$ est un π-système et $\mathcal{L}$ est un λ-système, alors $\mathcal{P} \subset \mathcal{L}$ entraîne $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

Démonstration. Voir par exemple Billingsley, théorème 3.2 □

Preuve de la proposition 1 :

Démonstration. Si on regarde l'ensemble $\mathcal{L}$ des parties de A de la tribu engendrée par le π-système qui sont telles que $\mathbb{P}(A) = \mathbb{Q}(A)$, il n'est pas difficile (en utilisant notamment le théorème de continuité séquentielle croissante) de voir que $\mathcal{L}$ est un λ-système ; il suffit alors d'appliquer le théorème $\lambda - \pi$. □

3.6 Premiers exercices de probabilité

1. Soient $A_1, \dots, A_n$ des événements indépendants. Montrer que les événements $A_1^c, \dots, A_n^c$ sont indépendants.
2. Soit $(A_n)_{n\geq 1}$ une suite d'événements indépendants, tous de probabilité non nulle. On pose
$$A = \overset{+\infty}{\underset{n=1}{\cap}} A_n$$
et
$$B = \underset{n\to+\infty}{\underline{\lim}} A_n = \underset{n=1}{\cup} \overset{+\infty}{\underset{k=n}{\cap}} A_k.$$
Montrer que $\mathbb{P}(A) = 0$ si et seulement si $\mathbb{P}(B) = 0$
3. Le but de cette exercice est de montrer qu'il est impossible de construire un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et une variable aléatoire X à valeurs entières sur cet espace tels que
$$\forall n \geq 1 \quad \mathbb{P}(n \text{ divise } X) = \frac{1}{n}.$$
On note $p_1, \dots, p_n, \dots$ la suite des nombres premiers.
   (a) Posons $E = \underset{n\to+\infty}{\underline{\lim}} \{p_n \text{ ne divise pas } X\}$. Montrer $E = \Omega$.

(b) On pose $D_n = \bigcap_{k=1}^{n} \{p_k \text{ ne divise pas } X\}$. Montrer

$$\frac{1}{\mathbb{P}(D_n)} \geq \sum_{i=1}^{p_n} \frac{1}{i}$$

(c) En déduire que $\mathbb{P}(D) = 0$, où on a posé $D = \bigcap_{k=1}^{+\infty} \{p_k \text{ ne divise pas } X\}$.

(d) Conclure.

(e) Montrer que la série de terme général $\frac{1}{p_i}$ diverge.

4. Une enquête effectuée parmi les nouveaux adhérents du parti socialiste français en 2002 a montré que les femmes représentaient $40,55\%$ des nouveaux adhérents. $20,4\%$ des nouvelles militantes socialistes sont enseignantes, tandis que seulement $12,81\%$ des nouveaux militants de sexe masculin sont enseignants. Parmi les enseignants qui militent nouvellement au parti socialiste, quelle est la proportion de femmes ?

5. On s'intéresse au *problème des dérangements* : n mathématiciens déposent leurs chapeaux au vestiaire au début d'un congrès et, à la fin du congrès, en reprennent un au hasard par distraction. On s'intéresse à la probabilité p_n qu'aucun ne retrouve son chapeau.

(a) Proposez un espace Ω convenable et une probabilité associée. En déduire que l'on doit avoir $p_n = \frac{d_n}{n!}$, où d_n est le nombre de permutations de $\mathcal{S}_n$ sans point fixe :

$$d_n = \text{Card}(\{\sigma \in \mathcal{S}_n; \forall i \in 1, \dots, n \quad \sigma(i) \neq i\}).$$

(On pose $d_0 = 1$.)

(b) Pour $0 \leq k \leq n$, on note A_k^n l'ensemble des permutations de $\mathcal{S}_n$ ayant exactement k points fixes :

$$A_k^n = \{\sigma \in \mathcal{S}_n; \text{Card}(\{i \in 1, \dots, d \quad | \quad \sigma(i) = i\}) = k\}.$$

Montrer $\text{Card}(A_k^n) = \binom{n}{k} d_{n-k}$. En déduire

$$\sum_{k=0}^{n} \binom{n}{k} d_k = n!$$

(c) Soit Φ l'endomorphisme de $\mathbb{R}_n[X]$ défini par $\Phi(P) = P(X+1)$ – on rappelle que $\mathbb{R}_n[X]$ désigne l'ensemble des polynômes réels de degré inférieur ou égal à n. Déterminer la matrice M de Φ dans la base $(1, X, \dots, X^n)$. Calculer M^{-1}.

(d) Montrer $(d_0, d_1, \dots, d_n).M = (0!, 1!, \dots, n!)$. En déduire

$$p_n = \frac{d_n}{n!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

(e) Montrer $\lim_{n\to\infty} p_n = \frac{1}{e}$. Montrer que pour $n \geq 2$, d_n est l'entier le plus proche de $\frac{n!}{e}$.

6. *Calcul probabiliste de la formule de l'indicatrice d'Euler*
On note Ω_n l'ensemble des entiers de 1 à n. On note $n = \prod_i p_i^{\alpha_i}$ la décomposition de n en produits de facteurs premiers. Le but de cet exercice est de déterminer le nombre $\varphi(n)$ qui est le cardinal de l'ensemble G_n des d'entiers entre 1 et n qui sont premiers avec n. On note $\mathbb{P}$ la loi uniforme sur Ω_n.
   (a) Pour d divisant n, on note $A_d = \{k \in \Omega_n; d|k\}$. Calculer $\mathbb{P}(A_d)$.
   (b) Soit $d_1, \dots, d_r$ des diviseurs de n premiers entre eux. Calculer $\mathbb{P}(\cap_{i=1}^r A_{d_i})$.
   (c) Montrer que $\mathbb{P}(G_n) = 1 - \mathbb{P}(\cup_i A_{p_i})$.
   (d) En déduire que $\varphi(n)/n = \prod_i (1 - 1/p_i)$.
7. On mélange $n (n \geq 6)$ paires de chaussetttes et l'on tire au hasard 6 chaussettes. On considère les événements suivants : $E_1 = \{$ obtenir trois paires $\}$, $E_2 = \{$ obtenir au moins une paire $\}$, $E_3 = \{$ obtenir une seule paire $\}$. En supposant que tous les ensembles de 6 chaussettes ont la même probabilité d'être tirés, calculer $\mathbb{P}(E_1), \mathbb{P}(E_2), \mathbb{P}(E_3)$.
8. On choisit au hasard, successivement et sans remise trois nombres parmi $\{1, ..., n\}$. Calculer la probabilité que le troisième nombre tiré se trouve entre les deux premiers.
9. Une élection a lieu entre deux candidats A et B. Le premier candidat A obtient a voix et le second B obtient b voix avec $a > b$.
   (a) Représenter le dépouillement des bulletins à l'aide d'un chemin dans $\mathbb{R}^2$ partant de $(0, 0)$ arrivant à (a, b) constitué uniquement de segments de longueur 1, parallèles à l'axe Ox ou Oy, orientés dans le sens croissant. En déduire un modèle équiprobable concernant le dépouillement.
   (b) Quelle est la probabilité pour qu'au cours du dépouillement,
      – le premier bulletin soit en faveur de B ?
      – A et B se retrouvent à un instant à égalité ? (indic. : distinguer suivant le premier bulletin)
      – A ait toujours strictement plus de voix que B ?

10. Donner un exemple de trois évènements A_1, A_2, A_3 qui ne sont pas indépendants et pour lesquels

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\,\mathbb{P}(A_2)\,\mathbb{P}(A_3)\ .$$

11. *Valeurs d'adhérence de la suite $\varphi(n)/n$*

    (a) Montrer que pour toute série divergente positive dont le terme général (u_n) tend vers 0, et pour tout $\ell > 0$, on peut extraire une sous-série (u_{n_k}) telle que $\displaystyle\sum_{k=1}^{+\infty} u_{n_k} = \ell$.

    (b) On note $\varphi(n)$ l'indicatrice d'Euler. Montrer que l'ensemble des valeurs d'adhérences de la suite $(\varphi(n)/n)_{n\geq 1}$ est l'intervalle $[0, 1]$ tout entier.

Chapitre 4

Intégrales

Jusqu'ici, on n'a parlé que de mesures et nullement d'intégrales. Le présent chapitre va pleinement compenser cela !

On va commencer par donner la définition de l'intégrale dite "de Lebesgue" et en énoncer les propriétés fondamentales. Vu le volume horaire du cours, certains résultats seront admis afin de se concentrer sur la pratique.

4.1 Définition de l'intégrale et propriétés de base

4.1.1 Définition

Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré. Pour toute fonction positive f, on définit l'intégrale de f, notée $\int f \, d\mu$ ou encore $\int f(x) \, d\mu(x)$ par

$$\int f \, d\mu = \sup \sum_i \inf\{f(\omega); \omega \in \Omega_i\} \mu(\Omega_i),$$

où le sup porte sur toutes les partitions finies de Ω.

Lorsque f prend des valeurs négatives, on écrit f comme différence de deux fonctions positives :

$$f = f^+ - f^-, \text{ où } f^+(\omega) = \max(f(\omega), 0) \text{ et } f^-(\omega) = \max(-f(\omega), 0)$$

Lorsque $\int f^+ \, d\mu$ et $\int f^- \, d\mu$ sont simultanément finies, on dit que f est intégrable et on peut définir

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

Lorsque $\int f^+ \, d\mu$ et $\int f^- \, d\mu$ sont, l'un fini, l'autre infini, on s'autorise toutefois à écrire

- $\int f \, d\mu = +\infty$ si $\int f^+ \, d\mu = +\infty$ et $\int f^- \, d\mu < +\infty$.
- $\int f \, d\mu = -\infty$ si $\int f^+ \, d\mu < +\infty$ et $\int f^- \, d\mu = +\infty$.

4.1.2 Propriétés de base de l'intégrale

Définition : on dit qu'une propriété $\mathcal{P}$ relative aux points de Ω est vérifiée μ-presque partout si il existe E mesurable avec $\mu(E) = 0$ tel que pour tout $x \in \Omega \backslash E$ $\mathcal{P}(x)$ est vérifié.

On donne d'emblée sans démonstration les propriétés de base de l'intégrale :

- **Lien avec la mesure** : Pour tout ensemble A mesurable, on a $\int \mathbb{1}_A \, d\mu = \mu(A)$.
- **Positivité** : Si f et g sont intégrables avec $f \le g$ μ-presque partout, alors $\int f \, d\mu \le \int g \, d\mu$, avec égalité si et seulement si $f = g$ μ-presque partout. En particularité, si $f \ge 0$ μ-presque partout et $\int f \, d\mu = 0$, alors $f = 0$ μ-presque partout.
- **Linéarité** : Si f et g sont intégrables, α et β des réels, alors $\int \alpha f + \beta \, d\mu = \alpha \int f + \beta \int g \, d\mu$
- **Convergence monotone (ou théorème de Beppo Levi[1])** : Si $(f_n)_{n \ge 1}$ est une suite croissante de fonctions positives convergeant presque partout vers f, alors la suite $\int f_n \, d\mu$ converge vers $\int f \, d\mu$. (la limite peut être infinie)

L'objectif prioritaire du lecteur est, nous semble-t'il, d'acquérir une bonne familiarité des propriétés de cette nouvelle intégrales. Aussi, afin de ne pas lasser par des preuves un peu techniques qui arriveraient avant que l'intérêt de l'outil soit réellement compris, nous rapportons les preuves à une section ultérieure qui viendra en fin de chapitre.

4.1.3 Conséquences importantes

Théorème 28 (Lemme de Fatou)**.** *Pour toute suite $(f_n)_{n \ge 1}$ de fonctions mesurables positives, on a*

$$\int \underline{\lim}_{n \to +\infty} f_n \, d\mu \le \underline{\lim}_{n \to +\infty} \int f_n \, d\mu$$

Démonstration. Il suffit de poser $g_n = \inf_{k \ge n} f_k$. (g_n) est une suite croissante,

[1]Beppo Levi (1875-1961) est un mathématicien italien. Pas de trait d'union donc entre Beppo et Levi !

dont la limite est, par définition, $\lim g_n = \underline{\lim} f_n$. On a pour tout n

$$\begin{aligned} f_n &\geq g_n \\ \int f_n \, d\mu &\geq \int g_n \, d\mu \end{aligned}$$

D'où

$$\underline{\lim}_{n \to +\infty} \int f_n \, d\mu \geq \lim_{n \to +\infty} \int g_n \, d\mu$$

Mais d'après le théorème de convergence monotone $\int g_n \, d\mu$ converge vers $\int g \, d\mu$, ce qui est le résultat voulu. □

Théorème 29 (Convergence dominée). *Si $(f_n)_{n \geq 1}$ est une suite de fonctions convergeant presque partout vers f, et telle qu'il existe une fonction g intégrable vérifiant pour tout n, $|f_n| \leq g$ alors la suite $\int f_n \, d\mu$ converge vers $\int f \, d\mu$.*

Démonstration. Les f_n sont intégrables car dominées par g : par suite les fonctions $g + f_n$ et $g - f_n$ sont intégrables et positives : on peut leur appliquer le lemme de Fatou :

$$\int \underline{\lim}_{n \to +\infty} (g + f_n) \, d\mu \leq \underline{\lim}_{n \to +\infty} \int (g + f_n) \, d\mu$$

et

$$\int \underline{\lim}_{n \to +\infty} (g - f_n) \, d\mu \leq \underline{\lim}_{n \to +\infty} \int (g - f_n) \, d\mu$$

soit

$$\int g \, d\mu + \int f \, d\mu \leq \int g \, d\mu + \underline{\lim}_{n \to +\infty} \int f_n \, d\mu$$

et

$$\int g \, d\mu - \int f \, d\mu \leq \int g \, d\mu - \overline{\lim}_{n \to +\infty} \int f_n \, d\mu$$

En simplifiant, on obtient

$$\int f \, d\mu \leq \underline{\lim}_{n \to +\infty} \int f_n \, d\mu$$

et

$$\overline{\lim}_{n \to +\infty} \int f_n \, d\mu \leq \int f \, d\mu$$

ce qui montre bien le résultat voulu. □

4.2 Intégration sur un ensemble

Pour tout ensemble mesurable A et toute fonction intégrable (ou positive) f, on note

$$\int_A f\ d\mu = \int f 1\!\!1_A\ d\mu.$$

Théorème 30. *Si f est intégrable et que $(A_n)_{n\geq 1}$ est une partition dénombrable de Ω, alors*

$$\int f\ d\mu = \sum_{k=1}^{+\infty} \int_{A_k} f\ d\mu.$$

Démonstration. On pose $f_n = f(\sum_{k=1}^n 1\!\!1_{A_k})$ et on applique le théorème de convergence dominée. □

4.3 Quelques cas particuliers importants

4.3.1 Intégration par rapport à une masse de Dirac

Théorème 31. *Soit $(\Omega, \mathcal{F})$ un espace mesuré. On suppose que le singleton $\{x\}$ est dans $\mathcal{F}$. Alors, pour toute fonction f mesurable, on a*

$$\int_\Omega f\ d\delta_x = f(x).$$

Démonstration. Par linéarité, comme $f = f^+ - f^-$, il suffit de traiter le cas où f est positive. Soit $(\Omega_i)_{i\in I}$ une partition mesurable finie. Posons $\mu = \delta_x$. Pour tout $i \in I$, on a

$$\mu(\Omega_i) \inf_{\Omega_i} f \leq \mu(\Omega_i) f(x)$$

En effet, si $x \notin \Omega_i$, les deux membres de l'égalité sont nuls, sinon l'inégalité $\inf\limits_{\Omega_i} f \leq f(x)$ est conséquence de $x \in \Omega_i$. En faisant la somme sur $i \in I$, on obtient

$$\sum_{i\in I} \mu(\Omega_i) \inf_{\Omega_i} f \leq (\sum_{i\in I} \mu(\Omega_i)) f(x) = f(x),$$

d'où en passant au sup sur toutes les partitions $\int f\ d\mu \leq f(x)$. Cependant, si l'on prend $I = \{1, 2\}$, $\Omega_1 = \{x\}$ et $\Omega_2 = \{x\}^c$, on a

$$\sum_{i\in I} \mu(\Omega_i) \inf_{\Omega_i} f = f(x),$$

d'où l'égalité voulue. □

4.3.2 Intégration par rapport à la mesure de comptage

Théorème 32. *Soit Ω un ensemble dénombrable. On note C la mesure de comptage sur $(\Omega, \mathcal{P}(\Omega))$. Toute fonction définie sur Ω est mesurable. Pour toute fonction f positive, on a*

$$\int_\Omega f(\omega)\ dC(\omega) = \sum_{k\in\Omega} f(k).$$

Dans le cas général, f est intégrable si et seulement si $\sum_{k\in\Omega} |f(k)| < +\infty$ et dans ce cas, on a encore l'égalité ci-dessus.

Démonstration. Soit f positive et $(\Omega_i)_{i\in I}$ une partition de Ω

$$\begin{aligned} C(\Omega_i)\ \inf_{\Omega_i} f &= \sum_{k\in\Omega} \left(1\!\!1_{\Omega_i}(k)\ \inf_{\Omega_i} f\right) \\ &\le \sum_{k\in\Omega} (1\!\!1_{\Omega_i}(k) f(k)) \end{aligned}$$

D'où en sommant sur I

$$\sum_{i\in I} C(\Omega_i)\ \inf_{\Omega_i} f \le \sum_{i\in I} \sum_{k\in\Omega} (1\!\!1_{\Omega_i}(k) f(k))$$

Cependant

$$\begin{aligned} \sum_{i\in I} \sum_{k\in\Omega} (1\!\!1_{\Omega_i}(k) f(k)) &= \sum_{k\in\Omega} \left(\sum_{i\in I} 1\!\!1_{\Omega_i}(k) f(k)\right) \\ &= \sum_{k\in\Omega} f(k) \end{aligned}$$

d'où en passant au sup

$$\int_\Omega f(\omega)\ dC(\omega) \le \sum_{k\in\Omega} f(k).$$

Réciproquement soit F une partie finie de Ω. On considère la partition de cardinal $|F|+1$, formée des $|F|$ singletons de F et de F^c : elle donne lieu à une somme

$$\sum_{k\in F} f(k) + \inf_{F^c} f \ge \sum_{k\in F} f(k)$$

On en déduit

$$\int_\Omega f(\omega)\, dC(\omega) \geq \sum_{k\in F} f(k)$$

En prenant la borne supérieure sur toutes les parties finies de Ω, on obtient

$$\int_\Omega f(\omega)\, dC(\omega) \geq \sum_{k\in \Omega} f(k),$$

d'où l'égalité voulue. Dans le cas où f est de signe quelconque, la formule précédente appliquée à $|f|$ donne

$$\int_\Omega |f|(\omega)\, dC(\omega) = \sum_{k\in \Omega} |f(k)|.$$

Dans le cas où la dernière somme est finie, en appliquant cette fois la formule à f^+ et f^-, on a

$$\int_\Omega f^+(\omega)\, dC(\omega) = \sum_{k\in \Omega} f^+(k).$$

et

$$\int_\Omega f^-(\omega)\, dC(\omega) = \sum_{k\in \Omega} f^-(k).$$

Ces deux quantités sont finies car $f^+ \leq |f|$ et $f^- \leq |f|$: en faisant la différence, on obtient alors par linéarité

$$\int_\Omega f(\omega)\, dC(\omega) = \sum_{k\in \Omega} f(k).$$

□

4.3.3 Fonctions simples (ou fonctions étagées)

On appelle fonction simple (ou fonction étagée) toute combinaison linéaire d'indicatrices d'ensembles mesurables.

On peut dire aussi qu'une fonction simple est une fonction mesurable qui ne prend qu'un nombre fini de valeurs.

Lemme 2. *Toute fonction mesurable positive f (éventuellement infinie) peut s'écrire comme limite simple d'une suite croissante de fonctions simples (f_n).*

Démonstration. On définit sur $[0, +\infty]$ une fonction φ_n par $\varphi_n(x) = 2^{-n} \operatorname{Int}(2^n x) \mathbb{1}_{[0,n]}(x)$ pour $x < +\infty$ et $\varphi_n(+\infty) = n$. Évidemment la suite $(\varphi_n(\infty))_{n \geq n1}$ tend en croissant vers $+\infty$. Soit $x \geq 0$. Évidemment $\mathbb{1}_{[0,n+1]}(x) \geq \mathbb{1}_{[0,n]}(x)$ Posons $y = 2^n x$. On a $y \geq \operatorname{Int}(y)$, donc $2y \geq 2 \operatorname{Int}(y)$. Mais $2 \operatorname{Int}(y)$ est entier, donc $\operatorname{Int}(2y) \geq 2 \operatorname{Int}(y)$, ce qui nous donne finalement $\varphi_{n+1}(x) \geq \varphi_n(x)$. D'autre part pour $n \geq x$, on a $x - 2^{-n} \leq \varphi_n(x) \leq x$, donc $\varphi_n(x)$ tend vers x. Il suffit alors de poser $f_n(x) = \varphi_n(f(x))$. □

4.3.4 Intégration par rapport à une somme de deux mesures

Théorème 33. *Soit $(\Omega, \mathcal{F})$ un espace mesuré, μ et ν deux mesures sur $(\Omega, \mathcal{F})$. Soit f une fonction $(\Omega, \mathcal{F})$-$(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable positive. On a*

$$\int_\Omega f \; d(\mu + \nu) = \int_\Omega f \; d\mu + \int_\Omega f \; d\nu. \tag{4.1}$$

Dans le cas où f est de signe quelconque, si $\int_\Omega |f| \; d\mu$ et $\int_\Omega |f| \; d\nu$ sont finis, f est intégrable par rapport à $\mu + \nu$ et on a encore (4.1).

Démonstration. Dans le cas où f est l'indicatrice d'un élément de $\mathcal{F}$, l'identité (4.1) découle de la définition de la mesure somme de deux mesures et de la valeur de l'intégrale d'une indicatrice. Par linéarité, la formule (4.1) est encore vraie si f est une combinaison linéaire d'indicatrices d'éléments de $\mathcal{F}$, autrement dit une fonction simple. En utilisant le lemme 2 et le théorème de convergence monotone, il s'ensuit que l'identité (4.1) est vraie pour toute fonction mesurable positive. Comme précédemment, le cas général s'en déduit en séparant partie positive et partie négative. □

4.4 Lien avec l'intégrale de Riemann

Théorème 34. *Soit f une fonction continue par morceaux sur un intervalle compact $[a, b]$. Alors*

$$\int_{[a,b]} f(x) \; d\lambda(x) = \int_a^b f(t) \; dt.$$

Démonstration. Avec la relation de Chasles et la linéarité de l'intégrale de Lebesgue, on peut se ramener au cas où f est continue sur $[a, b]$. Posons

$$f_n(x) = f\left(a + \operatorname{Ent}(\frac{n(x-a)}{b-a})\frac{b-a}{n}\right).$$

Comme f est continue sur $[a, b]$, elle y est uniformément bornée par une constante M. Comme f est continue sur $[a, b]$, $f_n(x)$ y converge partout vers $f(x)$. Comme $|f_n| \leq M\mathbb{1}_{[a,b]}$, le théorème de convergence dominée assure que $\int_{[a,b]} f_n(x)\ d\lambda(x)$ converge vers $\int_{[a,b]} f(x)\ d\lambda(x)$. Cependant

$$\int_{[a,b]} f_n(x)\ d\lambda(x) = \frac{b-a}{n} \sum_{k=0}^{n-1} f(a + k\frac{b-a}{n}).$$

On reconnait une somme de Riemann qui converge vers $\int_a^b f(x)\ dx$. Finalement, $\int_{[a,b]} f(x)\ d\lambda(x) = \int_a^b f(t)\ dt$. □

Théorème 35. *Soit f une fonction positive continue par morceaux sur un intervalle ouvert $[a, b[$ (b peut valoir $+\infty$). Alors l'intégrale impropre $\int_a^b f(t)\ dt$ est convergente si et seulement si $\int_{[a,b[} f(x)\ d\lambda(x) < +\infty$. Dans ce cas*

$$\int_{[a,b[} f(x)\ d\lambda(x) = \int_a^b f(t)\ dt.$$

Démonstration. Soit b_n une suite de réels de $[a, b[$ tendant vers b. $f\mathbb{1}_{[a,b_n]}$ converge en croissant vers $f\mathbb{1}_{[a,b[}$, donc d'après le théorème de convergence monotone $\int_{[a,b[} f(x)\ d\lambda(x) < +\infty$ est la limite de $\int_{[a,b_n]} f(x)\ d\lambda(x) < +\infty$. Par définition d'une intégrale impropre, $\int_a^b f(t)\ dt$ est la limite de $\int_a^{b_n} f(t)\ dt$, si elle est finie. Comme $\int_{[a,b_n]} f(x)\ d\lambda(x) = \int_a^{b_n} f(t)\ dt$, le résultat s'ensuit. □

Théorème 36. *Soit f une fonction continue par morceaux sur un intervalle ouvert $[a, b[$ (b peut valoir $+\infty$). Alors f est intégrable sur $[a, b[$ par rapport à la mesure de Lebesgue si et seulement si l'intégrale impropre $\int_a^b |f(t)|\ dt$ est convergente. Dans ce cas*

$$\int_{[a,b[} f(x)\ d\lambda(x) = \int_a^b f(t)\ dt.$$

Démonstration. Dire que f est intégrable sur $[a, b[$ par rapport à la mesure de Lebesgue, c'est dire que $\int_{[a,b[} f(x)\ d\lambda(x) < +\infty$. Le premier point découle donc du théorème précédent. Soit b_n une suite de réels de $[a, b[$ tendant vers b. $f\mathbb{1}_{[a,b_n]}$ converge vers $f\mathbb{1}_{[a,b[}$ et $|f\mathbb{1}_{[a,b_n]}| \leq |f|$, donc d'après le théorème de convergence dominée $\int_{[a,b[} f(x)\ d\lambda(x) < +\infty$ est la limite de $\int_{[a,b_n]} f(x)\ d\lambda(x)$, c'est à dire la limite de $\int_a^{b_n} f(x)\ dx$, qui est $\int_a^b f(x)\ dx$, par définition d'une intégrale impropre. □

Il est important de remarquer que la convergence de l'intégrale impropre $\int_a^b f(x)\, dx$ n'entraîne PAS l'intégrabilité de f.

Ainsi, on verra en exercice que l'intégrale de 0 à $+\infty$ de $\frac{\sin x}{x}$ est une intégrale de Riemann impropre convergente ; cependant $\frac{\sin x}{x}$ n'est pas intégrable sur $\mathbb{R}_+$ pour la mesure de Lebesgue.

Pour terminer, quelques remarques élémentaires : l'intérêt de l'intégrale de Riemann, c'est que l'on sait la calculer !

En particulier grâce au théorème fondamental de l'analyse : si F est une primitive de f sur $[a, b]$ (c'est à dire si $F' = f$), alors

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

En particulier, si φ est une fonction monotone strictement croissante, la dérivée de $F \circ \varphi$ est $\varphi'.(f \circ \varphi)$, ce qui nous dit que $F \circ \varphi$ est une primitive de $\varphi'.(f \circ \varphi)$, et donc

$$\int_a^b \varphi'(x).(f \circ \varphi)(x)\, dx = F(\varphi(b)) - F(\varphi(a)) = \int_{\varphi(a)}^{\varphi(b)} f(x)\, dx.$$

C'est la formule dite "de changement de variable".

4.5 Applications aux intégrales à paramètre

4.5.1 Continuité d'une intégrale dépendant d'un paramètre

Théorème 37. *Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré. Soit $f(x, t)$ une fonction de deux variables définie sur $\Omega \times T$, où T est un espace métrique. On suppose que pour tout $t \in T$, la fonction $x \mapsto f(x, t)$ est mesurable par rapport à $\mathcal{F}$. On suppose qu'il existe une fonction g intégrable par rapport à μ telle que pour tout $t \in T$.*

$$|f(x, t)| \leq g(x)\ \mu - p.p.$$

On suppose enfin que pour μ-presque tout x $t \mapsto f(x, t)$ est continue..

Alors

$$F(t) = \int_\Omega f(x, t)\, d\mu(x)$$

définit une fonction continue sur T

Démonstration. Que F soit bien définie découle de l'inégalité $|f(x, t)| \leq g(x)$ et de l'intégrabilité de g. Soit $t \in T$. Pour montrer que F est continue en

$t \in T$, il suffit de montrer que pour toute suite $(t_n)_{n\geq 1}$ d'éléments de T, $F(t_n)$ tend vers $F(t)$. Pour cela, il suffit d'appliquer le théorème de convergence dominée à la suite f_n définie par $f_n(x) = f(x, t_n)$: pour μ-presque tout x, $f_n(x)$ converge vers $f(x,t)$ grâce à la continuité de $t \mapsto f(x,t)$ et on a la domination $|f_n| \leq g$. □

4.5.2 Dérivabilité d'une intégrale dépendant d'un paramètre

Théorème 38. *Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré. Soit $f(x,t)$ une fonction de deux variables définie sur $\Omega \times T$, où T est un intervalle de $\mathbb{R}$. On suppose que pour tout $t \in T$, la fonction $x \mapsto f(x,t)$ est mesurable par rapport à $\mathcal{F}$ et intégrable par rapport à μ. On suppose enfin que pour μ-presque tout x $t \mapsto f(x,t)$ est dérivable par rapport à t, et qu'il existe une fonction g intégrable par rapport à μ telle que pour tout $t \in T$.*

$$|\frac{\partial f}{\partial t}(x,t)| \leq g(x) \ \mu - p.p.$$

Alors

$$F(t) = \int_\Omega f(x,t)\ d\mu(x)$$

définit une fonction de classe C^1 sur T, avec

$$F'(t) = \int_\Omega \frac{\partial f}{\partial t}(x,t)\ d\mu(x)$$

Démonstration. Il s'agit de montrer que pour toute suite t_n de points de T tendant vers t,

$$\lim_{n\to+\infty} \frac{F(t_n) - F(t)}{t_n - t} = \int_\Omega \frac{\partial f}{\partial t}(x,t)\ d\mu(x).$$

Posons $f_n(x) = \frac{f(x,t_n) - f(x,t)}{t_n - t}$. D'après l'inégalité des accroissements finis, on a $|f_n(x)| \leq g(x)$ μ presque partout. Ainsi, d'après le théorème de convergence dominée, $\int_\Omega f_n\ d\mu(x)$ converge vers $\int_\Omega \frac{\partial f}{\partial t}(x,t)\ d\mu(x)$, cependant $\int_\Omega f_n\ d\mu(x) = \frac{F(t_n) - F(t)}{t_n - t}$, ce qui donne donc le résultat voulu. □

Remarque importante : lorsque l'on veut démontrer la continuité (ou la dérivabilité) de $F(t)$ défini comme précédemment sur un intervalle T non compact, il est rare que l'on trouve une fonction majorante g qui convienne pour toutes les valeurs de T. Cependant, comme la continuité (ou la dérivabilité) est une propriété locale, il suffit de montrer que pour tout $t \in T$, il existe un voisinage V de t tel l'on ait une majoration uniforme pour les $t \in V$.

4.5.3 Holomorphie d'une intégrale dépendant d'un paramètre

Cette sous-section peut être sautée sans dommage par un lecteur ignorant l'analyse complexe.

Théorème 39. *Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré et O un ouvert de $\mathbb{C}$. Soit $f(x, z)$ une fonction de deux variables définie sur $\Omega \times O$. On suppose que pour tout $z \in O$, la fonction $x \mapsto f(x, z)$ est mesurable par rapport à $\mathcal{F}$. On suppose que pour tout compact K inclus dans O, il existe une fonction g_K intégrable par rapport à μ telle que pour tout $z \in K$.*

$$|f(x, z)| \le g_K(x) \ \mu - p.p.$$

On suppose enfin que pour μ-presque tout x $z \mapsto f(x, z)$ est homomorphe.

Alors

$$F(z) = \int_\Omega f(x, z) \ d\mu(x)$$

définit une fonction holomorphe sur O avec

$$\forall n \ge 1 \quad F^{(n)}(z) = \int_\Omega \frac{\partial^n f}{\partial^n z}(x, z) \ d\mu(x)$$

Remarquons que le contrôle repose sur $f(x, z)$, non pas sur sa dérivée. Si l'on laisse de côté l'argument standard de localisation, la preuve ressemble beaucoup à la preuve du théorème précédent, mais il y a un petit miracle lié à l'analycité : grâce aux inégalités de Cauchy, majorer localement $f(x, z)$ permet de majorer localement $\frac{\partial f}{\partial z}(x, z)$.

Démonstration. Comme annoncé, on va commencer par un argument d'analyse complexe : on va montrer que pour tout $n \ge 1$ et pour tout compact K inclus dans O, il existe une fonction $g_{n,K}$ intégrable par rapport à μ telle que

$$\forall z \in K \quad |\frac{\partial^n f}{\partial^n z}(x, z)| \le g_{n,K}(x) \ \mu - \text{p.p.}$$

Vu ce résultat, il suffira alors de montrer la formule pour $n = 1$, le résultat général venant aisément par récurrence.

Soit K un compact, $n \ge 1$. Un raisonnement classique de compacité donne l'existence d'un $r > 0$ tel que $K + B_F(0, r) \subset O$. Notons que $K + B_F(0, r) \subset O$ est également un compact. Pour tout z dans K, on a, pour μ-presque tout x, l'identité

$$\frac{\partial^n}{\partial^n z} f(x, z) = \frac{1}{2i\pi} \int_{C(z,r)} \frac{f(x, z')}{z'^{n+1}} \ dz' = \frac{1}{2\pi r^n} \int_0^{2\pi} f(x, z + re^{i\theta}) e^{-in\theta} \ d\theta.$$

Ainsi $|\frac{\partial^n}{\partial^n z} f(x,z)| \le \frac{1}{r^n} g_{K+B(0,r)}(x)$, ce qui donne le résultat voulu avec $g_{n,K} = r^{-n} g_{K+B(0,r)}$.

Passons maintenant à la preuve de l'identité et de la formule pour $n = 1$. Soit $z_0 \in O$. Prenons r tel que la boule fermée de centre z_0 et de rayon r soit inclue dans O. On prend $K = B_F(z_0, r)$.

Posons $F_{\theta,x}(r) = f(x, z_0 + re^{i\theta})$. On a, pour μ-presque tout x, l'égalité (vectorielle)

$$F_{\theta,x}(r) - F_{\theta,x}(0) = \int_0^r F'_{\theta,x}(u)\ du,$$

soit

$$\frac{f(x, z_0 + re^{i\theta}, x) - f(x, z_0)}{re^{i\theta}} = \frac{1}{r}\int_0^r \frac{\partial}{\partial z} f(x, z_0 + ue^{i\theta})\ du.$$

Ainsi pour tout z tel que $|z - z_0| \le r$, on a

$$|\frac{f(x,z) - f(x,z_0)}{z - z_0}| \le \sup_{z \in B_F(z_0,r)} |\frac{\partial}{\partial z} f(x,z)| \le g_{1,B(z_0,r)}(x)\ \mu - \text{p.p.}$$

On conclut alors comme précédemment avec le théorème de convergence dominée et une suite (z_n) quelconque de limite z_0 (à partir d'un certain rang, elle prend ses valeurs dans K). On peut remarquer que la fin de la preuve est presque identique à la preuve du théorème de dérivation sous le signe somme, à la différence près qu'on a redémontré à la main l'inégalité des accroissements dans le cadre du $\mathbb{R}$ espace vectoriel $\mathbb{C}$.

□

Exemple : posons $\varphi(z) = \int_{\mathbb{R}_+} \frac{e^{-x}}{x-z}\ d\lambda(x)$ et montrons que φ est holomorphe sur $\mathbb{C}\backslash\mathbb{R}_+$. Soit K un compact de $\mathbb{C}\backslash\mathbb{R}_+$: la fonction $z \mapsto d(z, \mathbb{R}_+)$ est continue sur K, donc y atteint son minimum, que l'on note ε_K. $\varepsilon_K > 0$ car K ne rencontre pas $\mathbb{R}_+$. On peut donc appliquer le théorème avec $g_K(x) = \frac{e^{-x}}{\varepsilon_K}$.

4.6 Mesures à densité

Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré Soit f une fonction positive mesurable de $(\Omega, \mathcal{F})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On peut définir une application ν de $(\Omega, \mathcal{F})$ dans $[0, +\infty]$ par

$$\nu(A) = \int_A f\ d\mu.$$

Il n'est pas difficile de démontrer que ν est une mesure (exercice laissé au lecteur)

On dit que ν est une mesure qui admet une densité par rapport à μ et que cette densité est f.

En réalité il y a ici au moins un abus de langage : en effet, une même mesure ne peut elle admettre plusieurs densités par rapport à μ ?

Proposition 2. *Soit f et g deux fonctions mesurables étant toutes deux des densités de ν par rapport à μ. Alors $f = g$ μ presque partout.*

Démonstration. Posons $A_+ = \{\omega : f(\omega) > g(\omega)\}$ $0 = \nu(A) - \nu(A) = \int_{A_+} f \ d\mu - \int_{A_+} g \ d\mu = \int_{A_+} (f - g) \ d\mu$. De même si l'on pose $A_- = \{\omega : f(\omega) < g(\omega)\}$ $0 = \nu(A) - \nu(A) = \int_{A_-} f \ d\mu - \int_{A_-} g \ d\mu = \int_{A_-} (f - g) \ d\mu$. Cependant $|f - g| = (f - g)1\!\!1_{A_+} - (f - g)1\!\!1_{A_-}$, donc

$$\begin{aligned}\int |f-g| \ d\mu &= \int (f-g)1\!\!1_{A_+} \ d\mu - \int (f-g)1\!\!1_{A_-} \ d\mu \\ &= \int_{A_+} (f-g) \ d\mu - \int_{A_-} (f-g) \ d\mu \\ &= 0 - 0 = 0\end{aligned}$$

Ce qui implique que $f = g$ μ presque partout. □

Théorème 40. *On suppose que ν est une mesure qui admet une densité f par rapport à μ. Alors, pour toute fonction mesurable g*

$$\int |g| \ d\nu = \int |g| f \ d\mu. \tag{4.2}$$

Si cette quantité est finie, on a alors

$$\int g \ d\nu = \int g f \ d\mu. \tag{4.3}$$

Démonstration. Si $g = 1\!\!1_A$ avec $A \in \mathcal{F}$ (4.3) est immédiat. Par linéarité, (4.3) est également vérifié lorsque g est une fonction simple positive. En utilisant le lemme 2 et le théorème de convergence monotone, il s'ensuit que (4.3) est vraie pour toute fonction mesurable positive, donc en particulier (4.2) est vraie pour toute fonction mesurable g. Supposons maintenant que $\int |g| \ d\nu = \int |g| f \ d\mu < +\infty$: on peut alors écrire $g = g_+ - g_-$ avec $\int g_+ \ d\nu < +\infty$ et $\int g_- \ d\nu < +\infty$. Comme g_+ et g_- sont des fonctions mesurables positives, on a $\int g_+ \ d\nu = \int g_+ f \ d\mu$ et $\int g_- \ d\nu = \int g_- f \ d\mu$. En faisant la différence, on obtient donc $\int (g_+ - g_-) \ d\nu = \int (g_+ - g_-) f \ d\mu$, soit (4.3). □

4.7 Intégration par rapport à une mesure image

Théorème 41. *Soit* $(\Omega, \mathcal{F}, \mu)$ *un espace mesuré,* T *une application mesurable de* $(\Omega, \mathcal{F})$ *dans* $(\Omega', \mathcal{F}')$. *Soit* f *une application mesurable de* $(\Omega', \mathcal{F}')$ *dans* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. *Alors* f *est intégrable par rapport à* μ_T *si et seulement si* $f \circ T$ *est intégrable par rapport à* μ. *Dans ce cas, on a*

$$\int_{\Omega'} f(y)\ d\mu_T(y) = \int_{\Omega} (f \circ T)(x)\ d\mu(x) \tag{4.4}$$

Démonstration. Prenons d'abord le cas où f est l'indicatrice d'un ensemble $A \in \mathcal{F}'$: on a $\int_{\Omega'} f\ d\mu_T = \int_{\Omega'} \mathbb{1}_A\ d\mu_T = \mu_T(A) = \mu(T^{-1}(A))$. D'un autre côté $\mathbb{1}_A \circ T = \mathbb{1}_{T^{-1}(A)}$, donc $\int_{\Omega} f \circ T\ d\mu = \int_{\Omega} \mathbb{1}_{T^{-1}(A)}\ d\mu = \mu(T^{-1}(A))$.L'égalité (4.4) est donc vérifiée lorsque lorsque f est l'indicatrice d'un ensemble $A \in \mathcal{F}'$. Par linéarité, elle est donc vérifiée pour toute fonction étagée mesurable.

Soit maintenant f une application mesurable positive de $(\Omega', \mathcal{F}')$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Il existe une suite croissante d'applications étagées (f_n) convergeant ponctuellement vers f. Pour tout n, on a

$$\int_{\Omega'} f_n(y)\ d\mu_T(y) = \int_{\Omega} (f_n \circ T)(x)\ d\mu(x)$$

En appliquant le théorème de convergence monotone, on obtient à la limite $\int_{\Omega'} f(y)\ d\mu_T(y) = \int_{\Omega} (f \circ T)(x)\ d\mu(x)$. En particulier, pour toute application mesurable de $(\Omega', \mathcal{F}')$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on a $\int_{\Omega'} |f|\ d\mu_T = \int_{\Omega} |f| \circ T\ d\mu(x)$, ce qui montre que bien f est intégrable par rapport à μ_T si et seulement si $f \circ T$ est intégrable par rapport à μ. Dans ce cas, f^+ et f^- sont intégrables, positives, et en soustrayant l'identité $\int_{\Omega'} f^-\ d\mu_T = \int_{\Omega} f^- \circ T\ d\mu(x)$ de l'identité $\int_{\Omega'} f^+\ d\mu_T = \int_{\Omega} f^+ \circ T\ d\mu(x)$, on obtient bien le résultat voulu. □

4.8 Mesure produit

On suppose que $(X, \mathcal{X}, \mu)$ et $(Y, \mathcal{Y}, \nu)$ sont des espaces mesurés. On rappelle que $\mathcal{X} \otimes \mathcal{Y}$ est la tribu engendrée par les ensembles de type $X \times Y$, où (X, Y) décrit $\mathcal{X} \times \mathcal{Y}$.

4.8.1 Construction de la mesure produit

Lemme 3. *Pour tout* $A \in \mathcal{X} \otimes \mathcal{Y}$, $x \in X$ *et* $y \in Y$, *on note*

$$A_y(x) = \{y \in Y : (x, y) \in A\}$$

et

$$A_x(y) = \{x \in X : (x, y) \in A\}.$$

Alors $A_y(x) \in \mathcal{Y}$ et $A_x(y) \in \mathcal{X}$. De plus, si une fonction f est mesurable de la tribu $\mathcal{X} \otimes \mathcal{Y}$ vers la tribu $\mathcal{C}$, alors pour chaque x fixé la fonction $y \mapsto f(x, y)$ est mesurable par rapport à la tribu $\mathcal{Y}$, et de même pour chaque y fixé la fonction $x \mapsto f(x, y)$ est mesurable par rapport à la tribu $\mathcal{X}$.

Démonstration. On va commencer par montrer la deuxième proposition. Fixons $x \in X$ et montrons que $f^1_x : y \mapsto f(x, y)$ est $(Y, \mathcal{Y}) - (C, \mathcal{C})$ mesurable. Notons $\pi^1_x : Y \to X \times Y$ qui à y associe (x, y). π^1_x est $(Y, \mathcal{Y}) - (X \times Y, \mathcal{X} \times \mathcal{Y})$ mesurable car chacune des composantes est mesurable. Maintenant, l'identité $f^1_x = f \circ \pi^1_x$ donne la mesurabilité voulue, par composition d'applications mesurables.

Revenons à la première proposition : la section verticale de niveau x : $A_y(x) = \{y \in Y : (x, y) \in A\}$ peut s'écrire $A_y(x) = (f^1_x)^{-1}(\{1\})$, avec $f^1_x(y) = 1_A(x, y)$. Or l'application de $X \times Y$ dans $\mathbb{R}$ qui à (x, y) associe $\mathbb{1}_A(x, y)$ est $(X \times Y, \mathcal{X} \times \mathcal{Y}) - (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable, donc d'après ce qui précède, f^1_x est $(Y, \mathcal{Y}) - (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable, d'où $A_y(x) \in \mathcal{Y}$.

On traite de la même manière $A_x(y)$ et $f^2_y : x \mapsto f(x, y)$. □

Théorème 42. *Soit $(X, \mathcal{X}, \mu)$ et $(Y, \mathcal{Y}, \nu)$ deux espaces mesurés dont les mesures μ et ν sont σ-finies. Il existe une unique mesure m sur $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$ telles que pour tous $X \in \mathcal{X}$ et $Y \in \mathcal{Y}$, on ait*

$$m(X \times Y) = \mu(X)\nu(Y).$$

On notera dans la suite $\mu \otimes \nu$ cette mesure. De plus, pour tout $E \in X \otimes \mathcal{Y}$, les fonctions $x \mapsto \nu(E_y(x))$ et $y \mapsto \mu(E_x(y))$ sont mesurables et l'on a

$$\int_X \nu(E_y(x))\ d\mu(x) = \int_Y \mu(E_x(y))\ d\nu(y) = (\mu \otimes \nu)(E).$$

Démonstration. Supposons d'abord que μ et ν sont finies. Soit E dans $\mathcal{X} \otimes \mathcal{Y}$; d'après le lemme précédent la fonction $x \mapsto \nu(E_y(x))$ est bien définie. Notons $\mathcal{T}'$ la famille des ensembles $E \in \mathcal{X} \otimes \mathcal{Y}$ tels que cette fonction soit mesurable de $\mathcal{X}$ dans $\mathcal{B}(\mathbb{R})$. Il n'est pas difficile de voir que $\mathcal{T}'$ est un λ-système (voir la dernière section du chapitre 3). Mais $\mathcal{T}'$ contient tous les pavés (les éléments de $\mathcal{X} \times \mathcal{Y}$) : en effet, soit prenons $E = A \times B$, avec $A \in \mathcal{X}$ et $B \in \mathcal{Y}$: on a $E_y(x) = \{y \in B(x, y) \in A \times B\}$. Ainsi $E_y(x) = B$ si $x \in A$ et $\varnothing$ sinon, et donc $\nu(E_y(x)) = \nu(B)$ si $x \in A$ et 0 sinon. Ainsi $\nu(E_y(x)) = \nu(B)\mathbb{1}_A(x)$, et $x \mapsto \nu(B)\mathbb{1}_A(x)$ est bien une fonction mesurable de $\mathcal{X}$ dans $\mathcal{B}(\mathbb{R})$. $\mathcal{T}'$ est donc un λ-système qui contient un π-système qui engendre $\mathcal{X} \otimes \mathcal{Y}$. Ainsi, d'après le théorème λ-π, $\mathcal{T}'$ est $\mathcal{X} \otimes \mathcal{Y}$ tout entier. Finalement, pour tout E dans $\mathcal{X} \otimes \mathcal{Y}$, on peut définir

$$m_1(E) = \int_X \nu(E_y(x))\, d\mu(x);$$

et de même on pourrait définir

$$m_2(E) = \int_Y \mu(E_x(y))\, d\nu(y).$$

Prenons à nouveau $E = A \times B$: $E_y(x) = \{y \in B (x, y) \in A \times B\}$. Ainsi $E_y(x) = B$ si $x \in A$ et $\varnothing$ sinon, et donc $\nu(E_y(x)) = \nu(B)$ si $x \in A$ et 0 sinon. Ainsi $m_1(E) = \int_X \mathbb{1}_A \nu(B)\, d\mu = \mu(A)\nu(B)$. En procédant de la même manière, on obtient $m_2(E) = \int_Y \mathbb{1}_B \mu(A)\, d\nu = \mu(A)\nu(B)$. m_1 et m_2 sont donc des mesures finies qui coïncident sur les pavés : elles sont donc égales.

Passons maintenant au cas où μ et ν sont σ-finies : on peut partitionner X (et Y) en une famille dénombrable d'ensembles de mesure finie : $X = \cup_{i=1}^{\infty} A_i$ et $Y = \cup_{j=1}^{\infty} B_j$. On peut alors noter $m^{i,j}$ la mesure associée comme précédemment aux mesures traces $\nu_{|A_i}$ et $\mu_{|B_j}$. En d'autres termes

$$m^{i,j}(E) = \int_X \nu_{|A_i}(E_y(x))\, d\mu_{|B_j}(x) = \int_Y \mu_{|B_j}(E_x(y))\, d\nu_{|A_i}(x)$$

Alors, il n'est pas difficile de voir que la mesure $m = \sum_i \sum_j m^{i,j}$. On a alors

$$\begin{aligned} m(A \times B) &= \sum_i \sum_j m^{i,j}(A \times B \cap (A_i \times B_j)) \\ &= \sum_i \sum_j m^{i,j}((A \cap A_i) \times (B \cap B_j)) \\ &= \sum_i \sum_j \mu(A \cap A_i)\nu(B \cap B_j) \\ &= (\sum_i \mu(A \cap A_i))(\sum_j \mu(B \cap B_j)) \\ &= \mu(A)\nu(B) \end{aligned}$$

□

Remarque : il n'est pas difficile de voir que si μ, ν sont des mesures σ-finies, a et b des réels, alors $(a\mu) \otimes (b\nu) = (ab)(\mu \otimes \nu)$ – utiliser la partie unicité du théorème.

Exemple : soit $(X, \mathcal{X}, \mu)$ un espace mesuré avec μ σ-finie, f une application mesurable de $(X, \mathcal{X})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On note T l'application de $X \times \mathbb{R}$ dans lui-même qui à (x, y) associe $(x, y + f(x))$. Alors, T est une application mesurable qui laisse invariante la mesure $\mu \otimes \lambda$.

Démonstration. T est mesurable car ses composantes sont mesurables. Notons $m = \mu \otimes \lambda$. Il suffit de montrer que pour tout $A \in \mathcal{X}$ et tout borélien de $\mathbb{R}$ B, l'ensemble $E = A \times B$ vérifie $m(E) = m(T^{-1}(E))$. On a

$$m(E) = \int_X \lambda(E_y(x)) \, d\mu(x) \text{ et } m(T^{-1}E) = \int_X \lambda((T^{-1}(E))_y(x)) \, d\mu(x).$$

Comme on l'a déjà vu, $E_y(x) = B$ si $x \in A$, $\varnothing$, sinon, de sorte que $\lambda(E_y(x)) = \lambda(B)\mathbb{1}_A(x)$. Cependant,

$$(T^{-1}(E))_y(x) = \{y \in \mathbb{R}; (x, y) \in T^{-1}(E)\} = \{y \in \mathbb{R}; (x, y + f(x)) \in E\},$$

qui est donc égal à $B - f(x)$ si $x \in A$, zéro sinon. Ainsi, $\lambda((T^{-1}(E))_y(x)) = \lambda(B - f(x))\mathbb{1}_A(x)$. Maisi on sait que la mesure de Lebesgue est invariante par translation : $\lambda(B - f(x)) = \lambda(B)$. Il n'y a plus qu'à intégrer et on a l'égalité voulue. $\square$

4.8.2 Théorèmes de Fubini et Tonelli

Théorème 43 (Tonelli)**.** *Soit $(X, \mathcal{X}, \mu)$ et $(Y, \mathcal{Y}, \nu)$ deux espaces mesurés dont les mesures μ et ν sont σ-finies et $f \in \mathcal{V}_+(X \times Y, \mathcal{X} \otimes \mathcal{Y})$.*

Pour tout $x \in X$, la fonction $y \mapsto f(x, y)$ est mesurable de $(Y, \mathcal{Y})$ dans $(\overline{\mathbb{R}}_+, \mathcal{B}(\overline{\mathbb{R}}_+))$ et la fonction

$$x \mapsto \int_Y f(x, y) \, d\nu(y)$$

est dans $\overline{\mathcal{V}}_+(X, \mathcal{X})$.
De même pour tout $y \in Y$, la fonction $x \mapsto f(x, y)$ est mesurable de $(X, \mathcal{X})$ dans $(\overline{\mathbb{R}}_+, \mathcal{B}(\overline{\mathbb{R}}_+))$

$$y \mapsto \int_X f(x, y) \, d\mu(x)$$

est dans $\overline{\mathcal{V}}_+(Y, \mathcal{Y})$.

Enfin, on a les égalités

$$\int_{X \times Y} f \, d\mu \otimes \nu = \int_X \left(\int_Y f(x, y) \, d\nu(y) \right) \, d\mu(x)$$

et

$$\int_{X \times Y} f \, d\mu \otimes \nu = \int_Y \left(\int_X f(x, y) \, d\mu(x) \right) \, d\nu(y)$$

Démonstration. La mesurabilité de $y \mapsto f(x,y)$ est une conséquence immédiate du lemme 3.

Supposons que f s'écrive comme l'indicatrice d'un ensemble $A \in \mathcal{X} \otimes \mathcal{Y}$: on a alors

$$\int_Y f(x,y)\ d\nu(y) = \nu(A_x)$$

D'après la deuxième partie du Théorème 42, l'application $x \mapsto \int_Y f(x,y)\,d\nu(y)$ est donc mesurable et l'on a

$$\begin{aligned} \int_X \int_Y f(x,y)\ d\nu(y)\ d\mu(x) &= \int_X \nu(A_x)\ d\mu(x) \\ &= \mu \otimes \nu(A) \\ &= \int_{X\times Y} f\ d\mu \otimes \nu. \end{aligned}$$

Le résultat s'étend aisément à la classe des fonctions simples par linéarité, puis à $f \in \mathcal{V}_+(X \times Y, \mathcal{X} \otimes \mathcal{Y})$ en utilisant le lemme 2 et le théorème de convergence monotone. □

Théorème 44 (Fubini)**.** *Soit $(X,\mathcal{X},\mu)$ et $(Y,\mathcal{Y},\nu)$ deux espaces mesurés dont les mesures μ et ν sont σ-finies et $f \in \mathcal{V}(\mathcal{X} \times \mathcal{Y}, \mathcal{X} \otimes \mathcal{Y})$. On suppose que*

$$\int_{X\times Y} |f|\ d\mu \otimes \nu < +\infty.$$

Alors, il existe $X' \in \mathcal{X}$ et $Y' \in \mathcal{Y}$ avec $\mu(X\backslash X') = \nu(Y\backslash Y') = 0$ et

$$\int_{X\times Y} f\ d\mu \otimes \nu = \int_{X'} \left(\int_Y f(x,y)\ d\nu(y) \right)\ d\mu(x)$$

et

$$\int_{X\times Y} f\ d\mu \otimes \nu = \int_{Y'} \left(\int_X f(x,y)\ d\mu(x) \right)\ d\nu(y)$$

Démonstration. On va juste montrer la première égalité. D'après le théorème de Tonelli,

$$\int_X \left(\int_Y |f(x,y)|\ d\nu(y) \right)\ d\mu(x) = \int_{X\times Y} |f|\ d\mu \otimes \nu < +\infty$$

Il s'ensuit que si l'on pose

$$X' = \{x \in X; \int_Y |f(x,y)|\ d\nu(y) < +\infty\},$$

on a $\mu(X\backslash X')=0$.
Par suite $\mu\otimes\nu(X\times Y\backslash X'\times Y)=\mu(X\backslash X')\nu(Y)=0$. (On rappelle que dans $\overline{\mathbb{R}}_+$, $0.\infty=0$.)
Ainsi, comme l'hypothèse $\int_{X\times Y}|f|\ d\mu\otimes\nu<+\infty$ entraîne l'existence de $\int_{X\times Y} f\ d\mu\otimes\nu$, on peut écrire

$$\begin{aligned}
\int_{X\times Y} f\ d\mu\otimes\nu &= \int_{X'\times Y} f\ d\mu\otimes\nu \\
&= \int_{X'\times Y} (f^+-f^-)\ d\mu\otimes\nu \\
&= \int_{X'\times Y} f^+\ d\mu\otimes\nu - \int_{X'\times Y} f^-\ d\mu\otimes\nu \\
&= \int_{X'}\left(\int_Y f^+(x,y)\ d\nu(y)\right)\ d\mu(x) - \int_{X'}\left(\int_Y f^-(x,y)\ d\nu(y)\right)\ d\mu(x) \\
&= \int_{X'}\left(\int_Y f^+(x,y)\ d\nu(y)\right) - \left(\int_Y f^-(x,y)\ d\nu(y)\right)\ d\mu(x) \\
&= \int_{X'}\left(\int_Y f(x,y)\ d\nu(y)\right)\ d\mu(x)
\end{aligned}$$

□

4.8.3 Associativité de la mesure produit

Soient $(X,\mathcal{X},\mu),(Y,\mathcal{Y},\nu),(Z,\mathcal{Z},\chi)$ trois espaces mesurés σ-finis. Comme précédemment, on note $\mathcal{X}\otimes\mathcal{Y}\otimes\mathcal{Z}$ la tribu engendrée par les ensembles de la forme $A\times B\times C$, où (A,B,C) décrit $\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$.

On note φ l'application de $(X\times Y)\times Z\to X\times Y\times Z : ((x,y),z)\mapsto(x,y,z)$ et ψ l'application de $X\times(Y\times Z)\to X\times Y\times Z : (x,(y,z))\mapsto(x,y,z)$. Alors la mesure image m_1 de $(\mu\otimes\nu)\otimes\chi$ par φ et la mesure image M_2 de $\mu\otimes(\nu\otimes\chi)$ par ψ sont égales : on note simplement cette mesure $\mu\otimes\nu\otimes\chi$.

Démonstration.

$$\begin{aligned}
m_1(A\times B\times C) &= (\mu\otimes\nu)\otimes\chi(\varphi^{-1}(A\times B\times C)) \\
&= (\mu\otimes\nu)\otimes\chi((A\times B)\times C) \\
&= \mu\otimes\nu(A\times B)\chi(C) \\
&= \mu(A)\nu(B)\chi(C)
\end{aligned}$$

$$\begin{aligned} m_2(A \times B \times C) &= \mu \otimes (\nu \otimes \chi)(\psi^{-1}(A \times B \times C)) \\ &= \mu \otimes (\nu \otimes \chi)(A \times (B \times C)) \\ &= \mu(A)(\nu \otimes \chi)(B \times C) \\ &= \mu(A)\nu(B)\chi(C) \end{aligned}$$

Ainsi, les mesures coïncident. □

4.8.4 Convolution de mesures

Soit μ et ν deux mesures σ-finies sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. On appelle convolée de μ et ν et on note $\mu * \nu$ la mesure image de $\mu \otimes \nu$ par l'application $(x, y) \mapsto x + y$.

On se contente pour l'instant dénoncer quelques propriétés simples : si μ, ν sont des mesures σ-finies, a et b des réels, alors

- $(a\mu) * b(\nu) = (ab)(\mu * \nu)$.
- $\mu * 0 = 0 * \mu = 0$.

Démonstration. Soit $f : (x, y) \mapsto x + y$.

$$\begin{aligned} (a\mu) * b(\nu)(A) &= ((a\mu) \otimes b(\nu))(f^{-1}(A)) \\ &= ab(\mu \otimes \nu)(f^{-1}(A)) \\ &= ab(\mu * \nu)(A) \end{aligned}$$

$\mu * 0(A) = (\mu \otimes 0)(f^{-1}(A)) = 0$ et de même $0 * \mu(A) = (0 \otimes \mu)(f^{-1}(A)) = 0$ □

4.9 Les théorèmes généraux et la mesure de comptage

Un certain nombre de théorèmes généraux donnent des résultats très pratiques lorsqu'ils sont utilisé avec la mesure de comptage. On va juste en donner deux, mais le lecteur aura intérêt à relire chaque théorème en se demandant quel résultat on obtient lorsqu'on prend pour une (ou toutes) les mesures en jeu la mesure de comptage. Bien sûr, il retrouvera parfois des résultats connus.

Théorème 45. *Soit donnée une famille de nombres réels $a(k,n)$ pour $k \geq 1, n \geq 1$ entiers. On suppose qu'il existe une suite de nombres réels positifs $(c_k)_{k\geq 1}$ avec les propriétés :*

$$\forall (n,k) \in \mathbb{N}^* \times \mathbb{N}^* \quad |a(k,n)| \leq c_k, \qquad \sum_{k=1}^{+\infty} c_k < +\infty.$$

On suppose que pour tout $k \geq 1$, la limite suivante existe :

$$\lim_{n\to+\infty} a(k,n) = a(k,\infty).$$

Montrer que les deux séries $s_n = \sum_{k=1}^{+\infty} a(k,n)$ et $s = \sum_{k=1}^{+\infty} a(k,\infty)$ convergent absolument et que l'on a $\lim_{n\to+\infty} s_n = s$, soit

$$\lim_{n\to+\infty} \sum_{k=1}^{+\infty} a(k,n) = \sum_{k=1}^{+\infty} a(k,\infty).$$

Démonstration. Ici, il s'agit d'appliquer le théorème de convergence dominée à la mesure de comptage. □

Théorème 46. *Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré, $(f_n)_{n\geq 1}$ une suite de fonctions mesurables positives de $(\Omega, \mathcal{F}, \mu)$ dans $(\overline{\mathbb{R}}_+, \mathcal{B}(\overline{\mathbb{R}}_+))$. On pose $f = \sum_{n=1}^{+\infty}$. Alors*

$$\int_\Omega f \, d\mu = \sum_{n=1}^{+\infty} \int_\Omega f_n \, d\mu.$$

Démonstration. On peut, au choix, appliquer le théorème de Tonelli à $f(n,x) = f_n(x)$ en intégrant sur $\mathbb{N}^* \times C$, ou encore appliquer le théorème de convergence monotone aux sommes partielles. □

Remarque : une conséquence de ce théorème, c'est que si la série de terme général $\int_\Omega f_n \, d\mu$ converge, alors f est intégrable et en particulier $f(\omega)$ est fini pour μ-presque tout ω. En particulier, pour une suite f_n de fonctions mesurables de signe quelconque, si $\int_\Omega |f_n| \, d\mu$ converge, la série de terme général $f_n(\omega)$ converge (absolument) pour μ-presque tout ω.

4.10 La mesure de Lebesgue sur $\mathbb{R}^d$.

On appelle mesure de Lebesgue sur $\mathbb{R}^d$ la mesure $\lambda^{\otimes d}$. On la note parfois λ^d, parfois même λ lorsqu'il n'y a pas de confusion possible (mais ce n'est pas une très bonne idée quand on débute).

4.10.1 Transformations affines

Théorème 47. *Soit $y \in \mathbb{R}^d$. La translation $x \mapsto x + y$ laisse invariant la mesure de Lebesgue sur $\mathbb{R}^d$.*

Démonstration. Il suffit de vérifier l'invariance pour un pavé, ce qui est immédiat. □

Théorème 48. *Soit $M \in SL_d(\mathbb{R})$ (c'est à dire que $\det M = 1$). L'application $x \mapsto Mx$ laisse invariant la mesure de Lebesgue sur $\mathbb{R}^d$.*

Démonstration. Un théorème d'algèbre linéaire dit que tout élément de $SL_d(\mathbb{R})$ peut s'écrire comme produit de matrices de transvections, c'est à dire de matrices de la forme

$$\begin{bmatrix} 1 & & O & \\ & 1 & \lambda & \\ & & \cdot & \\ & O & 1 & \\ & & & 1 \end{bmatrix} = I_n + \lambda E_{ij} \text{ avec } i \neq j,$$

où E_{ij} est la matrice donc tous les coefficients sont nul, sauf celui en (i, j) qui vaut 1. Ainsi, il suffit de montrer le résultat pour une matrice de transvection. Mais dans ce cas, c'est un cas particulier de l'exemple vu plus haut. □

Théorème 49. *Soit $M \in M_d(\mathbb{R})$. Pour tout borélien A, on a $\lambda^d(A) = |\det M|\lambda^d(A)$.*

Démonstration. Dans le cas où $M = \operatorname{diag}(\lambda, 1, \dots, 1)$, on vérifie facilement la formule lorsque A est un pavé : on a deux mesures qui coïncident sur un π-système qui engendre la tribu, elles sont donc égales. Passons au cas où M est inversible : on peut écrire $M = \operatorname{diag}(\det M, 1, \dots, 1)N$, et $N \in SL_d(\mathbb{R})$. Ainsi

$$\begin{aligned} \lambda^d(MA) &= \lambda^d(\operatorname{diag}(\det M, 1, \dots, 1)NA) = |\det M|\lambda^d(NA) \\ &= |\det M|\lambda^d(N^{-1}(NA)) = |\det M|\lambda^d(A). \end{aligned}$$

Reste le cas où M est non-inversible, dans ce cas $\det M = 0$, donc il faut montrer que $\lambda^d(MA) = 0$, pour cela il suffit de montrer que $\lambda^d(\text{Im } M) = 0$. Or $\text{Im } M$ est un espace vectoriel de dimension $n - 1$, il existe donc une application inversible qui envoie $\text{Im } M$ dans $\mathbb{R}^{d-1} \times \{0\}$. Comme $\mathbb{R}^{n-1} \times \{0\}$ est de mesure nulle, $\text{Im } M$ aussi. □

Corollaire 6. *Si M est inversible, la mesure image de λ^d par $x \mapsto Mx + b$ est $\frac{1}{|\det M|}\lambda^d$.*

Corollaire 7. *Soit M inversible et $b \in \mathbb{R}^d$. On pose $T(x) = Mx + b$.*

Soit maintenant μ_1 une mesure positive sur $\mathbb{R}^d$ admettant une densité f_1 par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. Alors, la mesure image de μ_1 par T admet comme densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ la fonction f_2 définie par

$$f_2(y) = \frac{1}{|\det M|} f_1(T^{-1}(y)).$$

Démonstration. Soit g une fonction mesurable positive sur O_2. Notons μ_2 la mesure image de μ_1 par T. D'après le théorème de transfert,

$$\begin{aligned}
\int_{\mathbb{R}^d} g\ d\mu_2 &= \int_{\mathbb{R}^d} (g \circ T)\ d\mu_1 \\
&= \int_{\mathbb{R}^d} (g \circ T) f_1\ d\lambda^d \\
&= \int_{\mathbb{R}^d} (g \circ T)(f_1 \circ T^{-1} \circ T)\ d\lambda^d \\
&= \int_{\mathbb{R}^d} (g \times f_1 \circ T^{-1}) \circ T\ d\lambda^d \\
&= \int_{\mathbb{R}^d} (g \times f_1 \circ T^{-1})\ d\lambda_T^d \\
&= \int_{\mathbb{R}^d} g \times (f_1 \circ T^{-1}) \frac{1}{|\det M|}\ d\lambda
\end{aligned}$$

ce qui donne le résultat voulu. □

Les théorèmes qui précèdent correspondent à des transformations affines, ou si l'on préfère, à des changement de variables affines. On va maintenant voir le cas général.

4.10.2 Changement de variables C^1

Théorème 50. *Soit U, U' deux ouverts de $\mathbb{R}^d$, φ un C^1-difféomorphisme de U dans U'. Soit f une application mesurable sur U'. f est intégrable sur U' si et seulement si $f(.) \times |\det D.\ \varphi|$ est intégrable sur U et dans ce cas*

$$\int_{U'} f(y)\ d\lambda(y) = \int_U f(\varphi(x)) \times |\det D_x\ \varphi|\ d\lambda(x).$$

Ce théorème est admis

Corollaire 8. *Soit O_1 et O_2 deux ouverts de $\mathbb{R}^d, d \geq 1$. On suppose que T est un C^1-difféomorphisme de O_1 dans O_2. Soit maintenant μ_1 une mesure positive sur $\mathbb{R}^d$ telle que $\mu_1(\mathbb{R}^d \backslash O_1) = 0$ et admettant une densité f_1 par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. Alors, la mesure image de μ_1 par T admet comme densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ la fonction f_2 définie par*

$$f_2(y) = \begin{cases} f_1(T^{-1}(y))|\det DT_y^{-1}| \text{ si } y \in O_2 \\ 0 \text{ si } y \notin O_2 \end{cases}$$

Démonstration. Soit g une fonction mesurable positive sur O_2. Notons μ_2 la mesure image de μ_1 par T. D'après le théorème de transfert,

$$\begin{aligned}
\int_{O_2} g \, d\mu_2 &= \int_{O_1} (g \circ T) \, d\mu_1 \\
&= \int_{O_1} (g \circ T) f_1 \, d\lambda \\
&= \int_{O_1} (g \circ T) f_1 \, |\det D_{T(x)}.T^{-1}|| \det D_x T| \, d\lambda \\
&= \int_{O_1} (g \times (f_1 \circ T^{-1}) \times |\det D_. T^{-1}|) \circ T \, |\det D_x T| \, d\lambda \\
&= \int_{O_2} g \times (f_1 \circ T^{-1}) \times |\det D_. T^{-1}| \, d\lambda
\end{aligned}$$

ce qui donne le résultat voulu. □

4.11 Preuve des propriétés de base de l'intégrale

4.11.1 Premiers résultats

L'implication $(f \leq g) \implies \int f \, d\mu \leq \int g \, d\mu$ découle évidemment de la définition, de même que le fait que l'intégrale d'une fonction nulle est nulle.

Ce qui est assez étonnant, c'est que la linéarité ne puisse être obtenue simplement ; nous l'obtiendrons en réalité comme corollaire du lemme de Beppo Levi.

Pour $(\Omega_i)_{i \in I}$ partition finie, notons

$$I((\Omega_i)_{i \in I}, f) = \sum_i \inf\{f(\omega); \omega \in \Omega_i\} \mu(\Omega_i)$$

Il est important de remarquer que si $(\Omega_i)_{i \in I}$ est une partition finie et $(\Omega'_j)_{j \in I}$ une autre partition finie, alors

$I((\Omega_i \cap \Omega'_j)_{(i,j)\in I\times J}, f) \geq I((\Omega_i)_{i\in I}, f)$ et $I((\Omega_i \cap \Omega'_j)_{(i,j)\in I\times J}, f) \geq I((\Omega'_j)_{j\in J}, f)$.

Lemme 4. *Si une fonction est μ-presque partout nulle, son intégrale par rapport à μ est nulle.*

Démonstration. Il suffit de considérer le cas d'une fonction positive, car si f est presque partout nulle, f^+ et f^- le sont aussi. Si f positive est μ-presque partout nulle, alors pour tout i tel que $\inf\{f(\omega); \omega \in \Omega_i\} > 0$, on a $\mu(\Omega_i) \leq \mu(f \geq \inf\{f(\omega); \omega \in \Omega_i\}) \leq \mu(f > 0) = 0$. Cela en entraîne que pour toute partition finie, $I((\Omega_i)_{i\in I}, f) = 0$, d'où la nullité de l'intégrale. □

Lemme 5. *Soit f mesurable positive $\alpha > 0$ et A mesurable. Alors*

$$\int (f + \alpha 1\!\!1_A)\ d\mu = \int f\ d\mu + \alpha\mu(A).$$

Démonstration. Soit $(\Omega_i)_{i\in\Omega_i}$ une partition quelconque. Posons $J = \{1, 2\}$ avec $\Omega'_1 = A$, $\Omega'_2 = A^c$.

Comme $\inf\{f(\omega) + \alpha 1\!\!1_A(\omega); \omega \in \Omega_i \cap \Omega'_j\} = \inf\{f(\omega); \omega \in \Omega_i \cap \Omega'_j\} + \alpha 1\!\!1_{i=1}$, on a en multipliant par $\mu(\Omega_i \cap \Omega'_j)$ et en faisant la somme sur les couples (i, j) :

$$I((\Omega_i \cap \Omega'_j)_{(i,j)\in I\times J}, f + \alpha 1\!\!1_A) = I((\Omega_i \cap \Omega'_j)_{(i,j)\in I\times J}, f) + \alpha\mu(A)$$

Ainsi

$$\begin{aligned} I((\Omega_i)_{i\in I}, f + \alpha 1\!\!1_A) &\leq I((\Omega_i \cap \Omega'_j)_{(i,j)\in I\times J}, f + \alpha 1\!\!1_A) \\ &= I((\Omega_i \cap \Omega'_j)_{(i,j)\in I\times J}, f) + \alpha\mu(A) \\ &\leq \inf f\ d\mu + \alpha\mu(A), \end{aligned}$$

d'où en passant au sup :

$$\int (f + \alpha 1\!\!1_A)\ d\mu \leq \int f\ d\mu + \alpha\mu(A).$$

Réciproquement, soit $M < \int f\ d\mu$. Il existe une partition finie $(\Omega_i)_{i\in I}$, avec $I((\Omega_i)_{i\in I}, f) > M$. On a alors

$$\int (f + \alpha 1\!\!1_A)\ d\mu \geq I((\Omega_i \cap \Omega'_j)_{(i,j)\in I\times J}, f + \alpha 1\!\!1_A) = I((\Omega_i \cap \Omega'_j)_{(i,j)\in I\times J}, f) + \alpha\mu(A) > M + \mu(A).$$

En passant au sup en M, on obtient l'égalité voulue. □

En particulier, si f s'écrit comme la somme finie $f = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}$, avec pour tout i $\alpha_i \geq 0$ et A_i mesurable, on a $\int f\ d\mu = \sum_{i=1}^{n} \alpha_i \mu(A_i)$.

Ainsi, on obtient la linéarité pour les fonctions simples positives.

Remarquons également que pour f positive, comme $f \geq \alpha \mathbb{1}_{\{f \geq \alpha\}}$, on a $\int f\ d\mu \geq \alpha\mu(f \geq \alpha)$.[2] Ainsi, si f positive est intégrale nulle, on a pour tout n : $0 = \int f\ d\mu \geq 1/n\mu(f \geq 1/n)$, soit $\mu(f \geq 1/n) = 0$, et donc d'après le théorème de continuité séquentielle décroissante, $\mu(f > 0) = 0$.

Passons à la preuve du théorème de Beppo Levi.

4.11.2 Preuve du théorème de Beppo Levi

On va d'abord on montrer une forme très faible : on va montrer que si f est une fonction simple positive et $(E_n)_{n\geq 1}$ une suite croissante d'ensembles de réunion E, alors $\lim_{n\to+\infty} \int \mathbb{1}_{E_n} f\ d\mu = \int \mathbb{1}_E f\ d\mu$. En effet, si $f = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}$, avec pour tout i $\alpha_i \geq 0$, on a $f\mathbb{1}_{E_n} = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i \cap E_n}$ et

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

La convergence vers $\sum_{i=1}^{n} \alpha_i \mu(A_i \cap E) = \int \mathbb{1}_E f\ d\mu$ découle alors du théorème de continuité séquentielle croissante.

Passons au cas général. On prends f_n une suite croissante de fonctions mesurables positives tendant vers f et on veut montrer que $\inf f_n\ d\mu$ tend vers $\int f\ d\mu$. Bien sûr, la suite $\int f_n\ d\mu$ est croissante, majorée par $\int f\ d\mu$, donc la limite existe et est majorée par $\int f\ d\mu$. Il suffit donc de montrer que $\lim \int f_n\ d\mu \geq \int f\ d\mu$ Par définition de l'intégrale, il suffit de montrer que pour toute partition finie $(\Omega_i)_{i\in I}$, on a $\lim \int f_n\ d\mu \geq I((\Omega_i)_{i\in I}, f)$ Posons $g = \sum_{i\in I} \inf\{f(\omega), \omega \in \Omega_i\} \mathbb{1}_{\Omega_i}$. g est une fonction simple, avec $0 \leq g \leq f$.

Fixons $\alpha \in]0,1[$ et posons $E_n = \{f_n \geq \alpha g\}$. Comme la suite (f_n) est croissante, la suite E_n est croissante. Si $g(\omega) = 0$, on a $\omega \in E_n$ pour tout n, sinon, comme $\lim f_n(\omega) = g(\omega) > \alpha g(\omega)$, on a $\omega \in E_n$ pour n assez grand. Finalement, la réunion des E_n est Ω tout entier. Ainsi

$$\lim_{n\to+\infty} \int \mathbb{1}_{E_n}\ \alpha g\ d\mu = \int \alpha g\ d\mu.$$

On a

$$f_n \geq \mathbb{1}_{E_n} f_n \geq \mathbb{1}_{E_n} \alpha g,$$

[2] On reverra cette inégalité plus tard dans le cas où μ est une mesure de probabilité. Elle aura alors le nom d'inégalité de Markov.

d'où

$$\lim \int f_n \, d\mu \geq \lim \int \mathbb{1}_{E_n} \alpha g \, d\mu = \int \alpha g \, d\mu,$$

Soit

$$\lim \int f_n \, d\mu \geq \alpha I((\Omega_i)_{i \in I}, f)$$

En faisant tendre α vers 1, on obtient

$$\lim \int f_n \, d\mu \geq I((\Omega_i)_{i \in I}, f),$$

ce qui était l'inégalité souhaitée.

4.11.3 Preuve de la linéarité

Soit f, g deux fonctions mesurables positives, $\alpha > 0$. On veut montrer que $\int (f + \alpha g) \, d\mu = \int f \, d\mu + \alpha \int g \, d\mu$. Soit $(f_n), (g_n)$ des suites croissantes de fonctions simples positives convergeant respectivement vers f et g. On a pour tout n

$$\int (f_n + \alpha g_n) \, d\mu = \int f_n \, d\mu + \alpha \int g_n \, d\mu$$

En appliquant trois fois le théorème de Beppo Levi, on obtient à la limite $\int (f + \alpha g) \, d\mu = \int f \, d\mu + \alpha \int g \, d\mu$.

Passons au cas où f et g sont intégrables, de signe quelconque. $\int |\alpha g| \, d\mu \leq \int |\alpha| f^+ + |\alpha| f^- | \, d\mu < +\infty$ en utilisant la linéarité pour les fonctions positives. Ainsi, pour $\alpha > 0$

$$\int \alpha g \, d\mu = \int (\alpha g)^+ \, d\mu - \int (\alpha g)^- \, d\mu = \alpha \int g^+ \, d\mu - \alpha \int g^- \, d\mu = \alpha \int g \, d\mu$$

et pour $\alpha < 0$

$$\int \alpha g \, d\mu = \int (\alpha g)^+ \, d\mu - \int (\alpha g)^- \, d\mu = (-\alpha) \int g^- \, d\mu + \alpha \int g^+ \, d\mu = \alpha \int g \, d\mu$$

On est ainsi ramené à étudier le cas $\alpha = 1$. $|f + g| \leq |f| + |g|$ donc $f + g$ est intégrable. Posons $f + g = h$. On a $f^+ - f^- + g^+ - g^- = h^+ - h^-$, soit $f^+ + g^+ + h^- = f^- + g^- + h^+$. D'où en intégrant $\int f^+ + \int g^+ + \int h^- = \int f^- + \int g^- + \int h^+$. En changeant de membre, on obtient le résultat voulu.

Maintenant qu'on a la linéarité, montrons que si $f \leq g$ μ-presque partout, alors $\int f \, d\mu \leq \int g \, d\mu$: on a $f = f\mathbb{1}_{f \leq g} + f\mathbb{1}_{f > g}$ et $g = g\mathbb{1}_{f \leq g} + g\mathbb{1}_{f > g}$. $f\mathbb{1}_{f > g}$ et $g\mathbb{1}_{f > g}$ sont d'intégrales nulle, donc par linéarité $f = f\mathbb{1}_{f \leq g}$ a même intégrale que f, et $g = g\mathbb{1}_{f \leq g}$ a même intégrale que g. Comme $f\mathbb{1}_{f \leq g} \leq g\mathbb{1}_{f \leq g}$, on a en intégrant le résultat voulu.

Ceci achève la preuve des propriétés de base de l'intégrale.

4.12 Premiers exercices d'intégration

1. Soit $\mathcal{C}_b$ l'ensemble des fonctions continues bornées sur $\mathbb{R}$. Soient μ et ν deux mesures finies sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Montrer que si pour toute $f \in \mathcal{C}_b$, $\int_{\mathbb{R}} f \, d\mu = \int_{\mathbb{R}} f \, d\nu$, alors $\mu = \nu$.
2. Soit μ une mesure finie sur $(\Omega, \mathcal{F})$ et f une application finie μ-presque partout. Montrer que les quatre conditions suivantes sont équivalentes :
   (a) f est intégrable par rapport à μ.
   (b) $\int_{\{|f|>n\}} |f| \, d\mu$ tend vers 0 quand n tend vers l'infini
   (c) $\sum_{n\geq 1} n\mu(n < |f| \leq n+1) < +\infty$.
   (d) $\sum_{n\geq 0} \mu(|f| > n) < +\infty$.

   Indication : montrer $a \iff b, a \iff c, d \Longrightarrow c, (c\&b) \Longrightarrow d$.
3. Soit μ une mesure sur $(\Omega, \mathcal{F})$ et f intégrable par rapport à μ . Montrer que la fonction
$$g = \sum_{n=1}^{+\infty} \frac{1}{n^2} f^2 \mathbb{1}_{\{n>|f|\}}$$
   est intégrable par rapport à μ.
4. Le but de cet exercice est de montrer le *théorème du retour* de Poincaré. Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré et T une transformation, c'est à dire une application mesurable de $(\Omega, \mathcal{F})$ dans lui-même. On suppose que μ est une mesure finie et qu'elle est invariante sous l'action de T, c'est à dire que la mesure image de μ par l'application T est μ elle-même. Alors, le théorème du retour dit que pour tout ensemble mesurable A de mesure non nulle, la suite des itérées $(T^n(x))_{n\geq 0}$ passe une infinité de fois dans A pour presque tout x appartenant à A.
   (a) On pose
$$N(x) = \sum_{n=0}^{+\infty} \mathbb{1}_A(T^n(x))$$
   ainsi que $Y(x) = \exp(-N(x))$, avec la convention $\exp(-(+\infty)) = 0$. Montrer que Y est une application mesurable intégrable par rapport à μ.
   (b) On pose $Z(x) = Y(Tx)$. Montrer que $Y(x) = e^{-1_A(x)} Z(x)$, puis que $\int Y(x) \, d\mu(x) = \int Z(x) \, d\mu(x)$.
   (c) Conclure.
5. *Intégration par rapport à une somme de mesures.*
   Soit $(\mu_i)_{i\in\mathbb{N}}$ une suite de mesure définies sur un espace $(\Omega, \mathcal{F})$.

(a) Montrer que $\mu = \sum_{i\in\mathbb{N}} \mu_i$ est une mesure sur $(\Omega, \mathcal{F})$.

(b) Montrer que pour f mesurable positive, puis pour f intégrable, on a

$$\int_\Omega f\ d\mu = \sum_{i\in\mathbb{N}} \int_\Omega f\ d\mu_i.$$

6. Soit $(f_n)_{n\geq 0}$ une suite croissante d'applications μ-intégrables convergeant μ presque partout vers f. On suppose qu'il existe une constante K telle que

$$\forall n \geq 0 \qquad \int f_n\ d\mu \leq K.$$

Montrer que f est μ-intégrable et que $\int |f_n - f|\ d\mu$ tend vers 0 lorsque n tend vers l'infini.

7. Soit $(f_n)_{n\geq 0}$ une suite décroissante d'applications μ-mesurables positives. On suppose que f_1 est μ-intégrable. Montrer que (f_n) converge simplement vers une fonction mesurable f, que f est μ-intégrable et que $\int |f_n - f|\ d\mu$ tend vers 0 lorsque n tend vers l'infini.

8. Soit f une fonction réelle intégrable sur $(\Omega, \mathcal{F}, \mu)$. Montrer que si $\int_A f\ d\mu = 0$ pour tout $A \in \mathcal{F}$, alors $f = 0$ presque partout.

9. Étudier la limite, lorsque n tend vers l'infini de

(a) $\int_0^{+\infty} \frac{\log(x+n)}{n} e^{-x} \cos x\ dx$

(b) $\int_0^1 \frac{1+nx}{(1+x)^n}\ dx$

(c) $\int_1^{+\infty} \frac{1+nx}{(1+x)^n}\ dx$

10. (a) Calculer $I_n = \int_0^1 x^n \ln x\ dx$ pour $n \in \mathbb{N}$

(b) En déduire la valeur de $\int_0^1 \frac{\ln x}{1-x}\ dx$, sachant que $\sum_{n=1}^{+\infty} n^{-2} = \pi^2/6$.

(c) En calculant de deux manières différentes l'intégrale $\displaystyle\int_0^1 (1-x)^n \ln(x)\ dx$, montrer que pour $n \geq 0$, on a

$$\sum_{k=0}^n \binom{n}{k} \frac{(-1)^k}{(k+1)^2} = \frac{H_{n+1}}{n+1}, \text{ où } H_n = \sum_{k=1}^n \frac{1}{k}.$$

11. Démontrer que

$$\int_0^1 (e^x - 1)(\log x + \frac{1}{x})\ d\lambda(x) = \sum_{n=1}^{+\infty} \frac{n^2+n+1}{n(n+1)(n+1)!}.$$

12. Calculer $\displaystyle\int_0^1 \frac{1}{1+x^2}\,dx$. En déduire $\displaystyle\sum_{n=0}^{+\infty} \frac{(-1)^n}{2n+1}$.

13. Démontrer que $\displaystyle\int_0^{+\infty} \frac{\sin ax}{e^x - 1}\, d\lambda(x) = \sum_{n=1}^{+\infty} \frac{a}{n^2 + a^2}$.

14. Démontrer que $\displaystyle\int_0^1 \frac{(x \log x)^2}{1 + x^2}\, d\lambda(x) = 2\sum_{n=0}^{+\infty} \frac{(-1)^n}{(2n+3)^3}$.

15. *Fonction* Γ
    On définit la fonction Γ sur $]0, +\infty[$ par

$$\Gamma(x) = \int_0^{+\infty} e^{-t} t^{x-1}\, d\lambda(t).$$

    Montrer que cette fonction est bien définie et qu'elle vérifie, pour tout réel $a > 0$, l'égalité $\Gamma(a+1) = a\cdot \Gamma(a)$. En particulier, vérifier que pour tout entier $n \geq 0$, on a :

$$\Gamma(n+1) = n\,! \quad \text{et} \quad \Gamma(n + \frac{1}{2}) = \Gamma(\frac{1}{2})\cdot \frac{(2n)\,!}{2^{2n}\cdot n\,!}.$$

16. *Intégrales de Wallis*
    (a) On pose

$$W_n = \int_0^{\pi/2} \cos^n \theta\; d\theta.$$

    Montrer que $W_n = \frac{n-1}{n} W_{n-2}$. En déduire que la suite $(nW_nW_{n-1})_{n\geq 1}$ est constante.
    (b) Montrer $W_nW_{n+1} \leq W_n^2 \leq W_nW_{n-1}$; en déduire l'équivalent $W_n \sim \sqrt{\frac{\pi}{2n}}$.

17. *Calcul de l'intégrale de Gauss*
    (a) On pose $J_n = \int_0^{\sqrt{n}} (1-\frac{t^2}{n})^n\, dt$. Montrer que $\lim\limits_{n\to+\infty} J_n = \int_0^{+\infty} e^{-t^2}\, dt$.
    (b) Exprimer J_n en fonction d'une intégrale de Wallis. En déduire la valeur de l'intégrale de Gauss :

$$\int_0^{+\infty} e^{-t^2/2}\, dt = \frac{1}{2}\sqrt{2\pi}.$$

18. *Calcul de* $\Gamma(1/2)$
    Connaissant la valeur de l'intégrale de Gauss, montrer que

$$\Gamma(1/2) = \sqrt{\pi}.$$

19. *Formule de Stirling*

   (a) Montrer que $\int_{2n}^{+\infty} x^n e^{-x}\, dx = o(\Gamma(n+1))$.

   (b) Montrer que $\frac{\Gamma(n+1)}{e^{-n}n^{n+1/2}} \sim \int_{-\sqrt{n}}^{\sqrt{n}} (1+\frac{u}{\sqrt{n}})^n e^{-u\sqrt{n}}\, du$.

   (c) Montrer que pour tout $x \in]-1,1[$, on a $\ln(1+x) - x \le -\frac{x^2}{6}$.

   (d) On rappelle que $\int_0^{+\infty} e^{-t^2/2}\, dt = \frac{1}{2}\sqrt{2\pi}$. Montrer que

$$\Gamma(n+1) \sim \sqrt{2\pi} e^{-n} n^{n+1/2}.$$

20. *Fonction* Γ *(suite)*

   (a) Démontrer que la fonction Γ est de classe $\mathcal{C}^\infty$ sur $]0,+\infty[$ et que pour tout entier $n \ge 1$,

$$\Gamma^{(n)}(x) = \int_0^{+\infty} e^{-t}(\log t)^n t^{x-1}\, d\lambda(t).$$

   (b) Pour tous réels $a, s > 0$, exprimer $\displaystyle\int_0^{+\infty} t^{s-1}e^{-at} d\lambda(t)$ à l'aide de la fonction Γ, , puis démontrer que pour tout réel $s > 1$, $\displaystyle\int_0^{+\infty} \frac{t^{s-1}}{e^t - 1}\, d\lambda(t) = \sum_{n=1}^{+\infty} \frac{\Gamma(s)}{n^s} < +\infty.$

   (c) Démontrer que pour tout entier $n \ge 0$, $\displaystyle\int_0^{+\infty} t^{2n} e^{-t^2}\, d\lambda(t) = \frac{\Gamma\left(n+\frac{1}{2}\right)}{2}.$

   (d) Démontrer que $\displaystyle\int_0^{+\infty} e^{-t^2} \cos(at)\, d\lambda(t) = \frac{\sqrt{\pi}}{2} \exp\left(-\frac{a^2}{4}\right)$, pour tout réel a.

21. (a) Montrer que l'intégrale $\int_0^{+\infty} \frac{\sin t}{t}\, dt$ est convergente, mais que la fonction $\frac{\sin t}{t}$ n'est pas intégrable par rapport à la mesure de Lebesgue sur $\mathbb{R}_+$.

   (b) On pose, pour $x \ge 0$, $F(x) = \int_0^{+\infty} \frac{\sin t}{t} e^{-xt}\, dt$. Montrer que F est définie sur $\mathbb{R}_+$, dérivable sur $]0,+\infty[$, avec $F'(x) = -\frac{1}{1+x^2}$.

   (c) Calculer F et en déduire la valeur de $\int_0^{+\infty} \frac{\sin t}{t}\, dt$.

22. *Calcul d'intégrales liées aux intégrales de Fresnel*
   Le but de cet exercice est le calcul des intégrales

$$\int_0^{+\infty} e^{iu} u^{\alpha-1}\, du, \quad 0 < \alpha < 1$$

   et d'intégrales liées.

(a) On pose, pour $\lambda \geq 0$,

$$\varphi(\lambda) = \int_0^{+\infty} e^{-\lambda u}\ e^{iu} u^{\alpha-1}\ du.$$

Montrer que φ définit une fonction continue sur $[0, +\infty[$.

(b) Montrer que φ est dérivable sur $]0, +\infty[$ et vérifie l'équation différentielle $\varphi'(\lambda) = \frac{\alpha}{i-\lambda}\varphi(\lambda)$. En déduire que pour tout $\lambda > 0$, on a

$$\varphi(\lambda) = \varphi(0)\frac{1}{(\lambda^2+1)^{\alpha/2}} \exp(-\alpha i \operatorname{atan} \lambda).$$

(c) Montrer que pour tout $\lambda > 0$, on a

$$\varphi(0) = (1+\lambda^{-2})^{\alpha/2} \exp(\alpha i \operatorname{atan} \lambda) \int_0^{+\infty} e^{-x} e^{i\frac{x}{\lambda}} x^{\alpha-1}\ dx.$$

(d) En déduire

$$\int_0^{+\infty} e^{iu} u^{\alpha-1}\ du = \exp(i\alpha\frac{\pi}{2})\Gamma(\alpha),$$

en particulier, comme $\Gamma(1/2) = \sqrt{\pi}$,

$$\int_0^{+\infty} \frac{\cos u}{\sqrt{u}}\ du = \int_0^{+\infty} \frac{\sin u}{\sqrt{u}}\ du = \sqrt{\frac{\pi}{2}}.$$

(e) Calculer les intégrales de Fresnel

$$\int_0^{+\infty} \cos(u^2)\ du \text{ et } \int_0^{+\infty} \sin(u^2)\ du.$$

23. *Calcul de l'intégrale de Dirichlet*
Soit $a > 0$. Montrer que la fonction, $f : (x, y) \mapsto e^{-xy} \sin x$, est intégrable sur $[0, a] \times [0, +\infty[$. On pose $I_a = \displaystyle\int_{[0,a]\times[0,+\infty[} f(x, y)\, dxdy$. Déterminer la limite de I_a quand a tend vers $+\infty$. En déduire la valeur de $\displaystyle\int_0^{+\infty} \frac{\sin x}{x}\ dx$.

24. Rappelons que la fonction *Gamma* $\Gamma : \mathbb{R}_+^* \to \mathbb{R}$ est définie pour tout $x > 0$ par $\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}\, dt$ et la fonction *Beta* $\beta : \mathbb{R}_+^* \times \mathbb{R}_+^* \to \mathbb{R}$ est définie pour tous $x > 0$, $y > 0$ par $\beta(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}\, dt$. Pour tous $x > 0$, $y > 0$, vérifier l'existence de $\beta(x, y)$ puis montrer que :

$$\beta(x, y) = \frac{\Gamma(x)\,\Gamma(y)}{\Gamma(x+y)}.$$

25. Pour $x \geq 0$, on pose $R(x) = \int_x^{+\infty} \frac{\sin u}{u}\, du$. Le but de l'exercice est de démontrer l'identité
$$\int_0^{+\infty} \frac{R(x)}{\sqrt{x}} = \sqrt{2\pi}.$$

   (a) Montrer que $R(x)$ est bien défini pour $x \geq 0$ et qu'on a en l'infini $R(x) = O(1/x)$. En déduire que R est bornée.

   (b) $\frac{R(x)}{\sqrt{x}}$ est-elle intégrable par rapport à la mesure de Lebesgue sur $\mathbb{R}_+$?

   (c) Pour $\lambda \geq 0$, On pose $R_\lambda(x) = \int_x^{+\infty} \frac{\sin u}{u} e^{-\lambda u}\, du$.

      i. Justifier l'existence de $R_\lambda(x)$, puis montrer que pour tout $x > 0$ et tout $\lambda > 0$, on a
$$|R_\lambda(x) - R(x)| \leq |\cos x| \frac{1 - e^{-\lambda x}}{x} + \int_x^{+\infty} \frac{\Psi(\lambda u)}{u^2}\, du,$$
      où on a posé $\Psi(x) = |e^{-x}(1 + x) - 1|$.

      ii. Montrer qu'il existe $M < +\infty$ telle que $\Psi(x) \leq M$ pour tout $x \geq 0$. En déduire que pour tout $x > 0$, $\lim_{\lambda \to 0} R_\lambda(x) = R(x)$.

      iii. Montrer qu'il existe $L > 0$ tel que $\Psi(x) \leq Lx^2$ pour tout $x \geq 0$. Établir les inégalités
$$\begin{aligned} \forall \lambda > 0 \quad \forall x \geq 1 \quad |R_\lambda(x) - R(x)| \quad &\leq \frac{1 + 2M}{x} \\ \forall \lambda > 0 \quad \forall x \leq 1 \quad |R_\lambda(x) - R(x)| \quad &\leq \lambda + L\lambda^2 + 2M. \end{aligned}$$

      iv. Montrer que pour tout $\lambda > 0$, $\frac{R_\lambda(x)}{\sqrt{x}}$ est intégrable par rapport à la mesure de Lebesgue sur $\mathbb{R}_+$, puis que
$$\int_0^{+\infty} \frac{R(x)}{\sqrt{x}} = \lim_{\lambda \to 0} \int_0^{+\infty} \frac{R_\lambda(x)}{\sqrt{x}}.$$

   (d) Montrer que pour tout $\lambda > 0$, on a
$$\int_0^{+\infty} \frac{R_\lambda(x)}{\sqrt{x}} = 2 \int_0^{+\infty} \frac{\sin u}{\sqrt{u}} e^{-\lambda u}\, du.$$

   (e) Montrer finalement que
$$\int_0^{+\infty} \frac{R(x)}{\sqrt{x}} = \sqrt{2\pi}.$$
   (On rappelle que $\int_0^{+\infty} \frac{\sin u}{\sqrt{u}}\, du = \sqrt{\frac{\pi}{2}}$.)

26. *Calcul de $\sum_{n\geq 1}\frac{1}{n^2}$ avec Fubini.*
Calculer de deux façons différentes :
$$J=\int_{-1}^{1}\int_{-1}^{1}\frac{1}{1+2xy+y^2}\,\mathrm{d}x\,\mathrm{d}y.$$

27. (a) Soit φ une application de $\mathbb{N}$ dans $\mathbb{R}_+$ Montrer que l'on a
$$\int_{\mathbb{R}^d}\varphi(\mathrm{Ent}(\|x\|_\infty))\;d\lambda^d(x)=\sum_{n=0}^{+\infty}\left((2n+2)^d-(2n)^d\right)\varphi(n).$$
    (b) Soit $\alpha>0$. À quelle condition la fonction $x\mapsto\frac{1}{\|x\|_2^\alpha}$ est elle intégrable sur le complémentaire de la boule unité ?
    (c) Montrer que l'application
$$f:x\mapsto\frac{x}{\|x\|^2}$$
réalise un C^1 difféomorphisme de $\{x\in\mathbb{R}^d;0<\|x\|_2<1\}$ sur $\{x\in\mathbb{R}^d;1<\|x\|_2\}$ et que sa différentielle est
$$h\mapsto Df_x.h=\frac{\|x\|_2^2h-2\langle x,h\rangle x}{\|x\|_2^4}.$$
    (d) Soit $\alpha>0$. À quelle condition la fonction $x\mapsto\frac{1}{\|x\|_2^\alpha}$ est elle intégrable sur la boule unité ?

28. Montrer que pour tout $z\in\mathbb{C}$, $(1+z/n)^n$ tend vers $\exp(z)$ lorsque n tend vers l'infini.

Chapitre 5

Lois des variables et des vecteurs aléatoires

Rappel : si X est un espace topologique (par exemple un espace métrique), on appelle *tribu borélienne* de X et on note $\mathcal{B}(X)$ la tribu engendrée par la famille des ouverts de X.

5.1 Définition

Si $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé, on appelle variable aléatoire toute application mesurable de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, où $\mathcal{B}(\mathbb{R})$ est la tribu borélienne de $\mathbb{R}$. De même, on appelle vecteur aléatoire toute application mesurable de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, où $\mathcal{B}(\mathbb{R}^d)$ est la tribu borélienne de $\mathbb{R}^d$.

On appelle loi d'une variable aléatoire X definie sur $(\Omega, \mathcal{F}, \mathbb{P})$ la loi image de $\mathbb{P}$ par X. Cette loi est notée $\mathbb{P}_X$ Rappelons que cette loi image est une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ définie par

$$\forall A \in \mathcal{B}(\mathbb{R}) \quad \mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A))$$

Par définition, $X^{-1}(A) = \{\omega \in \Omega; X(\omega) \in A\}$. Afin de simplifier les écritures, on écrit toujours $\{X \in A\}$ à la place de $X^{-1}(A)$. Ainsi, on écrira le plus souvent $\mathbb{P}(\{X \in A\})$ et même $\mathbb{P}(X \in A)$ pour désigner $\mathbb{P}_X(A)$.

Exemple : Soit $\mathbb{P}$ la mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ définie par $\mathbb{P} = \frac{1}{3}\delta_{-1} + \frac{1}{2}\delta_0 + \frac{1}{6}\delta_1$. $\mathbb{P}$ est une mesure positive, de masse totale 1 : c'est donc une probabilité. Considérons l'application $X : \mathbb{R} \to \mathbb{R}$ définie par $\forall \omega \in \mathbb{R} \quad X(\omega) = |\omega|$. Comme X est une application mesurable, X est une variable aléatoire. Pour

$\mathbb{P}$-presque tout $\omega, X(\omega) \in \{0, 1\}$. Ainsi, la loi de X sous $\mathbb{P}$ est

$$\begin{aligned} \mathbb{P}_X &= \mathbb{P}(X = 0)\delta_0 + \mathbb{P}(X = 1)\delta \\ &= \mathbb{P}(\{0\})\delta_0 + \mathbb{P}(\{-1, 1\})\delta_1 \\ &= \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1 \end{aligned}$$

Exemple : L'exemple qui suit ne paie pas de mine mais est cependant très instructif. Soit $\mathbb{P}$ la mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ définie par $\mathbb{P} = \frac{1}{3}\delta_{-1} + \frac{1}{2}\delta_0 + \frac{1}{6}\delta_1$. On a vu que $\mathbb{P}$ était une probabilité. Considérons l'application $Y : \mathbb{R} \to \mathbb{R}$ définie par $\forall \omega \in \mathbb{R} \quad Y(\omega) = \omega$. Comme Y est une application mesurable (!), Y est une variable aléatoire. Il est facile de voir que la loi de Y sous $\mathbb{P}$ est tout simplement $\mathbb{P}$. Ainsi, on voit que le problème de l'existence d'une variable aléatoire suivant une certaine loi se ramène à celui de l'existence de cette loi et relève donc des théories de la mesure.

5.1.1 Fonction de répartition

Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^d$. On appelle fonction de répartition de X et on note F_X la fonction définie sur $\mathbb{R}^d$ par

$$\begin{aligned} \forall t = (t_1, \dots, t_d) \in \mathbb{R}^d \quad F_X(t) &= \mathbb{P}_X(]-\infty, t_1] \times]-\infty, t_2] \times]-\infty, t_d]) \\ &= \mathbb{P}(X_1 \le t_1, X_2 \le t_2 \dots, X_d \le t_d) \end{aligned}$$

Théorème 51. *Si deux variables (ou vecteurs) aléatoires ont la même fonction de répartition,alors elles ont même loi.*

Démonstration. Si X et Y sont tels que $F_X = F_Y$, cela veut dire que $\mathbb{P}_X$ et $\mathbb{P}_Y$ coincident sur les ensembles de la forme $]-\infty, t_1] \times]-\infty, t_2] \times]-\infty, t_d]$. Or ces ensembles forment un π-système qui engendre $\mathcal{B}(\mathbb{R}^d)$, donc $\mathbb{P}_X$ et $\mathbb{P}_Y$ sont égales. □

C'est surtout en dimension 1 que la fonction de répartition est utile, car en dimension supérieure, ses propriétés sont plus difficiles à exprimer et les calculs sont souvent compliqués, voire infaisables. Nous allons juste nous contenter de donner quelques propriétés de la fonction caractéristique d'une variable aléatoire.

Propriétés de la fonction de répartition d'une variable aléatoire réelle

Théorème 52. *La fonction de répartition F_X d'une variable aléatoire vérifie les propriétés suivantes*

- *F_X est à valeurs dans* $[0,1]$
- *F_X est croissante sur* $\mathbb{R}$.
- $\lim_{t\to-\infty} F_X(t) = 0$.
- $\lim_{t\to+\infty} F_X(t) = 1$.
- *En tout point, F_X est continue à droite.*
- *En tout point, F_X admet une limite à gauche.*

Démonstration. – Le premier point découle du fait que $F_X(t)$ est la probabilité d'un événement.

- Si $s \le t$, on a $]-\infty, s] \subset]-\infty, t]$, d'où

$$F_X(s) = \mathbb{P}(]-\infty, s]) \le \mathbb{P}(]-\infty, t]) = F_X(t).$$

- Posons pour $n \ge 1$, $A_n =]-\infty, -n]$, on a $A_{n+1} \subset A_n$ et $\bigcap_{n\ge1} A_n = \varnothing$, d'où $\lim_{n\to+\infty} \mathbb{P}_X(A_n) = \mathbb{P}_X(\varnothing) = 0$. Soit $\varepsilon > 0$: d'après ce qui précède, il existe n tel que $\mathbb{P}_X(A_n) < \varepsilon$. Comme F_X est croissante et positive, on a

$$t \le -n \quad \Longrightarrow 0 \le F_X(t) \le F_X(-n) \le \varepsilon,$$

ce qui prouve que $\lim_{t\to-\infty} F_X(t) = 0$.

- Posons pour $n \ge 1$, $A_n =]-\infty, n]$, on a $A_n \subset A_{n+1}$ et $\bigcup_{n\ge1} A_n = \mathbb{R}$, d'où $\lim_{n\to+\infty} \mathbb{P}_X(A_n) = \mathbb{P}_X(\mathbb{R}) = 1$. Soit $\varepsilon > 0$: d'après ce qui précède, il existe n tel que $\mathbb{P}_X(A_n) \ge 1-\varepsilon$. Comme F_X est croissante et majorée par 1, on a

$$t \ge n \quad \Longrightarrow 1 \ge F_X(t) \ge F_X(n) \ge 1 - \varepsilon,$$

ce qui prouve que $\lim_{t\to+\infty} F_X(t) = 1$.

- Soit $t \in \mathbb{R}$, $(t_n)_{n\ge1}$ une suite décroissante convergeant vers t. Posons pour $n \ge 1$, $A_n =]-\infty, t_n]$, on a $A_{n+1} \subset A_n$ et $\bigcap_{n\ge1} A_n =]-\infty, t]$, d'où $\lim_{n\to+\infty} F_X(t_n) = \lim_{n\to+\infty} \mathbb{P}_X(A_n) = \mathbb{P}_X(]-\infty, t]) = F_X(t)$. Comme cette égalité est obtenue pour toute suite décroissante convergeant vers t, ceci prouve que la limite à droite de F_X au point t est $F_X(t)$ (critère de

continuité séquentiel). Remarquons qu'on aurait pu également utiliser des critères analogues pour les preuves des deux propriétés précédentes et éviter ainsi l'emploi de ε.
– Toute fonction croissante admet une limite à gauche en tout point de l'intérieur de l'ensemble de définition.

$\square$

5.1.2 Tribu engendrée par une ou plusieurs variables aléatoires

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, X une variable aléatoire (ou un vecteur aléatoire) sur cette espace. On note

$$\sigma(X) = \{X^{-1}(A); A \in \mathcal{B}(\mathbb{R})\}.$$

Cette famille est une tribu. On dit que c'est la tribu engendrée par une variable aléatoire X.

De la même manière, on appelle tribu engendrée par une famille de variables $(X_i)_{i \in I}$ et on note $\sigma((X_i)_{i \in I})$ la tribu

$$\bigwedge_{i \in I} \sigma(X_i).$$

Exemple : Soient X et Y deux variables aléatoires à valeurs dans $\mathbb{N}$. Alors, l'événement $\{X = Y\}$ est $\sigma(X, Y)$-mesurable.

Démonstration. On a

$$\begin{aligned}\{X = Y\} &= \bigcup_{k \in \mathbb{N}} \{X = Y\} \cap \{X = k\} \\ &= \bigcup_{k \in \mathbb{N}} \{X = k\} \cap \{Y = k\}\end{aligned}$$

Par définition de $\sigma(X)$, l'événement $\{X = k\}$ est $\sigma(X)$-mesurable. Comme $\sigma(X, Y)$ contient $\sigma(X)$, l'événement $\{X = k\}$ est $\sigma(X, Y)$-mesurable. De même, l'événement $\{Y = k\}$ est $\sigma(X, Y)$-mesurable. Comme $\sigma(X, Y)$ est une tribu, on en conclut que pour tout k, l'événement $\{X = k\} \cap \{Y = k\}$ est $\sigma(X, Y)$-mesurable, puis que l'événement $\{X = Y\}$ est $\sigma(X, Y)$-mesurable.

$\square$

5.2 Indépendance des variables aléatoires

Définition : On dit que des variables aléatoires $(X_i)_{i\in I}$ sont indépendantes si les tribus $(\sigma(X_i))_{i\in I}$qu'elles engendrent sont indépendantes.

Exemple : Si X et Y sont deux variables aléatoires indépendantes, alors pour tout couple de boréliens A et B, on a

$$\mathbb{P}(\{X\in A\}\cap\{Y\in B\}=\mathbb{P}(X\in A)\mathbb{P}(Y\in B).$$

Théorème 53. *Soient $(X_i)_{i\in I}$ une collection de vecteurs aléatoires aléatoires indépendants. On suppose que X_i est à valeurs dans $\mathbb{R}^{n_i}$. Soient $(f_i)_{i\in I}$ une famille d'applications telles que pour tout i, f_i soit une application mesurable de $\mathbb{R}^{n_i}$ dans $\mathbb{R}^{p_i}$. Alors, si on pose $Y_i=f_i(X_i)$ les variables aléatoires $(Y_i)_{i\in I}$ sont indépendantes.*

Démonstration. L'indépendance des variables aléatoires est en fait l'indépendance des tribus engendrées. Soit $B\in\mathcal{B}(\mathbb{R}^{p_i})$ un borélien. On a $\{Y_i\in B\}=\{X_i\in f_i^{-1}(B)\}$. Comme f_i est borélienne, $f_i^{-1}(B)\in\mathcal{B}(\mathbb{R}^{n_i})$, et donc $\{Y_i\in B\}$ est $\sigma(X_i)$-mesurable. Ceci prouve que $\sigma(Y_i)$ est une sous-tribu de $\sigma(X_i)$. Comme les tribus $(\sigma(X_i))_{i\in I}$ sont indépendantes, leur sous-tribus $(\sigma(Y_i))_{i\in I}$ le sont aussi. □

Exemple : Si X,Y et Z sont indépendantes, alors $\operatorname{ch}X,Y^2$ et Z^3 sont indépendantes.

Là, nous restons un peu sur notre faim. En effet, nous voudrions pouvoir dire aussi que $\operatorname{ch}X+Y^2$ est indépendante de Z^3. Pour cela, il faudrait que nous sachions que (X,Y) est indépendant de Z, auquel cas nous pourrions appliquer les fonctions $f(x,y)=\operatorname{ch}x+y^2$ et $g(z)=z^3$.

Par chance (!), ceci est vrai. En effet, on a le résultat suivant :

Théorème 54. *Soient $(\mathcal{A}_i)_{i\in I}$ une famille de sous-tribus de $(\Omega,\mathcal{F})$ indépendantes sous $\mathbb{P}$. Soient $J\subset I$ et $K\subset I$ disjoints.*

Alors les tribus $\underset{j\in J}{\wedge}\mathcal{A}_j$ et $\underset{k\in K}{\wedge}\mathcal{A}_k$ sont indépendantes.

Démonstration. On considère le π-système $\mathcal{C}$ défini par

$$\mathcal{C}=\underset{F\subseteq J}{\cup}\{\underset{x\in F}{\cap}A_x;\forall x\in F\quad A_x\in\mathcal{A}_x\}.$$

ainsi que le π-système $\mathcal{D}$ défini par

$$\mathcal{D}=\underset{F\subseteq K}{\cup}\{\underset{x\in F}{\cap}A_x;\forall x\in F\quad A_x\in\mathcal{A}_x\}.$$

Si $B_1 \in \mathcal{C}$, B_2 peut s'écrire sous la forme $B_1 = \underset{x\in F}{\cap} A_x$ où $F \subseteq J$ et où $\forall x \in F \quad A_x \in \mathcal{A}_x$. De même, si $B_2 \in \mathcal{D}$, B_2 peut s'écrire sous la forme $B_2 = \underset{x\in F'}{\cap} A_x$ où $F' \subseteq J$ et où $\forall x \in F \quad A_x \in \mathcal{A}_x$. Ainsi

$$\begin{aligned}
\mathbb{P}(B_1 \cap B_2) &= \mathbb{P}(\underset{x\in(F\cup F')}{\cap} A_x) \\
&= \prod_{x\in(F\cup F')} \mathbb{P}(A_x) \\
&= \big(\prod_{x\in F} \mathbb{P}(A_x)\big)\big(\prod_{x\in F'} \mathbb{P}(A_x)\big) \\
&= \mathbb{P}(B_1)\mathbb{P}(B_2)
\end{aligned}$$

Comme $\mathcal{C}$ est un π-système qui engendre $\underset{j\in J}{\wedge} \mathcal{A}_j$ et $\mathcal{D}$ un π-système qui engendre $\underset{j\in K}{\wedge} \mathcal{A}_j$, le théorème 25 permet de conclure.

□

Théorème 55. *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace probabilisé et* $X_1, \ldots, X_n$ *n vecteurs aléatoires. Les deux propositions suivantes sont équivalentes*

1. $X_1, \ldots, X_n$ *sont indépendantes.*
2. $\mathbb{P}_{X_1,\ldots,X_n} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$

Démonstration. – Preuve de 1 $\Longrightarrow$ 2. On pose $\mathcal{C} = \prod_{i=1}^{n} \mathcal{B}(\mathbb{R}^{n_i})$. Soit $A = A_1 \times \cdots \times A_n \in \mathcal{C}$. On a

$$\begin{aligned}
\mathbb{P}_{X_1,\ldots,X_n}(A) &= \mathbb{P}((X_1, \ldots, X_n) \in A) \\
&= \mathbb{P}(\bigcap_{i=1}^{n} \{X_i \in A_i\}) \\
&= \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i) \\
&= \prod_{i=1}^{n} \mathbb{P}_{X_i}(A_i) \\
&= (\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n})(A)
\end{aligned}$$

Ainsi $\mathbb{P}_{X_1,\ldots,X_n}$ et $\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$ coïncident sur un π-système qui engendre $\bigotimes_{i=1}^{n} \mathcal{B}(\mathbb{R}^{n_i})$. Il s'ensuit que ces deux mesures sont égales.

– Preuve de 2 $\Longrightarrow$ 1 Soient $B_1, \dots B_n$ quelconques tels que pour tout i B_i soit $\sigma(X_i)$-mesurable : alors, pour tout i, il existe un borélien A_i tel que $B_i = \{X_i \in A_i\}$. On pose $A = A_1 \times \dots \times A_n \in \mathcal{C}$.

$$\begin{aligned}
\mathbb{P}(\bigcap_{i=1}^{n} B_i) &= \mathbb{P}(\bigcap_{i=1}^{n} \{X_i \in A_i\}) \\
&= \mathbb{P}_{X_1,\dots,X_n}(A) \\
&= (\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n})(A) \\
&= \prod_{i=1}^{n} \mathbb{P}_{X_i}(A_i) \\
&= \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i) \\
&= \prod_{i=1}^{n} \mathbb{P}(B_i),
\end{aligned}$$

ce qui prouve l'indépendance des tribus.

□

Corollaire 9. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X_1, \dots, X_n$ n vecteurs aléatoires. On suppose qu'il existe des mesures de probabilités $\mu_1, \dots, \mu_n$ telles que $\mathbb{P}_{X_1,\dots,X_n} = \mu_1 \otimes \dots \otimes \mu_n$. Alors*

1. *$X_1, \dots, X_n$ sont indépendantes.*
2. *Pour tout i, la loi de X_i sous $\mathbb{P}$ est μ_i. ($\mathbb{P}_{X_i} = \mu_i$)*

Démonstration. Soit B un borélien.

$$\begin{aligned}
\mathbb{P}_{X_i}(B) &= \mathbb{P}(X_i \in B) \\
&= \mathbb{P}((X_1, \dots, X_n) \in \Omega \times \dots \times \Omega \times B \times \Omega \dots \Omega) \\
&= \mathbb{P}_{(X_1,\dots,X_n)}(\Omega \times \dots \times \Omega \times B \times \Omega \dots \Omega) \\
&= (\mu_1 \otimes \dots \otimes \mu_n)(\Omega \times \dots \times \Omega \times B \times \Omega \dots \Omega) \\
&= \mu_1(\Omega) \times \dots \times \mu_{i-1}(\Omega) \times \mu_i(B) \times \mu_{i+1}(\Omega) \dots \mu_n(\Omega) \\
&= \mu_i(B)
\end{aligned}$$

Ainsi $\mathbb{P}_{X_i} = \mu_i$. L'identité $\mathbb{P}_{X_1,\dots,X_n} = \mu_1 \otimes \dots \otimes \mu_n$. peut se réécrire $\mathbb{P}_{X_1,\dots,X_n} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$ et il suffit alors d'appliquer le théorème précédent. □

5.2.1 Application : loi $0-1$ de Kolmogorov

Théorème 56. *Soit S un ensemble infini et $(\mathcal{A}_i)_{i \in S}$ une famille de tribus indépendantes sous la loi $\mathbb{P}$. On pose*

$$\mathcal{T} = \bigcap_{\Lambda \subseteq S} \bigwedge_{k \in S \setminus \Lambda} \mathcal{A}_k.$$

$\mathcal{T}$ est appelée tribu de queue de la famille $(\mathcal{A}_i)_{i\in\Lambda}$. Alors

$$\forall A \in \mathcal{T} \quad \mathbb{P}(A) \in \{0,1\}.$$

Démonstration. Posons $I = S \cup \{\infty\}$ et $\mathcal{A}_\infty = \mathcal{T}$. Montrons que les tribus $(\mathcal{A}_i)_{i\in I}$ sont indépendantes. Soit $J \subseteq I$, et pour tout $i \in J$, $A_i \in \mathcal{A}_i$. Si $\{\infty\} \notin J$, l'indépendance des tribus $(\mathcal{A}_i)_{i\in S}$ assure que

$$\mathbb{P}(\bigcap_{i\in J} A_i) = \prod_{i\in J} \mathbb{P}(A_i).$$

Sinon, par définition de la tribu $\mathcal{T}$, on a $A_\infty \in \bigwedge_{k\in S\backslash J} \mathcal{A}_k$.

Comme $\bigcap_{i\in J\backslash\{\infty\}} A_i \in \bigwedge_{k\in J\backslash\{\infty\}} \mathcal{A}_k$, le théorème 54 implique que A_∞ et $\bigcap_{k\in J\backslash\{\infty\}} A_k$ sont indépendants. Il s'ensuit que

$$\begin{aligned}
\mathbb{P}(\bigcap_{i\in J} A_i) &= \mathbb{P}(A_\infty \cap \bigcap_{i\in J\backslash\{\infty\}} A_i) \\
&= \mathbb{P}(A_\infty)\mathbb{P}(\bigcap_{i\in J\backslash\{\infty\}} A_i) \\
&= \mathbb{P}(A_\infty) \prod_{i\in J\backslash\{\infty\}} \mathbb{P}(A_i) \\
&= \prod_{i\in J} \mathbb{P}(A_i).
\end{aligned}$$

Ainsi, les tribus $(\mathcal{A}_i)_{i\in I}$ sont indépendantes. En particulier, la tribu $\mathcal{A}_\infty = \mathcal{T}$ est indépendante de la tribu $\bigwedge_{k\in S} \mathcal{A}_k$. Mais $\mathcal{T}$ est une sous-tribu de $\bigwedge_{k\in S} \mathcal{A}_k$, donc $\mathcal{T}$ est indépendante d'elle même. Soit donc $A \in \mathcal{T}$: on a

$$0 = \mathbb{P}(\varnothing) = \mathbb{P}(A \cap A^c) = \mathbb{P}(A)\mathbb{P}(A^c) = \mathbb{P}(A)(1 - \mathbb{P}(A)),$$

donc $\mathbb{P}(A) \in \{0,1\}$. □

5.2.2 Variables aléatoires indépendantes et convolutions

Théorème 57. *Si deux variables aléatoires X et Y sont indépendantes sous $\mathbb{P}$, alors $\mathbb{P}_X * \mathbb{P}_Y = \mathbb{P}_{X+Y}$.*

Démonstration. Si X et Y sont indépendantes, alors $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$. $\mathbb{P}_X * \mathbb{P}_Y$ est donc la loi image de $\mathbb{P}_{(X,Y)}$ par $(x, y) \mapsto x + y$. Mais la loi image de $\mathbb{P}_{(X,Y)}$ par $(x, y) \mapsto x + y$, ce n'est rien d'autre que $\mathbb{P}_{X+Y}$. □

Théorème 58. *Soit μ, ν et χ trois mesures finies sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. On a*

$$\mu * \nu = \nu * \mu$$

et

$$(\mu * \nu) * \chi = \mu * (\nu * \chi).$$

Démonstration. Si une des trois mesures de la deuxième formule est nulle, chacun des produits est nulle car la mesure nulle est absorbante pour le produit de convolution. Idem pour la première formule si l'une des deux est nulle Sinon, posons Notons $\mathbb{P} = \frac{\nu}{\nu(\mathbb{R}^d)} \otimes \frac{\mu}{\mu(\mathbb{R}^d)} \otimes \frac{\chi}{\chi(\mathbb{R}^d)}$. $\mathbb{P}$ est une mesure de probabilité. X,Y, Z Définissons sur $(\mathbb{R}^d)^3$: $X(x, y, z) = x; Y(x, y, z) = y; Z(x, y, z) = z, S = X + Y; T = Y + Z$. D'après l'associativité de l'indépendance, X et T sont indépendantes, de même S et Z sont indépendantes. On a donc

$$\mathbb{P}_X * \mathbb{P}_T = \mathbb{P}_{X+T} = \mathbb{P}_{S+Z} = \mathbb{P}_S * \mathbb{P}_Z.$$

Maintenant, il est facile de voir que $\mathbb{P}_T = \frac{1}{\nu(\mathbb{R}^d)\chi(\mathbb{R}^d)}\nu * \chi$ et que $\mathbb{P}_S = \frac{1}{\mu(\mathbb{R}^d)\nu(\mathbb{R}^d)}\mu * \nu$, d'où $(\mu * \nu) * \chi = \mu * (\nu * \chi)$. De même $\mathbb{P}_X * \mathbb{P}_Y = \mathbb{P}_{X+Y} = \mathbb{P}_{Y+X} = \mathbb{P}_Y * \mathbb{P}_X$ permet de montrer la première formule. □

Ainsi l'ensemble des mesures finies munies de (+,*) forme un anneau commutatif.

5.3 Variables aléatoires discrètes

On dit qu'une loi μ est discrète s'il existe un ensemble D fini ou dénombrable inclus dans $\mathbb{R}^d$ tel que $\mu(D) = 1$.

De même, on dit qu'une variable aléatoire X définie sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ est discrète si sa loi $\mathbb{P}_X$ est discrète.

Ainsi, si D est un ensemble dénombrable tel que $\mathbb{P}(X \in D) = \mathbb{P}_X(D) = 1$ et l'on pose

$$\forall i \in D \quad p_i = \mathbb{P}(X = i)$$

La famille $(p_i)_{i\in D}$ est une famille de réels positifs vérifiant

$$\sum_{i\in D} p_i = 1.$$

La connaissance de D et des p_i permet de reconstituer la loi de X. En effet, on a le théorème suivant :

Théorème 59. *Soit X une variable aléatoire discrète, et D un ensemble D fini ou dénombrable inclus dans $\mathbb{R}$ tel que $X(\Omega) = D$. Pour $i \in D$, on pose $p_i = \mathbb{P}(X = i)$. Alors,*

1. *pour tout $A \in \mathcal{B}(\mathbb{R})$, on a*

$$\mathbb{P}_X(A) = \sum_{i \in D \cap A} p_i.$$

2. $\mathbb{P}_X = \sum_{i \in D} p_i \delta_i$.

3. $\mathbb{P}_X$ *admet comme densité par rapport à la mesure de comptage sur D la fonction $f(x)$ définie par*

$$f(x) = \begin{cases} p_x & si\ x \in D \\ 0\ si\ x \notin D \end{cases}$$

Démonstration. On note m la mesure de comptage sur D Soit A un borélien. $\{X \in A \cap D\}$ est réunion dénombrable disjointe des événements $\{X = i\}$, où i décrit $A \cap D$. On peut donc écrire

$$\begin{aligned} \mathbb{P}_X(A) &= \mathbb{P}(X \in A) \\ &= \mathbb{P}(X \in A \backslash D) + \mathbb{P}(X \in A \cap D) \\ &= 0 + \mathbb{P}(X \in A \cap D) \\ &= \sum_{i \in A \cap D} \mathbb{P}(X = i) \\ &= \sum_{i \in A \cap D} p_i \end{aligned}$$

Posons $\mu = \sum_{i \in D} p_i \delta_i$.

$$\begin{aligned} \mu(A) &= \sum_{i \in D} p_i \delta_i(A) \\ &= \sum_{i \in D} p_i \mathbb{1}_A(i) \end{aligned}$$

On a d'une part

$$\begin{aligned}\mu(A) &= \sum_{i\in D} p_i \mathbb{1}_A(i) \\ &= \sum_{i\in D\cap A} p_i \mathbb{1}_A(i) + \sum_{i\in D\setminus A} p_i \mathbb{1}_A(i) \\ &= \sum_{i\in D\cap A} p_i + 0\end{aligned}$$

Comme A est quelconque, on en déduit que $\mu = \mathbb{P}_X$. D'autre part

$$\begin{aligned}\mu(A) &= \sum_{i\in D} p_i \mathbb{1}_A(i) \\ &= \int_D p_i \mathbb{1}_A(i)\ dm(i) \\ &= \int_{\mathbb{R}} f(i)\mathbb{1}_A(i)\ dm(i) \\ &= \int_A f(x)\ dm(x),\end{aligned}$$

ce qui signifie que μ (c'est à dire $\mathbb{P}_X$) admet f comme densité par rapport à la mesure de comptage.

□

On a la réciproque suivante :

Théorème 60. *Soit D un ensemble fini ou dénombrable, $(p_i)_{i\in D}$ une famille de réels positifs vérifiant*

$$\sum_{i\in D} p_i = 1.$$

Alors, on peut construire un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et une variable aléatoire X sur cet espace telle que

$$\forall i \in D \quad p_i = \mathbb{P}(X = i)$$

Démonstration. Comme on l'a déjà remarqué, le problème d'existence d'une variable aléatoire se ramène souvent à l'existence d'une loi. Ici,on peut prendre $\Omega = D$, $\mathcal{F} = \mathcal{P}(\Omega)$ et $X(\omega) = \omega$ avec

$$\mathbb{P} = \sum_{i\in D} p_i \delta_i.$$

□

5.3.1 Fonction d'une variable aléatoire discrète

Théorème 61. *La loi image μ_f d'une loi discrète μ par une application mesurable f est une loi discrète.*

Démonstration. Soit D un ensemble fini ou dénombrable tel que $\mu(D) = 1$. $f(D)$ est un ensemble fini ou dénombrable. Il est donc mesurable par rapport à la mesure de Lebesgue. On a $\mu_f(f(D)) = \mu(f^{-1}(f(D)) \geq \mu(D) = 1$ car $f^{-1}(f(D) \supset D$. Il s'ensuit que $f(D)$ est un ensemble fini ou dénombrable dont la mesure sous μ_f est 1. $\square$

Corollaire 10. *Soit X une variable aléatoire discrète définie sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et f une fonction quelconque définie sur $X(\Omega)$. Alors, la fonction Y définie par*

$$\forall \omega \in \Omega \quad Y = f(X(\omega))$$

est une variable aléatoire discrète sur $(\Omega, \mathcal{F}, \mathbb{P})$.

De manière plus concise, on écrit $Y = f(X)$.

Exemple : Soit X une variable aléatoire vérifiant $X(\Omega) = \{-1; 0; 1\}$, avec $\mathbb{P}(X = -1) = \mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \frac{1}{3}$.

On pose $Y = X^2$.
On a $Y(\Omega) = \{0; 1\}$, avec

$$\{Y = 0\} = \{X = 0\}$$

et

$$\{Y = 1\} = \{X = 1\} \cup \{X = -1\},$$

d'où

$$\mathbb{P}(Y = 0) = \mathbb{P}(X = 0) = \frac{1}{3}$$

et

$$\mathbb{P}(Y = 1) = \mathbb{P}(X = 1) + \mathbb{P}(X = -1) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}.$$

5.4 Variables et vecteurs aléatoires à densité

On dit qu'une loi μ est à densité (sous-entendu par rapport à la mesure de Lebesgue) s'il existe une fonction mesurable f qui soit une densité de μ par rapport à la mesure de Lebesgue.

Ainsi, si f est une densité de la loi μ, on a pour tout borélien A

$$\mu(A) = \int_A f(\omega) d\lambda(\omega).$$

Exercice : Montrer que si f est la densité d'une loi, alors $\lambda(f < 0) = 0$.
Évidemment, si f est la densité d'une loi, on a

$$1 = \mu(\mathbb{R}^d) = \int_{\mathbb{R}^d} f(\omega) d\lambda(\omega).$$

Réciproquement, si f est une fonction mesurable, positive λ presque partout et d'intégrale 1, $\mu = f.\lambda$ est une mesure de probabilité admettant f pour densité.

Ainsi, on dit qu'une variable (ou un vecteur) aléatoire X est à densité si sa loi $\mathbb{P}_X$ est à densité.

5.4.1 Premières propriétés

Soit X une variable aléatoire de densité f. f a les propriétés suivantes :

- $\forall a, b \in \mathbb{R} \quad a \leq b \Longrightarrow \quad \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a < X < b) = \int_{[a,b]} f(x)\ d\lambda(x)$.
- $\forall a \in \mathbb{R} \quad \mathbb{P}(a \leq X) = \mathbb{P}(a < X) = \int_{[a,+\infty[} f(x)\ d\lambda(x)$.
- $\forall a \in \mathbb{R} \quad \mathbb{P}(a \geq X) = \mathbb{P}(a > X) = \int_{]-\infty,a]} f(x)\ d\lambda(x)$.
- $\int_{\mathbb{R}} f(x)\ d\lambda(x) = 1$.

Remarquons que pour montrer qu'une fonction f positive est une densité de la variable aléatoire X, il suffit de montrer que pour tout t réel, on a

$$F_X(t) = \int_{]-\infty,t]} f(x)\ d\lambda(x).$$

En effet, en faisant tendre t vers l'infini, on obtient que $1 = \int_{\mathbb{R}} f(x)\ d\lambda(x)$, et donc f est la densité d'une variable aléatoire Y. Mais X et Y ont même fonction de répartition, donc même loi, f est la densité de la loi de Y, c'est donc aussi la densité de la loi de X.

5.4.2 Densités et lois marginales

Théorème 62. *Soit $(\Omega, \mathcal{F}, \mu)$ et $(\Omega', \mathcal{F}', \mu')$ deux espaces mesurés. On suppose que la loi ν sur $(\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}')$ admet une densité h par rapport à $\mu \otimes \mu'$. Alors la loi image ν_π de ν par l'application*

$$\begin{array}{rcl} \pi : \Omega \times \Omega' & \to & \Omega \\ (x, x') & \mapsto & x \end{array}$$

admet comme densité par rapport à μ la fonction f définie par $f(x) = \int_{\Omega'} h(x,x')d\mu'(x')$.

Démonstration. Soit $B \in \mathcal{F}$

$$\begin{aligned}
\nu_\pi(B) &= \nu(\pi^{-1}(B)) \\
&= \nu(B \times \Omega') \\
&= \int_{B\times\Omega'} d\nu(x,x') \\
&= \int_{B\times\Omega'} h(x,x')d(\mu\otimes\mu')(x,x') \\
&= \int_B (\int_{\Omega'} h(x,x')d\mu'(x'))d\mu(x) \\
&= \int_B f(x)d\mu(x),
\end{aligned}$$

ce qui prouve le résultat. □

Théorème 63. *Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^n$ et Y un vecteur aléatoire à valeurs dans $\mathbb{R}^p$ définis sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. Si $h(x,y)$ est une densité de (X,Y) par rapport à la mesure de Lebesgue sur $\mathbb{R}^{n+p}$, alors X admet la densité $f(x) = \int_{\mathbb{R}^p} h(x,y)d\lambda^p(y)$, tandis que*
Y admet la densité $g(y) = \int_{\mathbb{R}^n} h(x,y)d\lambda^n(x)$.

Démonstration. Le diagramme commutatif ci-dessous traduit que
$X = \pi \circ (X,Y)$.

$$\begin{array}{ccc}
(\Omega, \mathcal{F}, \mathbb{P}) & \xrightarrow{(X,Y)} & (\mathbb{R}^n \times \mathbb{R}^p, \mathcal{B}(\mathbb{R}^n)\otimes\mathcal{B}(\mathbb{R}^p), \mathbb{P}_{(X,Y)}) \\
 & X \searrow & \downarrow \pi \\
 & & (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbb{P}_X)
\end{array}$$

Il s'ensuit que $\mathbb{P}_X$ est la mesure image de $\mathbb{P}_{(X,Y)}$ par π. Comme $h(x,y)$ est la densité de (X,Y) par rapport à $\lambda^n \otimes \lambda^p = \lambda^{n+p}$, la densité de X est bien $f(x) = \int_{\mathbb{R}^p} h(x,y)d\lambda^p(y)$, On procède de même pour calculer la densité de Y. □

5.4.3 Indépendance et densités

Théorème 64. *Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^n$ et Y un vecteur aléatoire à valeurs dans $\mathbb{R}^p$ définis sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. On suppose que X admet une densité f et Y une densité g. Si X et Y sont* *indépendants, alors le vecteur aléatoire* (X,Y) *admet la fonction* $h(x,y) = f(x)g(y)$ *comme densité.*

Démonstration. Comme X et Y sont indépendantes, sous $\mathbb{P}$, on a $\mathbb{P}_{X,Y} = \mathbb{P}_X \otimes \mathbb{P}_Y$. Ainsi

$$\begin{aligned}
\mathbb{P}_{X,Y} &= \mathbb{P}_X \otimes \mathbb{P}_Y \\
&= f\lambda^n \otimes g\lambda^p \\
&= ((x,y) \mapsto f(x)g(y))\lambda^n \otimes \lambda^p \\
&= h\lambda^{n+p},
\end{aligned}$$

ce qui montre bien que (la loi de) (X,Y) admet h comme densité. □

Théorème 65. *Soit* X *un vecteur aléatoire à valeurs dans* $\mathbb{R}^n$ *et* Y *un vecteur aléatoire à valeurs dans* $\mathbb{R}^p$ *définis sur le même espace probabilisé* $(\Omega, \mathcal{F}, \mathbb{P})$. *On suppose que* (X,Y) *admet une densité* h_1 *qui s'écrive sous la forme* $h_1(x,y) = f_1(x)g_1(y)$, *où* f *et* g *sont des fonctions positives. Alors,* X *et* Y *sont indépendantes ;* X *admet comme densité par rapport à la mesure de Lebesgue sur* $\mathbb{R}^n$ *la fonction*

$$f(x) = \frac{f_1(x)}{\int_{\mathbb{R}^n} f_1(x')\ d\lambda^n(x')}$$

et Y *admet comme densité par rapport à la mesure de Lebesgue sur* $\mathbb{R}^n$ *la fonction*

$$g(y) = \frac{g_1(y)}{\int_{\mathbb{R}^p} g_1(y')\ d\lambda^p(y')}$$

Démonstration. Posons $A = \int_{\mathbb{R}^n} f_1(x) d\lambda^n(x)$ et $B = \int_{\mathbb{R}^p} g_1(y) d(\lambda^p)(y)$. Comme f_1 et f_2 sont positives, on a $A \ge 0$ et $B \ge 0$. Comme h est une densité, on a

$$\begin{aligned}
1 &= \int_{\mathbb{R}^n \times \mathbb{R}^p} h(x,y) d(\lambda^n \otimes \lambda^p)(x,y) \\
&= \int_{\mathbb{R}^n \times \mathbb{R}^p} f_1(x) g_1(y) d(\lambda^n \otimes \lambda^p)(x,y) \\
&= (\int_{\mathbb{R}^n} f_1(x) d\lambda^n(x))(\int_{\mathbb{R}^p} g_1(y) d(\lambda^p)(y)) \\
&= AB
\end{aligned}$$

On en peut donc dire que A et B sont tous deux strictement positifs. Ainsi les fonctions f et g sont bien définies. Elles sont positives, et, par construction, chacune d'entre elle admet 1 comme intégrale par rapport à la mesure de Lebesgue. Ainsi $\mu = f\lambda^n$ et $\nu = g\lambda^p$ sont des mesures de probabilité.

Comme dans le théorème précédent, la densité de $\mu \otimes \nu =$ est $h(x,y) = f(x)g(y)$. Donc $h(x,y) = f(x)g(y) = \frac{f_1(x)}{A}\frac{g_1(y)}{B} = \frac{f_1(x)g_1(y)}{AB} = h_1(x,y)$.

On en déduit $\mathbb{P}_{(X,Y)} = \mu \otimes \nu$.

Il suffit d'appliquer le corollaire 9 pour conclure.

□

5.5 Variables et lois discrètes classiques

5.5.1 Indicatrice d'un évévenement

On rappelle que pour $A \subset \Omega$, l'application $\mathbb{1}_A$ (appelée indicatrice de A) est définie sur Ω par

$$\mathbb{1}_A(x) = \begin{cases} 1 \text{ si } x \in A \\ 0 \text{ si } x \notin A \end{cases}$$

1_A est une variable aléatoire à valeurs dans $\{0;1\}$.

5.5.2 Masse de Dirac

On appelle masse de Dirac en un point $x \in \Omega$ la mesure δ_x définie par

$$\delta_x(A) = \begin{cases} 1 \text{ si } x \in A \\ 0 \text{ si } x \notin A \end{cases}$$

C'est bien une loi car elle est positive et $\delta_x(\Omega) = 1$.

Remarque : Si Ω est un groupe abélien $\delta_{x+y} = \delta_x * \delta_y = \delta_y * \delta_x$.

5.5.3 Loi de Bernoulli

On appelle loi de Bernoulli de paramètre p la loi $\mu = (1-p)\delta_0 + p\delta_1$.

Ainsi, on dit qu'une variable aléatoire X suit la loi de Bernoulli de paramètre p si on a $\mathbb{P}(X = 1) = p$ et $\mathbb{P}(X = 0) = 1 - p$.

Remarques importantes :

- Pour tout évènement A, $\mathbb{1}_A$ suit la loi de Bernoulli de paramètre $\mathbb{P}(A)$.
- Réciproquement, si une variable aléatoire X suit une loi de Bernoulli, elle vérifie $X = \mathbb{1}_{X=1}$.(Réfléchir un peu...)

Ainsi les variables aléatoires qui suivent des lois de Bernoulli sont exactement les indicatrices d'événements.

5.5.4 Loi uniforme sur un ensemble

Soit $E \subset \Omega$ un ensemble fini. On appelle loi uniforme sur E la loi définie sur $\mathcal{P}(\Omega)$ par

$$\begin{array}{rcl} \mathcal{P}(\Omega) & \to & [0,1] \\ A \mapsto \dfrac{|A \cap E|}{|E|} & & \end{array}$$

Ainsi, une variable aléatoire X suit la loi uniforme sur E si l'on a

$$\forall x \in E \quad \mathbb{P}(X = x) = \frac{1}{|E|}.$$

Exemple : La variable aléatoire X représentant le résultat du lancer d'un dé non truqué suit la loi uniforme sur l'ensemble $\{1, 2, 3, 4, 5, 6\}$.

5.5.5 Loi binomiale

On appelle loi binomiale de paramètres n et p et on note $\mathcal{B}(n, p)$ la loi de la somme de n variables de Bernoulli indépendantes de même probabilité p.

Ainsi $\mathcal{B}(n, p) = (\mathrm{Ber}(p))^{*n}$.

Théorème 66. *$\mathcal{B}(n, p)$ charge les entiers $\{0, \ldots, n\}$. Plus précisément, on a*

$$\forall k \in \{0, \ldots, n\} \quad \mathcal{B}(n, p)(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Démonstration. Posons $\mu = \mathcal{B}(n, p)$. On a

$$\begin{aligned} \mu &= (\mathrm{Ber}(p))^{*n} \\ &= ((1-p)\delta_0 + (p\delta_1)^{*n} \end{aligned}$$

Comme l'ensemble des mesures positives munie de (+,*) est un anneau commutatif, la formule du binôme de Newton s'applique et l'on a

$$\begin{aligned}
\mu &= ((1-p)\delta_0 + p\delta_1)^{*n} \\
&= \sum_{k=0}^{n} \binom{n}{k} ((1-p)\delta_0)^{*(n-k)} * (p\delta_1)^{*k} \\
&= \sum_{k=0}^{n} \binom{n}{k} ((1-p)^{n-k}(\delta_0)^{*(n-k)}) * (p^k(\delta_1)^{*k}) \\
&= \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k}\delta_0^{*(n-k)} * \delta_1^{*k} \\
&= \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k}\delta_0 * \delta_k \\
&= \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k}\delta_k
\end{aligned}$$

□

Corollaire 11. *Soit $A_1, \dots, A_n$ n événements indépendants de même probabilité p. On pose*

$$X = \sum_{k=1}^{n} \mathbb{1}_{A_k}.$$

Alors X suit la loi binomiale de paramètres n et p .

Remarque : X est le nombre de A_k qui sont réalisés.

Exemple : On lance n fois une pièce de monnaie équilibrée. Le nombre X de "pile" obtenus suit une loi binomiale de paramètres n et $\frac{1}{2}$.

5.5.6 Loi géométrique

On dit qu'une variable aléatoire X suit la loi géométrique de paramètre p si l'on a

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

Théorème 67. *Soit $(A_i)_{i\in\mathbb{N}^*}$ une suite d'événements indépendants de même probabilité $p > 0$. On pose*

$$X(\omega) = \inf\{k \in \mathbb{N}^* \quad \omega \in A_k\}.$$

Alors X suit la loi géométrique de paramètre p . De plus

$$\forall k \in \mathbb{N} \qquad F_X(k) = 1 - (1-p)^k.$$

Démonstration.

$$\{X > k\} = \underset{i\in\{1,\dots,k\}}{\cap} A_i^c.$$

Comme les A_i sont dans $\mathcal{F}$, les A_i^c le sont aussi, et donc comme on peut l'écrire comme intersection finie d'éléments de $\mathcal{F}$,$\{X > k\}$ est dans $\mathcal{F}$, et donc, par passage au complémentaire

$$\{X \le k\} \in \mathcal{F}.$$

En utilisant l'indépendance des (A_i), on obtient

$$\mathbb{P}(X > k) = (1-p)^k.$$

Comme

$$\{X = +\infty\} = \underset{k\ge 1}{\cap} \{X > k\},$$

on obtient, par continuité séquentielle décroissante

$$\mathbb{P}(X = +\infty) = \lim_{k\to+\infty} \mathbb{P}(\{X > k\}) = \lim_{k\to+\infty} (1-p)^k = 0.$$

Ainsi X est bien une variable aléatoire, et l'on a

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

De plus

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

□

Exemple : On lance une pièce de monnaie équilibrée jusqu'à obtention de "pile". Le nombre de lancers effectués suit une loi géométrique de paramètre $\frac{1}{2}$.

5.5.7 Loi de Poisson

On dit qu'une variable aléatoire X suit la loi de Poisson de paramètre $\lambda > 0$ si l'on a

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

La construction de telles variables est bien possible car $e^{-\lambda}\frac{\lambda^k}{k!} \ge 0$ et

$$\sum_{k=0}^{+\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1.$$

5.5.8 Loi hypergéométrique

La loi hypergéométrique $\mathcal{H}(N,n,k)$ modélise le phénomène suivant : on tire au hasard k individus dans une population de N individus, et l'on compte le nombre d'individus possédant une certaine particularité, sachant qu'il y a exactement n personnes dans la population totale qui possédaient cette particularité.

De manière théorique, la loi hypergéométrique est la loi image de la loi uniforme sur $\Omega = \mathcal{B}(N,k)$ par l'application

$$\begin{array}{rcl} X : \mathcal{B}(N,k) & \to & \mathbb{N} \\ \omega & \mapsto & X(\omega) = |\{1,\dots,n\}) \cap \omega| \end{array}$$

Ainsi pour $i \in \{0,\dots,\min(n,k)\}$, on a

$$\mathcal{H}(N,n,k)(i) = \frac{\binom{n}{i}\binom{N-n}{k-i}}{\binom{N}{k}}$$

Démonstration. Notons $\mathbb{P}$ la loi uniforme sur Ω. On a

$$\mathcal{H}(N,n,k)(i) = \mathbb{P}(\omega \in S),$$

où $S = \{\omega \in \mathcal{B}(N,k); |\{1,\dots,n\} \cap \omega| = i$. L'application

$$\begin{array}{rcl} \mathcal{B}(\{1,\dots,n\}N,i) \times \mathcal{B}(\{n+1,\dots,N\},k-i) & \to & S \\ (A,B) \mapsto A \cup B & & \end{array}$$

est une bijection, donc

$$|S| = |\mathcal{B}(\{1,\dots,n\}N,i) \times \mathcal{B}(\{n+1,\dots,N\},k-i)| = \binom{n}{i}\binom{N-n}{k-i}$$

Comme $\mathbb{P}$ est la loi uniforme sur Ω, et que $|\Omega| = \binom{N}{k}$, le résultat s'ensuit. □

5.6 Lois à densité usuelles

5.6.1 Loi uniforme sur un compact de $\mathbb{R}^d$

On dit qu'une variable aléatoire X suit la loi uniforme sur un compact K de $\mathbb{R}^d$ $[a,b]$ si elle admet la densité

$$x \mapsto \frac{1}{\lambda(K)} \mathbb{1}_K(x)$$

5.6.2 Loi uniforme sur un intervalle

On dit qu'une variable aléatoire X suit la loi uniforme sur l'intervalle $[a, b]$ si elle admet la densité

$$x \mapsto \frac{1}{b-a} 1\!\!1_{[a,b]}(x)$$

5.6.3 Loi gaussienne de paramètres m et σ^2

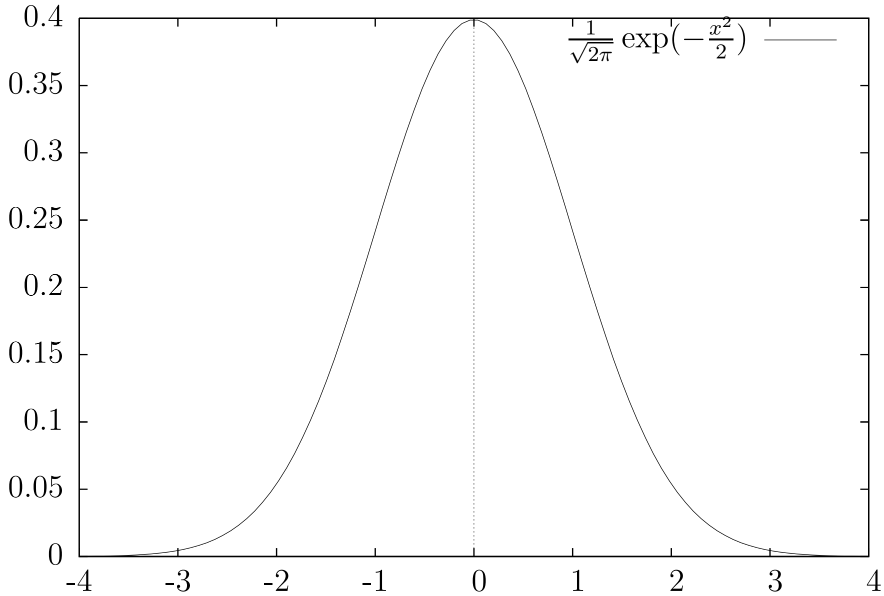
Soit $m \in \mathbb{R}$ et $\sigma^2 > 0$. On dit qu'une variable aléatoire X suit la loi gaussienne $\mathcal{N}(m, \sigma^2)$ de paramètres m et σ^2 si elle admet la densité

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

On emploie également parfois le mot "normale" à la place de "gaussienne" : ces deux mots signifient exactement la même chose. gamma On dit qu'une variable gaussienne est centrée lorsque $m = 0$.
On dit qu'une variable gaussienne est réduite lorsque $\sigma^2 = 1$.

Quelques résultats qui seront prouvés ultérieurement : si $X \sim \mathcal{N}(m, \sigma^2)$, alors $aX + b \sim \mathcal{N}(am + b, a^2\sigma^2)$. En particulier, si $X \sim \mathcal{N}(m, \sigma^2)$, alors $\frac{X-m}{\sigma} \sim \mathcal{N}(0, 1)$; et si $X \sim \mathcal{N}(0, 1)$, alors $\sigma X + m \sim \mathcal{N}(m, \sigma^2)$.

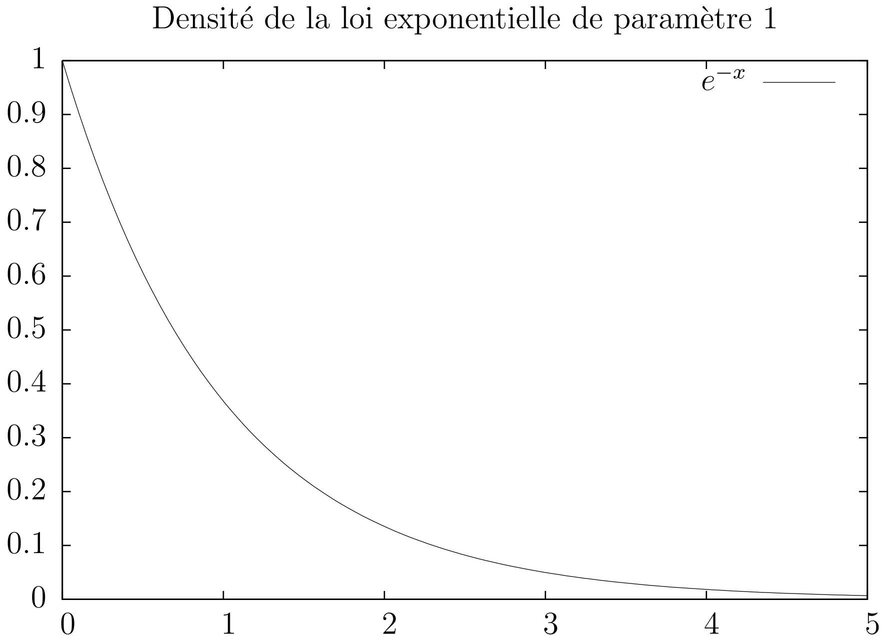
Densité de la loi normale $\mathcal{N}(0, 1)$



5.6.4 Loi exponentielle de paramètres a

Soit $a > 0$. On dit qu'une variable aléatoire X suit la loi exponentielle de paramètres a si elle admet la densité

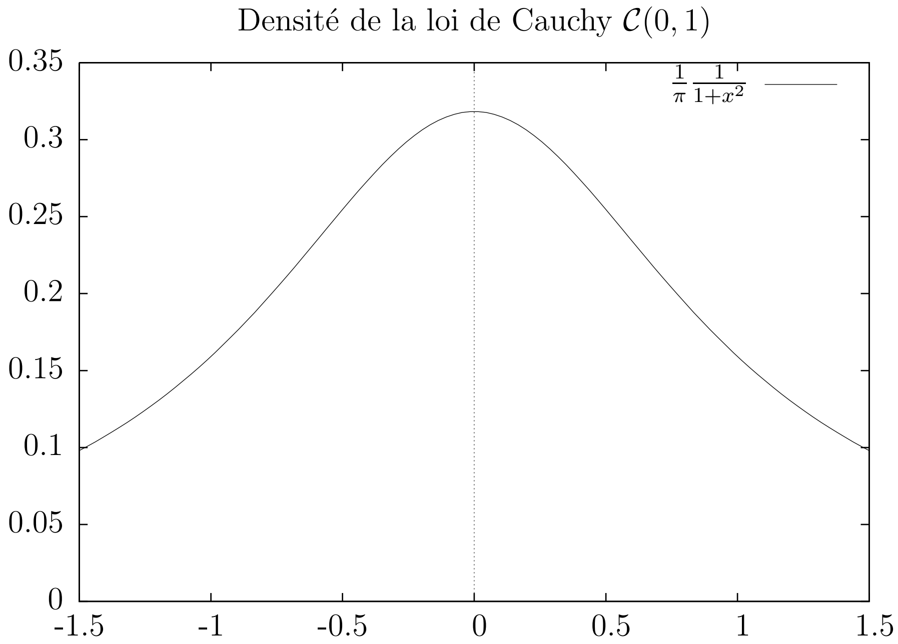
$$x \mapsto a \exp(-ax) \mathbb{1}_{\mathbb{R}^+}(x).$$



5.6.5 Lois de Cauchy

Soient $a \in \mathbb{R}, b > 0$. La loi de Cauchy $\mathcal{C}(a, b)$ admet comme densité par rapport à la mesure de Lebesgue :

$$x \mapsto \frac{1}{\pi} \frac{b}{(x-a)^2 + b^2}.$$



5.6.6 Lois Gamma

Soient a et λ des réels strictement positifs. On appelle loi Gamma $\Gamma(a,\lambda)$ la loi dont la densité par rapport à la mesure de Lebesgue est

$$x \mapsto \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \mathbb{1}_{]0,+\infty[}(x),$$

où $\Gamma(a)$ est la valeur au point a de la fonction Γ, définie par

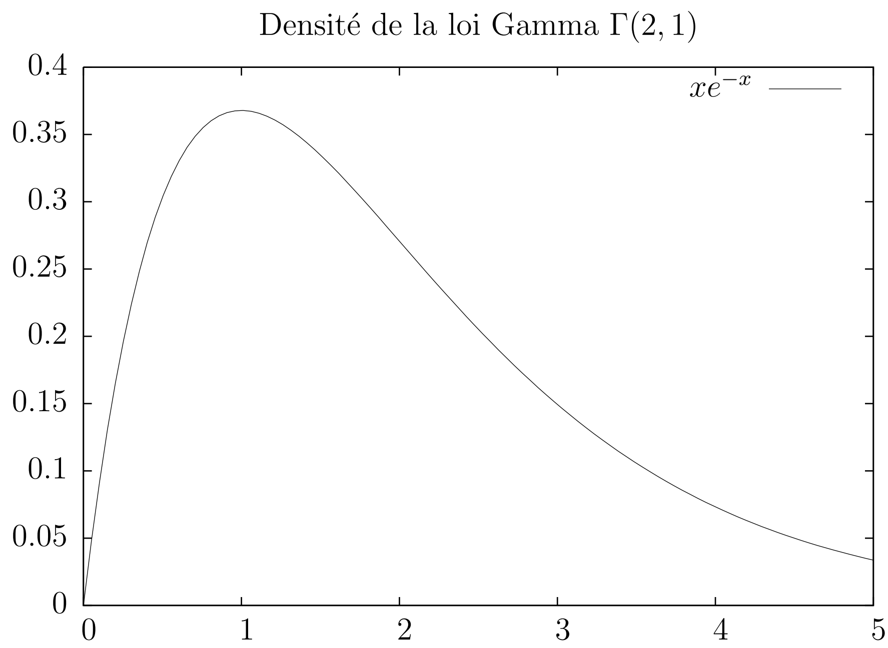
$$\Gamma(a) = \int_{\mathbb{R}+} x^{a-1} e^{-x} \, dx.$$

On dit parfois que a est le paramètre de forme et λ le paramètre d'échelle de la loi. En effet, on montrera plus loin que si $X \sim \Gamma(a,\lambda)$, alors pour tout $\mu > 0$, on a $\frac{1}{\mu}X \sim \Gamma(a, \lambda\mu)$.

On rappelle quelques propriétés classiques de la fonction Γ qui seront utiles dans la suite :

- $\forall a > 0 \quad \Gamma(a+1) = a\Gamma(a)$.
- $\forall n \in \mathbb{N} \quad \Gamma(n+1) = n!$

La preuve de ces deux propriétés sera vue en exercice.



5.7 Exercices sur les lois des variables aléatoires

1. Soit $s>1$. On dit que X suit une loi ζ de paramètre si l'on a

$$\forall n \in \mathbb{N}^* \quad \mathbb{P}(X=n) = \frac{1}{\zeta(s)}\frac{1}{n^s},$$

où l'on a posé

$$\zeta(s) = \sum_{n=1}^{+\infty} \frac{1}{n^s}.$$

Soit donc X suivant une loi ζ de paramètre s. On tire Y au hasard – c'est à dire avec équiprobabilité – entre 1 et X.

(a) Pour $n,k \in \mathbb{N}^*$, calculer $\mathbb{P}(Y=k|X=n)$.

(b) On pose $Z = \frac{Y}{X}$. Montrer que la fonction de répartition F_Z est strictement croissante sur $[0,1]$.

(c) Soient p,q deux entiers positifs premiers entre eux, avec $p \le q$. Calculer $\mathbb{P}(Z = \frac{p}{q})$.

(d) On rappelle que $\varphi(n)$ désigne le nombre d'entiers entre 1 et n qui sont premiers avec n. Déduire de ce qui précède une preuve probabiliste de l'identité

$$\zeta(s+1)\sum_{n=1}^{+\infty}\frac{\varphi(n)}{n^{s+1}} = \zeta(s).$$

(e) Montrer

$$(\sum_{n=1}^{+\infty}\frac{1}{n^2})(\sum_{n=1}^{+\infty}\frac{\varphi(n)}{n^3})(\sum_{n=1}^{+\infty}\frac{\varphi(n)}{n^4}) = \int_2^3 x\ dx.$$

(On pourra admettre que $5\zeta(4) = 2\zeta(2)^2$.)

2. Donner un exemple de familles d'événements $\mathcal{C}$ et $\mathcal{D}$ telles que
   – $\forall A \in \mathcal{C} \quad \forall B \in \mathcal{D} \quad \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
   – les tribus $\sigma(\mathcal{C})$ et $\sigma(\mathcal{D})$ ne sont pas indépendantes.
3. Donner un exemple de deux lois distinctes sur $(\Omega, \mathcal{F})$ coïncidant sur un système $\mathcal{C}$ engendrant $\mathcal{F}$.
4. On choisit de manière uniforme sur $[0,1]$ un réel Y. Quelle est la probabilité pour que le polynome

$$p(x) = x^2 + x + Y$$

   ait des racines complexes ? des racines distinctes ?
5. Dans le segment $[AB]$ de longueur 1, on choisit au hasard un point M. Quelle est la probabilité pour que l'on ait $AM.MB \geq \frac{2}{9}$?
6. Soient $X_1, \dots, X_n, \dots$ des variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On pose $M_n = \max(X_1, \dots, X_n)$. Déterminer la fonction de répartition de M_n. Montrer que M_n admet une densité que l'on déterminera.
7. Soient $X_1, \dots, X_n, \dots$ des variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On pose $M_n = \max(X_1, \dots, X_n)$ et $m_n = \min(X_1, \dots, X_n)$. Montrer que M_n et $1 - m_n$ ont même loi.
8. Soit X une variable aléatoire suivant la loi $\mathcal{B}(2n, 1/2)$. On pose $Y = |X - n|$. Déterminer la loi de Y.
9. Soit (X,Y) un vecteur aléatoire suivant la loi uniforme sur le rectangle $[-1,2] \times [-1,1]$. Montrer que

$$\mathbb{P}(1 - Y \geq 2|X|) = \frac{1}{3}.$$

10. Pour n entier strictement positif, on note $A_n = n\mathbb{N}^*$. Notons $\mathcal{P}$ l'ensemble des nombres premiers positifs et $\mathcal{T}$ la sous-tribu de $(\mathbb{N}^*, \mathcal{B}(\mathbb{N}^*))$ engendrée par les $(A_p)_{p\in\mathcal{P}}$. Pour $\omega \in \mathbb{N}^*$, on pose

$$F(\omega) = \prod_{p\in\mathcal{P};d \text{ divise } \omega} p.$$

   (a) Montrer que $\mathcal{T} = \sigma(X)$.
   (b) Déterminer le plus petit ensemble $\mathcal{T}$-mesurable contenant 1980.
   (c) Montrer que A_n est $\mathcal{T}$-mesurable si et seulement si n n'est divisible par aucun carré.
   (d) On munit $(\mathbb{N}^*, \mathcal{B}(\mathbb{N}^*))$ de la mesure de probabilité ζ de paramètre s, c'est à dire que

$$\forall n \in \mathbb{N}^* \quad \mathbb{P}(\{n\}) = \frac{1}{\zeta(s)}\frac{1}{n^s},$$

   où l'on a posé

$$\zeta(s) = \sum_{n=1}^{+\infty} \frac{1}{n^s}.$$

   Montrer que $\mathbb{P}(A_n) = \frac{1}{n^s}$. À quelle condition les événements A_n et A_m sont-ils indépendants sous la loi $\mathbb{P}$?
   (e) Soit $\mathcal{N} = \{\omega \in \mathbb{N}^*; \quad A_\omega \text{est } \mathcal{T} - \text{mesurable}\}$ Montrer que $\mathcal{N} = \underset{p\in\mathcal{P}}{\cap} A^c_{p^2}$, puis que $0 < \mathbb{P}(\mathcal{N})$.

11. (a) Soit $(\mathcal{A}_n)_{n\geq 1}$ une suite de tribus indépendantes. Montrer que la tribu $\mathcal{Q} = \overset{+\infty}{\underset{k=1}{\cap}} \underset{i\geq k}{\wedge} \mathcal{A}_i$ est triviale.
   (b) Soit $(A_n)_{n\geq 1}$ une suite d'événements indépendants. Soit A l'événement " une infinité de A_i se produisent. Montrer que $\mathbb{P}(A)$ ne peut valoir que 0 ou 1

12. Démontrer les propriétés de la fonction Γ laissées en exercice :
   – $\forall a > 0 \quad \Gamma(a+1) = a\Gamma(a)$.
   – $\forall n \in \mathbb{N} \quad \Gamma(n+1) = n!$

13. *Queue de la gaussienne*
   Soit X une variable aléatoire suivant la loi normale $\mathcal{N}(0,1)$. On pose $\Psi(x) = \mathbb{P}(X > x) = 1 - F_X(x)$. Montrer l''équivalent en l'infini

$$\Psi(x) \sim \frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-\frac{x^2}{2}}.$$

14. La tradition veut que l'Épiphanie soit l'occasion de « tirer les rois » : une fève est cachée dans une galette, découpée entre les convives et la personne qui obtient cette fève devient le roi de la journée. Lorsque le premier coup de couteau est porté sur la fève, c'est la consternation ! Qu'elle est la probabilité de cette malheureuse issue ?
    Hypothèses et simplifications : on admet que la galette est circulaire, de rayon unité, et que la fève est aussi circulaire, le rayon r. Enfin, on suppose que
    - la position du centre de la fève suit la loi uniforme sur le disque de rayon $1 - r$ ayant le même centre que la galette
    - le coup de couteau est un rayon du disque représentant la galette

    Application numérique avec une une fève de 2,7 centimètres de diamètre dans une galette de 23 centimètres de diamètre achetée ce matin.
15. Soient $(X_1, \dots X_n)$ des variables aléatoires indépendantes telles que pour tout i entre 1 et n X_i suit la loi exponentielle $\mathcal{E}(\lambda_i)$. On note $m = \inf(X_1, \dots, X_n)$ et $N = \inf\{i \geq 1; X_i = m\}$.
    (a) Montrer que $\mathbb{P}(\exists(i,j) \in \mathbb{N}^2 \quad 1 \leq i < j \leq n; X_i = X_j) = 0$.
    (b) Montrer que pour tout i entre 1 et n
    $$\mathbb{P}(T > t, N = i) = \mathbb{P}\left(\bigcap_{\substack{1 \leq j \leq n \\ j \neq i}} \{t < X_i < X_j\} \right)$$
    $$= \int_{\mathbb{R}^n} \mathbb{1}_{]t,+\infty[}(x_i) \prod_{\substack{1 \leq j \leq n \\ j \neq i}} \mathbb{1}_{]x_i,+\infty[}(x_j) \prod_{j=1}^{n} \lambda_j e^{-\lambda_j x_j} \, d\lambda^{\otimes n}(x_1, \dots, x_n)$$
    (c) On pose $\lambda = \sum_{j=1}^{n} \lambda_j$. Montrer que $\mathbb{P}(T > t, N = i) = \frac{\lambda_i}{\lambda} \exp(-\lambda t)$.
    (d) Montrer que T et N sont indépendantes et préciser leurs lois.
16. Soit n un entier naturel. On considère X une variable aléatoire exponentielle de paramètre 1 et Y une binomiale $B(n, \frac{1}{2})$. On suppose que X et Y sont indépendantes.
    Montrer que $Z = \frac{X}{Y+1}$ est une variable à densité et déterminer sa densité.

Chapitre 6

Espérances et calculs

6.1 Quelques rappels sur la construction de l'espérance

Définition Si X est une variable aléatoire intégrable définie sur $(\Omega, \mathcal{F}, \mathbb{P})$, on appelle espérance de X et on note $\mathbb{E}X$ le réel défini par

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

Remarque : En toute rigueur, il faudrait écrire $\mathbb{E}_P X$.

Définition On note $\mathcal{L}^1((\Omega, \mathcal{F}, \mathbb{P}))$ l'ensemble des variables aléatoires intégrables sur $(\Omega, \mathcal{F}, \mathbb{P})$.

Définition Si $X = (X_1, \dots, X_n)$ est un vecteur aléatoire dont toutes les composantes sont intégrables, on note $\mathbb{E}X$ le vecteur $(\mathbb{E}X_1, \dots, \mathbb{E}X_n)$.

6.2 Quelques propriétés

- $\mathcal{L}^1$ est un espace vectoriel.
- $\forall X, Y \in \mathcal{L}^1 \quad \mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$.
- $\forall X \in \mathcal{L}^1, \forall a \in \mathbb{R} \quad \mathbb{E}aX = a\mathbb{E}X$.
- $\forall A \in \mathcal{F} \quad \mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$.
- La variable aléatoire X est intégrable si et seulement si $|X|$ est intégrable.
- $\forall X \in \mathcal{L}^1 \quad \mathbb{P}(X \ge 0) = 1 \Longrightarrow \mathbb{E}X \ge 0$
- $\forall X \in \mathcal{L}^1 \quad \mathbb{P}(X \le a) = 1 \Longrightarrow \mathbb{E}X \le a$.
- $\forall X \in \mathcal{L}^1 \quad \mathbb{P}(X \ge b) = 1 \Longrightarrow \mathbb{E}X \ge b$.
- $\forall X \in \mathcal{L}^1 \quad \mathbb{P}(X = a) = 1 \Longrightarrow \mathbb{E}X = a$.
- $\forall X \in \mathcal{L}^1 \quad \mathbb{P}(|X| \le a) = 1 \Longrightarrow \mathbb{E}|X| \le a$.

– Soient X, Y deux variables aléatoires vérifiant $0 \leq X \leq Y$. Si Y est intégrable, alors X est intégrable.
– $\forall X, Y \in \mathcal{L}^1 \quad \mathbb{P}(X \leq Y) = 1 \Longrightarrow \mathbb{E}X \leq \mathbb{E}Y$.

Vocabulaire : on dit qu'une variable aléatoire X est centrée si $\mathbb{E}X = 0$. On définit de même ce qu'est un vecteur aléatoire centré.

6.3 Application : Formule de Poincaré et inégalités de Bonferroni

La formule de Poincaré est l'analogue de la formule du même nom du cours de dénombrement. On peut considérer que c'en est une généralisation.

Théorème 68 (Formule de Poincaré)**.** *Pour tous événements $A_1, A_2, ..., A_n$ sous la probabilité $\mathbb{P}$*

$$\begin{aligned}
\mathbb{P}\Big(\bigcup_{i=1}^{n} A_i\Big) &= \sum_{B\in\mathcal{P}(\{1,\dots,n\})\backslash\varnothing} (-1)^{1+|B|}\mathbb{P}(\cap_{j\in B}A_j) && (6.1)\\
&= \sum_{i=1}^{n}\mathbb{P}(A_i) - \sum_{1\leq i_1<i_2\leq n}\mathbb{P}(A_{i_1}\cap A_{i_2}) + \cdots && (6.2)\\
&\quad \cdots + (-1)^{k+1}\sum_{1\leq i_1<i_2<\dots<i_k\leq n}\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k}) + \cdots && (6.3)\\
&\quad \cdots + (-1)^{n+1}\mathbb{P}(A_1\cap\cdots\cap A_n). && (6.4)
\end{aligned}$$

Exemple : Pour $n = 3$, on a

$$\begin{aligned}
\mathbb{P}(A_1\cup A_2\cup A_3) &= \mathbb{P}(A_1)+\mathbb{P}(A_2)+\mathbb{P}(A_3)\\
&\quad -\mathbb{P}(A_1\cap A_2)-\mathbb{P}(A_2\cap A_3)-\mathbb{P}(A_1\cap A_3)\\
&\quad +\mathbb{P}(A_1\cap A_2\cap A_3).
\end{aligned}$$

Pour prouver la formule de Poincaré, on va utiliser un lemme qui va nous permettre d'obtenir des encadrements de la probabilité d'une réunion.

Lemme 6. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé ; $(A_x)_{x\in I}$ des événements.*

Pour $n \geq 1$, on pose

$$V_n = P_n\Big(\sum_{x\in I}\mathbb{1}_{A_x} - 1\Big)\mathbb{1}_{\cup_{x\in I}A_x},$$

où $(P_k)_{k\geq 0}$ est la suile de polynômes définie par

$$\begin{cases} P_0 = 1 \\ P_1 = X \\ P_2 = \frac{X(X-1)}{2} \\ \dots \\ P_k = \frac{X(X-1)\dots(X-k+1)}{k!} \end{cases}$$

Ainsi pour $n \geq k \quad P_k(n) = \binom{n}{k}$ *tandis que* $P_k(n) = 0$ *pour* $0 \leq n < k$.
Alors

$$\forall n \in \mathbb{N}^* \quad \mathbb{P}(\bigcup_{x\in I} A_x) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} P(\bigcap_{x\in J} A_x) + (-1)^n \mathbb{E}V_n \quad (6.5)$$

Démonstration. Il suffit de montrer

$$\forall n \in \mathbb{N}^* \quad \mathbb{1}_{\bigcup\limits_{x\in I} A_x} = \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} \prod_{x\in J} \mathbb{1}_{A_x} + (-1)^n V_n \qquad (6.6)$$

Soit $\omega \in \Omega$. Si $\omega \notin \cup_{x\in I} A_x$, les deux membres de l'égalité sont nuls. Sinon, posons $N = \sum_{x\in I} \mathbb{1}_{A_x}(\omega) = |\{x \in I; \omega \in A_x\}|$.

On a $\prod\limits_{x\in J} \mathbb{1}_{A_x} = 1$ si et seulement si $J \subset \{x \in I; \omega \in A_x\}$.

Ainsi $\sum\limits_{J\in\mathcal{B}_k(I)} \prod\limits_{x\in J} \mathbb{1}_{A_x} = P_k(N)$

On doit donc montrer

$$1 = \sum_{k=1}^{n} (-1)^{k+1} P_k(N) + (-1)^n P_n(N-1),$$

ce qui est équivalent à

$$\sum_{k=0}^{n} (-1)^k P_k(N) = (-1)^n P_n(N-1).$$

Pour conclure, deux preuves possibles :
Méthode 1 : On va montrer par récurrence sur n :

$$\sum_{k=0}^{n} (-1)^k P_k(X) = (-1)^n P_n(X-1)$$

Pour $n = 0$, c'est vérifié. Pour passer de n à $n+1$, on a

$$\begin{aligned}
\sum_{k=0}^{n+1}(-1)^k P_k(X) &= \sum_{k=0}^{n}(-1)^k P_k(X) + (-1)^{n+1}P_{n+1}(X) \\
&= (-1)^n P_n(X-1) + (-1)^{n+1}P_{n+1}(X) \\
&= (-1)^{n+1}(P_{n+1}(X) - P_n(X-1)) \\
&= (-1)^{n+1}(\frac{1}{n+1}XP_n(X-1) - P_n(X-1)) \\
&= (-1)^{n+1}\frac{(X-(n+1))P_n(X-1)}{n+1} \\
&= (-1)^{n+1}P_{n+1}(X-1)
\end{aligned}$$

Méthode 2 : De manière équivalente, il faut montrer que

$$\sum_{k=0}^{n}(-1)^{n-k}P_k(N) = P_n(N-1).$$

Mais on reconnaît en $\sum_{k=0}^{n}(-1)^{n-k}P_k(N)$ le coefficient en x^n de la série entière sur $B(0,1)$:

$$\begin{aligned}
&\left(\sum_{k=0}^{+\infty} P_k(N)x^k\right)\left(\sum_{k=0}^{+\infty}(-1)^k x^k\right) \\
&= \left(\sum_{k=0}^{N}\binom{N}{k}x^k\right)\frac{1}{1+x} \\
&= (1+x)^N\frac{1}{1+x} \\
&= (1+x)^{N-1}
\end{aligned}$$

dont le coefficient en x^n est précisément $P_n(N-1)$, d'où l'identification des termes. $\square$

Ensuite, il suffit d'intégrer (6.6) pour obtenir (6.5).

Démonstration. En prenant $I = J = \{1,\ldots,n\}$ dans le lemme précédent, on obtient

$$\mathbb{P}(\underset{x\in I}{\cup} A_x) = \sum_{k=1}^{|I|}(-1)^{k+1}\sum_{J\in\mathcal{B}_k(I)}\mathbb{P}(\underset{x\in J}{\cap} A_x),$$

car V_n est identiquement nulle. Cela démontre la formule de Poincaré. $\square$

Théorème 69 (Inégalités de Bonferroni). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé ; $(A_x)_{x\in I}$ des événements. Soit $n \in \mathbb{N}^*$.*

Alors,

- *Si n est impair, on a*

$$\mathbb{P}(\underset{x\in I}{\cup} A_x) \leq \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} \mathbb{P}(\underset{x\in J}{\cap} A_x) \tag{6.7}$$

- *Si n est pair, on a*

$$\mathbb{P}(\underset{x\in I}{\cup} A_x) \geq \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} \mathbb{P}(\underset{x\in J}{\cap} A_x) \tag{6.8}$$

Démonstration. Il suffit d'appliquer le lemme 6 en remarquant que V_n est une variable aléatoire positive, et que donc son espérance l'est aussi. □

6.3.1 Application aux problème des dérangements

On reprend le problème des dérangements, qui avait été étudié en exercice au chapitre 3 par des méthodes d'algèbre linéaire. On note $\Omega = \mathcal{S}_n$ l'ensemble des permutations de $I = \{1, \dots, n\}$, que l'on munit de la loi uniforme. On cherche donc à calculer $p_n = \frac{d_n}{n!}$, où d_n est le nombre de permutations de $\mathcal{S}_n$ sans point fixe :

$$d_n = \mathrm{Card}(\{\sigma \in \mathcal{S}_n ; \forall i \in 1, \dots, n \quad \sigma(i) \neq i\}).$$

(On pose $d_0 = 1$.) Pour i entre 1 et n, posons $A_i = \{\sigma(i) = i\}$. Ce que nous cherchons, c'est

$$p_n = \mathbb{P}(\overset{n}{\underset{i=1}{\cap}} A_i^c) = 1 - \mathbb{P}(\overset{n}{\underset{i=1}{\cup}} A_i),$$

d'où, avec la formule de Poincaré

$$p_n = 1 - \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} \mathbb{P}(\underset{x\in J}{\cap} A_x).$$

Mais il est facile de déterminer $\underset{x\in J}{\cap} A_x$: ce sont les permutations de I qui fixent les points de J et qui permutent les points de $I\backslash J$: ce sous-ensemble de permutations de $\mathcal{S}_n$ est en bijection avec $\mathcal{S}(I\backslash J)$; il est donc de cardinal $(n-k)!$, d'où, avec l'hypothèse d'équiprobabilité, $\mathbb{P}(\underset{x\in J}{\cap} A_x) = \frac{(n-k)!}{k!}$.

Finalement,

$$\begin{aligned} p_n &= 1 - \sum_{k=1}^{n} (-1)^{k+1} \sum_{J \in \mathcal{B}_k(I)} \frac{(n-k)!}{k!} \\ &= 1 + \sum_{k=1}^{n} (-1)^k \binom{n}{k} \frac{(n-k)!}{k!} \\ &= \sum_{k=0}^{n} \frac{(-1)^k}{k!} \end{aligned}$$

6.4 Théorèmes de transfert

Théorème 70. *Soit X est une variable aléatoire dont la loi $\mathbb{P}_X$ admet une densité f par rapport à la mesure m. Soit g une fonction mesurable.*

Alors, $g(X)$ est intégrable si et seulement si

$$\int_{\mathbb{R}} |g(x)| f(x)\ dm(x) < +\infty.$$

Si cette intégrale est finie, on a alors

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} |g(x)| f(x)\ dm(x) < +\infty.$$

Démonstration. D'après le théorème de transfert

$$\begin{aligned} \int_{\Omega} |g(X(\omega))| d\mathbb{P}(\omega) &= \int_{\mathbb{R}} |g(x)|\ d\mathbb{P}_X(x) \\ &= \int_{\mathbb{R}} |g(x)| f(x)\ dm(x) \end{aligned}$$

De même, si cette quantité est finie, le théorème de transfert nous dit encore que

$$\begin{aligned} \int_{\Omega} g(X(\omega)) d\mathbb{P}(\omega) &= \int_{\mathbb{R}} g(x)\ d\mathbb{P}_X(x) \\ &= \int_{\mathbb{R}} g(x) f(x)\ dm(x) \end{aligned}$$

□

6.4.1 Calcul de l'espérance d'une variable aléatoire discrète

Théorème 71. *Soit X est une variable aléatoire discrète, et D un ensemble fini ou dénombrable inclus dans $\mathbb{R}$ tel que $X(\Omega) = D$. Soit g une fonction quelconque de D dans $\mathbb{R}$. Alors, la variable aléatoire $Y = g(X)$ est intégrable si et seulement si*

$$\sum_{i \in D} |g(i)| p_i < +\infty,$$

où l'on a posé $p_i = \mathbb{P}(X = i)$. Si cette somme est finie, on a alors

$$\mathbb{E}Y = \mathbb{E}g(X) = \sum_{i \in D} g(i) p_i.$$

Démonstration. D'après nos hypothèses, $\mathbb{P}_X$ admet une densité par rapport à la mesure de comptage de support D : c'est la fonction $i \mapsto p_i$. Il suffit donc d'appliquer le théorème 70 en prenant pour m la mesure de comptage sur D et pour f :

$$f(x) = \begin{cases} p_x & \text{si } x \in D \\ 0 \text{ si } x \notin D \end{cases}$$

On a alors

$$\int_{\mathbb{R}} |g(x)| f(x)\ dm(x) = \sum_{x \in D} |g(x)| f(x),$$

et, si cette somme est finie :

$$\int_{\mathbb{R}} g(x) f(x)\ dm(x) = \sum_{x \in D} g(x) f(x).$$

□

Corollaire 12. *Soit X est une variable aléatoire discrète, et D un ensemble fini ou dénombrable inclus dans $\mathbb{R}$ tel que $X(\Omega) = D$. Alors, X est intégrable si et seulement si*

$$\sum_{i \in D} |i| p_i < +\infty,$$

où l'on a posé $p_i = \mathbb{P}(X = i)$. Si cette somme est finie, on a alors

$$\mathbb{E}X = \sum_{i \in D} i p_i.$$

Démonstration. Il suffit d'appliquer le thèorème précédent avec $g(x) = x$. □

6.4.2 Calcul de l'espérance d'une variable aléatoire à densité

Voici maintenant le théorème de transfert pour les variables à densité

Théorème 72. *Soit X est une variable aléatoire admettant la fonction f comme densité, et g une fonction continue par morceaux définie sur $X(\Omega)$. Alors, la variable aléatoire $Y = g(X)$ est intégrable si et seulement si*

$$\int_{\mathbb{R}} |g(x)|f(x)\ d\lambda(x) < +\infty.$$

Si cette intégrale est convergente, on a alors

$$\mathbb{E}Y = \mathbb{E}g(X) = \int_{\mathbb{R}} g(x)f(x)\ d\lambda(x).$$

Démonstration. Il suffit d'appliquer le théorème 70 avec pour m la mesure de Lebesgue. □

Corollaire 13. *Soit X est une variable aléatoire admettant la fonction f comme densité. Alors, X est intégrable si et seulement si*

$$\int_{\mathbb{R}} |x|f(x)\ d\lambda(x) < +\infty.$$

Si cette intégrale est convergente, on a alors

$$\mathbb{E}X = \int_{\mathbb{R}} xf(x)\ dx.$$

Démonstration. Il suffit d'appliquer le thèorème précédent avec $g(x) = x$. □

Remarque importante : Si la densité de X est paire et que X est intégrable, alors $\mathbb{E}X = 0$.

Démonstration. On a

$$\mathbb{E}X = \int_{\mathbb{R}} xf(x)\ d\lambda(x),$$

et avec le théorème de changement de variable,

$$\mathbb{E}X = \int_{\mathbb{R}} -xf(-x)\ d\lambda(x) = \int_{\mathbb{R}} -xf(x)\ d\lambda(x),$$

par parité de f. D'où finalement $\mathbb{E}X = -\mathbb{E}X$, c'est à dire $\mathbb{E}X = 0$. □

6.5 Convexité

6.5.1 Rappels sur la convexité

On dit qu'une fonction est convexe sur l'intervalle I si pour tous x, y dans I et $\theta \in [0,1]$, on a $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$.

Lemme 7. *Si $x < z < y$, alors il existe $\theta \in]0,1[$ avec $z = \theta x + (1-\theta)y$.*

Démonstration. Facile. □

Lemme 8. *Les pentes d'une fonction convexe sont croissantes : si $h_1 < h_2$ avec $x, x+h_1, x+h_2$ dans I, avec h_1 et h_2 non nuls, alors*

$$p_x(h_1) = \frac{f(x+h_1) - f(x)}{h_1} \leq p_x(h_2) = \frac{f(x+h_2) - f(x)}{h_2}.$$

Si x n'est pas la borne droite (resp. gauche) de I, p_x admet une limite à droite en 0 : c'est la dérivée à droite (resp. à gauche) de f en x.

Démonstration. Premier cas : h_1 et h_2 sont positifs. On applique le lemme avec $z = x + h_1$ et $y = x + h_2$ et on utilise la définition de la convexité. Deuxième cas : h_1 et h_2 sont négatifs. Même chose que dans le cas précédent, avec $x' = x + h_1$, $z' = x + h_2$, $y' = x$. Dernier cas : $h_1 < 0 < h_2$ dans ce cas $p_x(h_1) \leq p_x(\max(h_1, -h_2))$ et $p_x(\min(h_2, -h_1)) \leq ph(h_2)$ d'après ce qui précède, donc il suffit de vérifier que $p_x(-h) \leq p_x(h)$, ce qui découle immédiatement de la convexité. Ceci achève la preuve de la croissance. L'existence de la limite est une conséquence du théorème sur les fonctions croissantes minorées à droite pour la limite à droite, et croissantes majorées à gauche pour la limite à gauche. □

Lemme 9. *Soit f convexe avec $f(x) = f(y) = 0$. f est négative sur $[x,y]$, positive à l'extérieur.*

Démonstration. Si z est entre x et y avec $z = \theta f(x) + (1-\theta)f(y)$, alors $f(z) \leq \theta f(x) + (1-\theta)f(y) \leq 0$. Sinon, l'inégalité

$$\theta f(a) + (1-\theta)f(b) \geq f(\theta a + (1-\theta)b)$$

entraine que si de trois nombres, le terme médian et un autre ont une image nulle, le troisième doit avoir une image positive. □

Lemme 10. *On appelle corde portée par x et y la droite passant par $(x, f(x))$ et $(y, f(y))$, d'équation $\ell(t) = \frac{f(y)-f(x)}{y-x}(t-x) + f(x)$. La fonction f est en-dessous de la corde entre x et y, au-dessus à l'extérieur.*

Démonstration. Il suffit d'appliquer le lemme précédent à $t \mapsto f(t) - \ell(t)$ qui est convexe. □

Lemme 11. *Si x n'est pas la borne droite de I, alors pour tout t dans I, $f(t) \geq f(x) + f'_d(x)(t-x)$.*

Démonstration. Soit $h > 0$ Pour tout t qui n'est pas dans $]x, x+h[$, la propriété de la corde donne

$$f(t) \geq \frac{f(x+h) - f(x)}{h}(t-x) + f(x) \geq f'_d(x)(t-x) + f(x),$$

ce qui est l'inégalité que l'on veut. Mais pour tout t il existe $h > 0$ tel que t qui n'est pas dans $]x, x+h[$, d'où le résultat. □

6.5.2 Inégalité de Jensen

Théorème 73. *Soit X une variable aléatoire intégrable à valeurs dans l'intervalle I. Soit f une fonction convexe de I dans $\mathbb{R}$. Alors*

$$f(\mathbb{E}X) \leq \mathbb{E}f(X).$$

Démonstration. Soit Φ l'ensemble des fonctions affines φ telles que

$$\forall x \in I \quad \varphi(x) \leq f(x).$$

Soit $\varphi \in \Phi$. On a presque sûrement

$$\varphi(X) \leq f(X).$$

On a donc $\mathbb{E}\varphi(X) \leq \mathbb{E}f(X)$. Mais comme φ est une fonction affine, $\mathbb{E}\varphi(X) = \varphi(\mathbb{E}X)$. Ainsi,

$$\forall \varphi \in \Phi \quad \varphi(\mathbb{E}X) \leq \mathbb{E}f(X),$$

Prenons alors $\varphi(t) = f(\mathbb{E}X) + f'_d(\mathbb{E}X)(t-x)$ (ou $\varphi(t) = f(\mathbb{E}X) + f'_g(\mathbb{E}X)(t-x)$ si $\mathbb{E}X$ est l'extrémité à droite de I). D'après le lemme 11, $\varphi \in \Phi$, ce qui donne

$$f(\mathbb{E}X) \leq \mathbb{E}f(X).$$

□

Pour retenir quel est le sens de l'égalité, prendre la fonction convexe $\varphi(x) = |x|$.

Corollaire 14. *Soient f une fonction convexe sur l'intervalle I, $\theta_1, \dots \theta_n$ des réels positifs de somme 1, $x_1, \dots, x_n$ des éléments de I. Alors*

$$f\left(\sum_{k=1}^{n} \theta_k x_k\right) \leq \sum_{k=1}^{n} \theta_k f(x_k).$$

Démonstration. Il suffit de considérer unve rariable aléatoire discrète X telle que $\mathbb{P}(X = x_i) = \theta_i$ pour tout i entre 1 et n. Le théorème de transfert pour une variable aléatoire discrète et l'inégalité de Jensen donnent

$$f\left(\sum_{k=1}^{n} \theta_k x_k\right) = f(\mathbb{E}X) \leq \mathbb{E}[f(x)] = \sum_{k=1}^{n} \theta_k f(x_k).$$

□

6.6 Intégrale et queue de distribution

Théorème 74. *Soit X une variable aléatoire positive. On a*

$$\mathbb{E}X = \int_{\mathbb{R}_+} \mathbb{P}(X > t)\ d\lambda(t).$$

Démonstration.

$$\begin{aligned}
\int_{\mathbb{R}_+} \mathbb{P}(X > t)\ d\lambda(t) &= \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} 1\!\!1_{]t,+\infty[}(s)\ d\mathbb{P}_X(s)\ d\lambda(t) \\
&= \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} 1\!\!1_{[0,s[}(t)\ d\mathbb{P}_X(s)\ d\lambda(t) \\
&= \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} 1\!\!1_{[0,s[}(t)\ d\lambda(t)\ d\mathbb{P}_X(s) \\
&= \int_{\mathbb{R}_+} s\ d\mathbb{P}_X(s) \\
&= \mathbb{E}X
\end{aligned}$$

□

Corollaire 15. *Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. On a*

$$\mathbb{E}X = \sum_{k=0}^{+\infty} \mathbb{P}(X > k).$$

Démonstration. D'après le théorème précédent, $\mathbb{E}X = \int_{\mathbb{R}_+} \mathbb{P}(X > t)\ d\lambda(t)$. Comme $t \mapsto \mathbb{P}(X > t)$ est une fonction positive, on a

$$\int_{\mathbb{R}_+} \mathbb{P}(X > t)\ d\lambda(t) = \sum_{k=0}^{+\infty} \int_{[k,k+1[} \mathbb{P}(X > t)\ d\lambda(t).$$

Mais comme X est à valeurs entières, on a

$$\forall t \in [k, k+1[\quad \mathbb{P}(X > t) = \mathbb{P}(X > k),$$

d'où le résultat. □

6.7 Moments d'ordre 2

On dit qu'une variable aléatoire X admet un moment d'ordre 2 si elle est de carré intégrable, c'est à dire si $X^2 \in \mathcal{L}^1$.

On note $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ (ou encore $\mathcal{L}^2$) l'ensemble des variables aléatoires de carré intégrable.

Lemme 12. *Soient $X, Y \in \mathcal{L}^2$. Alors la variable aléatoire XY est intégrable.*

Démonstration. Pour tous les réels a, b, on a $|ab| \leq \frac{1}{2}(a^2 + b^2)$.
(En effet, $a^2+b^2+2ab = (a+b)^2 \geq 0$, d'où $(a^2+b^2)/2 \geq -ab$ et $a^2+b^2-2ab = (a-b)^2 \geq 0$, d'où $(a^2+b^2)/2 \geq ab$.)
On a donc

$$0 \leq |XY| \leq \frac{1}{2}(X^2 + Y^2),$$

Comme $X^2 + Y^2$ est intégrable, on en déduit que $|XY|$ est intégrable, ce qui est ce que l'on voulait montrer □

Corollaire 16. $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ *est un espace vectoriel.*

Démonstration. La stabilité par multiplication ne pose pas de problème. Pour la stabilité par addition, il faut remarquer que

$$(X + Y)^2 = X^2 + Y^2 + 2XY,$$

puis utiliser le lemme précédent et le fait que $\mathcal{L}^1$ est un espace vectoriel. □

6.7.1 Covariance et variance

Soient X et Y deux variables aléatoires admettant chacune un moment d'ordre 2. On appelle covariance du couple (X, Y) le nombre

$$\text{Covar}(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)$$

On appelle variance de X le nombre

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

On appelle écart-type de X le nombre

$$\sigma(X) = (\text{Var}\, X)^{1/2}.$$

Vocabulaire : On dit qu'une variable aléatoire est réduite si on a $\text{Var}\, X = 1$ (ou de manière équivalente si $\sigma(X) = 1$).

On a les propriétés suivantes

1. $(X, Y) \mapsto \text{Covar}(X, Y)$ est une forme bilinéaire symétrique positive.
2. $\forall a, b \in \mathbb{R} \quad \text{Covar}(X - a, Y - b) = \text{Covar}(X, Y)$.
3. $\text{Var}(X + Y) = \text{Var}\, X + \text{Var}\, Y + 2\, \text{Covar}(X, Y)$.
4. $\text{Covar}(X, Y) = \mathbb{E}XY - (\mathbb{E}X)(\mathbb{E}Y)$.
5. $\text{Var}\, X = \mathbb{E}X^2 - (\mathbb{E}X)^2$.
6. $|\mathbb{E}XY|^2 \leq \mathbb{E}X^2 \mathbb{E}Y^2$ (inégalité de Cauchy-Schwarz)
7. $|\text{Covar}(X, Y)| \leq \sigma(X)\sigma(Y)$.

Démonstration.

1. Notons $\langle X, Y \rangle = \mathbb{E}X, Y$. Il est facile de voir que $(X, Y) \mapsto \langle X, Y \rangle$ est une forme bilinéaire symétrique positive. Posons $L(X) = X - \mathbb{E}X$. $X \mapsto L(X)$ est une application linéaire de $\mathcal{L}^2$ dans lui-même. On a $\text{Covar}(X, Y) = \langle L(X), L(Y) \rangle$. Les deux observations faites ci-dessus permettent de dire que $(X, Y) \mapsto \text{Covar}(X, Y)$ est une forme bilinéaire symétrique positive.

2.
$$\begin{aligned} \text{Covar}(X - a, Y - b) &= \langle L(X - a), L(Y - b) \rangle \\ &= \langle L(X) - L(a), L(Y) - L(b) \rangle \\ &= \langle L(X) - 0, L(Y) - 0 \rangle \\ &= \text{Covar}(X, Y) \end{aligned}$$

3.
$$\begin{aligned}\operatorname{Var}(X+Y) &= \operatorname{Covar}(X+Y, X+Y) \\ &= \operatorname{Covar}(X,X) + 2\operatorname{Covar}(X,Y) + \operatorname{Covar}(Y,Y) \\ &= \operatorname{Var} X + 2\operatorname{Covar}(X,Y) + \operatorname{Var} Y\end{aligned}$$

   Pour passer de la ligne 1 à la ligne 2, on utilise le fait que la covariance est bilinéaire symétrique.

4. $(X - \mathbb{E}X)(Y - \mathbb{E}Y) = XY + \mathbb{E}X\mathbb{E}Y - (\mathbb{E}X)Y - (\mathbb{E}Y)X$.
   D'où
$$\begin{aligned}\mathbb{E}(X-\mathbb{E}X)(Y-\mathbb{E}Y) &= \mathbb{E}XY + \mathbb{E}(EX\mathbb{E}Y) - \mathbb{E}X\mathbb{E}Y - \mathbb{E}Y\mathbb{E}X \\ &= \mathbb{E}XY + \mathbb{E}X\mathbb{E}Y - 2\mathbb{E}X\mathbb{E}Y \\ &= \mathbb{E}XY - (\mathbb{E}X)(\mathbb{E}Y).\end{aligned}$$

5. Il suffit d'appliquer la formule précédente avec $X = Y$.
6. Comme $\langle ., . \rangle$ est une forme bilinéaire symétrique positive, l'inégalité de Cauchy-Schwarz s'applique.
7. Il suffit d'appliquer l'inégalité de Cauchy-Schwarz à la forme bilinéaire symétrique positive Covar , puis de prendre la racine carrée.

$\square$

Lorsque $\sigma(X)$ et $\sigma(Y)$ sont non nuls, on définit le coefficient de corrélation de X et Y par

$$\operatorname{Corr}(X,Y) = \frac{\operatorname{Covar}(X,Y)}{\sigma(X)\sigma(Y)}.$$

D'après ce qui précède, $\operatorname{Covar}(X,Y) \in [-1;1]$. Lorsque $\operatorname{Covar}(X,Y) = 0$ (ce qui implique $\operatorname{Corr}(X,Y) = 0$ si $\sigma(X)$ et $\sigma(Y)$ sont non nuls), on dit que X et Y ne sont pas corrélées.
Lorsque $\operatorname{Covar}(X,Y) \geq 0$ (ce qui implique $\operatorname{Corr}(X,Y) \geq 0$ si $\sigma(X)$ et $\sigma(Y)$ sont non nuls), on dit que X et Y sont positivement corrélées.
Lorsque $\operatorname{Covar}(X,Y) \leq 0$ (ce qui implique $\operatorname{Corr}(X,Y) \leq 0$ si $\sigma(X)$ et $\sigma(Y)$ sont non nuls), on dit que X et Y sont négativement corrélées.

6.7.2 Matrice de covariance

Si $X = (X_1, \dots, X_n)$ est un vecteur aléatoire dont toutes les composantes admettent un moment d'ordre deux, on convient de dire que le vecteur a un moment d'ordre deux et on appelle matrice de covariance de X la matrice $n \times n$ dont les coefficients sont $(\operatorname{Covar}(X_i, X_j))_{1 \leq i,j \leq n}$.

Théorème 75. *Si $X = (X_1, \dots, X_n)$ est un vecteur aléatoire admettant un moment d'ordre deux , la matrice de covariance de X est la matrice dans la base canonique de l'application bilinéaire positive*

$$\begin{array}{rcl} \mathbb{R}^n \times \mathbb{R}^n & \to & \mathbb{R} \\ (a,b) & \mapsto & \mathrm{Covar}(\langle X, a\rangle, \langle X, b\rangle) \end{array}$$

C'est une matrice symétrique positive.

Démonstration. À X fixé, l'application $X \mapsto \langle X, a\rangle$ est une application linéaire. Comme on a déjà montré que Covar était une forme bilinéaire symétrique positive, il s'ensuit que l'application considérée ici est une forme bilinéaire symétrique positive. Cette application envoie le couple (e_i, e_j) sur $\mathrm{Covar}(\langle X, e_i\rangle, \langle X, e_j\rangle) = \mathrm{Covar}(X_i, X_j)$. La matrice d'une forme bilinéaire symétrique positive est une matrice symétrique positive. □

Théorème 76. *Soit $X = (X_1, \dots, X_n)$ est un vecteur aléatoire admettant un moment d'ordre deux et de matrice de covariance C_X et d'espérance m_X. Soit A une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$, et b un vecteur de $\mathbb{R}^p$. Alors $Y = AX + b$ admet $C_Y = AC_XA^*$ comme matrice de covariance et l'espérance de Y vaut $Am_X + b$.*

Démonstration.

$$\begin{aligned} \mathbb{E}Y_i &= \mathbb{E}(\sum_{k=1}^n a_{i,k}X_k + b_i) \\ &= \sum_{k=1}^n a_{i,k}\mathbb{E}X_k + b_i \\ &= \sum_{k=1}^n a_{i,k}m_k + b_i \\ &= (Am + b)_i \end{aligned}$$

$$\begin{aligned} \mathrm{Covar}(\langle Y, a\rangle, \langle Y, b\rangle) &= \mathrm{Covar}(\langle AX + c, a\rangle, \langle AX + c, b\rangle) \\ &= \mathrm{Covar}(\langle AX, a\rangle, \langle AX, b\rangle) \\ &= \mathrm{Covar}(\langle X, A^*a\rangle, \langle X, A^*b\rangle) \\ &= \langle C_XA^*a, A^*b\rangle \\ &= \langle AC_XA^*a, b\rangle \end{aligned}$$

□

6.7.3 Espérance et indépendance

Le théorème suivant est très important :

Théorème 77. *Soient X, Y deux variables aléatoires intégrables indépendantes. Alors, leur produit XY est une variable aléatoire intégrable et l'on a*

$$\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y.$$

Démonstration. D'après le théorème de transfert, on a

$$\mathbb{E}|XY| = \int_{\mathbb{R}^2} |xy| \, dP_{(X,Y)}$$

Il vient

$$\begin{aligned}
\mathbb{E}|XY| &= \int_{\mathbb{R}^2} |xy| \, dP_{(X,Y)} \\
&= \int_{\mathbb{R}^2} |x|.|y| \, d(P_X \otimes P_Y)(x,y) \\
&= \int_{\mathbb{R}} |x| dP_X(x). \int_{\mathbb{R}} |y| \, dP_Y(y) \\
&= \mathbb{E}|X|.\mathbb{E}|Y| \\
&< +\infty
\end{aligned}$$

Ainsi, le théorème de Fubini nous a permis de montrer que XY était intégrable. Maintenant, et on a

$$\begin{aligned}
\mathbb{E}XY &= \int_{\mathbb{R}^2} xy \, dP_{(X,Y)} \\
&= \int_{\mathbb{R}^2} xy \, d(P_X \otimes P_Y)(x,y) \\
&= \int_{\mathbb{R}} x dP_X(x). \int_{\mathbb{R}} y \, dP_Y(y) \\
&= \mathbb{E}X\mathbb{E}Y
\end{aligned}$$

□

Corollaire 17. *Soient X, Y deux variables aléatoires intégrables indépendantes. Alors X et Y ne sont pas corrélées.*

Démonstration. On a $\mathrm{Covar}(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = 0$. □

Remarque importante : Des variables aléatoires peuvent être non corrélées sans être indépendantes.
Exemple : soient deux variables aléatoires vérifiant

$$\mathbb{P}(\{X=1\}\cap\{Y=1\}) = \mathbb{P}(\{X=1\}\cap\{Y=-1\}) = \mathbb{P}(\{X=-1\}\cap\{Y=0\}) = 1/3.$$

La matrice M associée à la loi du couple est

$$\begin{pmatrix} 0 & 1/3 & 0 \\ 1/3 & 0 & 1/3 \end{pmatrix}$$

La loi de Y s'obtient en faisant la somme des lignes : on obtient

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 \end{pmatrix}$$

On a donc $\mathbb{E}Y = \frac{1}{3}\times(-1) + \frac{1}{3}\times(0) + \frac{1}{3}\times(1) = 0$.
D'autre part $\mathbb{E}XY = \sum_{i\in\{-1;1\}}\sum_{j\in\{-1;0;1\}} ij\mathbb{P}(\{X=i\}\cap\{Y=j\}) = 1/3 - 1/3 = 0$.
On a donc $\operatorname{Covar}(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = 0$.
Cependant

$$0 = \mathbb{P}(\{X=1\}\cap\{Y=0\}) \neq \mathbb{P}(X=1)\mathbb{P}(Y=0) = \frac{2}{3}\times\frac{1}{3}.$$

Corollaire 18. *Soient X,Y deux variables aléatoires indépendantes de carré intégrable. Alors on a*

$$\operatorname{Var}(X+Y) = \operatorname{Var} X + \operatorname{Var} Y.$$

Démonstration. On a toujours $\operatorname{Var} X+Y = \operatorname{Var} X + \operatorname{Var} Y + 2\operatorname{Covar}(X,Y)$.Comme X et Y sont indépendantes, elles ne sont pas corrélées, d'où le résultat. □

6.8 Lois images par des transformations affines

6.8.1 Exemple fondamental

Théorème 78. *Soit $A \in Gl_d(\mathbb{R})$ et $b \in \mathbb{R}^d$. On suppose que le vecteur aléatoire X admet la densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. Alors, le vecteur aléatoire $Y = AX + b$ admet la densité*

$$g(y) = \frac{1}{|\det A|} f(A^{-1}(y-b)).$$

Démonstration. C'est essentiellement une reformulation du corollaire 7. □

Applications :

1. Si X suit la loi uniforme sur un compact K de $\mathbb{R}^d$, alors $Y = AX + b$ suit la loi uniforme sur l'image de K par $x \mapsto Ax+b$. Cette application est particulièrement intéressante en dimension 1.
2. Si $X \sim \Gamma(a, \lambda)$, alors pour tout $\mu > 0$, on a $\frac{1}{\mu}X \sim \Gamma(a, \lambda\mu)$.
3. Si X suit la loi exponentielle de paramètre e et que $a > 0$ alors $\frac{1}{\mu}X$ suit la loi exponentielle de paramètre μa. (Remarquer que ceci constitue un cas particulier de la remarque prédente.)
4. Pour $a \in \mathbb{R}$ et $b > 0$, X suit la loi de Cauchy $\mathcal{C}(0, 1)$ si et seulement si $Y = bX + a$ suit la loi de Cauchy $\mathcal{C}(a, b)$.
5. Soit $\sigma > 0$, $m \in \mathbb{R}$. On a $X \sim \mathcal{N}(m, \sigma^2) \iff \frac{X-m}{\sigma} \sim \mathcal{N}(0, 1)$.

6.8.2 Application aux lois gaussiennes

Lemme 13. *Soient $X = (X_1, X_2)$ un vecteur aléatoire formé de deux variables aléatoires indépendantes suivant la loi normale $\mathcal{N}(0, 1)$. On pose $Y_1 = \cos\theta X_1 + \sin\theta X_2$ et $Y_2 = -\sin\theta X_1 + \cos\theta X_2$. Alors Y_1 et Y_2 sont deux variables aléatoires indépendantes suivant la loi normale $\mathcal{N}(0, 1)$.*

Démonstration. Si l'on note, pour $x \in \mathbb{R}^2$, $x = (x_1, x_2)$, la densité de X est

$$\frac{1}{\sqrt{2\pi}}\exp(-\frac{x_1^2}{2}) \times \frac{1}{\sqrt{2\pi}}\exp(-\frac{x_2^2}{2}) = \frac{1}{2\pi}\exp(-\frac{x_1^2+x_2^2}{2}) = \frac{1}{2\pi}\exp(-\frac{\|x\|_2^2}{2}).$$

On a donc $Y = MX$, avec

$$M = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Ainsi, le vecteur $Y = MX$ admet pour densité

$$y \mapsto \frac{1}{\det M}\frac{1}{2\pi}\exp(-\frac{\|M^{-1}y\|_2^2}{2}).$$

M est une matrice de rotation, donc son déterminant vaut 1 et c'est une isométrie pour la norme euclidienne, ce qui implique que pour tout $x \in \mathbb{R}^2$ on a $\|M^{-1}y\|_2 = \|y\|_2$: la densité de Y est donc

$$y \mapsto \frac{1}{2\pi}\exp(-\frac{\|y\|_2^2}{2}),$$

ce qui est précisément la densité de X : Y a donc même loi que X, donc ses composantes Y_1 et Y_2 sont indépendantes et suivent la loi normale $\mathcal{N}(0, 1)$.

□

Théorème 79. *Soient U_1 et U_2 deux variables aléatoires indépendantes, avec $U_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ et $U_2 \sim \mathcal{N}(m_2, \sigma_2^2)$. Alors $U_1 + U_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.*

Démonstration. Si $\sigma_1 = 0$ ou $\sigma_2 = 0$, la variable aléatoire associée est constante est donc le résultat provient de la remarque faite plus haut – l'application affine est une translation.

Supposons donc $\sigma_1 > 0$ et $\sigma_2 > 0$. On pose $X_1 = \frac{U_1 - m_1}{\sigma_1}$ et $X_2 = \frac{U_2 - m_2}{\sigma_2}$. On peut trouver θ tel que $\cos\theta = \frac{\sigma_1}{\sqrt{\sigma_1^2+\sigma_2^2}}$ et $\sin\theta = \frac{\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}$. Alors, si on pose $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$, on a

$$U_1 + U_2 \quad = \quad m_1 + m_2 + \sigma(\cos\theta X_1 + \sin\theta X_2).$$

D'après le lemme, $\cos\theta X_1 + \sin\theta X_2 \sim \mathcal{N}(0, 1)$, donc $U_1 + U_2 \sim \mathcal{N}(m_1 + m_2, \sigma^2)$. □

6.8.3 Application : convolution de deux lois à densité

Théorème 80. *Soient X et Y deux variables aléatoires indépendantes sur $(\Omega, \mathcal{F}, \mathbb{P})$, de densité f et g. Alors $Z = X + Y$ admet comme densité la fonction*

$$x \mapsto \int_{-\infty}^{+\infty} f(x-t)g(t)\ dt.$$

Si, de plus, X et Y sont à valeurs positives, alors la densité est simplement

$$x \mapsto \mathbb{1}_{\mathbb{R}_+}(x) \int_0^x f(x-t)g(t)\ dt.$$

Démonstration. On pose

$$\begin{pmatrix} Z \\ T \end{pmatrix} = A \begin{pmatrix} X \\ Y \end{pmatrix}, \text{ avec } A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

La densité de (X, Y) est $h(x, y) = f(x)g(y)$. D'après l'exemple fondamental, la densité de (Z, T) est

$$g(z, t) = \frac{1}{|\det A|} f(A^{-1} \begin{pmatrix} z \\ t \end{pmatrix}).$$

On a

$$\det A = 1 \text{ et } A^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix},$$

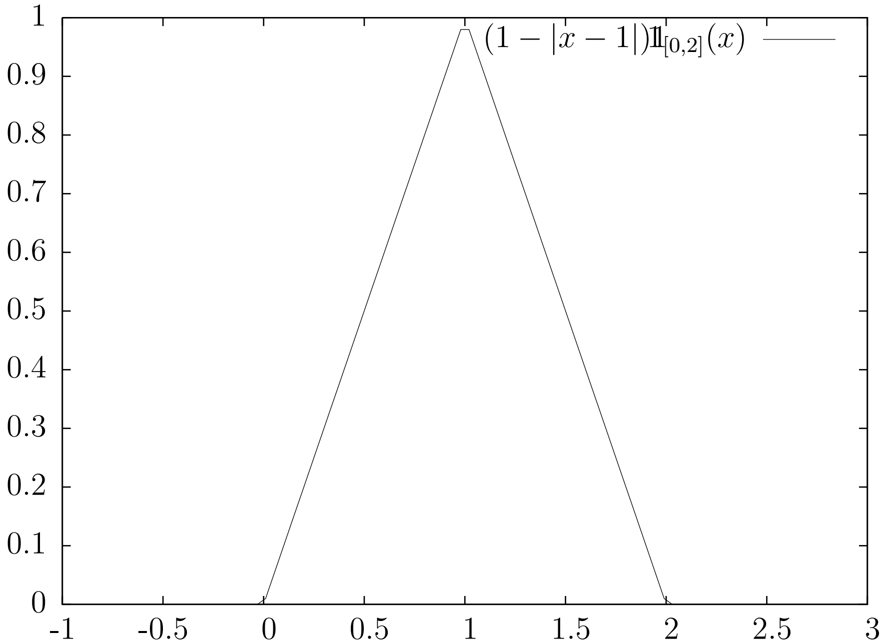
donc $g(z, t) = h(z - t, t) = f(z - t)g(t)$.

D'après le théorème 62, Z admet comme densité par rapport à la mesure de Lebesgue :

$$z \mapsto \int_{\mathbb{R}} g(z,t)\ d\lambda(t) = \int_{\mathbb{R}} f(z-t)g(t)\ d\lambda(t).$$

Dans le cas où X et Y sont à valeurs positives, il suffit de remarquer que $f(z-t)$ est nul si z dépasse t et que $g(t)$ est nul si t est négatif. Ainsi, $f(z-t)g(t)$ ne peut être non nul que pour z vérifiant $0 \le t \le z$, ce qui n'est évidemment jamais vérifié si z est négatif. $\square$

Exemple : ci-dessous, le graphe de la densité de Z lorsque X et Y suivent toutes les deux la loi uniforme sur $[0,1]$.



Application : $\Gamma(a,\lambda) * \Gamma(b,\lambda) = \Gamma(a+b,\lambda)$

Théorème 81. *Soit a, b, λ strictement positifs, X et Y deux variables aléatoires indépendantes, X suivant la loi $\Gamma(a,\lambda)$ et Y la loi $\Gamma(b,\lambda)$. Alors $Z = X+Y$ suit la loi $\Gamma(a+b,\lambda)$.*

Démonstration. Pour tous a et λ strictement positifs, on note $f_{a,\lambda}$ la densité de la loi $\Gamma(a,\Lambda)$, soit (rappel)

$$f_{a,\lambda}(x) = \mathbb{1}_{\mathbb{R}_+}(x)\frac{\lambda^a}{\Gamma(a)}x^{a-1}e^{-\lambda x}.$$

D'après le théorème précédent, Z admet une densité f_Z. Cette densité est nulle sur $\mathbb{R}_-$ tandis que pour x positif, on a

$$\begin{aligned} f_Z(x) &= \int_0^x f_{a,\lambda}(x-t)f_{b,\lambda}(t)\ dt \\ &= \int_0^x \frac{\lambda^a}{\Gamma(a)}t^{a-1}e^{-\lambda t}\frac{\lambda^b}{\Gamma(b)}(x-t)^{b-1}e^{-\lambda(x-t)}\ dt \\ &= \frac{\lambda^{a+b}e^{-\lambda x}}{\Gamma(a)\Gamma(b)}\int_0^x t^{a-1}(x-t)^{b-1}\ dt \end{aligned}$$

On fait le changement de variable $t = \theta x$. On obtient

$$\begin{aligned} f_Z(x) &= \frac{\lambda^{a+b}e^{-\lambda x}}{\Gamma(a)\Gamma(b)}x^{a-1}x^{b-1}x\int_0^1 \theta^{a-1}(1-\theta)^{b-1}\ d\theta \\ &= \frac{\lambda^{a+b}e^{-\lambda x}}{\Gamma(a)\Gamma(b)}x^{a+b-1}\int_0^1 \theta^{a-1}(1-\theta)^{b-1}\ d\theta \\ &= K_{a,b}f_{a+b,\lambda}(x), \end{aligned}$$

où $K_{a,b} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 \theta^{a-1}(1-\theta)^{b-1}\ d\theta$. Evidemment f_Z et $x \mapsto K_{a,b}f_{a+b,\lambda}(x)$ coincident également sur $\mathbb{R}_-$ où elles sont nulles. On a donc

$$\int_{\mathbb{R}} f_Z(x)\ d\lambda(x) = \int_{\mathbb{R}} K_{a,b}f_{a+b,\lambda}(x)\ d\lambda(x) = K_{a,b}\int_{\mathbb{R}} f_{a+b,\lambda}(x).$$

Mais f_Z et $f_{a+b,\lambda}$ sont des densités donc leur intégrale sur $\mathbb{R}$ vaut un. On en déduit $K_{a,b} = 1$, d'où

$$f_Z(x) = f_{a+b,\lambda}(x).$$

La densité de Z est la densité de $\Gamma(a+b,\Lambda)$, donc Z suit la loi $\Gamma(a+b,\Lambda)$.

Remarque : comme sous-produit de cette démonstration, on a obtenu le résultat non trivial suivant :

$$\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 \theta^{a-1}(1-\theta)^{b-1}\ d\theta.$$

□

6.9 Loi image par un C^1-difféomorphisme

Le premier réflexe à avoir est d'utiliser le corollaire 8, qui est bien sûr très utile. Le paragraphe suivant décrit des stratégies lorsque l'application n'est pas immédiate.

6.9.1 Compléments méthodologiques

Cette sous-section peut être omise en première lecture.

Il arrive que l'on ait besoin de calculer l'image d'une mesure de probabilité μ à densité par une transformation qui est C^1 (ou presque partout C^1), mais qui ne réalise pas un C^1-difféomorphisme de l'ensemble de départ vers l'ensemble d'arrivée. Il y a deux raisons qui peuvent empêcher une application T de classe C^1 de réaliser un C^1-difféomorphisme entre l'ensemble de départ O et l'ensemble d'arrivée :

- $\exists x \in O; D_x T$ non inversible.
- T non injectif

Dans tous les cas, on commence par enlever les points où T n'est pas C^1 – si du moins cet ensemble est de mesure nulle, sinon ce texte ne peut pas grand chose pour vous. Ainsi pour une loi à densité sur $\mathbb{R}^2$ la transformation

$$(x, y) \mapsto (\min(x, y), |x - y|)$$

est tout à fait recevable puisqu'elle est C^1 sur l'ouvert $\mathbb{R}^2 \backslash \{(x, x); x \in \mathbb{R}^2\}$.

- Le premier écueil n'est pas très dangereux : en effet, dans la plupart des cas $\{x \in O; \det D_x T = 0\}$ est de mesure nulle, dans ce cas il suffit de remplacer O par $O \backslash \{x \in O; \det D_x T = 0\}$ (en effet $\{x \in O; \det D_x T = 0\}$ est toujours un fermé donc $O \backslash \{x \in O; \det D_x T = 0\}$ est bien un ouvert) , et on peut appliquer le théorème du cours.
  - Un exemple dans $\mathbb{R}$
    Soit $X \sim \mathcal{N}(0, 1)$. On veut calculer la loi de $Y = X^3$.
    On a ici $O = \mathbb{R}$ et $T(x) = x^3$. On a $\{x \in O; \det D_x T = 0\} = \{x \in \mathbb{R}; 3x^2 = 0\} = \{0\}$ On prend donc $\mathbb{R}\backslash\{0\}$ comme ouvert de départ : T est C^1 sur $\mathbb{R}\backslash\{0\}$, sa différentielle y est partout inversible et T est injective sur $\mathbb{R}\backslash\{0\}$ (puisqu'elle l'est sur $\mathbb{R}$) : T réalise donc un C^1 difféomorphisme de T sur son image, qui est $\mathbb{R}\backslash\{0\}$. On trouve ainsi qu'une densité de Y est

$$y \mapsto \frac{1}{\sqrt{2\pi}} \frac{1}{3} \frac{1}{|y|^{\frac{2}{3}}} e^{-\frac{|y|^{\frac{2}{3}}}{2}} \mathbb{1}_{\mathbb{R}\backslash\{0\}}(y),$$

qui est presque partout égale à

$$y \mapsto \frac{1}{\sqrt{2\pi}} \frac{1}{3} \frac{1}{|y|^{\frac{2}{3}}} e^{-\frac{|y|^{\frac{2}{3}}}{2}},$$

dont l'écriture est tout de même plus simple.

- Un exemple dans $\mathbb{R}^2$
  Soit X et Y indépendantes suivant toutes deux $\mathcal{N}(0,1)$. On veut calculer la loi du couple $(U,V) = (X^3, XY)$.
  On a ici $O = \mathbb{R}^2$ et $T(x) = (x^3, xy)$. On a $\{x \in O; \det D_x T = 0\} = \{(x,y) \in \mathbb{R}; 3x^3 = 0\} = \{0\} \times \mathbb{R}$. On prend donc $\mathbb{R}^2 \backslash (\{0\} \times \mathbb{R})$ comme ouvert de départ : T est C^1 sur $\mathbb{R}^2 \backslash (\{0\} \times \mathbb{R})$, sa différentielle y est partout inversible et T est injective sur $\mathbb{R}\backslash\{0\}$, de réciproque $(u,v) \mapsto T^{-1}(u,v) = (\sqrt[3]{u}, \frac{v}{\sqrt[3]{u}})$. T réalise donc un C^1 difféomorphisme de T sur son image, qui est $(\mathbb{R}\backslash\{0\})^2$.
  On trouve ainsi qu'une densité de (U,V) est
  $$(u,v) \mapsto \frac{1}{2\pi}\frac{1}{3|u|} e^{-\frac{|u|^{\frac{2}{3}} + v^2 |u|^{-\frac{2}{3}}}{2}} \mathbb{1}_{(\mathbb{R}\backslash\{0\})^2}(u,v),$$
  qui est presque partout égale à
  $$(u,v) \mapsto \frac{1}{2\pi}\frac{1}{3|u|} e^{-\frac{|u|^{\frac{2}{3}} + v^2 |u|^{-\frac{2}{3}}}{2}}.$$
- Le cas où T est C^1, mais non injective sur O est plus délicat. Voilà comment procéder.
  On commence par déterminer $N = \{x \in O; \det D_x T = 0\}$. Si N est de mesure nulle, on peut procéder comme suit. On commence, comme précédemment, par remplacer O par $O' = O\backslash N$. Ensuite on découpe O' suivant ses composantes connexes. Dans la suite, je supposerai que
  $$O' = \underset{i \in I}{\cup} O_i,$$
  où les O_i' sont des connexes et où I est dénombrable. (Je n'ai jamais rencontré d'exemple où $O'\backslash N$ a une infinité non dénombrable de composantes connexes, en fait en général $|I| = 1$ ou 2.)
  Je dis que si pour tout $i \in I$, T est injective sur O_i, alors on a gagné. Notons ν la mesure image de μ par T. On a pour tout borélien A
  $$\nu(A) = \mu(T^{-1}(A)) = \sum_{i \in I} \mu(T^{-1}(A) \cap O_i) = \sum_{i \in I} \mu(O_i) \mu_i(T^{-1}(A)),$$
  où l'on a posé
  $$\mu_i(B) = \frac{\mu(B \cap O_i)}{\mu(O_i)} = \mu(B|O_i)$$
  Ainsi, notre problème est maintenant, pour tout i, de trouver la loi image ν_i de μ_i par T. Mais maintenant, nous pouvons appliquer le théorème classique car

- μ_i admet la densité $x \mapsto \frac{1}{\mu(O_i)} f(x) 1\!\!1_{O_i}(x)$, où f est la densité de μ.
- f est C^1 sur O_i et sa différentielle y est partout inversible (par définition de O')
- T est injective sur O_i

Ainsi, on obtient que la loi image de μ par T admet la densité

$$y \mapsto \sum_{i \in I} f(T_i^{-1} y) |\det(DT_i^{-1})_y| 1\!\!1_{f(O_i)}(y)$$

où T_i^{-1} représente la réciproque de la restriction de T_i à O_i.
Remarque : en dimension 1, l'injectivité de O_i sur les composantes connexes est gratuite, en vertu du théorème de Rolle

- Un exemple dans $\mathbb{R}$
  X suit la loi uniforme sur $[-1,2]$ et $Y = X^2$. On a $O =]-1,2[$ et $N = \{0\}$, de sorte que $O_1 =]-1,0[$ et $O_2 =]0,2[$ On trouve la densité

$$y \mapsto \frac{1}{2\sqrt{y}} \frac{1}{3} 1\!\!1_{]0,1[}(y) + \frac{1}{2\sqrt{y}} \frac{1}{3} 1\!\!1_{]0,4[}(y)$$

- Un exemple dans $\mathbb{R}^2$
  Soient X et Y deux variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On pose $S = X + Y$ et $P = XY$. Déterminer la loi de (S, P).
  Ici $N = \{(x,y) \in \mathbb{R}^2; x = y\}$, donc

  $O_1 = \{(x,y) \in \mathbb{R}^2; 0 < x < y < 1\}$ et $O_2 = \{(x,y) \in \mathbb{R}^2; 0 < y < x < 1\}$.

  Ici la symétrie axiale $s(x,y) = (y,x)$ d'axe $x = y$ envoie laisse μ invariante – c'est à dire que μ est la mesure image de μ par s – , envoie O_2 sur O_2, et vérifie de plus $T \circ s = T$, on a

$$\begin{aligned}
\nu_2(A) &= \frac{\mu(T^{-1}A \cap O_2)}{\mu(O_2)} \\
&= \frac{\mu((T \circ s)^{-1}(A) \cap s^{-1}(O_1))}{\mu(O_2)} \\
&= \frac{\mu(s^{-1}(T^{-1}(A)) \cap s^{-1}(O_1))}{\mu(s^{-1}(O_1))} \\
&= \frac{\mu_s(T^{-1}(A) \cap O_1)}{\mu_s(O_1)} \\
&= \frac{\mu(T^{-1}(A) \cap O_1)}{\mu(O_1)} \\
&= \nu_1(A)
\end{aligned}$$

Il s'ensuit que $\nu = \mu(O_1)\nu_1 + \mu(O_2)\nu_2 = \nu_1 = \nu_2$.
Dans les problèmes comme celui-ci où une symétrie entre en jeu, on peut parfois rédiger la solution d'une manière un peu différente en factorisant l'application T. Ici, si on pose $(X', Y') = (\max(X, Y), \min(X, Y))$, on remarque sans peine que $(S, P) = T((X', Y'))$. On peut donc décomposer le problème en deux étape. D'abord calculer la loi de (X', Y') – on trouvera que c'est la loi uniforme sur O_2 – , puis calculer la loi image de la loi de (X', Y') par l'application T. Ce dernier calcul peut être fait sans difficulté théorique, puisqu'on a vu que T réalise un C^1-difféomorphisme de O_2 sur son image. Notons qu'on échappe pas à l'emploi d'arguments de symétrie, puisqu'on en a besoin pour calculer la loi de (X', Y'). Cette deuxième rédaction a comme avantage de mieux séparer les difficultés du problème. Gardons toutefois à l'esprit que ceci ne peut être érigé en méthode : ici les choses se passent très bien car la symétrie s laisse invariante à la fois l'application T et a mesure de départ.
Note : si on n'a pas de chance, T peut ne pas être injective sur les composantes connexes de O'. Dans ce cas, il faut essayer de trouver un fermé de mesure nulle telle que T soit injective sur chaque composante de $O'' = O' \backslash F$ et remplacer O' par O''. Là, il n'y a pas vraiment de méthode générale. Exemple : $T(x, y) = (x^2+y^2, 2xy)$. Le seul point où la différentielle n'est pas inversible est l'origine, mais bien sûr T n'est pas injectif sur l'ouvert connexe $\mathbb{R}^2 \backslash \{(0, 0)\}$, puisque $T(-x, -y) = T(x, y)$. On constatera que $F = \{(x, 0); x \in \mathbb{R}\}$ convient.

6.10 Calcul des premiers moments des lois discrètes usuelles

6.10.1 Indicatrice d'un événement

On rappelle que pour $A \subset \Omega$, l'application $\mathbb{1}_A$ (appelée indicatrice de A) est définie sur Ω par

$$\mathbb{1}_A(x) = \begin{cases} 1 \text{ si } x \in A \\ 0 \text{ si } x \notin A \end{cases}$$

1_A est une variable aléatoire à valeurs dans $\{0; 1\}$. Il est important de remarquer que, comme $\forall x \in \{0; 1\} \quad x^2 = x$, on a $\mathbb{1}_A^2 = \mathbb{1}_A$. Maintenant, on a

- $\mathbb{E}\mathbb{1}_A = \mathbb{P}(A)$.

–

$$\begin{aligned}
\operatorname{Var} \mathbb{1}_A &= \mathbb{E}\mathbb{1}_A^2 - (\mathbb{E}\mathbb{1}_A)^2 \\
&= \mathbb{E}\mathbb{1}_A - (\mathbb{E}\mathbb{1}_A)^2 \\
&= \mathbb{P}(A) - \mathbb{P}(A)^2 \\
&= \mathbb{P}(A)(1 - \mathbb{P}(A)).
\end{aligned}$$

6.10.2 Loi binomiale

On a vu que la loi binomiale de paramètres n et p était la loi de

$$X = \sum_{k=1}^{n} \mathbb{1}_{A_k},$$

où $A_1, \ldots, A_n$ sont n événements indépendants de même probabilité p. On a donc

$$\mathbb{E}X = \sum_{k=1}^{n} \mathbb{E}\mathbb{1}_{A_k} = \sum_{k=1}^{n} \mathbb{P}(A_k) = np,$$

et comme les variables aléatoires sont indépendantes

$$\operatorname{Var} X = \sum_{k=1}^{n} \operatorname{Var} \mathbb{1}_{A_k} = \sum_{k=1}^{n} \mathbb{P}(A_k)(1 - \mathbb{P}(A_k)) = np(1-p).$$

6.10.3 Loi géométrique

Soit X une variable aléatoire suivant une loi géométrique de paramètre $p \in]0,1]$. On a

$$\begin{aligned}
\mathbb{E}X &= \sum_{k=0}^{+\infty} k\mathbb{P}(X=k) \\
&= \sum_{k=1}^{+\infty} k\mathbb{P}(X=k) \\
&= \sum_{k=1}^{+\infty} kp(1-p)^{k-1} \\
&= p\sum_{k=1}^{+\infty} k(1-p)^{k-1} \\
&= p\frac{1}{(1-(1-p))^2} \\
&= \frac{1}{p}.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}X(X-1) &= \sum_{k=0}^{+\infty} k(k-1)\mathbb{P}(X=k) \\
&= \sum_{k=2}^{+\infty} k(k-1)\mathbb{P}(X=k) \\
&= \sum_{k=2}^{+\infty} k(k-1)p(1-p)^{k-1} \\
&= p(1-p)\sum_{k=2}^{+\infty} k(k-1)(1-p)^{k-2} \\
&= p(1-p)\frac{2}{(1-(1-p))^3} \\
&= \frac{2(1-p)}{p^2}.
\end{aligned}$$

On a alors $\mathbb{E}X^2 = \mathbb{E}X(X-1) + \mathbb{E}X = \frac{1}{p} + \frac{2(1-p)}{p^2}$ et $\mathrm{Var}\, X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{1}{p} + \frac{2(1-p)}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$.

6.10.4 Loi de Poisson

$$\begin{aligned}
\mathbb{E}X &= \sum_{k=0}^{+\infty} k\mathbb{P}(X=k) \\
&= \sum_{k=1}^{+\infty} k\mathbb{P}(X=k) \\
&= \sum_{k=1}^{+\infty} k e^{-\lambda}\frac{\lambda^k}{k!} \\
&= e^{-\lambda}\lambda\sum_{k=1}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!} \\
&= e^{-\lambda}\lambda e^{\lambda} \\
&= \lambda
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}X(X-1) &= \sum_{k=0}^{+\infty} k(k-1)\mathbb{P}(X=k) \\
&= \sum_{k=2}^{+\infty} k(k-1)\mathbb{P}(X=k) \\
&= \sum_{k=2}^{+\infty} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} \\
&= e^{-\lambda}\lambda^2\sum_{k=2}^{+\infty}\frac{\lambda^{k-2}}{(k-2)!} \\
&= e^{-\lambda}\lambda^2 e^{\lambda} \\
&= \lambda^2
\end{aligned}$$

On a alors $\mathbb{E}X^2 = \mathbb{E}X(X-1) + \mathbb{E}X = \lambda^2 + \lambda$ et $\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

6.10.5 Loi hypergéométrique

On rappelle que la loi hypergéométrique $\mathcal{H}(N,n,k)$ est la loi image de la loi uniforme sur $\Omega = \mathcal{B}(N,k)$ par l'application

$$\begin{array}{rcl} X : \mathcal{B}(N,k) & \to & \mathbb{N} \\ \omega & \mapsto & X(\omega) = |\{1,\dots,n\}) \cap \omega| \end{array}$$

On va montrer que $\mathbb{E}X = k\frac{n}{N}$ et $\operatorname{Var} X = k\frac{n}{N}(1-\frac{n}{N})\frac{N-1}{N}$.

Démonstration. Notons $\mathbb{P}$ la loi uniforme sur Ω. Par souci de lisibilité, on définit l'ensemble aléatoire A par $A(\omega) = \omega$. Ainsi

$$\begin{array}{rcl} X & = & |\{1,\dots,n\}) \cap A| \\ & = & \sum_{i=1}^{n} 1\!\!1_{\{i\in A\}} \end{array}$$

Pour tout $k \in \{1,\dots,n\}$, on a

$$\mathbb{E}1\!\!1_{\{i\in A\}} = \mathbb{P}(i\in A) = 1 - \frac{\binom{N-1}{k}}{\binom{N}{k}} = 1 - \frac{(N-1)!(N-k)!}{N!(N-k-1)!} = 1 - \frac{N-k}{N} = \frac{k}{N}.$$

Ainsi

$$\mathbb{E}X = \sum_{i=1}^{n} 1\!\!1_{\{i\in A\}} = \frac{nk}{N} = k\frac{n}{N}.$$

Maintenant, on a

$$\operatorname{Var} X = \sum_{i=1}^{n}\sum_{j=1}^{n} \operatorname{Covar}(1\!\!1_{\{i\in A\}}, 1\!\!1_{\{j\in A\}}).$$

Pour $i=j$, on a

$$\operatorname{Covar}(1\!\!1_{\{i\in A\}}, 1\!\!1_{\{j\in A\}}) = \operatorname{Var} 1\!\!1_{\{i\in A\}} = \mathbb{P}(j\in A)(1-\mathbb{P}(j\in A)) = \frac{k}{N}(1-\frac{k}{N}).$$

Pour $i \neq j$, on a

$$\begin{array}{rcl} \operatorname{Covar}(1\!\!1_{\{i\in A\}}, 1\!\!1_{\{j\in A\}}) & = & \operatorname{Covar}(1-1\!\!1_{\{i\in A\}}, 1-1\!\!1_{\{j\in A\}}) \\ & = & \operatorname{Covar}(1\!\!1_{\{i\notin A\}}, 1\!\!1_{\{j\notin A\}}) \\ & = & \mathbb{P}(i\notin A, j\notin A) - \mathbb{P}(i\notin A)\mathbb{P}(j\notin A) \\ & = & \dfrac{\binom{N-2}{k}}{\binom{N}{k}} - (\dfrac{N-k}{N})^2 \\ & = & \dfrac{(N-k)(N-k-1)}{N(N-1)} - (\dfrac{N-k}{N})^2 \\ & = & -\dfrac{1}{N-1}\dfrac{k}{N}(1-\dfrac{k}{N}) \end{array}$$

On en déduit

$$\begin{aligned}
\operatorname{Var} X &= n \times \frac{k}{N}(1-\frac{k}{N}) + n(n-1)\times(-\frac{1}{N-1}\frac{k}{N}(1-\frac{k}{N})) \\
&= n\frac{k}{N}(1-\frac{k}{N})(1-\frac{n-1}{N-1}) \\
&= n\frac{k}{N}(1-\frac{k}{N})\frac{N-n}{N-1} \\
&= k\frac{n}{N}(1-\frac{n}{N})\frac{N-1}{N}
\end{aligned}$$

On remarque qu'une loi hypergéométrique a la même espérance qu'une loi binomiale $\mathcal{B}(k, \frac{n}{N})$ et que sa variance ne diffère de celle de cette binomiale que d'un facteur $\frac{N-1}{N}$.

$\square$

6.11 Calcul des premiers moments des lois à densité usuelles

6.11.1 Loi uniforme sur un segment

Soit X une variable aléatoire suivant la loi uniforme sur $[-1, 1]$. La densité de X est donc

$$x \mapsto \frac{1}{2}\mathbb{1}_{[-1,1]}(x).$$

On a donc

$$\mathbb{E}X = \int_{-1}^{1} \frac{1}{2}x\ dx = 0$$

et

$$\mathbb{E}X^2 = \int_{-1}^{1} \frac{1}{2}x^2\ dx = \frac{1}{3}.$$

Comme X est centrée, on a $\operatorname{Var} X = \mathbb{E}X^2$.

Passons au cas général : on pose $Y = \frac{a+b}{2} + \frac{b-a}{2}X$. X suit la loi uniforme sur $[-1, 1]$ si et seulement Y suit la loi uniforme sur $[a, b]$. On a alors

- $\mathbb{E}Y = \mathbb{E}\frac{a+b}{2} + \frac{b-a}{2}\mathbb{E}X = \frac{a+b}{2}$.
- $\operatorname{Var} Y = (\frac{b-a}{2})^2 \operatorname{Var} X = \frac{(b-a)^2}{12}$

6.11.2 Loi gaussienne

Soit X une variable aléatoire suivant la loi $\mathcal{N}(0,1)$. On rappelle que la densité de X est

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Lemme 14. *Soit g une fonction dérivable sur $\mathbb{R}$ telle qu'il existe A et c tels que*

$$\forall x \in \mathbb{R} \quad |g(x)| + |g'(x)| \leq A \exp(-c|x|).$$

Alors, si $X \sim \mathcal{N}(0,1)$, alors $g'(X)$ et $Xg(X)$ sont intégrables et on a

$$\mathbb{E}[g'(X)] = \mathbb{E}[Xg(X)].$$

Démonstration. Il est facile de vérifier que

$$\frac{\mathrm{d}}{\mathrm{dx}}(g(x)f(x)) = (g'(x) - xg(x))f(x).$$

On a donc

$$\forall a, b \in \mathbb{R} \quad g(b)f(b) - g(a)f(a) = \int_a^b g'(x)f(x)\ dx - \int_a^b xg(x)f(x)\ dx$$

Les hypothèses faites sur g et g' assurent l'intégrabilité sur $\mathbb{R}$ de $g'f$ et gf. Comme, de plus $\lim_{a\to-\infty} g(a)f(a) = \lim_{b\to+\infty} g(b)f(b) = 0$, on en déduit que

$$0 = \int_{\mathbb{R}} g'(x)f(x)\ d\lambda(x) - \int_{\mathbb{R}} xg(x)f(x)\ d\lambda(x),$$

soit $\mathbb{E}[g'(X)] = \mathbb{E}[Xg(X)]$. □

En prenant $g(x) = x$, on obtient l'existence d'un moment d'ordre 2 avec $\mathbb{E}[X^2] = 1$. D'autre part, l'existence d'un moment d'ordre 2 implique celle d'un moment d'ordre 1. Comme la densité de X est paire, on en déduit que $\mathbb{E}X = 0$. On a donc $\operatorname{Var} X = \mathbb{E}X^2 = 1$.

Passons au cas général. Si l'on a $Y = m + \sigma X$, on sait que Y suit la loi $\mathcal{N}(m, \sigma^2)$. On a alors $\mathbb{E}Y = m + \sigma\mathbb{E}X = m$ et $\operatorname{Var} Y = \sigma^2 \operatorname{Var} X = \sigma^2$.

Exercice laissé au lecteur : pour $X \sim \mathcal{N}(0,1)$, exprimer $\mathbb{E}[X^{2n}]$ en fonction de n.

6.11.3 Lois Gamma

Soit X une variable aléatoire suivant la loi $\Gamma(a,\lambda)$. Alors, X admet des moments de tout ordre, avec pour tout $\alpha \geq 0$, on a

$$\mathbb{E}X^\alpha = \lambda^{-\alpha}\frac{\Gamma(a+\alpha)}{\Gamma(a)}.$$

En particulier $\mathbb{E}X = \frac{a}{\lambda}$ et $\operatorname{Var} X = \frac{a}{\lambda^2}$.

Démonstration. Pour tous a et λ strictement positifs, on note $f_{a,\lambda}$ la densité de la loi $\Gamma(a,\Lambda)$, soit (rappel)

$$f_{a,\lambda}(x) = 1\!\!1_{\mathbb{R}_+}(x)\frac{\lambda^a}{\Gamma(a)}x^{a-1}e^{-\lambda x}.$$

D'après le théorème de transfert,

$$\begin{aligned}
\mathbb{E}X^\alpha &= \int_{\mathbb{R}_+} x^\alpha f_{a,\lambda}(x)\ dx \\
&= \int_{\mathbb{R}_+} \lambda^{-\alpha}\frac{\Gamma(a+\alpha)}{\Gamma(a)} f_{a+\alpha,\lambda}(x)\ dx \\
&= \lambda^{-\alpha}\frac{\Gamma(a+\alpha)}{\Gamma(a)}
\end{aligned}$$

Ainsi $\mathbb{E}X = \lambda^{-1}\frac{\Gamma(a+1)}{\Gamma(a)} = \frac{a}{\lambda}$, $\mathbb{E}X^2 = \lambda^{-2}\frac{\Gamma(a+2)}{\Gamma(a)} = \frac{a(a+1)}{\lambda^2}$ $\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{a(a+1)}{\lambda^2} - \frac{a^2}{\lambda^2} = \frac{a}{\lambda^2}$. □

6.11.4 Lois exponentielles

Soit X suivant une loi exponentielle de paramètre $\lambda > 0$. La loi exponentielle est un cas particulier de la loi Gamma : on a $\mathcal{E}(\lambda) = \Gamma(1,\lambda)$. On déduit du calcul précédent que $\mathbb{E}X = \frac{1}{\lambda}$ et $\operatorname{Var} X = \frac{1}{\lambda^2}$.

6.11.5 Lois de Cauchy

Soient $a \in \mathbb{R}, b > 0$. La loi de Cauchy $\mathcal{C}(a,b)$ admet comme densité par rapport à la mesure de Lebesgue :

$$x \mapsto \frac{1}{\pi}\frac{b}{(x-a)^2+b^2},$$

donc pour $k \geq 1$, on a

$$\mathbb{E}|X|^k = \frac{1}{\pi}\frac{b|x|^k}{(x-a)^2+b^2} = +\infty,$$

donc les lois de Cauchy n'admettent pas de moment d'ordre 1, ni , a fortiori, d'ordre supérieur.

6.12 Exercice sur les espérances

1. Un jeu consiste à effectuer une mise en choisissant un nombre entre 1 et 6, puis à lancer simultanément trois dés. Si le numéro choisi sort une fois, le joueur récupère sa mise plus une somme égale à sa mise. Si le numéro choisi sort deux fois, le joueur récupère sa mise plus une somme égale à deux fois sa mise. Enfin, si le numéro choisi sort trois fois, le joueur récupère sa mise plus une somme égale à trois fois sa mise. Quelle est l'espérance de gain à ce jeu ?
2. Soient A, B deux éléments observables. On note

$$A\Delta B = \{x \in A; x \notin B\} \cup \{x \in B; x \notin A\}.$$

   Ce sont donc les éléments qui sont dans A ou dans B, mais pas dans les deux. Montrer $\mathbb{1}_{A\Delta B} = (\mathbb{1}_A - \mathbb{1}_B)^2$. En déduire

$$\mathbb{P}(A\Delta B) = \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A\cap B).$$

3. Soient A, B deux éléments observables. Montrer que

$$|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| \le \sqrt{\mathbb{P}(A)\mathbb{P}(B)}.$$

4. (a) Soit X une variable aléatoire de carré intégrable. On note σ^2 sa variance et m son espérance. Montrer que pour tout a réel, $\sigma^2 \le \mathbb{E}(X-a)^2$.

   (b) Soient $a_1 \le a_2 \le \cdots \le a_n$. On pose

$$m = \frac{a_1 + \cdots + a_n}{n}.$$

   Montrer que

$$\frac{1}{n}\sum_{k=1}^{n}(a_k - m)^2 \le \frac{(a_n - a_1)^2}{4}.$$

5. Soit X une variable aléatoire à valeurs dans $\mathbb{N}^*$ telle la suite $(p_n)_{n\geq 1}$ définie par $p_n = \mathbb{P}(X = n)$ soit décroissante. Montrer que pour toute injection σ de $\mathbb{N}^*$ dans lui-même, on a

$$\mathbb{E}\sigma(X) \geq \mathbb{E}X.$$

6. On suppose que $Y = \ln X$ vérifie $Y \sim \mathcal{N}(m, \sigma^2)$ (on dit alors que X est log-normale). Calculer $\mathbb{E}X$ et $\operatorname{Var} X$.

7. Calculer $\mathbb{E}\sin X$, où $\mathbb{P}(X = \frac{\pi}{2}) = \frac{1}{6}, \mathbb{P}(X = \frac{\pi}{3}) = \frac{1}{3}, \mathbb{P}(X = \frac{\pi}{6}) = \frac{1}{2}$.

8. Soient X, Y deux variables aléatoires suivant chacune une loi uniforme sur $[a, b]$. Montrer que $\mathbb{E}|X - Y| \leq \frac{b-a}{2}$. Que vaut $\mathbb{E}|X - Y|$ lorsque X et Y sont indépendantes ?

9. Soit X une variable aléatoire suivant la loi uniforme sur $[0, 1]$. Déterminer la loi de $Z = -\ln(1 - X)$.

10. Soit X une variable aléatoire de densité f. Montrer que la variable aléatoire $|X|$ admet comme densité

$$x \mapsto \mathbb{1}_{R_+}(x)(f(x) + f(-x)).$$

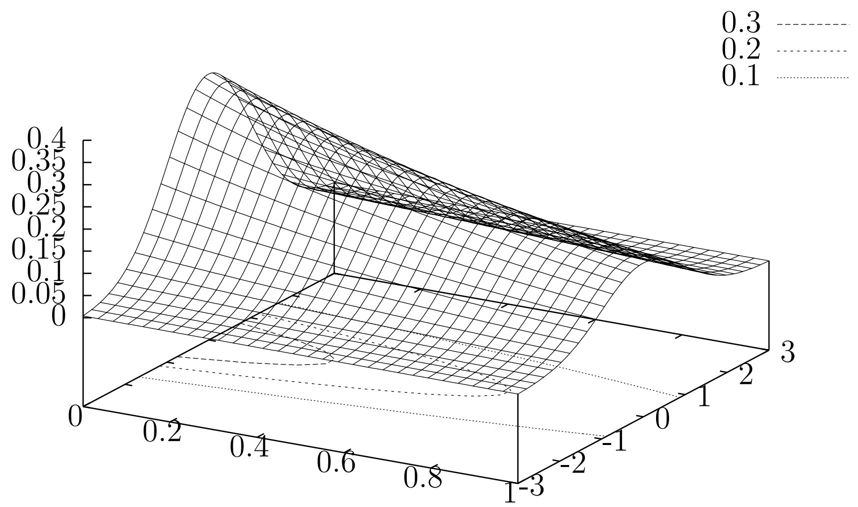
11. Soit X une variable aléatoire positive de densité f. Montrer que la variable aléatoire $X^{1/2}$ admet comme densité

$$x \mapsto \mathbb{1}_{R_+}(x)2x(f(x^2)).$$

12. Soit X une variable aléatoire normale centrée réduite. Montrer que la variable aléatoire X^2 est à densité et la déterminer.

13. La figure ci-dessous représente la densité $f(x, y)$ d'un couple de variables aléatoires indépendantes X et Y. X suit une loi exponentielle de paramètre 1 et Y une loi normale centrée réduite.

    On a tracé quelques isoclines, c'est à dire des courbes reliant des points de même densité : $f(x, y) =$ constante. Quelle est la nature géométrique de ces isoclines ?



14. Soit X et Y deux variables aléatoires indépendantes. On suppose que X suit la loi normale $\mathcal{N}(0,1)$ et Y la loi gamma $\gamma(\frac{n}{2}, \frac{1}{2})$. Calculer la loi de
$$Z = \frac{Y}{\sqrt{Y/n}}.$$
La loi de Z est appelée loi de Student à n degrés de libertés.

15. Soient X et Y deux variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On pose $S = X + Y$ et $P = XY$.
Déterminer la loi de (S, P).

16. Soit f une fonction réelle continue sur l'intervalle fermé $[0,1]$. Pour $n \in \mathbb{N}^*$, on note B_n le polynôme de Bernstein
$$B_n(x) = \sum_{k=0}^{n} f(\frac{k}{n}) \binom{n}{k} x^k (1-x)^{n-k} \ .$$
Pour tout $x \in]0,1[$ on se donne une suite (X_k) de variables de Bernoulli indépendantes de même paramètre x. On note $S_n = \sum_{k=1}^{n} X_k$.
    (a) Déterminer la moyenne $\mathbb{E}[f(\frac{S_n}{n})]$.
    (b) Soit pour tout $\varepsilon > 0$, le réel $\delta(\varepsilon)$ défini par
$$\delta(\varepsilon) = \sup\{|f(x) - f(y)| : x, y \in [0,1] \text{ et } |x - y| \le \varepsilon\} \ .$$

i. Démontrer que $\delta(\varepsilon)$ tend vers 0 avec ε.

ii. Démontrer que

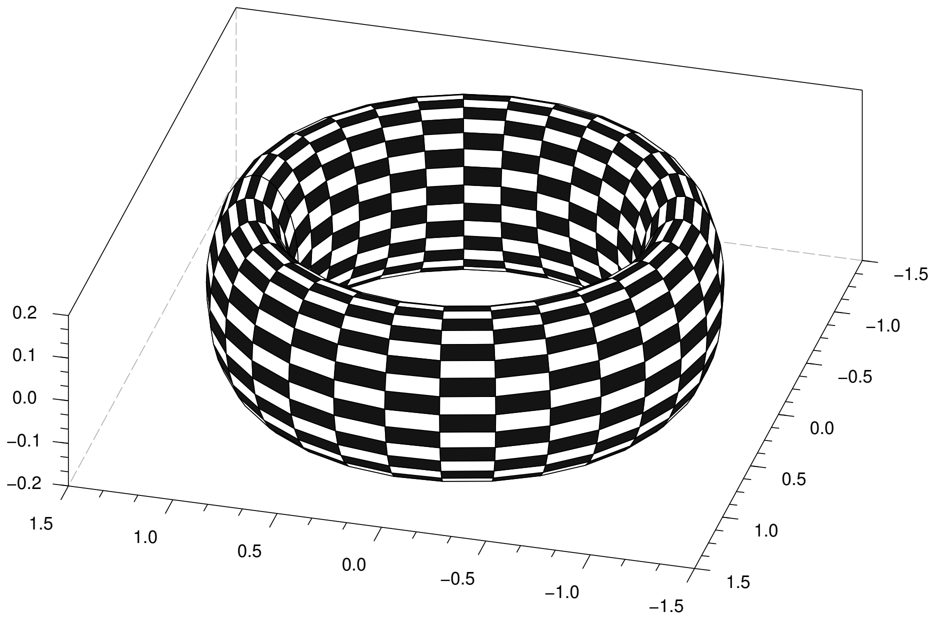
$$\sup_{x\in[0,1]} |B_n(x) - f(x)| \leq \delta(\varepsilon) + \frac{2\|f\|_\infty}{n\varepsilon^2} .$$

En déduire que la suite des polynômes B_n converge vers f uniformément sur $[0, 1]$.

17. On place 7 dames sur un échiquier torique 41×41 de telle manière qu'aucune dame ne puisse en prendre une autre.

Soit φ une permutation des cases de l'échiquier.

Montrer qu' il existe x tel que x et $\varphi(x)$ puissent chacun être pris par au moins une des dames.



On rappelle que les dames peuvent prendre les pièces qui sont sur la même ligne, sur la même colonne, ou sur une même diagonale.

18. Soient n, r deux entiers tels que $1 \leq r \leq n$. On prend r nombres distincts au hasard dans $\{1, \ldots, n\}$ et on note X le plus petit de ces r nombres.

(a) Quelles valeurs peut prendre X ? Montrer que pour $k \in \{0, n-r\}$, on a

$$\mathbb{P}(X > k) = \frac{\binom{n-k}{r}}{\binom{n}{r}}.$$

(b) En déduire que

$$\mathbb{E}X = \frac{\binom{n+1}{r+1}}{\binom{n}{r}} = \frac{n+1}{r+1}.$$

19. Soient X et Y deux variables aléatoires indépendantes suivant la loi uniforme sur $[0, 1]$.

    (a) Soit r la rotation dans $\mathbb{R}^2$ de centre $(0, 0)$ et d'angle $-\frac{\pi}{4}$. On pose $(U, V) = r(X, Y)$. Montrer que la loi du vecteur (U, V) est la loi uniforme sur un ensemble que l'on déterminera.

    (b) Pour quelles valeurs de α la variable aléatoire $\frac{1}{|X-Y|^\alpha}$ est-elle intégrable ? Lorsqu'elle l'est, calculer sa valeur.

20. Soient $(X_n)_{n\geq 1}$ une suite de variables aléatoires centrées de carré intégrable. On suppose qu'il existe une fonction b de $\mathbb{Z}$ dans $\mathbb{R}$ telle que $\mathbb{E}[X_i X_j] = b(i-j)$ quels que soient i et j dans $\mathbb{N}$.

    (a) Montrer que b est paire et exprimer simplement la variance de $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ en fonction des $b(i)$.

    (b) Montrer que si $b(i) < 0$ pour tout i non nul, alors la série de terme général $b(i)$ est convergente, avec $\sum_{i=1}^{+\infty}(-b(i)) \leq \frac{b(0)}{2}$.

21. *Probabilité de retour en zéro au temps n d'une marche aléatoire symétrique*

    (a) Soit X une variable aléatoire à valeurs dans $\mathbb{Z}$. Montrer que

    $$\mathbb{P}(X = 0) = \frac{1}{2\pi}\int_0^{2\pi} \mathbb{E}[e^{i\theta X}]\, d\theta.$$

    (b) Soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes identiquement distribuées, avec $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$. Montrer que $\mathbb{P}(S_{2n} = 0) = \frac{2}{\pi}W_{2n}$, où W_k désigne l'intégrale de Wallis (voir exercices du chapitre 4). En déduire que

    $$\mathbb{P}(S_{2n} = 0) \sim \frac{1}{\sqrt{\pi n}}.$$

Chapitre 7

Espaces $\mathcal{L}^p$ et L^p

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Pour $p \in [1, +\infty)$, On note $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ l'ensemble des applications mesurables de $(\Omega, \mathcal{A}, \mu)$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telles que

$$\int_\Omega |f(x)|^p \, d\mu(x) < +\infty.$$

On dit que des nombres p et q de $]1, +\infty[$ sont des exposants conjugués si ils vérifient.

$$\frac{1}{p} + \frac{1}{q} = 1.$$

On convient également parfois que 1 et l'infini sont des exposants conjugués, mais cela ne sera pas utilisé ici.

7.1 De $\mathcal{L}^p$ à L^p

7.1.1 Inégalité de Hölder

Théorème 82 (Inégalité de Hölder)**.** *Soient p et q des exposants conjugués de $]1, +\infty[$, $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, f et g deux éléments de $\overline{\mathcal{V}}(\Omega, \mathcal{A}, \mu)$. On a*

$$\int_\Omega fg \, d\mu \leq \left(\int_\Omega |f(x)|^p \, d\mu(x) \right)^{1/p} \left(\int_\Omega |g(x)|^q \, d\mu(x) \right)^{1/q}.$$

Démonstration. Si f est nulle μ presque partout, alors l'inégalité est évidente (c'est en fait une égalité). Idem pour g. Dans le cas inverse, on a

$$\left(\int_\Omega |f(x)|^p \, d\mu(x) \right)^{1/p} > 0 \text{ et } \left(\int_\Omega |g(x)|^q \, d\mu(x) \right)^{1/q} > 0.$$

Bien sûr, $\int_\Omega fg\,d\mu \le \int_\Omega |f|.|g|\,d\mu$, donc remplaçant f par $f/\left(\int_\Omega |f(x)|^p\ d\mu(x)\right)^{1/p}$ et g par $g/\left(\int_\Omega |g(x)|^q\ d\mu(x)\right)^{1/q}$, on peut se ramener au cas où f et g sont positives avec $\left(\int_\Omega f(x)^p\ d\mu(x)\right)^{1/p} = \left(\int_\Omega g(x)^q\ d\mu(x)\right)^{1/q} = 1$.

Or pour tous x, y dans $\overline{\mathbb{R}}_+$, on a

$$xy \le \frac{x^p}{p} + \frac{y^q}{q}.$$

Si x ou y est infini, c'est évident. Sinon, on peut écrire $x = e^{a/p}$, $y = e^{y/q}$ et appliquer la convexité de la fonction exponentielle.

Ainsi

$$f(x)g(x) \le \frac{f(x)^p}{p} + \frac{g(x)^q}{q},$$

d'où

$$\int_\Omega f(x)g(x)\ d\mu(x) \le \int_\Omega \frac{f(x)^p}{p}\ d\mu(x) + \int_\Omega \frac{g(x)^q}{q}\ d\mu(x) = 1/p + 1/q = 1.$$

□

7.1.2 Inégalité triangulaire

Théorème 83 (Inégalité triangulaire)**.** *Soient $p \in [1, +\infty[$, $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, f et g deux éléments de $\overline{\mathcal{V}}(\Omega, \mathcal{A}, \mu)$. On a*

$$\left(\int_\Omega |f(x) + g(x)|^p\ d\mu(x)\right)^{1/p} \le \left(\int_\Omega |f(x)|^p\ d\mu(x)\right)^{1/p} + \left(\int_\Omega |g(x)|^p\ d\mu(x)\right)^{1/p}.$$

Démonstration. Dans le cas où $p = 1$, c'est une conséquence immédiate de l'inégalité triangulaire sur $\mathbb{R}$ et de la positivité de l'intégrale. Supposons donc $p \in]1, +\infty[$ et notons q l'exposant conjugué de p. Comme précédemment, on peut supposer que f et g ne sont pas presque partout nulles. Par ailleurs, si $\int_\Omega |f(x)|^p\ d\mu(x) = +\infty$ ou que $\int_\Omega |g(x)|^p\ d\mu(x) = +\infty$, l'inégalité est évidente. On suppose donc que ces deux quantités sont finies. Comme $|f + g|^p \le (|f| + |g|)^p$, on peut supposer sans perte de généralité que f et g sont positives. Maintenant, comme $\left(\frac{f+g}{2}\right)^p \le \frac{f^p + g^p}{2}$ par convexité de $x \mapsto x^p$, il s'ensuit qu'on a également $\int_\Omega |f(x) + g(x)|^p\ d\mu(x) < +\infty$.

On écrit alors

$$(f + g)^p = f(f + g)^{p-1} + g(f + g)^{p-1}.$$

L'inégalité de Hölder donne

$$\int_\Omega f(f+g)^{p-1}\,d\mu \le \left(\int_\Omega f^p\,d\mu\right)^{1/p}\left(\int_\Omega (f+g)^{(p-1)q}\,d\mu\right)^{1/q},$$

soit

$$\int_\Omega f(f+g)^{p-1}\,d\mu \le \left(\int_\Omega f^p\,d\mu\right)^{1/p}\left(\int_\Omega (f+g)^{p}\,d\mu\right)^{1/q}.$$

De même,

$$\int_\Omega g(f+g)^{p-1}\,d\mu \le \left(\int_\Omega g^p\,d\mu\right)^{1/p}\left(\int_\Omega (f+g)^{p}\,d\mu\right)^{1/q}.$$

En additionnant, on obtient

$$\int_\Omega (f+g)^p\,d\mu \le \left(\left(\int_\Omega f^p\,d\mu\right)^{1/p}+\left(\int_\Omega g^p\,d\mu\right)^{1/p}\right)\left(\int_\Omega (f+g)^{p}\,d\mu\right)^{1/q}.$$

D'où

$$\left(\int_\Omega (f+g)^p\,d\mu\right)^{1/p} \le \left(\int_\Omega f^p\,d\mu\right)^{1/p}+\left(\int_\Omega g^p\,d\mu\right)^{1/p}.$$

□

Ainsi, il est maintenant simple de constater que si l'on pose

$$\|f\|_p = \left(\int_\Omega |f|^p\,d\mu\right)^{1/p},$$

on définit ainsi une semi-norme sur l'espace vectoriel $\mathcal{L}^p(\Omega,\mathcal{A},\mu)$.

Remarquons bien qu'en général, $\|.\|_p$ ne définit pas une norme sur $\mathcal{L}^p(\Omega,\mathcal{A},\mu)$ car l'axiome de séparation peut être pris en défaut. Ainsi, sur $\mathcal{L}^p(\mathbb{R},\mathcal{B}(\mathbb{R}),\lambda)$, on a bien $\|\mathbb{1}_{\mathbb{Q}}\|_p = 0$, mais bien sûr, $\mathbb{1}_{\mathbb{Q}} \neq 0$.

Notons $V = \{v \in \mathcal{L}^p; \|v\|_p = 0\}$. D'après l'inégalité triangulaire, V est un sous-espace vectoriel de $\mathcal{L}^p$. Un raisonnement simple (exercice !) permet en fait de montrer que $V = \{v \in \mathcal{L}^p; v = 0 \quad \mu \text{ p.p.}\}$

Notons L^p le quotient de l'espace vectoriel $\mathcal{L}^p$ par son sous-espace vectoriel V.

Soit f et g deux éléments de la même classe : $k = f - g \in V$. D'après l'inégalité triangulaire $\|f\|_p \le \|g\|_p + \|k\|_p = \|g\|_p$. De même $\|g\|_p \le \|f\|_p + \|k\|_p = \|f\|_p$, d'où $\|f\|_p = \|g\|_p$. La semi-norme passe donc au quotient :

pour $f \in L^p$, on note $\|f\|_p = \|g\|_p$ où g est un quelconque représentant de la classe f. Evidemment, $f \mapsto \|f\|_p$ est encore une semi-norme sur L^p.
Mais en réalité, $f \mapsto \|f\|_p$ est une norme sur L^p. En effet, supposons $\|f\|_p = 0$. Soit g un représentant de f : on a $\|g\|_p = 0$, donc $g \in V$, ce qui signifie que g est dans la classe de 0, donc f est le zéro de L^p.

Bien que $\mathcal{L}^p$ ne soit pas un espace vectoriel normé, on pourra lire fréquemment pour des fonctions $(f_n)_{n\ge1}, f$ de $\mathcal{L}^p$: $(f_n)_{n\ge1}$ converge dans $\mathcal{L}^p$ (ou parfois $(f_n)_{n\ge1}$ converge dans L^p) vers f. Cela signifie que $\lim\limits_{n\to+\infty} \|f_n - f\|_p = 0$, ou de manière équivalente, que la suite des classes dans L^p des éléments de $(f_n)_{n\ge1}$ converge dans L^p vers la classe de f dans L^p.

7.2 Complétude de L^p

Théorème 84. *Pour tout $p \in [1, +\infty[$, L^p est complet*

Lemme 15. *Soit f_n une suite d'éléments de L^p avec*

$$\sum_{n\ge1} \|f_n\|_< + \infty.$$

Alors la suite
$\sum_{k=1}^n f_k$ converge dans L^p quand n tend vers l'infini.

Démonstration. On note g_n un représentant de f_n. On va montrer qu'il existe une fonction g dans $\mathcal{L}^p$ telle que $\| \sum_{k=1}^n g_k - g\|_p$ tend vers 0, ce qui donnera la convergence de la suite $\sum_{k=1}^n f_k$ vers la classe de g.

Supposons d'abord que les (g_k) sont positives : dans ce cas la suite de fonctions $S_n = \sum_{k=1}^n g_k$ converge simplement vers une fonction g mesurable (éventuellement infinie en certains points) Cependant d'après l'inégalité triangulaire

$$\int_\Omega S_n^p \, d\mu \le (\sum_{k=1}^n \|g_k\|_p)^p,$$

et donc d'après le théorème de convergence monotone

$$\int_\Omega g^p \, d\mu \le (\sum_{k=1}^{+\infty} \|g_k\|_p)^p < +\infty.$$

Ainsi g est dans $\mathcal{L}^p$. Soient n et n' avec $n' \ge n$. On a $(S_{n'} - S_n)^p = (\sum_{k=n+1}^{n'} g_k)^p$. Faisons tendre n' vers $+\infty$: d'après le théorème de convergence dominée, on a

$$\int_\Omega (g - S_n)^p \, d\mu = \lim_{n' \to +\infty} \int_\Omega (S_{n'} - S_n)^p \, d\mu,$$

d'où

$$\|g - S_n\|_p = \lim_{n' \to +\infty} \|S_{n'} - S_n\|_p.$$

Cependant, d'après l'inégalité triangulaire

$$\begin{aligned} \|S_{n'} - S_n\|_p &\leq \sum_{k=n+1}^{n'} \|g_k\|_p \\ &\leq \sum_{k=n+1}^{+\infty} \|g_k\|_p, \end{aligned}$$

d'où

$$\|S_n - g\|_p \leq \sum_{k=n+1}^{+\infty} \|g_k\|_p.$$

Mais on reconnait là le reste d'une série convergente, donc $\|S_n - g\|_p$ tend bien vers 0.

Dans le cas général, écrivons $g_k = g_k^+ - g_k^-$. On définit évidemment $g^+ = \sum g_k^+$ et $g^- = \sum g_k^-$, $S_n^+ = \sum_{k=1}^n g_k^+$, $S_n^- = \sum_{k=1}^n g_k^-$. La série de terme général $\|g_k^+\|_p$ est convergente car $\|g_k^+\|_p \leq \|g_k\|_p$. On montre ainsi que $\|S_n^+ - g^+\|_p$ tend bien vers 0, et de même que $\|S_n^- - g^-\|_p$ tend bien vers 0. Enfin, l'inégalité triangulaire permet de conclure que $\|S_n - g\|_p$ tend bien vers 0. □

Ainsi, on a montré que dans L^p, toute série absolument convergente est convergente. Pour conclure, il suffit de s'appuyer sur le résultat d'analyse suivant.

Lemme 16. *Un espace vectoriel normé où toute série absolument convergente converge est complet.*

Démonstration. Remarquons d'abord que si une suite de Cauchy admet une sous-suite convergente, elle converge. En effet supposons (x_n) de Cauchy avec x_{n_k} qui converge vers l. Soit k_0 tel que $\|x_{n_k} - l\| \leq \varepsilon/2$ pour $k \geq k_0$ et b_0 tel que $\|x_k - x_{k'}\| \leq \varepsilon/2$ lorsque k et k' dépassent b_0. Alors $\|x_n - l\| \leq \varepsilon$ dès que n dépasse $\max(n_0, n_{b_0})$.

Soit maintenant x_n une suite de Cauchy dans un espace où toute série absolument convergente converge. On pose $n_0 = 1$, puis pour $k \geq 1$:

$$n_k = \inf\{n > n_{k-1} : i, i' \geq n \Longrightarrow \|x_i - x_{i'}\| \leq 2^{-k}\}.$$

Cette suite d'indices est strictement croissante et est bien définie car (x_k) est de Cauchy. Par construction $\|x_{n_k} - x_{n_{k+1}}\| \le 2^{-k}$ pour $k \ge 1$, donc la série de terme général $x_{n_k} - x_{n_{k+1}}$ est absolument convergente. Mais on a fait l'hypothèse ici qu'une série absolument convergente est convergente, donc elle est convergente, ce qui veut dire que x_{n_k} est convergente. (x_n) est donc une suite de Cauchy qui admet une sous-suite convergente, elle est donc convergente. □

Théorème 85. *Soient $f, (f_n)_{n\ge1}$ des fonctions dans $\mathcal{L}^p$ telles que $(f_n)_{n\ge1}$ converge dans $\mathcal{L}^p$ vers f. Alors, il existe une suite strictement croissante d'indices $(n_k)_{k\ge1}$ telle que $(f_{n_k})_{k\ge1}$ converge presque partout vers f.*

Démonstration. On pose $g_n = |f - f_n|^p$. (g_n) converge dans $\mathcal{L}^1$ vers 0 et nous devons montrer l'existence d'une suite strictement croissante d'indices $(n_k)_{k\ge1}$ telle que $(g_{n_k})_{k\ge1}$ converge presque partout vers 0.

On pose $n_0 = 1$, puis pour $k \ge 1$:

$$n_k = \inf\{n > n_{k-1} : i, i' \ge n \Longrightarrow \|g_i - g_{i'}\|_1 \le 2^{-k}\}.$$

Cette suite d'indices est strictement croissante et est bien définie car (g_k) est de Cauchy dans $\mathcal{L}^1$. Par construction $\|g_{n_k} - g_{n_{k+1}}\| \le 2^{-k}$ pour $k \ge 1$, donc la série de terme général $\|g_{n_k} - g_{n_{k+1}}\|_1$ est convergente. Mais

$$\begin{aligned}\sum_{n=1}^{+\infty} \|g_{n_k} - g_{n_{k+1}}\|_1 &= \sum_{n=1}^{+\infty} \int |g_{n_k} - g_{n_{k+1}}| \, d\mu \\ &= \int \sum_{n=1}^{+\infty} |g_{n_k} - g_{n_{k+1}}| \, d\mu\end{aligned}$$

La fonction positive

$$\sum_{n=1}^{+\infty} |g_{n_k} - g_{n_{k+1}}|$$

est intégrable, elle est donc en particulier finie presque partout. En un point x tel que

$$\sum_{n=1}^{+\infty} |g_{n_k}(x) - g_{n_{k+1}}(x)| < +\infty,$$

la suite $(g_{n_k}(x))_{k\ge1}$ converge. $(g_{n_k})_{k\ge1}$ converge presque partout vers une fonction g^*. Mais d'après le lemme de Fatou,

$$\int g^* d\mu = \int \underline{\lim}\; g_{n_k} d\mu \le \underline{\lim} \int g_{n_k} \, d\mu = 0,$$

donc g^* est nulle presque partout, ce qui achève la preuve.

□

7.3 Théorèmes d'approximation

Théorème 86. *Soit $\mathcal{S}$ l'ensemble des fonctions simples s sur $(\Omega, \mathcal{F})$ telles que*

$$\mu(\{x \in \Omega; s(x) \neq 0\}) < +\infty.$$

Pour tout $p \in [1, +\infty[$, $\mathcal{S}$ est dense dans $L^p(\mu)$ (et donc les classes de ces fonctions sont denses dans $L^p(\mu)$).

Démonstration. D'abord, il est facile de voir que $\mathcal{S}$ est dans $L^p(\mu)$. Soit $f \in \mathcal{L}^p$. Supposons $f \geq 0$ et prenons f_n comme dans le lemme 2. On a

$$1\!\!1_{\{f_n>0\}} 2^{-np} \leq 1\!\!1_{\{f_n>0\}} f_n^p \leq f^p,$$

d'où

$$\mu(f_n > 0) 2^{-np} \leq \int_\Omega f^p \, d\mu,$$

et donc $f_n \in \mathcal{S}$. On a $|f_n - f|^p \leq f^p$, donc d'après le théorème de convergence dominée, $\int_\Omega |f_n - f|^p \, d\mu$ tend vers 0, c'est à dire que f_n tend vers f dans $\mathcal{L}^p$. Le cas général s'ensuit en séparant partie positive et partie négative, comme dans la preuve du théorème 84. $\square$

Théorème 87. *Soit $p \in [1, +\infty[$. Les classes des fonctions continues à support compact forment une partie dense dans $L^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$*

Ce théorème est admis.

7.4 Exercices sur les espaces $\mathcal{L}^p$ et L^p

1. Étudier l'appartenance à $\mathcal{L}^1(\mathbb{R})$ et à $\mathcal{L}^2(\mathbb{R})$ des fonction suivantes :
   (a) $f(t) = e^{-|t|}$.
   (b) $g(t) = \frac{\sin t}{t}$.
   (c) $g(t) = \frac{1}{\sqrt{|t|(1+t^2)}}$.
2. Étudier la convergence dans $\mathcal{L}^1(\mathbb{R})$ et dans $\mathcal{L}^2(\mathbb{R})$ des suites suivantes :
   (a) $f_n(t) = \sqrt{n} \exp(-n^2 t^2)$.
   (b) $g_n(t) = \frac{n^2 \sin nt}{2\pi} 1\!\!1_{[-\pi/n, \pi/n]}(t)$.
   (c) $h_n(t) = \frac{2}{\pi n^2} \sqrt{n^2 - t^2} 1\!\!1_{[-/n,n]}(t)$.
3. Soit f une fonction de $\mathbb{R}$ dans $\mathbb{R}$ intégrable et soit $\tilde{f}$ la classe de f dans $L^1(\mathbb{R}, \lambda)$. Montrer que $\tilde{f}$ contient au plus une fonction continue.

4. Soit $E = \{f \in \mathcal{L}^1(\mathbb{R}, \lambda) \ ; \ |f| \leq 1 \ \lambda - p.p\}$. Montrer que E est un sous-ensemble fermé de $\mathcal{L}^1(\mathbb{R}, \lambda)$.
5. Montrer que la fonction $f \ : \ x \mapsto \dfrac{1}{\sqrt{x}(1 + |\ln(x)|)}$ est dans $\mathcal{L}^p(]0, +\infty[, \lambda)$ si et seulement si $p = 2$.
6. Montrer que si f et g appartiennent à $\mathcal{L}^1(X, \mu)$ alors $\sqrt{|f^2 + g^2|}$ appartient aussi à $\mathcal{L}^1(X, \mu)$.
7. Soient $\alpha \in \mathbb{R}$ et $p \in [1, \ +\infty[$, on note f_α l'application de $]0, \ +\infty[$ dans $]0, \ +\infty[$, définie par $f_\alpha(x) = x^\alpha$.
   (a) Pour quelle(s) valeur(s) de α, la fonction f_α appartient-elle à $\mathcal{L}^p(]0, 1], \lambda)$? Calculer alors les normes de f_α dans chacun de ces espaces.
   (b) Même question avec les espaces $\mathcal{L}^p([1, \infty[, \lambda)$
8. Donner un exemple de suite (f_n) dans $\mathcal{L}^1(X, \mu)$ telle que
   (a) (f_n) converge vers f presque partout mais (f_n) ne converge pas vers f au sens de la norme $\mathcal{L}^1$;
   (b) (f_n) converge vers f dans $\mathcal{L}^1$ mais (f_n) ne converge pas vers f presque partout ;
   (c) (f_n) converge vers f presque partout, $(\int f_n d\mu)$ converge vers $\int f d\mu$, mais (f_n) ne converge pas vers f au sens de la norme $\mathcal{L}^1$.
9. Soient $(X, \ \mathcal{M}, \ \mu)$ un espace mesuré avec $\mu(X) = 1$ et f, g des fonctions mesurables sur X à valeurs dans $[0, \ +\infty]$ telles que $fg \geq 1$. Montrer que l'on a $\left(\int_X f d\mu\right)\left(\int_X g d\mu\right) \ \geq \ 1$.
10. Soient $f \in \mathcal{L}^p(X, \mu)$, $g \in \mathcal{L}^q(X, \mu)$ et r tel que $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$. Montrer que $fg \in \mathcal{L}^r(X, \mu)$ et que $||fg||_r \ \leq \ ||f||_p ||g||_q$.
11. Soit $p \in]1, +\infty[$. Pour f dans $\mathcal{L}^p(]0, +\infty[)$ et pour $x > 0$, on pose
$$T(f)(x) \ = \ \frac{1}{x} \int_{]0,x[} f d\lambda \ .$$
   (a) Montrer que $T(f)$ est bien définie sur $]0, +\infty[$.
   (b) On suppose dans cette question que f est positive continue à support compact.
      i. Montrer que $T(f)$ est dérivable sur $]0, +\infty[$ et calculer sa dérivée.
      ii. Montrer que $T(f) \in \mathcal{L}^p(]0, +\infty[)$.

iii. Montrer que $\displaystyle\int_{]0,\infty[} T(f)^p d\lambda = \frac{p}{p-1}\int_{]0,\infty[} T(f)^{p-1} f d\lambda$.

iv. En déduire que $||T(f)||_p \leq \dfrac{p}{p-1}||f||_p$.

v. Montrer que cette inégalité reste vraie pour f de signe quelconque.

(c) Soit $f \in \mathcal{L}^p(]0,+\infty[)$.

i. Montrer que si (f_n) est une suite de fonctions continues à support compact qui converge vers f dans $\mathcal{L}^p(]0,+\infty[)$, alors $T(f_n)$ converge vers $T(f)$ $\lambda-$presque partout, puis que la suite $(T(f_n))$ est de Cauchy dans $L^p(]0,+\infty[)$ et enfin que $(T(f_n))$ converge vers $T(f)$ dans $L^p(]0,+\infty[)$.

ii. En déduire que $||T(f)||_p \leq \dfrac{p}{p-1}||f||_p$.

12. Soit $(X, \mathcal{T}, \mu)$ un espace mesuré tel que $\mu(X) < \infty$, $1 < p < \infty$ et $f : X \to \mathbb{R}$ une application borélienne. On suppose que pour toute fonction $g \in \mathcal{L}^p(X,\mu)$, la fonction fg est intégrable et il existe $C > 0$ telle que pour toute fonction $g \in \mathcal{L}^p(X,\mu)$ on ait $\left|\int fg\, d\mu\right| \leq C||g||_p$.
Montrer que $f \in \mathcal{L}^q(X,\mu)$ où q est défini par $\frac{1}{p} + \frac{1}{q} = 1$

Chapitre 8

Convolution et transformation de Fourier

8.1 Produit de convolution

Remarques :

- Si f_1, f_2 sont deux fonctions de $\mathcal{L}^1$ qui représentent le même élément de L^1, alors $\int f_1 \, d\mu$ et $\int f_2 \, d\mu$ sont égales, donc on peut se permettre d'écrire $\int f \, d\mu$ pour $f \in L^1$.
- L'application $T_t : f \mapsto (x \mapsto f(x-t))$ passe au quotient dans $\mathcal{L}^1(\mathbb{R}^d)$, car si $f_1 = f_2$ presque sûrement, alors $f_1(. - t) = f_2(. - t)$ presque sûrement.

Théorème 88. *approxsimple Pour tout f dans L^p, l'application*

$$t \mapsto T_t f$$

est continue sur $\mathbb{R}$.

Démonstration. $\|T_{t+h}f - T_t f\|_p = T_h(T_t f) - (T_t f)\|_p$, donc il suffit de montrer la continuité en 0. Traitons d'abord le cas où f est une fonction continue à support compact : comme f est continue $T_h f$ tend simplement vers f. En utilisant le théorème de convergence dominée, on optient alors la convergence dans $\mathcal{L}^p$ de $T_h f$ vers f. Passons au cas général. D'après le théorème 87, on peut trouver g et h, avec $f = g + h$, g continue à support compact et $\|h\|_p \leq \varepsilon$. On a

$$(T_t f - f) = (T_t g - g) + (T_t h - h),$$

d'où

$$\begin{aligned}\|T_t f - f\|_p &\leq \|T_t g - g\|_p + \|T_t h\|_p + \|h\|_p \\ &\leq \|T_t g - g\|_p + 2\|h\|_p,\end{aligned}$$

ce qui entraîne, en faisant tendre t vers 0

$$\overline{\lim_{t\to 0}}\ \|T_t f - f\|_p \leq 2\varepsilon.$$

Comme c'est vrai pour tout ε, on en déduit que $\overline{\lim\limits_{t\to 0}}\ \|T_t f - f\|_p = 0$, ce qui est bien ce qu'on voulait montrer. □

8.1.1 convolution dans $\mathcal{L}^1$

Soient f, g deux éléments de $\mathcal{L}^1(\mu)$.

$$\begin{aligned}\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} |f(x-t)||g(t)|\ d\lambda^d(t)\right) d\lambda^d(x) &= \int_{\mathbb{R}^d\times\mathbb{R}^d} |f(x-t)||g(t)|\ d\lambda^d\otimes\lambda^d(t,x) \\ &= \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} |f(x-t)||g(t)|\ d\lambda^d(x)\ \right) d\lambda^d(t) \\ &= \int_{\mathbb{R}^d} |g(t)| \left(\int_{\mathbb{R}^d} |f(x-t)|\ d\lambda^d(x)\ \right) d\lambda^d(t) \\ &= \int_{\mathbb{R}^d} |g(t)| \left(\int_{\mathbb{R}^d} |f(x)|\ d\lambda^d(x)\ \right) d\lambda^d(t) \\ &= \left(\int_{\mathbb{R}^d} |f(x)|\ d\lambda^d(x)\ \right)\left(\int_{\mathbb{R}^d} |g(t)|\ d\lambda^d(t)\ \right) \\ &< +\infty.\end{aligned}$$

Ainsi, la fonction $f * g$ définie par

$$x \mapsto f * g(x) = \int_{\mathbb{R}^d} f(x-t)g(t)\ d\lambda^d(t)$$

est définie en presque tout point x et elle est dans $\mathcal{L}^1$: cette fonction est le produit de convolution de f par g

Les arguments évoqués plus haut fonctionnent encore : le produit de convolution "passe au quotient" et définit ainsi une application de $L^1 \times L^1$ dans L^1.

Au passage, notons qu'on a démontré

$$\|f * g\|_1 \leq \|f\|_1 \|g\|_1.$$

En reprenant le calcul précédant et en supposant que f et g sont dans $\mathcal{L}^1$, le théorème de Fubini permet alors d'écrire

$$\begin{aligned}
\int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(x-t)g(t)\ d\lambda^d(t) \right) d\lambda^d(x) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x-t)g(t)\ d\lambda^d \otimes \lambda^d(t,x) \\
&= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(x-t)g(t)\ d\lambda^d(x) \right) d\lambda^d(t) \\
&= \int_{\mathbb{R}^d} g(t) \left(\int_{\mathbb{R}^d} f(x-t)\ d\lambda^d(x) \right) d\lambda^d(t) \\
&= \int_{\mathbb{R}^d} g(t) \left(\int_{\mathbb{R}^d} f(x)\ d\lambda^d(x) \right) d\lambda^d(t) \\
&= \left(\int_{\mathbb{R}^d} f(x)\ d\lambda^d(x) \right) \left(\int_{\mathbb{R}^d} g(t)\ d\lambda^d(t) \right),
\end{aligned}$$

soit

$$\int_{\mathbb{R}^d} (f * g)(x)\ d\lambda^d(x) = \left(\int_{\mathbb{R}^d} f(x)\ d\lambda^d(x) \right) \left(\int_{\mathbb{R}^d} g(t)\ d\lambda^d(t) \right). \tag{8.1}$$

8.1.2 autres produits

Supposons maintenant que $g \in \mathcal{L}^1$ et que $f \in \mathcal{L}^p$. On a

$$\begin{aligned}
\int |f(x-t)||g(t)|\ d\lambda^d(t) &= \int |f(x-t)||g(t)|^{1/p}|g(t)|^{1/q}\ d\lambda^d(t) \\
&\leq \left(\int |f(x-t)|^p |g(t)|\ d\lambda^d(t) \right)^{1/p} \left(\int |g(t)|\ d\lambda^d(t) \right)^{1/q}
\end{aligned}$$

D'où

$$\int \left(\int |f(x-t)||g(t)|\ d\lambda^d(t) \right)^p d\lambda^d(x) \leq \int \left(\int |f(x-t)|^p |g(t)|\ d\lambda^d(t) \right) d\lambda^d(x) \|g\|_1^{p/q}$$

Cependant,

$$\begin{aligned}
\int \left(\int |f(x-t)|^p |g(t)|\ d\lambda^d(t)\right) d\lambda^d(x) &= \int \left(\int |f(x-t)|^p |g(t)|\ d\lambda^d(x)\right) d\lambda^d(t) \\
&= \int \left(\int |f(x-t)|^p\ d\lambda^d(x)\right) |g(t)| d\lambda^d(t) \\
&= \int \|f\|_p^p |g(t)| d\lambda^d(t) \\
&= \|f\|_p^p \|g\|_1,
\end{aligned}$$

donc finalement

$$\int \left(\int |f(x-t)||g(t)|\ d\lambda^d(t)\right)^p\ d\lambda^d(x) \quad \leq \|f\|_p^p \|g\|_1^{1+p/q}$$

Ainsi, l'intégrale

$$\int f(x-t)g(t)\ d\lambda^d(t)$$

converge pour presque que tout x et l'application

$$x \mapsto f * g(t) = \int f(x-t)g(t)\ d\lambda^d(t)$$

représente un élément de $\mathcal{L}^p$ avec

$$\int |f * g(t)|^p\ d\lambda^d(t) \leq \|f\|_p^p \|g\|_1^{1+p/q},$$

soit

$$\|f * g\|_p \leq \|f\|_p \|g\|_1.$$

Remarque importante : quel que soit l'espace où on définit les choses, on a toujours

$$\int f(x-t)g(t)\ d\lambda^d(t) = \int g(x-t)f(t)\ d\lambda^d(t)$$

pour les x tels que

$$\int |f(x-t)g(t)|\ d\lambda^d(t) < +\infty,$$

de sorte que

$$f * g = g * f$$

toutes les fois où cela a un sens.

8.1.3 Approximations de l'unité

Théorème 89. *Soit φ une fonction positive avec*

$$\int_{\mathbb{R}^d} \varphi(x)\ d\lambda^d(x) = 1.$$

Pour tout $k \geq 1$ posons $\varphi_k(x) = k^d \varphi(kx)$.

Alors, pour tout f dans $\mathcal{L}^p$ la suite $f * \varphi_k$ converge vers f dans $\mathcal{L}^p$.

Démonstration. Notons M_k l'application qui a f associe $f * \varphi_k$. C'est une application linéaire continue de $\mathcal{L}^p$ dans lui même. C'est même une contraction car

$$\forall f \in \mathcal{L}^p \quad \|M_k f\|_p \leq \|f\|_p.$$

(C'est ce qu'on a montré dans la sous-section précédente.)

Soit $x \in \mathbb{R}^d$:

$$\begin{aligned}
f * \varphi_k(x) - f(x) &= \int f(x-t)\varphi(kt)k^d\ d\lambda^d(t) - f(x) \\
&= \int f(x-t/k)\varphi(t)\ d\lambda^d(t) - f(x) \\
&= \int f(x-t/k)\varphi(t)\ d\lambda^d(t) - \int f(x)\varphi(t)\ d\lambda^d(t) \\
&= \int (f(x-t/k) - f(x))\varphi(t)\ d\lambda^d(t)
\end{aligned}$$

Ainsi

$$\begin{aligned}
|(M_k f - f)(x)| &= |\int (T_{t/k} f - f)(x)\varphi(t)\ d\lambda^d(t)| \\
&\leq \left(\int |(T_{t/k} f - f)(x)|^p \varphi(t)\ d\lambda^d(t)\right)^{1/p},
\end{aligned}$$

ce qui donne

$$\|M_k f - f\|_p \leq \left(\int \|T_{t/k} f - f\|_p^p \varphi(t)\ d\lambda^d(t)\right)^{1/p}$$

D'après de théorème 88, $\|T_{t/k} f - f\|_p^p$ tend vers 0 lorsque k tend vers l'infini. Comme

$$|\|T_{t/k} f - f\|_p^p \varphi(t)| \leq (2\|f\|_p)^p \varphi(t),$$

le théorème de convergence dominée permet de conclure. □

8.1.4 Régularisation

Théorème 90. *Soit $f \in \mathcal{L}^p(\mathbb{R}^d)$, g C^1 à support compact. Alors $f * g$ est C^1 sur $\mathbb{R}^d$, avec*

$$D_x(f * g) = \int f(t) D_{x-t} g.$$

Démonstration. Soit M tel que $g(x) = 0$ pour $\|x\| \geq M$. Soit $R > 0$. Par définition

$$f * g(x) = \int f(x-t)g(t)\ d\lambda^d(t) = \int g(x-t)f(t)\ d\lambda^d(t).$$

Ici, c'est bien sûr la deuxième écriture qui va nous intéresser. Supposons $\|x\| \leq R$. La différentielle de $g(x-t)f(t)$, vue comme une fonction de x, est $f(t)D_{x-t}g$. Bien entendu

$$|f(t)D_{x-t}g| \leq |f(t)| \|Dg\|_\infty \mathbb{1}_{B(0,R+M)}(t).$$

$f \in \mathcal{L}^p$ et $\|Dg\|_\infty \mathbb{1}_{B(0,R+M)} \in \mathcal{L}^q$, donc $|f| \|Dg\|_\infty \mathbb{1}_{B(0,R+M)}$ est dans $\mathcal{L}^1$. Le théorème de convergence dominée pour la différenciation sous le signe somme donne alors le résultat voulu. □

Corollaire 19. *Soit $f \in \mathcal{L}^p(\mathbb{R}^d)$, g C^k à support compact. Alors $f * g$ est C^k sur $\mathbb{R}^d$, avec*

$$D_x^\alpha(f * g) = \int f(t) D_{x-t}^\alpha g,$$

où on a supposé que $|\alpha_1| + |\alpha_2| + \cdots + |\alpha_d| \leq k$.

Démonstration. Par récurrence sur k. □

Corollaire 20. *Les fonctions C^∞ à support compacts sont denses dans $\mathcal{L}^p$.*

Démonstration. Cela provient immédiatement du Théorème 89 et du corollaire précédent. □

8.2 transformée de Fourier

Soit $f \in \mathcal{L}^1(\mathbb{R}^d)$. On appelle transformée de Fourier de f, et l'on note $\hat{f}$ la fonction définie sur $\mathbb{R}^d$ par

$$\hat{f}(x) = \int e^{i\langle x,t\rangle} f(t)\ d\lambda^d(t).$$

Evidemment, $f \mapsto \hat{f}$ est linéraire, et comme $|e^{i\langle x,t\rangle} f(t)| \leq |f(t)|$, est on a

$$\forall f \in \mathcal{L}^1 \quad \|\hat{f}\|_\infty \leq \|f\|_1.$$

8.2.1 propriétés élémentaires

Pour $f, g \in \mathcal{L}^1$, on a
- $\widehat{f * g} = \hat{f}.\hat{g}$.
- $\widehat{T_t f}(x) = e^{i\langle x,t\rangle}\hat{f}(x)$
- Si $g(x) = f(x/\lambda)$, alors $\hat{g}(x) = \lambda^d \hat{f}(\lambda x)$.
- Si $g(x) = f(x)e^{i\langle x,\theta\rangle}$, alors $\hat{g}(x) = \lambda^d \hat{f}(x - \theta)$.
- $\int f(x)\ d\lambda^d(u) = \hat{f}(0)$

La première propriété mérite qu'on consacre quelques lignes à sa preuve :

$$\begin{aligned}
\widehat{f * g}(x) &= \int e^{i\langle x,t\rangle}\left(\int f(t-u)g(u)\ d\lambda^d(u)\right)\ d\lambda^d(t) \\
&= \int\left(\int e^{i\langle x,t\rangle}f(t-u)g(u)\ d\lambda^d(u)\right)\ d\lambda^d(t) \\
&= \int\left(\int f(t-u)e^{i\langle x,t-u\rangle}g(u)e^{i\langle x,u\rangle}\ d\lambda^d(u)\right)\ d\lambda^d(t) \\
&= \int (F * G)(t)\ d\lambda^d(t),
\end{aligned}$$

où $F(t) = f(t)e^{i\langle x,t\rangle}$ et $G(t) = g(t)e^{i\langle x,t\rangle}$. Mais d'après l'équation (8.1)

$$\int (F * G)(t)\ d\lambda^d(t) = \left(\int F(t)d\lambda^d(t)\right)\left(\int G(t)d\lambda^d(t)\right),$$

d'où le résultat voulu.

8.2.2 Théorème d'inversion

Théorème 91. *Soit $f \in \mathcal{L}^1(\mathbb{R}^d)$ telle que $\hat{f} \in \mathcal{L}^1(\mathbb{R}^d)$, alors on a*

$$f(x) = \frac{1}{(2\pi)^d}\int e^{-i\langle x,t\rangle}\hat{f}(t)\ d\lambda^d(t)\ p.p.$$

Démonstration. On aura besoin du lemme suivant, qui sera démontré (au moins une fois) en exercice.

Lemme 17. *Soit $G(x) = \frac{1}{(2\pi)^{d/2}}e^{-\|x\|_2^2}$. Alors $\hat{G}(x) = e^{-\|x\|_2^2} = (2\pi)^{d/2}G(x)$.*

Pour $k \geq 1$, posons $G_k(x) = k^d G(kx)$. On a

$$\hat{G}_k(x) = k^d k^{-d}\hat{G}(x/k) = (2\pi)^{d/2}G(x/k).$$

On recommence :

$$\widehat{\widehat{G_k}}(x) = (2\pi)^{d/2} k^d \hat{G}(kx) = (2\pi)^d k^d G(kx) = (2\pi)^d G_k(x),$$

et comme G_k est paire

$$\widehat{\widehat{G_k}}(-x) = (2\pi)^d G_k(x),$$

soit

$$\frac{1}{(2\pi)^d} \int e^{-i\langle x,t\rangle} \hat{G}_k(t)\ d\lambda^d(t) \quad = \quad G_k(x)$$

On a donc

$$\begin{aligned} f * G_k(x) &= \frac{1}{(2\pi)^d} \int \int e^{-i\langle x-y,t\rangle} \hat{G}_k(t) f(y)\ d\lambda^d(t)\ d\lambda^d(y) \\ &= \frac{1}{(2\pi)^d} \int e^{-i\langle x,t\rangle} \hat{G}_k(t) \hat{f}(t)\ d\lambda^d(t) \end{aligned}$$

En utilisant le théorème de convergence dominée, on voit que le terme de droite tend vers $\frac{1}{(2\pi)^d} \int e^{-i\langle x,t\rangle} \hat{f}(t)\ d\lambda^d(t)$ lorsque k tend vers l'infini. Mais d'après le théorème 89, le membre de gauche converge dans $\mathcal{L}^1$ vers f. Comme la convergence dans $\mathcal{L}^1$ entraîne la convergence d'une sous-suite presque partout, l'unicité de la limite donne l'égalité voulue. □

8.3 Exercices sur la convolution et la transformée de Fourier

1. Soient f et $g \in \mathcal{L}^1(\mathbb{R}^n)$. Montrer que si f (resp. g) est nulle presque-partout en dehors d'un ensemble A (resp. B) alors $f * g$ est nulle presque-partout en dehors de $A + B = \{a + b; a \in A, b \in B\}$.
2. Calculer le produit de convolution $f * g$ des fonctions suivantes définies sur $\mathbb{R}$ $(a > 0, b > 0)$:
   (a) $f(x) = \exp(-\frac{x^2}{2a^2})$ et $g(x) = \exp(-\frac{x^2}{2b^2})$.
   (On admettra que $\int \exp(-\frac{x^2}{2})\ dx = \sqrt{2\pi}$.)
   (b) $f(x) = \mathbb{1}_{[-a,a]}(x)$ et $g(x) = \mathbb{1}_{[-b,b]}(x)$
3. Pour tout entier n on définit la fonction $g_n(x) = (1 - x^2)^n \mathbb{1}_{[-1,1]}(x)$. On pose $a_n = \displaystyle\int_{\mathbb{R}} g_n(x)\, dx$ et $k_n = a_n^{-1} g_n$.

(a) Montrer que la suite (a_n) tend vers 0 et que $a_n \geq \frac{2}{n+1}$ pour tout entier n.

(b) Soit f une fonction uniformément continue sur $\mathbb{R}$ et bornée. Montrer que $f * k_n$ converge uniformément vers f.

(c) Soit f une fonction continue à support dans $[-\frac{1}{2}, \frac{1}{2}]$. Montrer que la restriction de $f * k_n$ à $[-\frac{1}{2}, \frac{1}{2}]$ est un polynôme de degré $\leq 2n$.

(d) En déduire le théorème de Weierstrass : Toute fonction continue d'un intervalle $[a, b]$ dans $\mathbb{R}$ est limite uniforme sur $[a, b]$ d'une suite de polynômes.

4. Soit $f = 1\!\!1_{[-\frac{1}{2},\frac{1}{2}]}$.
   (a) Déterminer $f * f$ et $f * f * f$.
   (b) On note $f^{(*)1} = f$ et pour $n \geq 2$, $f^{(*)n} = f^{(*)(n-1)} * f$. Vérifier que pour tout $n \geq 1$, $f^{(*)n} \in \mathcal{L}^1(\mathbb{R})$ et que $\|f^{(*)n}\|_1 = 1$.
   (c) Montrer que pour tout $n \geq 2$, $f^{(*)n}$ est de classe $\mathcal{C}^{n-2}$.
5. Soit $E \in \mathcal{B}(\mathbb{R})$ tel que $0 < \lambda(E) < +\infty$.
   (a) Montrer que $1\!\!1_E * 1\!\!1_{-E}$ est continue sur $\mathbb{R}$.
   (b) En déduire que $E - E = \{x - y \,/\, x \in E,\, y \in E\}$ est un voisinage de 0.
6. Soit f la fonction de $\mathbb{R}$ dans lui-même définie par $f(x) = e^{-\frac{x^2}{2}}$ pour $x \in \mathbb{R}$.
   (a) Déterminer la transformée de Fourier de f en remarquant que $\widehat{f}$ est solution d'une équation différentielle linéaire.
   (b) Soit A une matrice carrée réelle symétrique d'ordre n définie positive. Déterminer la transformée de Fourier de la fonction de $\mathbb{R}^n$ dans $\mathbb{R}$ définie par $f(x) = e^{-\langle Ax,x\rangle}$ pour $x \in \mathbb{R}^n$.
7. (a) Soit $f \in L^1(\mathbb{R}^n)$ telle que $f * f = 0$. Montrer que $f = 0$.
   (b) Montrer que $\mathcal{L}^1(\mathbb{R})$ n'a pas d'unité pour la convolution.
8. Déterminer la transformée de Fourier de la fonction indicatrice d'un intervalle $[a, b]$. Montrer que $1\!\!1_{[-1,1]} * 1\!\!1_{[-1,1]}$ est la transformée de Fourier d'une fonction de $L^1(\mathbb{R})$ qu'on déterminera.
9. Calculer la transformée de Fourier de la fonction f de $\mathbb{R}$ dans lui-même définie par $f(x) = e^{-a|x|}$, pour $x \in \mathbb{R}$ (où $a > 0$). En déduire la transformée de Fourier de la fonction $g : x \mapsto \dfrac{1}{a^2 + x^2}$.

Chapitre 9

Convergence presque sûre, loi des grands nombres

9.1 Inégalités classiques

9.1.1 Inégalité de Markov

Théorème 92. *Soit X une variable aléatoire positive, intégrable. Alors, on a*

$$\forall a > 0 \quad \mathbb{P}(X \geq a) \leq \frac{\mathbb{E}X}{a}.$$

Démonstration. Comme X est positive, on a

$$\mathbb{E}X = \int_{\Omega} x \ d\mathbb{P}(\omega) \geq \int_{\{X \geq a\}} x \ d\mathbb{P}(\omega) \geq \int_{\{X \geq a\}} a \ d\mathbb{P}(\omega) = a\mathbb{P}(X \geq a).$$

□

9.1.2 Inégalité de Tchebytchef

Théorème 93. *Soit X une variable aléatoire admettant un moment d'ordre 2. Alors, on a*

$$\forall a > 0 \quad \mathbb{P}(|X - \mathbb{E}X| \geq a) \leq \frac{\operatorname{Var} X}{a^2}.$$

Démonstration.

$$\mathbb{P}(|X - \mathbb{E}X| \geq a) = \mathbb{P}(|X - \mathbb{E}X|^2 \geq a^2)$$

Il suffit alors d'appliquer l'inégalité de Markov à la variable aléatoire $Y = |X - \mathbb{E}X|^2$. Comme $\mathbb{E}Y = \operatorname{Var} X$, l'inégalité s'ensuit. □

9.2 Convergence presque sûre

Définition : on dit qu'une suite de variables (ou de vecteurs) aléatoires $(X_n)_{n\geq 0}$ converge presque sûrement vers une variable (ou un vecteur) aléatoire X lorsqu'il existe un ensemble mesurable $\Omega' \subset \Omega$ tel que $\mathbb{P}(\Omega') = 1$ et que

$$\forall \omega \in \Omega' \subset \Omega \quad X_n(\omega) \to X(\omega).$$

On écrit alors $X_n \xrightarrow[n\to+\infty]{\text{p.s.}} X$.

La convergence presque sûre n'est autre que la convergence presque partout relativement à une mesure de probabilité. On a alors les résultats classiques suivants : si $X_n \xrightarrow[n\to+\infty]{\text{p.s.}} X$ et $Y_n \xrightarrow[n\to+\infty]{\text{p.s.}} Y$, (avec X et Y dans $\mathbb{R}^d$, $d \geq 1$) alors

- $\forall a \in \mathbb{R} \quad aX_n \xrightarrow[n\to+\infty]{\text{p.s.}} aX$.
- $X_n + Y_n \xrightarrow[n\to+\infty]{\text{p.s.}} X + Y$.
- $\langle X_n, Y_n\rangle \xrightarrow[n\to+\infty]{\text{p.s.}} \langle X, Y\rangle$.

Plus généralement, si $X_1, \dots, X_n, \dots, X$ sont à valeurs dans un ouvert O et que $X_n \xrightarrow[n\to+\infty]{\text{p.s.}} X$, alors pour toute fonction f continue définie sur O, on a $f(X_n) \xrightarrow[n\to+\infty]{\text{p.s.}} f(X)$.

Il peut être intéressant de remarquer que la convergence presque sûre d'une suite de vecteurs aléatoires est équivalente à la convergence presque sûre de chacune des composantes.

9.2.1 Rappels d'analyse

En probabilités, le retour aux ε est très fréquent. Si l'on ne veut pas que cela devienne trop compliqué, il importe de bien connaître les outils d'analyse permettant de simplifier les choses.

Pour toute suite $(x_n)_{n\geq 0}$ de nombres réels, on peut définir

$$\overline{\lim_{n\to+\infty}}\ x_n = \lim_{n\to+\infty} \sup_{k\geq n} x_k$$

et

$$\underline{\lim_{n\to+\infty}}\ x_n = \lim_{n\to+\infty} \inf_{k\geq n} x_k.$$

Ces deux limites existent toujours dans $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty; +\infty\}$. La suite $(x_n)_n$ converge dans $\overline{\mathbb{R}}$ si et seulement si ces deux limites sont égales. Rappelons quelques propriétés des limites supérieures.

– Pour $x \in \mathbb{R}$

$$\lim_{n\to+\infty} x_n = x \iff \overline{\lim_{n\to+\infty}} |x_n - x| = 0$$

–

$$\overline{\lim_{n\to+\infty}} \; x_n \leq M \iff \forall \varepsilon > 0 \quad \{n : x_n \geq M + \varepsilon\} \text{ est fini} \tag{9.1}$$

$$\overline{\lim_{n\to+\infty}} \; x_n \geq M \iff \forall \varepsilon > 0 \quad \{n : x_n \geq M - \varepsilon\} \text{ est infini} \tag{9.2}$$

– En prenant la contraposée de (9.1) et(9.2) , on a

$$\overline{\lim_{n\to+\infty}} \; x_n > M \iff \exists \varepsilon > 0 \quad \{n : x_n \geq M + \varepsilon\} \text{ est infini}$$

$$\overline{\lim_{n\to+\infty}} \; x_n < M \iff \exists \varepsilon > 0 \quad \{n : x_n \geq M - \varepsilon\} \text{ est fini}$$

–

$$\overline{\lim_{n\to+\infty}} \; (x_n + y_n) \leq \overline{\lim_{n\to+\infty}} \; x_n + \overline{\lim_{n\to+\infty}} \; y_n.$$

Nous verrons en exercice un exemple où l'inégalité est stricte

9.2.2 Limites supérieures, inférieures d'ensembles

Si $(A_n)_{n\geq 0}$ est une suite d'ensembles, on note

$$\overline{\lim_{n\to+\infty}} \; A_n = \bigcap_{n\geq 1} \bigcup_{k\geq n} A_k$$

et

$$\underline{\lim_{n\to+\infty}} \; A_n = \bigcup_{n\geq 1} \bigcap_{k\geq n} A_k.$$

Ainsi la limite supérieure d'une suite d'ensembles est l'ensemble des points qui appartiennent à une infinité de ces ensembles, tandis que la limite inférieure d'une suite d'ensembles est l'ensemble des points qui appartient à tous ces ensembles à partir d'un certain rang.

Ainsi, dire que $\{n : X_n(\omega) \geq M - \varepsilon\}$ est infini , c'est dire que ω appartient à $\overline{\lim_{n\to+\infty}} \; \{\omega : X_n(\omega) \geq M - \varepsilon\}$.

On en déduit

$$\{ \overline{\lim_{n\to+\infty}} X_n \geq M\} = \cap_{\varepsilon>0} \overline{\lim_{n\to+\infty}} \{X_n \geq M - \varepsilon\} \tag{9.3}$$

Par ailleurs, dire que

$$\{n : x_n \geq M + \varepsilon\} \text{ est fini },$$

c'est dire qu'à partir d'un certain rang , on a $x_n < M + \varepsilon$. Donc si ω est tel que $\{n : X_n(\omega) \geq M + \varepsilon\}$ est fini , c'est que $\omega \in \underline{\lim_{n\to+\infty}} \{X_n < M + \varepsilon\}$. On en déduit que

$$\{ \overline{\lim_{n\to+\infty}} X_n \leq M\} = \cap_{\varepsilon>0} \underline{\lim_{n\to+\infty}} \{X_n < M + \varepsilon\} \tag{9.4}$$

Si on remplace X_n par $-X_n$ et M par $-M$ dans (9.4), on obtient :

$$\{ \overline{\lim_{n\to+\infty}} - X_n \leq -M\} = \cap_{\varepsilon>0} \underline{\lim_{n\to+\infty}} \{-X_n < -M + \varepsilon\}$$

Soit

$$\{ \underline{\lim_{n\to+\infty}} X_n \geq M\} = \cap_{\varepsilon>0} \underline{\lim_{n\to+\infty}} \{X_n > M - \varepsilon\} \tag{9.5}$$

Et en passant aux complémentaires dans (9.4) et (9.5), on a

$$\{ \overline{\lim_{n\to+\infty}} X_n > M\} = \cup_{\varepsilon>0} \overline{\lim_{n\to+\infty}} \{X_n \geq M + \varepsilon\} \tag{9.6}$$

et

$$\{ \underline{\lim_{n\to+\infty}} X_n < M\} = \cup_{\varepsilon>0} \overline{\lim_{n\to+\infty}} \{X_n \leq M - \varepsilon\} \tag{9.7}$$

Si on fait subir à la formule (9.3) les mêmes transformations qu'à (9.4), on peut obtenir 3 autres formules.

Dans la pratique, comment fait-on si l'on veut montrer que $\overline{\lim_{n\to+\infty}} X_n = M$ presque sûrement ? Comme vous l'avez deviné, on montre $\overline{\lim_{n\to+\infty}} X_n \geq M$ presque sûrement, puis $\overline{\lim_{n\to+\infty}} X_n \leq M$ presque sûrement Comme la suite $\overline{\lim_{n\to+\infty}} \{X_n \geq M - \varepsilon\}$ est monotone en ε, on a

$$\{ \overline{\lim_{n\to+\infty}} X_n \geq M\} = \cap_{\varepsilon>\in\mathbb{Q}_+^*} \overline{\lim_{n\to+\infty}} \{X_n \geq M - \varepsilon\} \tag{9.8}$$

L'avantage est que l'intersection est maintenant dénombrable. Or, on a le résultat classique très utile suivant :

Théorème 94. *L'intersection d'une famille dénombrable d'événements est de probabilité 1 si et seulement si chacun des événements est de probabilité 1.*

Démonstration. Soit D un ensemble d'indexs dénombrable. $(A_n)_{n\in D}$ une famille d'événements indexée par D. On pose $A = \bigcap_{n\in D} A_n$. Pour tout n, $A \subset A_n$, donc $\mathbb{P}(A) \leq \mathbb{P}(A_n)$. Ainsi si $\mathbb{P}(A) = 1$, on a pour tout $n \in D$ $\mathbb{P}(A_n) = 1$. Réciproquement, on a

$$\begin{aligned} \mathbb{P}(A^c) &= P(\bigcup_{n\in D} A_n^c) \\ &\leq \sum_{n\in D} \mathbb{P}(A_n^c) \\ &\leq \sum_{n\in D} 0 \end{aligned}$$

Donc $\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - 0 = 1$. □

Pour prouver que $\overline{\lim_{n\to+\infty}} X_n \geq M$ presque sûrement, il suffit donc de prouver que $\forall a < M \quad \mathbb{P}(\overline{\lim_{n\to+\infty}} \{X_n \geq a\}) = 1$

De la même manière, on voit que pour avoir $\overline{\lim_{n\to+\infty}} X_n \leq M$ presque sûrement, il suffit donc de prouver que $\forall a > M \quad \mathbb{P}(\underline{\lim_{n\to+\infty}} \{X_n < a\}) = 1$, on de manière équivalente que $\forall a > M \quad \mathbb{P}(\overline{\lim_{n\to+\infty}} \{X_n \geq a\}) = 0$.

On peut donc énoncer le théorème suivant

Théorème 95. *Soit X_n une suite de variables aléatoires et M un réel. On suppose que*

1. $\forall \quad a < M \quad \mathbb{P}(\overline{\lim_{n\to+\infty}} \{X_n \geq a\}) = 1$

2. $\forall \quad a > M \quad \mathbb{P}(\overline{\lim_{n\to+\infty}} \{X_n \geq a\}) = 0$

Alors

$$\overline{\lim_{n\to+\infty}} X_n = M \text{ presque sûrement.}$$

Le théorème suivant très important en est une application directe

Théorème 96 (Critère fondamental de convergence presque-sûre)**.** *La suite de variables aléatoires X_n converge presque sûrement vers la variable aléatoire X si et seulement si*

$$\forall \varepsilon > 0 \quad \mathbb{P}(\overline{\lim_{n\to+\infty}} \{|X_n - X| \geq \varepsilon\}) = 0.$$

Démonstration. Il suffit d'appliquer le lemme précédent à la suite de variables aléatoires $(|X_n - X|)_{n\geq 0}$, avec $M = 0$ et a joue le rôle de ε. □

9.3 Convergence en probabilité

Définition : On dit que (X_n) converge en probabilité vers X si

$$\forall \varepsilon > 0 \qquad \lim_{n\to+\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0.$$

9.3.1 Comparaison avec les autres modes de convergence

Convergence dans L^p et convergence en probabilité

Théorème 97. *La convergence dans L^p ($p \geq 1$) implique la convergence en probabilité*

Démonstration.

$$\mathbb{P}(\|X_n - X\| \geq \varepsilon) = \mathbb{P}(\|X_n - X\|^p \geq \varepsilon^p) \leq \frac{\mathbb{E}\|X_n - X\|^p}{\varepsilon^p}.$$

□

Convergence presque sûre et convergence en probabilité

Théorème 98. *La convergence presque sûre implique la convergence en probabilité.*

Démonstration. Soit $\varepsilon > 0$. D'après le théorème 96, on a

$$\mathbb{P}(\overline{\lim_{n\to+\infty}} \{|X_n - X| \geq \varepsilon\}) = 0.$$

Or, d'après le théorème de continuité séquencielle décroissante, on a

$$\mathbb{P}(\overline{\lim_{n\to+\infty}} \{|X_n - X| \geq \varepsilon\}) = \lim_{n\to+\infty} \mathbb{P}(\bigcup_{k\geq n} \{|X_k - X| \geq \varepsilon\})$$

Comme

$$0 \leq \mathbb{P}(|X_n - X| \geq \varepsilon) \leq \mathbb{P}(\bigcup_{k\geq n} \{|X_k - X| \geq \varepsilon\}),$$

on en déduit que

$$\lim_{n\to+\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0.$$

Comme ε est quelconque, on peut dire que X_n converge en probabilité vers X. □

9.3.2 Loi faible des grands nombres

Théorème 99. *Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires de même loi, admettant un moment d'ordre 2 et deux à deux non corrélées.*
On pose

$$S_n = \sum_{k=1}^{n} X_k \text{ et } M_n = \frac{1}{n} S_n.$$

Alors

1. $M_n \xrightarrow[n\to+\infty]{\mathrm{L}^2} \mathbb{E}X_0$. *On dit que M_n converge en moyenne quadratique vers $\mathbb{E}X_0$.*
2. *Et donc* $M_n \xrightarrow[n\to+\infty]{\mathbb{P}} \mathbb{E}X_0$.

Démonstration. $\mathbb{E}M_n = \frac{1}{n}\mathbb{E}S_n = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}X_k = \frac{1}{n} n\mathbb{E}X_0 = \mathbb{E}X_0$. Par conséquent $\mathbb{E}|M_n - \mathbb{E}X_0|^2 = \operatorname{Var} M_n = \frac{1}{n^2} \operatorname{Var} S_n$. Comme les X_k sont 2 à 2 non corrélées, on a

$$\operatorname{Var} S_n = \sum_{k=1}^{n} \operatorname{Var} X_k = n \operatorname{Var} X_1.$$

On a donc

$$\mathbb{E}|M_n - \mathbb{E}X_0|^2 = \operatorname{Var} M_n = \frac{\operatorname{Var} X_1}{n}, \tag{9.9}$$

qui tend bien vers zéro.

□

9.4 Lemmes de Borel-Cantelli

9.4.1 Premier lemme de Borel-Cantelli

Théorème 100. *Soit $(A_n)_{n\geq 1}$ une suite d'événements observables. Si la série de terme général $\mathbb{P}(A_n)$ est convergente, alors $\mathbb{P}(\overline{\lim\limits_{n\to+\infty}} A_n) = 0$.*

Démonstration. On pose $B_n = \underset{k\geq n}{\cup} A_k$. la suite (B_n) est décroissante, et l'intersection des (B_n) est, par définition, $\overline{\lim\limits_{n\to+\infty}} A_n$. D'après le théorème de continuité séquentielle décroissante, on a donc

$$0 \leq \mathbb{P}(\overline{\lim_{n\to+\infty}} A_n) = \lim_{n\to+\infty} \mathbb{P}(B_n).$$

Or

$$\mathbb{P}(B_n) = \mathbb{P}(\underset{k\geq n}{\cup} A_k) \leq \sum_{k\geq n} \mathbb{P}(A_k) = r_n$$

Comme r_n est le reste d'ordre n d'une série convergente, r_n est de limite nulle, et donc, par comparaison $\mathbb{P}(\overline{\lim\limits_{n\to+\infty}} A_n) = 0$. □

9.4.2 Deuxième lemme de Borel-Cantelli

Le deuxième lemme de Borel-Cantelli est une espèce de réciproque du premier, dans le cas où les événements considérés sont indépendante. Ici, on choisit de présenter d'emblée une généralisation du deuxième lemme de Borel-Cantelli, dûe à Erdös et Renyi (1959). Le théorème classique s'en déduira aisément.

Théorème 101 (Erdös-Renyi)**.** *Soient $(A_n)_{n\geq 1}$ une suite d'évévements. On pose*

$$N_n = \sum_{k=1}^{n} \mathbb{1}_{A_k} \text{ et } N = \sum_{k=1}^{+\infty} \mathbb{1}_{A_k}$$

$$m_n = \sum_{k=1}^{n} \mathbb{P}(A_k) = \mathbb{E}N_n$$

$$\text{On a } \overline{\lim_{n\to+\infty}} A_n = \{N = +\infty\}.$$

Si

$$\lim_{n\to+\infty} m_n = +\infty \text{ et } \lim_{n\to+\infty} \frac{\operatorname{Var} N_n}{m_n^2} = 0,$$

alors

$$\mathbb{P}(\overline{\lim_{n\to\infty}} A_n) = 1.$$

Démonstration. Pour $m_n > a$, on a

$$\begin{aligned}\mathbb{P}(N \le a) &\le \mathbb{P}(N_n \le a)\\ &\le \mathbb{P}(|N_n - m_n| \ge m_n - a) \le \frac{\operatorname{Var} N_n}{(m_n - a)^2}\end{aligned}$$

En faisant tendre n vers $+\infty$, on en déduit que

$$\forall a \in \mathbb{N} \quad \mathbb{P}(N \le a) = 0.$$

Ainsi

$$\mathbb{P}(N < +\infty) = \mathbb{P}(\lim_{a\to+\infty}\uparrow \{N \le a\}) = \lim_{a\to+\infty} \mathbb{P}(N \le a) = \lim_{a\to+\infty} 0 = 0.$$

□

Théorème 102 (2[ème] lemme de Borel-Cantelli)**.** *Soit $(A_n)_{n\ge1}$ une suite d'événements indépendants. Si la série de terme général $\mathbb{P}(A_n)$ est divergente, alors $\mathbb{P}(\overline{\lim_{n\to+\infty}} A_n) = 1$.*

Démonstration. On va appliquer le théorème précédent : comme les $(A_k)_{k\ge1}$ sont indépendants, leurs indicatrices sont des variables aléatoires indépendantes, et donc

$$\operatorname{Var} N_n = \sum_{k=1}^{n} \operatorname{Var} 1\!\!1_{A_k} = \sum_{k=1}^{n} \mathbb{P}(A_k)(1 - \mathbb{P}(A_k) \le \sum_{k=1}^{n} \mathbb{P}(A_k) = m_n$$

Ainsi $\frac{\operatorname{Var} N_n}{m_n^2} = \frac{1}{m_n}$. Comme $\lim_{n\to+\infty} m_n = \sum_{k=1}^{+\infty} \mathbb{P}(A_k) = +\infty$, le résultat s'ensuit. □

Exercice : La conclusion du 2[ème] lemme de Borel-Cantelli reste-t-elle vraie si l'on suppose seulement que les $(A_k)_{k\ge1}$ sont deux à deux indépendants ? Que les $(A_k)_{k\ge1}$ sont négativement corrélés[1] ?

[1] c'est à dire que $\mathbb{P}(A_i \cap A_j) \le \mathbb{P}(A_i)\mathbb{P}(A_j)$ pour $i \ne j$

Théorème 103. *Soit $(X_n)_{n\geq 0}$ une suite convergeant en probabilité vers X. Alors, il existe une sous-suite X_{n_k} telle que $X_{n_k} \xrightarrow[k\to\infty]{p.s.} X$..*

Démonstration. On pose $n_0 = 0$, puis, pour $k \geq 1$:

$$n_k = \inf\{n > n_{k-1}; \mathbb{P}(|X_n - X| \geq \frac{1}{k}) \leq \frac{1}{2^k}\}$$

À k fixé, $\mathbb{P}(|X_n - X| \geq \frac{1}{k})$ tend vers 0 quand n tend vers l'infini, donc on a bien pour tout k : $n_k < +\infty$.

Maintenant, on a pour tout $k \geq 0$:

$$\mathbb{P}(|X_{n_k} - X| \geq \frac{1}{k}) \leq \frac{1}{2^k}$$

Comme la série de terme général converge $\frac{1}{2^k}$, le premier lemme de Borel-Cantelli nous permet d'affirmer que

$$\mathbb{P}(\overline{\lim_{k\to\infty}} \{|X_{n_k} - X| \geq \frac{1}{k})\}) = 0,$$

ce qui est équivalent

$$\mathbb{P}(\underline{\lim_{k\to\infty}} \{|X_{n_k} - X| < \frac{1}{k})\}) = 1,$$

ce qui veut dire que pour presque tout ω, il existe un $k_0(\omega)$ tel que

$$k \geq k_0(\omega) \Longrightarrow |X_{n_k}(\omega) - X(\omega)| < \frac{1}{k},$$

ce qui implique bien sûr que $X_{n_k}(\omega)$ tend vers $X(\omega)$ pour $\mathbb{P}$-presque tout ω. □

9.5 Loi forte des grands nombres

9.5.1 La loi forte des grands nombres

Théorème 104. *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires deux à deux indépendantes, de même loi μ. On suppose que μ admet un moment d'ordre 1. Alors*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to+\infty]{p.s.} \mathbb{E}X_1.$$

9.5.2 Probabilités et fréquences asymptotiques

Théorème 105. *Soit $(A_n)_{n\geq 0}$ une suite d'événements observables indépendants de même probabilité p. Pour ω dans l'univers Ω On note $N_n(\omega)$ le nombre d'événements qui sont réalisés parmi $A_1, \dots, A_n$. Ainsi, on a*

$$N_n = \sum_{k=1}^{n} 1\!\!1_{A_k} \text{ et } f_n = \frac{1}{n} N_n.$$

Alors il existe un événement observable $\tilde{\Omega} \subset \Omega$ avec $\mathbb{P}(\tilde{\Omega} \subset \Omega) = 1$ et

$$\forall \omega \in \tilde{\Omega} \subset \Omega \quad f_n(\omega) \to p.$$

Démonstration. Il suffit de poser $X_k = 1\!\!1_{A_k}$ et d'appliquer le théorème 104. X_k admet bien un moment d'ordre 1 car $0 \leq X_k \leq 1$ et l'on a $\mathbb{E}X_1 = \mathbb{P}(A_1) = p$. □

9.6 Exercices sur la convergence presque sûre

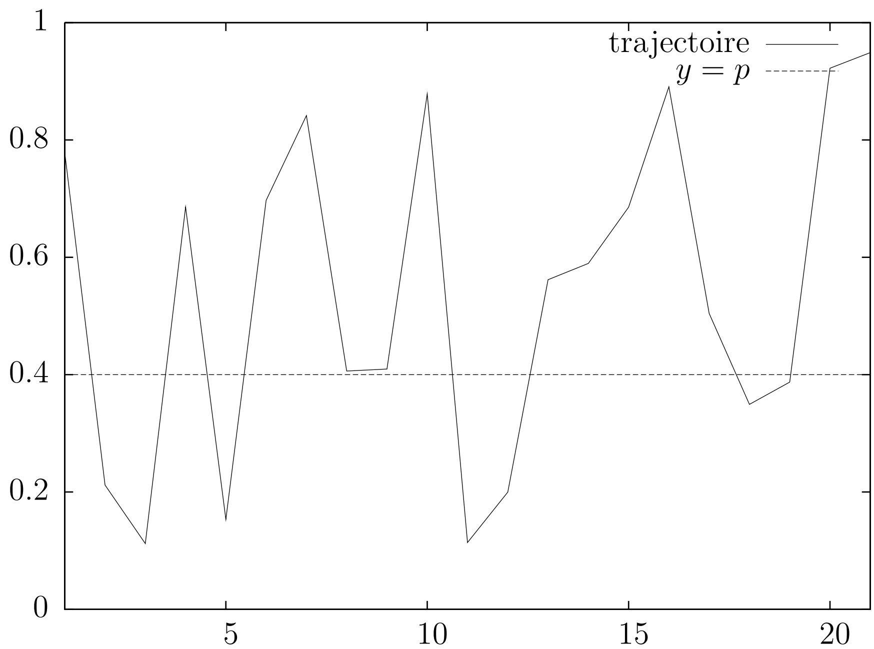
1. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi $\mathcal{N}(m, \sigma^2)$. Montrer que la suite $\frac{X_1^2 + \cdots + X_n^2}{n}$ convergence presque sûrement et déterminer sa limite.
2. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires identiquement distribuées telle qu'il existe $\alpha > 0$ avec $\mathbb{E}\exp(\alpha |X_1|) < +\infty$.
   Montrer que
   $$\lim_{n\to+\infty} \frac{X_n}{(\ln n)^{\frac{3}{2}}} = 0.$$
3. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant une loi exponentielle de paramètre 1.

   Calculer $\overline{\lim\limits_{n\to+\infty}} \frac{X_n}{\ln n}$.
4. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi $\mathcal{N}(0, 1)$.

   (a) Calculer $\overline{\lim\limits_{n\to+\infty}} \frac{X_n}{\sqrt{2\ln n}}$.

   (b) On se donne maintenant une deuxième suite $(Y_n)_{n\geq 1}$ de variables aléatoires indépendantes suivant la loi $\mathcal{N}(0, 1)$, cette deuxième suite étant indépendante de la première. Comparer
   $$\overline{\lim_{n\to+\infty}} \frac{X_n}{\sqrt{2\ln n}} + \overline{\lim_{n\to+\infty}} \frac{Y_n}{\sqrt{2\ln n}} \text{ et } \overline{\lim_{n\to+\infty}} \frac{(X_n + Y_n)}{\sqrt{2\ln n}}.$$

5. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes telles que pour tout n, X_n suive une loi binomiale $\mathcal{B}(n, \frac{1}{n^{1,01789}})$. Déterminer l'ensemble des valeurs d'adhérences de la suite (X_n)
6. Soit $p \in [0,1]$ et $(U_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On note T_n le nombre de fois où le graphe associé à (U_n) coupe la droite d'équation $y = p$ avant le temps n. Dans notre exemple, $p = 0.4$ et $T_{20} = 8$.



Montrer que $\frac{T_n}{n}$ converge presque sûrement et déterminer la limite.

7. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires de Bernoulli indépendantes de paramètre $1/2$. On pose

$$M_n = \begin{pmatrix} 2 + X_n & 1 \\ 1 & 2 + X_n \end{pmatrix}$$

et

$$A_n = M_n \times M_{n-1} \times \dots M_2 \times M_1.$$

(a) Montrer que la suite $(\det A_n)^{1/n}$ converge presque sûrement et déterminer la limite.

(b) Soit $(x,y) \in \mathbb{R}^2 \backslash \{(0,0)\}$. On pose

$$Y_n = A_n \begin{pmatrix} x \\ y \end{pmatrix}$$

Montrer que

$$\|Y_n\|^{1/n} \xrightarrow[n\to+\infty]{\text{p.s.}} \begin{cases} \sqrt{2} & \text{si } x+y=0 \\ \sqrt{12} & \text{si } x+y\neq 0 \end{cases}$$

8. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires telles que pour tout n, X_n suive une loi de Poisson de paramètre λ_n, où $(\lambda_n)_{n\geq 1}$ est une suite tendant vers 0 en l'infini. Montrer que la suite $(Y_n)_{n\geq 1}$ définie par $Y_n = X_1 X_2 \dots X_n$ est nulle à partir d'un certain rang.
9. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires telles que pour tout n, X_n suive une loi de Poisson de paramètre λ_n, avec

$$\sum_{n=1}^{+\infty} \lambda_n^2 < +\infty$$

Montrer que la suite $(X_n)_{n\geq 1}$ est presque sûrement bornée.

10. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes telles que pour tout n, X_n suive une loi de Poisson de paramètre λ_n, avec

$$\lambda_n = o(\ln n)$$

Montrer que la suite $(Y_n)_{n\geq 1}$ définie par $Y_n = X_1 X_2 \dots X_n$ est nulle à partir d'un certain rang.

11. Soit $(X_n)_{n\geq 2}$ une suite de variables aléatoires indépendantes telle que pour tout n, X_n suive une loi de Poisson de paramètre $2\ln n$. Montrer que

$$X_2 X_3 \dots X_n \xrightarrow[n\to+\infty]{\text{p.s.}} \begin{cases} 0 & \text{avec probabilité } p \\ +\infty & \text{avec probabilité } 1-p \end{cases},$$

où $p \in]0,1[$.

12. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes identiquement distribuées avec $\mathbb{E}|X_1| = +\infty$.

    (a) Soit $a > 0$. Montrer que $\overline{\lim\limits_{n\to+\infty}} \frac{|X_n|}{n} \geq a$ p.s.

    (b) On pose $S_n = \sum_{k=1}^n X_k$. Montrer que $\sup\limits_{n\geq 1} \frac{|S_n|}{n} = +\infty$ p.s.

13. *Un raffinement de Erdös-Renyi*

(a) Soit $(A_n)_{n\geq 1}$ une famille d'événements telle que $\sum_{n=1}^{+\infty} \mathbb{P}(A_n) = +\infty$. On pose $S_n = \sum_{k=1}^n \mathbb{1}_{A_k}$. Montrer que

$$\mathbb{P}(\overline{\lim_{n\to+\infty}} A_n) \geq 1-\theta, \text{ où } \theta = \underline{\lim_{n\to+\infty}} \frac{\operatorname{Var} S_n}{(\mathbb{E}S_n)^2}.$$

(b) Soient $\alpha > 0$ et $(n_k)_{k\geq 1}$ une suite d'entiers strictement positifs tels que $\lim_{k\to+\infty} \frac{n_k}{k} = \alpha$. Soit $(B_k)_{k\geq 1}$ une suite d'événements indépendants, avec $\mathbb{P}(B_k) = 1/n_k$. On pose $A_k = B_1$ si k est entre 1 et n_1, $A_k = B_2$ si k est entre $n_1 + 1$ et $n_1 + n_2$, $A_k = B_{i+1}$ si k est entre $n_1 + n_2 + \cdots + n_i$ et $n_1 + n_2 + \cdots + n_{i+1}$. Montrer que dans ce cas $\theta = \alpha/2$. En déduire que θ peut prendre des valeurs finies arbitrairement grandes.

14. *Lemme de Kochen-Stone*

(a) Soit X une variable aléatoire de carré intégrable et d'espérance strictement positive. Montrer que pour tout $\lambda \in]0,1[$,

$$\mathbb{P}(X > \lambda\mathbb{E}[X]) \geq (1-\lambda)^2 \frac{(\mathbb{E}X)^2}{\mathbb{E}[X^2]}.$$

(Ce n'est pas une conséquence de l'inégalité de Tchebitchev.)

(b) Soit B_n une famille d'événements. Montrer que

$$\mathbb{P}(\overline{\lim_{n\to+\infty}} B_n) \geq \overline{\lim_{n\to+\infty}} \mathbb{P}(B_n).$$

(c) Soit $(A_n)_{n\geq 1}$ une famille d'événements telle que $\sum_{n=1}^{+\infty} \mathbb{P}(A_n) = +\infty$. On pose $S_n = \sum_{k=1}^n \mathbb{1}_{A_k}$. Montrer que

$$\mathbb{P}(\overline{\lim_{n\to+\infty}} A_n) \geq \overline{\lim_{n\to+\infty}} \frac{(\mathbb{E}S_n)^2}{\mathbb{E}[S_n^2]}.$$

(d) Comparer avec le résultat de l'exercice précédent.

15. Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes identiquement distribuées, avec $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$. On pose $S_n = X_1 + \cdots + X_n$. On rappelle les résultats suivants qui ont été vu en cours ou en exercice :
– On a l'équivalent en l'infini $\mathbb{P}(S_{2n} = 0) \sim \frac{1}{\sqrt{\pi n}}$.

– pour tout $A \in \mathcal{Q} = \bigcap_{n \geq 1} \sigma((S_i)_{i \geq n})$, on a $\mathbb{P}(A) = 0$ ou $\mathbb{P}(A) = 1$.

À la lumière de ces résultats et du lemme de Kochen-Stone vu à l'exercice précédent, montrer que

$$\mathbb{P}(S_{2n^2} = 0 \text{ pour une infinité de valeurs de } n) = 1.$$

Annexe A

Rappels de dénombrement

A.1 Rappels de vocabulaire ensembliste

Un ensemble Ω est constitué de points, tous distincts. On dit qu'un ensemble A est inclus dans Ω, et l'on écrit $A \subset \Omega$ lorsque tous les éléments de A appartiennent à Ω.

On rappelle que l'ensemble vide (noté $\varnothing$) ne contient aucun élément et est inclus dans tous les ensembles

Pratiquement, si l'on veut montrer le résultat $A \subset \Omega$, la preuve ressemblera donc à "Soit $x \in A$...(raisonnement)...donc $x \in \Omega$. Comme on a choisi x quelconque dans A, on peut conclure que $A \subset \Omega$."

Si A est inclus dans Ω, on dit que A est un sous-ensemble, ou encore une partie de Ω.

Si A et B sont des parties de Ω, l'ensemble $A \cup B$ est constitué des éléments de Ω qui sont dans A ou dans B, éventuellement dans les deux. Plus généralement, si I est un ensemble quelconque et $(A_i)_{i \in I}$ une famille de parties de Ω indexé par I, $\underset{i \in I}{\cup} A_i$ est constitué des points de Ω qui sont dans au moins un des A_i.

Pratiquement, si l'on veut montrer le résultat $x \in \underset{i \in I}{\cup} A_i$, , la preuve ressemblera donc à "...(raisonnement)...Il existe donc $i_0 \in I$ tel que $x \in A_{i_0}$. Donc $x \in \underset{i \in I}{\cup} A_i$."

Si A et B sont des parties de Ω, l'ensemble $A \cap B$ est constitué des éléments de Ω qui sont dans A et dans B. Plus généralement, si I est un ensemble quelconque et $(A_i)_{i \in I}$ une famille de parties de Ω indexé par I, $\underset{i \in I}{\cap} A_i$ est constitué des points de Ω qui sont dans tous les A_i.

Pratiquement, si l'on veut montrer le résultat $x \in \bigcap_{i\in I} A_i$, , la preuve ressemblera donc à "Soit $i \in I$...(raisonnement)...Donc $x \in A_i$. Comme i est quelconque, on a donc $x \in \bigcap_{i\in I} A_i$."

A.2 Applications et cardinaux : définitions et notations

Pour A, D deux ensembles non vides quelconques, on note A^D ou $\mathcal{F}(D, A)$ l'ensemble des fonctions de D (ensemble de départ) vers A (ensemble d'arrivée). Soit f une application de A dans D. On dit que f est

- *injective* si $\forall x, y \in D \quad x \neq y \Longrightarrow f(x) \neq f(y)$
- *surjective* si $\forall z \in A \quad \exists x \in D \quad f(x) = z$.
- *bijective* si elle elle à fois injective et surjective.

Une application injective (resp. surjective, bijective) est une injection (resp. surjection, bijection).

Une bijection d'un ensemble Ω dans lui-même est appelée *permutation* de Ω.

Un ensemble est Ω dit *fini* si

- ou bien c'est l'ensemble vide. ($\varnothing$)
- ou bien il existe un entier n tel qu'il existe une bijection entre Ω et $\{1, \ldots, n\}$.

Cet entier n est unique : on l'appelle le *cardinal* de l'ensemble Ω. On le note $|\Omega|$. De manière intuitive, c'est le nombre d'éléments de Ω.

Le cardinal de l'ensemble vide est zéro.

Pour Ω fini de cardinal n, et $p \in \{0, \ldots, n\}$, on note $\mathcal{B}_p(\Omega)$ l'ensemble des parties de Ω de cardinal p. Par exemple $\mathcal{B}_2(\{a, b, c\}) = \{\{a, b\}, \{b, c\}, \{a, c\}\}$. De même, on note $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω, quelque cardinal qu'elles aient. Par exemple $\mathcal{P}(\{a, b, c\}) = \{\varnothing, \{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{a, c\}, \{a, b, c\}\}$.

Soit A et D deux ensembles finis. On admettra les résultats suivants :

- Il existe (au moins) une bijection de D dans A si et seulement si $|A| = |D|$.
- Il existe (au moins) une injection de D dans A si et seulement si $|A| \geq |D|$.
- Il existe (au moins) une surjection de D dans A si et seulement si $|A| \leq |D|$.

Le premier des trois résultats énoncés est évidemment le plus utilisé lorsque l'on veut des dénombrements exacts, alors que les deux autres sont plutôt utilisés dans les cas trop complexes où l'on peut juste espérer des encadrements.

A.3 Principes de base du dénombrement

A.3.1 Principe de bijection

Dans la pratique, lorsque l'on veut compter les éléments d'un ensemble, on montre que cet ensemble est en bijection avec un ensemble dont on connaît (par coeur !) le nombre d'éléments. La section suivante énoncera un certain nombre de résultats qu'il faut connaître.

A.3.2 Principe d'indépendance

C'est juste la formule

$$|A \times B| = |A|.|B| \quad .$$

Considérée isolément, elle peut paraître sans intérêt mais elle est souvent utilisée en association avec le principe de bijection.

A.3.3 Principe de partition

On dit que les ensembles $(A_i)_{i\in I}$ forment une partition de A si l'on a $A = \underset{i\in I}{\cup} A_i$ et $i \neq j \Longrightarrow A_i \cap A_j = \varnothing$. On a alors

$$|A| = \sum_{i\in I} |A_i|.$$

Le résultat élémentaire suivant peut souvent être utile.

Théorème 106. *Soit Ω un ensemble quelconque, I un ensemble d'index fini ou dénombrable et $(\Omega_i)_{i\in I}$ une partition de Ω. Alors, si l'on pose $A_i = A\cap\Omega_i$, les ensembles $(A_i)_{i\in I}$ forment une partition de A.*

Démonstration. Comme $\Omega = \underset{i\in I}{\cup} \Omega_i$, on a $A = A\cap\Omega = \underset{i\in I}{\cup} A\cap\Omega_i = \underset{i\in I}{\cup} A_i$. D'autre part, pour $i \neq j$, on a $A_i \cap A_j \subset \Omega_i \cap \Omega_j = \varnothing$, d'où $A_i \cap A_j = \varnothing$. $\square$

A.3.4 Lemme des bergers

Le lemme suivant peut également être utile

Lemme 18 (Lemme des bergers)**.** *Soit φ une application* surjective *de A dans D. On suppose qu'il existe un entier $a \geq 1$ tel que*

$$\forall y \in A \quad |\{x \in D; \varphi(x) = y\}| = a$$

(autrement dit si tout élément de A a exactement a antécédents), on a

$$|A| = \frac{|D|}{a}.$$

Démonstration. On applique le principe de partition avec $I = A$: si l'on pose, pour $y \in A$: $D_y = \{x \in D; \varphi(x) = y\}$, les D_y forment clairement une partition de D, d'où

$$|D| = \sum_{y \in A} |D_y| = \sum_{y \in A} a = |A|a.$$

□

Le nom du lemme est du à la procédure prétendûment employée par les bergers chaldéens pour compter le nombre de leurs moutons : il s'agit de compter le nombre de pattes et de diviser par 4. Dans cet exemple A est l'ensemble des moutons, D l'ensemble des pattes de mouton, et φ l'application qui a une patte associe le mouton auquel elle appartient.

A.4 Quelques résultats incontournables

A.4.1 Nombre d'applications de D dans A

Il existe exactement $|A|^{|D|}$ applications de D dans A, ce qui peut s'écrire

$$|A^D| = |A|^{|D|}.$$

On pose $|A| = n$ et $|D| = p$. Un cas particulier important est celui où l'on a $D = \{1, \dots, p\}$. Or, un p-uplet $(x_1, x_2, \dots, x_p)$ dont les composantes sont des éléments de A peut être considéré comme la donnée d'une application de $\{1, \dots, p\}$ dans A. Le nombre de p-uplets $(x_1, x_2, \dots, x_p)$ dont les composantes sont des éléments de A est donc n^p.

Exemple : Un professeur note chaque étudiant d'une classe de 30 étudiants par une note entière de 0 à 20. Le nombre de résultats possibles est le nombre de fonctions de l'ensemble D des étudiants dans l'ensemble $A = \{0, \dots, 20\}$ des notes possibles. Comme $|A| = 21$ et $|D| = 30$, il y a donc 21^{30} résultats possibles.

Remarque : : au lycée, vous avez vu ce résultat sous la dénomination "choix indépendant (avec remise) de p objets dans un ensemble de cardinal $|A| = n$."

A.4.2 Nombre de permutations de Ω

On pose $|\Omega| = n$. Le nombre de permutations de Ω est

$$n(n-1)\dots 1.$$

On note $n! = n(n-1)\dots 1$.

Remarque : $n!$ se lit "*factorielle n*" .

Exemple : Un professeur doit faire passer dans la journée 5 étudiants à l'oral de contrôle. Il a 5 !=120 manières de choisir l'ordre dans lequel il va les interroger.

A.4.3 Nombre d'injections de D dans A

On pose $|A| = n$ et $|D| = p$. En vertu de la remarque faite en A.2, il existe une injection de D si et seulement si $p \leq n$. Alors, le nombre d'injections de A est

$$n(n-1)\dots(n-p+1).$$

Démonstration. Soit n un entier. On pose $A = \{1, \dots, n\}$ et on note $\mathcal{I}_p$ l'ensemble des injections de $\{1, \dots, p\}$ dans A. On va montrer par récurrence sur $p \in \{1, \dots, n\}$ que $|\mathcal{I}_p| = \frac{n!}{(n-p)!}$. Il est évident que $|\mathcal{I}_1| = 1 = \frac{n!}{(n-1)!}$. Considérons l'application

$$R_p : \mathcal{I}_{p+1} \to \mathcal{I}_p$$

qui à chaque injection de $\{1, \dots, p+1\}$ dans A associe sa restriction à $\{1, \dots, p\}$. Un peu de réflexion montre que

$$\forall f \in \mathcal{I}_p \quad |\{g \in \mathcal{I}_{p+1}; R_p(g) = f\}| = n - p$$

D'après le lemme des bergers, on a donc

$$|\mathcal{I}_{p+1}| = (n-p)|\mathcal{I}_p|.$$

Cette identité permet d'achever la preuve par récurrence. □

Remarques :

– Comme on l'a vu dans la preuve, ce nombre peut s'écrire aussi $\frac{n!}{(n-p)!}$.
– Lorsque $n = p$, on trouve $n!$. En fait, une injection entre deux ensembles de même cardinaux est une bijection.

Exemple : 3500 personnes se présentent au concours de l'Agrégation de Mathématiques. 300 places sont mises aux concours. Combien y a-t-il de palmarès possibles, en supposant qu'il n'y ait pas d'ex-æquos ?
Réponse : $3500 \times 3499 \times \cdots \times 3202 \times 3201$. Ici D est l'ensemble des rangs, on a donc $D = \{1, \dots, 300\}$ et A l'ensemble des candidats (donc $|A| = 300$). On compte bien le nombre d'applications injectives puisqu'une même personne ne peut avoir deux rangs différents.

A.4.4 Nombre de parties de Ω possédant p éléments

On pose $|\Omega| = n$. Par définition, on note $\binom{n}{p}$ le nombre de parties à p éléments d'un ensemble de n éléments. Il s'agit donc de calculer $|\mathcal{B}_p(\Omega)|$. On va montrer que

$$\binom{n}{p} = \frac{n(n-1)\dots(n-p+1)}{p(p-1)\dots 1} = \frac{n!}{p!(n-p)!}.$$

Démonstration. Il suffit d'appliquer le lemme des bergers à
– D : ensemble des injections de $\{1, \dots, p\}$ dans Ω
– $A = \mathcal{B}_p(\Omega)$
– φ définie par $\varphi(f) = \{f(k); k \in \{1, \dots, p\}\}$

On a vu précédemment que $|A| = n(n-1)\dots(n-p+1)$. Il n'est pas difficile de voir que φ est surjective. Une partie $\{e_1, \dots, e_p\}$ de Ω étant donnée, combien existe-t-il d'injections (en fait de bijections) de $\{1, \dots, p\}$ dans Ω telles que $\{f(1), \dots, f(p)\} = \{e_1, \dots, e_p\}$? C'est évidemment le nombre d'injections de $\{1, \dots, n\}$ dans $\{e_1, \dots, e_p\}$, c'est à dire $p!$. Le lemme des bergers s'applique donc avec $a = p!$, d'où le résultat. □

Exemple : 3500 personnes se présentent au concours de l'Agrégation de Mathématiques. 300 places sont mises aux concours. Combien y a-t-il de listes alphabétiques des reçus possibles ? Réponse : $\binom{3500}{300}$. Ici Ω est l'ensemble des candidats et $p = 300$ le nombre de reçus.

A.4.5 Nombre total de parties de Ω

Le nombre total de parties de Ω est $|\mathcal{P}(\Omega)| = 2^{|\Omega|}$.

Démonstration. Il suffit de remarquer que l'application

$$\begin{array}{rcl} \mathcal{P}(\Omega) & \to & \{0; 1\}^{\Omega} \\ A & \mapsto & \mathbb{1}_A \end{array}$$

est une bijection. On rappelle que pour $A \subset \Omega$, l'application $\mathbb{1}_A$ (appelée indicatrice de A) est définie sur Ω par

$$\mathbb{1}_A(x) = \begin{cases} 1 \text{ si } x \in A \\ 0 \text{ si } x \notin A \end{cases}$$

□

Exemple : 200 étudiants se présentent à un examen. Combien y a-t-il de listes alphabétique des reçus possibles ? Réponse : 2^{200}. Ici Ω est l'ensemble des candidats. La grande différence avec l' exemple précédent est qu'ici, le nombre de reçus n'est pas fixé à l'avance.

A.5 Équations et inéquations en entiers

Lemme 19. *Soit n, p entiers. Si $n \geq p$, il existe exactement $\binom{n}{p}$ applications strictement croissantes de $\{1, \ldots, p\}$ dans $\{1, \ldots, n\}$. Sinon, il n'en existe aucune.*

Démonstration. Une application strictement croissante étant injective, il est nécessaire que $n \geq p$. Mais se donner une suite strictement croissantes de p éléments pris dans $\{1, \ldots, n\}$ revient à choisir une partie de $\{1, \ldots, n\}$ possédant p éléments, puis à les ordonner dans l'ordre naturel. Or $\binom{n}{p}$ est, par définition, le nombre de parties de $\{1, \ldots, n\}$ possédant p éléments, d'où le résultat. □

Exemple : Un enseignant devrait faire un cours de 70 pages en 7 séances. Combien y a-t-il de progressions possibles, en admettant qu'à chaque séance, l'enseignant progresse d' un nombre entier strictement positif de pages, mais sans être astreint à terminer le programme ?
Réponse : une progression correspond donc à une fonction strictement croissante de $\{1, \ldots, 7\}$ dans $\{1, \ldots, 70\}$ qui au numéro de chaque cours associe le numéro de la dernière page étudiée à ce cours : il y a donc $\binom{70}{7}$ progressions possibles.

Théorème 107. *Pour n, p entiers vérifiant $n \geq p$, il existe exactement $\binom{n}{p}$ p-uplets $(x_1, \ldots, x_p) \in (\mathbb{N}\backslash\{0\})^p$ solutions à l'inéquation :*

$$x_1 + x_2 + \cdots + x_p \leq n \tag{A.1}$$

Démonstration. Il suffit de remarquer que l'application

$$(x_1, \ldots, x_p) \quad \mapsto \quad (x_1, x_1 + x_2, x_1 + x_2 + x_3, \ldots, x_1 + x_2 + \cdots + x_p)$$

réalise une bijection entre les solutions recherchées de l'inéquation et l'ensemble des suites croissantes de p éléments à valeurs dans $\{1, \ldots, n\}$. □

Théorème 108. *Pour n, p entiers vérifiant $n \geq p$, il existe exactement $\binom{n-1}{p-1}$ p-uplets $(x_1, \ldots, x_p) \in (\mathbb{N}\backslash\{0\})^p$ solutions à l'équation :*

$$x_1 + x_2 + \cdots + x_p = n \tag{A.2}$$

Démonstration. Il suffit de remarquer que les solutions $(x_1, \ldots, x_p) \in (\mathbb{N}\backslash\{0\})^p$ de l'équation (A.2) sont exactement les solutions de l'inéquation (A.2) qui ne sont pas solutions de l'inéquation

$$x_1 + x_2 + \cdots + x_p \leq n - 1 \tag{A.3}$$

Il y en a donc $\binom{n}{p} - \binom{n-1}{p} = \binom{n-1}{p-1}$.

□

Théorème 109. *Pour n, p entiers positifs, il existe exactement $\binom{n+p-1}{p-1}$ p-uplets $(x_1, \ldots, x_p) \in \mathbb{N}^p$ solutions à l'équation :*

$$x_1 + x_2 + \cdots + x_p = n \tag{A.4}$$

Démonstration. Il suffit de remarquer que l'application

$$(x_1, \ldots, x_p) \quad \mapsto (x_1 + 1, \ldots, x_p + 1)$$

réalise une bijection entre les solutions $(x_1, \ldots, x_p) \in \mathbb{N}^p$ de l'équation (A.4) et les solutions $(x_1, \ldots, x_p) \in (\mathbb{N}\backslash\{0\})^p$ de l'équation

$$x_1 + x_2 + \cdots + x_p = n + p \tag{A.5}$$

□

Exemple : 4 listes se présentent aux élections syndicales étudiantes où 9 sièges sont à pourvoir. Combien y a-t-il de répartitions des sièges possibles. Réponse : il s'agit de compter les solutions en entiers positifs ou nuls à l'équation

$$x_1 + x_2 + x_3 + x_4 = 9,$$

x_k représente le nombre d'élus de la liste k. Il y a donc $\binom{9+4-1}{4-1} = \binom{12}{3} = \frac{12\times11\times10}{1\times2\times3} = 220$ répartitions possibles.

Théorème 110. *Pour n, p entiers positifs, il existe exactement $\binom{n+p}{p}$ p-uplets $(x_1, \ldots, x_p) \in \mathbb{N}^p$ solutions à l'inéquation :*

$$x_1 + x_2 + \cdots + x_p \leq n \tag{A.6}$$

Démonstration. La preuve, analogue à celle du théorème précédent, est laissée en exercice. □

A.6 Formule de Poincaré (aussi appelée formule du crible)

Cette formule est très utile en combinatoire, son application la plus classique étant le calcul du nombre de permutations sans point fixe (nombre de dérangements).

Pour tous ensembles $A_1, A_2, ..., A_n$, on a

$$\begin{aligned}
|\bigcup_{i=1}^{n} A_i| &= \sum_{B\in\mathcal{P}(\{1,\dots,n\})\setminus\varnothing} (-1)^{1+\mathrm{Card}(B)} |\cap_{j\in B} A_j| && \text{(A.7)}\\
&= \sum_{i=1}^{n} |A_i| - \sum_{1\leq i_1<i_2\leq n} |A_{i_1}\cap A_{i_2}| + \cdots && \text{(A.8)}\\
&\quad \cdots + (-1)^{k+1} \sum_{1\leq i_1<i_2<\ldots<i_k\leq n} |A_{i_1}\cap\cdots\cap A_{i_k}| + \cdots && \text{(A.9)}\\
&\quad \cdots + (-1)^{n+1}|A_1\cap\cdots\cap A_n|. && \text{(A.10)}
\end{aligned}$$

Exemple : Pour $n = 3$, on a

$$|A_1\cup A_2\cup A_3| = |A_1|+|A_2|+|A_3|-|A_1\cap A_2|-|A_2\cap A_3|-|A_1\cap A_3|+|A_1\cap A_2\cap A_3|.$$

Démonstration. On pourrait prouver la formule par récurrence sur n, mais c'est plutôt lourd. On préférera une preuve probabiliste (voir plus loin). □

A.7 Exercices

1. Combien existe-t-il de mots de n lettres construits avec l'alphabet $\{a; b\}$ et ne comportant pas deux "a" consécutifs ?
   Indication : montrer que si u_n est le nombre de tels mots se terminant par "a" et v_n est le nombre de tels mots se terminant par "b", on a la récurrence

$$\begin{pmatrix} u_{n+1} \\ v_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} u_n \\ v_n \end{pmatrix}$$

2. On donne un ensemble Ω de 15 entiers compris entre 1000 et 2000. Montrer que l'on peut en extraire deux sous-ensembles disjoints non-vides A et B tels que la somme des éléments de A soit égale à la somme des éléments de B.

Indication : montrer que l'application

$$\begin{array}{rcl} \mathcal{P}(\Omega) & \to \mathbb{N} & \\ A & \mapsto & \sum_{x \in A} x \end{array}$$

n'est pas injective.

Raffiner le raisonnement pour montrer que si l'on remplace 15 par 14, le résultat est encore vrai.

Indication : remplacer $\mathcal{P}(\Omega)$ par $\bigcup_{i=5}^{8} \mathcal{B}_i(\Omega)$.

3. Au tapis vert, il faut choisir 4 cartes (as,roi, dame, valet, dix, neuf, huit ou sept) par couleur (pique, coeur, carreau ou trèfle). On appelle grille l'un des résultats possibles.
   exemple de grille : as de pique, dame de cœur, 7 de carreau, 7 de trèfle.
   Dénombrer les grilles contenant :
   (a) la dame de cœur
   (b) une dame et une seulement
   (c) deux dames et deux seulement
   (d) aucune dame
   (e) au moins une dame
   (f) l'as de pique et une dame seulement
4. 5 prévenus sont amenés à choisir un avocat dans une liste de 10 avocats commis d'office.
   (a) Combien y a-t-il de choix possibles ?
   (b) Combien y a-t-il de choix tels que les 5 prévenus choisissent le même avocats ?
   (c) Combien y a-t-il de choix tels que 2 avocats soient appelés ?
5. De combien de manières peut-on disposer n personnes autour d'une table ronde ? Précision : seule la position relative à la maîtresse de maison importe, une rotation ne change donc rien.
   Si $n = 2k$, avec k hommes et k femmes, combien de dispositions permettent une alternance des sexes ?
6. Dans une assemblée de n personnes, on note la date d'anniversaire de chacun.
   (a) Donner le cardinal de toutes les répartitions possibles.
   (b) Calculer le cardinal des ensembles suivants :

   i. tous les participants ont la même date d'anniversaire.
   ii. Pierre et Paul ont la même date d'anniversaire.
   iii. deux personnes ont la même date d'anniversaire.
   iv. toutes les personnes ont des dates d'anniversaires différentes.

7. On considère l'ensemble Ω des suites de n chiffres (les chiffres sont pris dans $0, 1, \dots, 9$). Combien vaut $|\Omega|$? Combien y-a-il de chiffres comportant un nombre pair de zéros ?
8. Combien y a-t-il de nombres à trois chiffres (les chiffres sont pris dans $0, 1, \dots, 9$) dont la somme des chiffres est égale à 22 ?
   Indication : noter

$$\Omega_{22} = \{(x_1, x_2, x_3) \in \mathbb{N}^3 ; x_1 + x_2 + x_3 = 22\},$$

$$\begin{aligned} B = \{(x_1, x_2, x_3) \in \{0, \dots, 9\}^3 ; x_1 + x_2 + x_3 = 22\}, \\ M_1 = \{x \in \Omega_{22} ; x_1 \geq 10\}, \\ M_2 = \{x \in \Omega_{22} ; x_2 \geq 10\}, \\ M_3 = \{x \in \Omega_{22} ; x_3 \geq 10\}. \end{aligned}$$

   Remarquer que $\Omega \backslash B = M_1 \cup M_2 \cup M_3$. Ensuite, si l'on définit Ω_{12} et Ω_2 de manière analogue à Ω_{22}, on pourra exhiber une bijection entre Ω_{12} et M_1 et entre Ω_2 et $M_1 \cap M_2$.

Annexe B

Indications

B.1 Exercices sur les limites supérieures et inférieures

1. On rappelle que tout sous-groupe additif de $\mathbb{R}$ est dense dans $\mathbb{R}$ ou de la forme $a\mathbb{Z}$, $a \geq 0$. Il en résulte notamment que $\{n + 2m\pi \ ; \ n, m \in \mathbb{Z}\}$ est dense dans $\mathbb{R}$.
2. On peut faire la remarque simple suivante, très utile dans les problèmes d'inf et de sup : pour a, b dans $\overline{\mathbb{R}}$
$$(a = b) \iff (\forall x \in \mathbb{R} \quad (x > a) \iff (y > a)).$$
3. On peut utiliser le résultat de la première question de l'exercice précédent.
4. (a) À partir d'un certain rang, a_n ...et b_n ...donc $a_n + b_n$ ...
   (b) On pourra comparer l'ensemble des valeurs d'adhérence de (b_n) et l'ensemble des valeurs d'adhérences de $(a_n + b_n)$.
5. *Suites sous-additives (lemme de Fekete)*

   (a) Commencer par remarquer que $u_{kn+r} \leq u_{kn} + u_r$, puis procéder par récurrence sur n.

   (b) Commencer par montrer que pour tout k, $\overline{\lim\limits_{n\to+\infty}} \frac{u_n}{n} \leq \frac{u_k}{k}$.

   (c) $\inf\limits_{n\to+\infty} \frac{u_n}{n} \leq \underline{\lim\limits_{n\to+\infty}} \frac{u_n}{n} \leq \overline{\lim\limits_{n\to+\infty}} \frac{u_n}{n}$.

   (d) On pourra considérer $u_n = \log |||A^n|||$.

   (e) On pourra considérer $u_n = \log |A_n|$ et construire une injection de A_{n+p} dans $\subset A_n \times A_p$.

6. Pour n assez grand, $(1-\varepsilon)a \leq (a_n)^n \leq (1+\varepsilon)a$.
7. Commencer par montrer que la série de terme général u_k diverge. A partir de là, on peut remarquer que pour tout i_0

$$1 = \lim_{N\to+\infty} \sum_{i=i_0}^{N} (\sum_{k=i}^{N} u_k)^{-1/n},$$

ce qui donnera les inégalités sur les limites inférieure et supérieure de la suite.

B.2 Exercices de théorie de la mesure

1. (a) Remarquer que $\{f < g\} = \cup_x \{f < x < g\}$ et utiliser la séparabilité de $\mathbb{R}$.
   (b) Que signifie "$A = B$" presque partout ?
2. Pour les deux premières questions, relire le cours, le résultat est $\frac{11}{6}$. Pour la deuxième question, penser aux sommes de Riemann.
3. Pour la deuxième question, on pourra d'abord observer que certaines des conditions des axiomes sont vérifiées sans hypothèse supplémentaire sur f.
4. On procèdera par double inclusion. On rappelle que si $\mathcal{C} \subset \mathcal{D}$, $\sigma(\mathcal{C}) \subset \sigma(\mathcal{D})$.
5. $\mathbb{Q}$ est dense dans $\mathbb{R}$, ainsi chaque réel est limite d'une suite croissante et d'une suite décroissante.
6. (a) Exprimer C en fonction de A_2 et A_5.
   (b) Exprimer B en fonction des $(A_p)_{p\in\mathcal{P}}$.
7. Pour F fermé de E, on pourra considérer la fonction $d_F(x) = \inf\{d(x,y); y \in F\}$.
8. Traiter le cas où g est une indicatrice, puis une fonction étagée, puis une fonction positive,...
9. *Support d'une mesure sur* $\mathbb{R}^d$
   (a) On rappelle que si $x \in \overline{A}$, tout ouvert contenant x contient un élément de A.
   (b) On pourra montrer que pour tout $y \in O$, in existe $x \in \mathbb{Q}^x d$ et $r \in \mathbb{Q}_*^+$ avec $\{x\} \subset B(y,r) \subset O$.
   (c) Considérer la réunion de tous les ouverts de mesure nulle.

B.3 Premiers exercices de probabilité

1. On pourra commencer par remplacer un seul élément de la famille par son complémentaire.
2. Remarquer qu'une union dénombrable d'événements est de probabilité nulle si et seulement si chacun est de probabilité nulle.
3. (a) Remarquer qu'aucun entier n'a une infinité de diviseurs premiers.
   (b) On pourra remarquer que
   $$(1-\frac{1}{p_i})^{-1} \geq \sum_{k=0}^{N} p_i^{-k},$$
   où N est choisi tel que $p_i^{N+1} > p_n$.
   (c) Utiliser la continuité séquentielle décroissante.
   (d) Utiliser le résultat de l'exercice précédent.
   (e) Reprendre la preuve de (c).
4. Utiliser la formule de Bayes.
5. (a) On numérote les mathématiciens, et à chaque mathématicien, on asocie le numéro du propriétaire du chapeau qu'il prend
   (b) On pourra utiliser le principe de partition.
   (c) Quel est l'endomorphisme réciproque ?
   (d) Relire les résultats précédents
   (e) On pourra remarquer que la série est alternée.
6. *Calcul probabiliste de la formule de l'indicatrice d'Euler*

   (a) $A_d = d\mathbb{N} \cap \Omega_n$.
   (b) On déterminera un entier d tel que $\cap_{i=1}^{r} A_{d_i} = A_d$.
   (c) On remarquera que deux nombres sont premiers entre eux si et seulement si ils n'ont pas de diviseur premier commun.
   (d) Deux méthodes sont possibles : utiliser la formule du crible (formule de Poincaré) ou utiliser le résultat de l'exercice 1. Cette dernière méthode est utilisée dans le sujet du capes 2003.
7. Commencer par choisir clairement l'espace Ω.
8. On pourra conditionner par la valeur prise par l'*ensemble* des trois nombres tirés au sort.
9. (a) $\Omega = \{(x_1, \dots, x_{a+b}) \in \{\overrightarrow{i}\,;\overrightarrow{j}\}^{a+b}\ x_1 + \dots + x_{a+b} = a\overrightarrow{i} + b\overrightarrow{j}\}$ est un choix possible.

(b) Si l'on note $I = \{$ le graphe coupe la diagonale.$\}$, on pourra montrer que

$$\mathbb{P}(I \cap \{\text{A gagne le premier échange}\}) = \mathbb{P}(I \cap \{\text{B gagne le premier échange}\}).$$

10. Faire en sorte que le résultat soit nul.
11. *Valeurs d'adhérence de la suite $\varphi(n)/n$*
    (a) On pose $n_0 = 0$, $s_0 = 0$, puis pour $k \geq 0$:
    $n_{k+1} = \inf\{n > n_k; s_k + u_n < \ell\}$ et $s_{k+1} = s_k + u_{n_{k+1}}$.
    (b) Utiliser la formule donnant $\varphi(n)/n$ en fonction de ses facteurs premiers et utiliser la divergence de la série des inverses des nombres premiers.

B.4 Premiers exercices d'intégration

1. On note $f_{a,b}$ la fonction affine par morceaux, valant 1 avant a, 0 après b. On peut remarquer que $\mathbb{1}_{]-\infty,a]} \leq f_{a,a+1/n} \leq \mathbb{1}_{]-\infty,a+1/n]}$.
2. Utiliser la transformation d'Abel.
3. On pourra montrer qu'il existe une constante A telle que $g \leq A|f|$.
4. (a) Pour tout t dans $\mathbb{R}$, on pourra montrer que l'événement $\{Y \leq t\}$ est mesurable.
   (b) Utiliser le fait que T préserve la mesure μ.
   (c) On pourra montrer que $Y(x) = Z(x)$ pour μ presque tout x.
5. (a) Vérifier les axiomes.
   (b) Commencer par le cas où f est étagée.
6. Utiliser le théorème de convergence monotone.
7. Utiliser le théorème de convergence dominée.
8. On pourra prendre $A = \{f > 0\}$.
9. (a) On pourra démontrer que pour $n \geq 1$, $0 \leq \frac{\log(x+n)}{n} \leq 1 + x$.
   (b) $(1+x)^n \geq \dots$
   (c) Pour $n \geq 3$, $\frac{1+nx}{(1+x)^n} \leq \frac{3}{(1+x)^2}$.
10. (a) Procéder par récurrence.
    (b) Faire un développement en série.

(c) On peut développer avec la formule du binôme d'une part, d'autre part remarquer que $\displaystyle\int_0^1 (1-x)^n \ln(x)\ dx = \int_0^1 x^n \ln(1-x)\ dx$. Une intégration par partie donne alors le résultat, pourvu que la primitive soit choisie judicieusement.

11. Remarquer que $\log x + \frac{1}{x} \geq 0$ et développer $e^x - 1$ en série.
12. Développer en série et utiliser le théorème de convergence dominée.
13. On pourra remarquer que $\frac{1}{e^x-1} = \frac{e^{-x}}{1-e^{-x}}$.
14. On pourra remarquer que $|\frac{(x\log x)^2}{1+x^2}| \leq 2(x\log x)^2$ sur $[0,1]$, développer en série. Le calcul des sommes requiert des intégrations par parties.
15. *Fonction* Γ
    Pour l'existence, séparer les problèmes en 0 et l'infini. Ensuite, intégrer par parties.
16. *Intégrales de Wallis*
    (a) $\cos^2 = 1 - \sin^2$ et intégrer par parties.
    (b) Calculer $W_0 W_1$.
17. *Calcul de l'intégrale de Gauss*
    (a) Appliquer le théorème de convergence dominée.
    (b) Faire (au moins)un changement de variable
18. *Calcul de* $\Gamma(1/2)$
    Faire (au moins)un changement de variable.
19. *Formule de Stirling*
    (a) On pourra écrire $\int_{2n}^{+\infty} x^n e^{-x}\ dx = \int_n^{+\infty} (x+n)^n e^{-(n+x)}\ dx$.
    (b) Remarquer que $\Gamma(n+1) \sim \int_0^{2n} x^n e^{-x}\ dx$, puis faire un changement de variables affine.
    (c) On traitera séparément suivant le cas où $x > 0$ ou $x < 0$. Dans les deux, on pourra penser à un développement en série.
    (d) Utiliser le théorème de convergence dominée.
20. *Fonction* Γ *(suite)*
    (a) Procéder par récurrence
    (b) On pourra remarquer que $\frac{1}{e^x-1} = \frac{e^{-x}}{1-e^{-x}}$.
    (c) Faire un changement de variable
    (d) Développer le cosinus en série entière. Faire attention pour intervertir la somme et l'intégrale.

21. (a) La convergence se montre classiquement à l'aide d'une intégration par partie. Pour la divergence, on pourra par exemple noter que

$$\int_{n\pi}^{(n+1)\pi} \frac{|\sin t|}{t}|dt \geq \int_{(n+1/4)\pi}^{(n+3/4)\pi} \frac{|\sin t|}{t}|dt \geq \frac{\sqrt{2}}{4n+3}.$$

(b) Voir les théorèmes généraux de régularité du cours

(c) Il faut montrer que F est continue en 0. En réalité, ça ne coute pas plus cher de montrer que F est continue sur $\mathbb{R}_+$. On pouura écrire

$$\begin{aligned} F(x) &= \int_0^1 e^{-xt}\frac{\sin t}{t}\,dt + \int_1^{+\infty} e^{-xt}\frac{\sin t}{t}dt \\ &= \int_0^1 e^{-xt}\frac{\sin t}{t}\,dt + \operatorname{Im}\int_1^{+\infty} e^{(i-x)t}\frac{1}{t}\,dt. \end{aligned}$$

Penser à une intégration par parties pour le 2e morceau.

22. *Calcul d'intégrales liées aux intégrales de Fresnel*

(a) Couper l'intégrale en 2. La partie entre 0 et 1 ne pose pas de difficulté ; pour le reste on pourra faire une intégration par parties.

(b) On pourra commencer par montrer que pour $\delta > 0$,

$$\varphi(\lambda) = \varphi(\delta)\exp(\int_\delta^\lambda \frac{\alpha}{i-u}\,du).$$

(c) Poser $\alpha u = x$.

(d) Utiliser un théorème de convergence dominée.

(e) Faire un changement de variables

23. Fubini est ton ami.

24. Tonelli aussi.

25. (a) Faire une intégration par parties.

(b) oui !

(c) i. Faire une intégration par parties.

ii. Pour le premier point, noter que Ψ est continue, de limite 1 en l'infini. Enfin, appliquer le théorème de convergence dominée.

iii. On peut noter que $\Psi(x)/x^2$ est continue sur $\mathbb{R}_+^*$, avec des limites finies en $+\infty$ et en 0. Pour la dernière inégalité, on pourra écrire

$$\int_x^{+\infty} \frac{\Psi(\lambda u)}{u^2}\,du = \int_x^1 \frac{\Psi(\lambda u)}{u^2}\,du + \int_1^{+\infty} \frac{\Psi(\lambda u)}{u^2}\,du.$$

iv. Dans un premier temps, dominer $|R_\lambda(x) - R(x)|$ par une fonction intégrable. Dans un second temps, pour $\lambda \leq 1$, dominer $|R_\lambda(x) - R(x)|$ par une fonction intégrable ne dépendant pas de λ.

(d) Utiliser le théorème de Fubini.

(e) On pourra écrire

$$\int_0^{+\infty} \frac{R_\lambda(x)}{\sqrt{x}} = \int_0^1 \frac{R_\lambda(x)}{\sqrt{x}} + \int_1^{+\infty} \frac{R_\lambda(x)}{\sqrt{x}},$$

le premier morceau se traite aisément par convergence dominée, le deuxième nécessite une intégration par parties comme précédemment.

26. *Calcul de* $\sum_{n\geq 1} \frac{1}{n^2}$ *avec Fubini.*
Une intégration en x pour commencer donnera

$$J = 4\sum_{k=1}^{+\infty} \frac{1}{(2k+1)^2}.$$

Pour intégrer en y, noter que $1 + 2xy + y^2 = (y + x)^2 + (1 - y^2)$ et faire le changement de variable $u = \dfrac{y+x}{\sqrt{1-x^2}}$.

27. (a) Utiliser une partition de l'espace $\mathbb{R}^d$.

(b) On pourra montrer qu'il existe des constantes positives C et D telles que pour tout x avec $\|x\|_2 \geq 1$, on ait

$$\frac{C}{\|x\|_2^\alpha} \leq \frac{1}{\mathrm{Ent}(\|x\|_\infty)^\alpha} \leq \frac{D}{\|x\|_2^\alpha}.$$

(c) Le plus simple ici est de montrer que f est C^1 par des théorèmes généraux, puis de calculer la différentielle dans une direction donnée.

(d) On pourra utiliser le théorème de C^1-difféomorphisme. Pour le calcul de la différentielle, il pourra être intéressant d'écrire la matrice de Df_x dans une base orthogonale bien choisie.

28. On pourra intégrer des fonctions par rapport à la mesure de comptage et utiliser le théorème de convergence dominée.

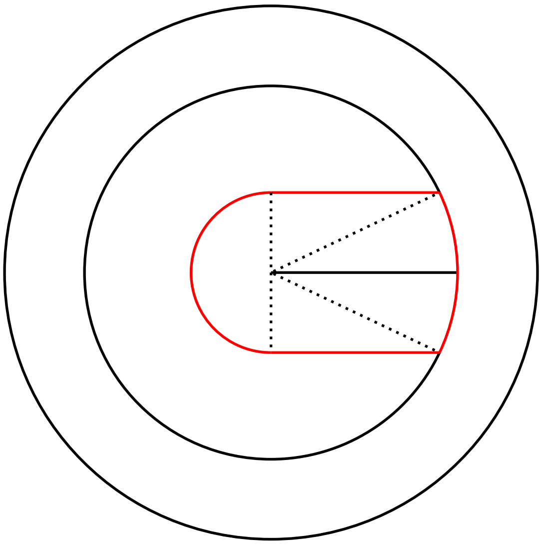
B.5 Exercices sur les lois des variables aléatoires

1. (a) Pensez à discuter suivant les positions relatives de n et k.
   (b) $\mathbb{Q}$ est dense dans $\mathbb{R}$.
   (c) Utiliser le principe de partition
   (d) Écrire $\mathbb{P}(D) = 1$, avec D bien choisi.
   (e) Trivialiser la question !
2. Prendre X et Y deux variables indépendantes suivant chacune une loi de Bernoulli de paramètre $1/2$ et poser $Z = |X - Y|$.
3. On prend $\Omega = \{0,1\}\times\{0,1\}$, $\mathcal{F} = \mathcal{P}(\omega)$ $\mathcal{C} = \{(0,0),(0,1)\},\{(0,0),(1,0)\}\}$. $\mathbb{P} = \frac{1}{2}(\delta_0 + \delta_1) \otimes \frac{1}{2}(\delta_0 + \delta_1)$, $Q = \frac{1}{2}(\delta_{(0,0)} + \delta_{(1,1)})$
4. Traduire les événements considérés en fonction de Y
5. Poser $x = AM$ et résoudre l'inéquation.
6. Dire que le maximum de n nombres ne dépasse pas x, c'est dire que chacun ne dépasse pas x.
7. Remarquer que $1 - m_n = \max(1 - X_1, \dots, 1 - X_n)$.
8. À k fixé, il faut déterminer les valeurs de X qui sont telles que $Y = k$.
9. Représenter graphiquement les domaines considérés.
10. (a) On pourra montrer que X est $\mathcal{T}$-mesurable, puis que les A_p sont $\sigma(X)$-mesurables.
    (b) On montrera que l'ensemble cherché est $\{\omega \in \mathbb{N}; X(\omega) = 2 \times 3 \times 5 \times 11\}$.
    (c) Le sens direct est le plus simple. Pour la réciproque, remarquer que si A_n est $\mathcal{T}$-mesurable, il doit contenir $\{\omega \in \mathbb{N}^*; X(\omega) = X(n)\}$.
    (d) On montrera que $A_m \cap A_n = A_{m \wedge n}$.
    (e) Remarquer que pour p assez grand $1 - \frac{1}{p^{2s}} \geq e^{-\frac{2}{p^{2s}}}$.
11. (a) On pourra montrer que $\mathcal{Q}$ est la tribu de queue associée à la famille $(\mathcal{A}_n)_{n\geq 1}$.
    (b) Un ensemble d'entiers est infini si et seulement si il contient au moins un entier plus grand que n'importe quel entier fixé à l'avance.

    Ainsi, on pourra montrer que pour tout n_0, $A = \bigcap_{n\geq n_0} \bigcup_{k\geq n} A_k$.
12. Pour la première formule, faire une intégration par parties ; pour la deuxième, procéder par récurrence.
13. On pourra écrire $e^{-\frac{x^2}{2}} = \frac{1}{x}(xe^{-\frac{x^2}{2}})$ afin de procéder à une intégration par parties.

14. Si on pose $\alpha = 2\arcsin\frac{r}{1-r}$, on doit trouver par exemple

$$p = \frac{1-\cos\alpha}{4} + \frac{\alpha+\sin\alpha}{2\pi}.$$



15. (a) On pourra remarquer que pour $i \neq j$, $X_i - X_j$ est une variable à densité.
    (b) Pour la première égalité, utiliser la question précédente ; ensuite utiliser le théorème de transfert.
    (c) Appliquer le théorème de Fubini.
    (d) On pourra par exemple calculer $\mathbb{P}_{(T,N)}$ sur des ensembles de type $]a,+\infty[\times]b,+\infty[$.
16. Commencer par calculer la fonction de répartition

B.6 Exercice sur les espérances

1. Si X désigne le nombre de fois où l'on a obtenu le nombre choisi, le gain est $X - \mathbb{1}_{X=0}$.
2. Une probabilité est l'indicatrice d'une espérance.
3. Interpréter le membre de gauche comme la valeur absolue d'une covariance.
4. (a) On montrera que $\mathbb{E}(X-a)^2 - \sigma^2 = (m-a)^2$.

   (b) Prendre $a = \frac{a_1+a_n}{2}$.

5. Effectuer une transformation d'Abel.
6. Appliquer le théorème de transfert, et penser à la forme canonique des polynômes du second degré.
7. Appliquer le théorème de transfert.
8. On pourra commencer par supposer que la loi est centrée (c'est à dire que $a+b=0$) et remarquer que $|X-Y| \le |X|+|Y|$). On s'y ramènera dans le cas général.
9. Fonction de répartition, ou transformation : vous avez le choix !
10. On peut raisonner en termes de loi.
11. S'inspirer de l'exercice précédent et utiliser une transformation.
12. Appliquer l'exercice précédent.
13. Bien observer que X et Y sont indépendantes.
14. On pourra observer que l'application $(x,y) \mapsto (\frac{x}{\sqrt{y/n}}, y)$ réalise un C^1-difféomorphisme de $\mathbb{R}^* \times \mathbb{R}_+^*$ dans lui même.
15. Remarquer que tout se passe comme si (X,Y) suivait la loi uniforme sur $T = \{(x,y) \in \mathbb{R}^2; 0 < y < x < 1\}$.
    Si x et y sont solutions réelles de $x^2 - Sx + P = 0$, alors $|x-y| = \sqrt{S^2-4P}$
16. (a) Identifier la loi de S_n et appliquer le théorème de transfert.
    (b) i. Remarquer que $[0,1]$ est compact.
        ii. Remarquer que
        $$B_n(x)-f(x) = \int_{|\frac{S_n}{n}-x|\le\varepsilon} f(\frac{S_n}{n})-f(x)\,dP + \int_{|\frac{S_n}{n}-x|>\varepsilon} f(\frac{S_n}{n})-f(x)\,dP$$
        En déduire que la suite des polynômes B_n converge vers f uniformément sur $[0,1]$.
17. Si on note C_x l'ensemble des cases controlées par la dame i, et $C = \bigcup_{i=1}^{d} C_i$, on peut minorer $|C|$ grâce aux inégalités de Bonferroni. Dans un second temps, on remarquera que si φ est une bijecttion de l'échiquier dans lui-même, les ensembles C et $\varphi(C)$ se coupent dès que $|C| > n^2/2$.
18. (a) On prendra $\Omega = \mathcal{B}_r(\{1,\dots,n\})$.
    (b) Utiliser les relations entre l'espérance et la queue de distribution, puis utiliser la relation de récurrence du triangle de Pascal.

19. (a) On pourra remarquer qu'une rotation est une application linéaire isométrique.
    (b) Remarquer que $|X - Y| = \sqrt{2}U$.
20. (a) On trouvera comme variance $b(0) + \sum_{i=1}^{n-1} 2b(i)(1 - i/n)$.
    (b) Remarquer qu'une variance est toujours positive. On peut alors par exemple appliquer le lemme de Fatou ou procéder de manière plus élémentaire.
21. *Probabilité de retour en zéro au temps n d'une marche aléatoire symétrique*
    (a) Appliquer le théorème de Tonelli et le théorème de Fubini.
    (b) Appliquer la question précédente à $X = 2n$ et utiliser l'indépendance des X_i.

B.7 Exercices sur les espaces $\mathcal{L}^p$ et L^p

1. (a) Cherchez un peu plus...
   (b) Découper $\mathbb{R}$ en intervalles de longeur 2π.
   (c) Utiliser des équivalents.
2. Étudier d'abord la convergence ponctuelle.
3. On pourra raisonner par l'absurde
4. On pourra utiliser des sous-suites.
5. Retrousser ses manches (ou équivalents).
6. Majorer $\sqrt{|f^2 + g^2|}$ par une fonction intégrable.
7. Passer à l'intégrale de Riemann.
8. (a) Prendre $X = [0, 1]$ et pour μ la mesure de Lebesgue sur $[0, 1]$, choisir ensuite f_n telle que $f_n(x) \to 0$ pour tout $x \in [0, 1[$ et que l'on ait $\int f_n = 1$.
   (b) Pour p entier s'écrivant $p = 2^n + k$, avec $0 \leq k < 2^n$, poser $u_p = \frac{k}{2^n}$. Ensuite, poser $\varphi_n(x) = \max(1 - \frac{n}{4}|x|, 0)$ et finalement $f_n(x) = \varphi_n(x - u_n)$.
   (c) Symétriser l'exemple trouvé à la première question.
9. Utiliser l'inégalité de Cauchy-Schwartz.
10. Utiliser l'inégalité de Hölder.
11. (a) Utiliser l'inégalité de Holder.
    (b) i. Considérer l'intégrale $\int_{]0,x[} f d\lambda$ comme intégrale de Riemann.

ii. Remarquer que $T(f)(x)$ est bornée et décroit suffisamment vite à l'infini.

iii. Remarquer que $f(x) = T(f)(x)+xT(f)'(x)$ et faire une intégration par parties.

iv. Cherchez un peu plus...

v. $f = f^+ - f^-$.

(c) i. Pour le premier point, on pourra utiliser l'inégalité de Hölder.

ii. Utiliser la densité des fonctions continues à support compact dans $\mathcal{L}^p$.

12. Considérer la suite (g_n) définie par $g_n = \frac{|f|^q}{f}\mathbb{1}_{\{|f|\leq n\}}$ sur $\{|f| > 0\}$ et $g_n = 0$ sur $\{f = 0\})$.

B.8 Exercices sur la convolution et la transformée de Fourier

1. Si $a = 0$ ou $b = 0$, alors $ab = 0$.
2. (a) Pleins de calculs en perspective. On conseille de commencer par exprimer $\int \exp(-p(x))\, dx$ en fonction de A, B, C, lorsque $p(x) = Ax^2 + Bx + C$ et $A > 0$.

   (b) Commencer par identifier les points où la convolée est nulle
3. (a) $1 - x^2 \geq 1 - x$ pour $0 \leq x \leq 1$.

   (b) On pourra montrer que pour tout $\delta \in]0, 1[$,

$$f * k_n(x) - f(x) \leq \frac{4(1-\delta^2)^n}{a_n}\|f\|_\infty + \omega_f(\delta).$$

   (c) On pourra montrer que

$$f * k_n(x) = \frac{1}{a_n}\int_{-1/2}^{1/2} \left(1 - (x-t)^2\right)^n f(t)\, dt.$$

   (d) On pourra commencer par le cas où $a = -1/4$ et $b = 1/4$.
4. (a) Pour éviter d'oublier des cas, se souvenir que le support de la convolée est inclus dans la “somme” des supports ; la parité peut également permettre de simplifier des choses...

   (b) Remarquer que $f^{(*)n}$ est positive.

   (c) Procéder par récurrence.
5. (a) Si l'on pose $g = \mathbb{1}_E * \mathbb{1}_{-E}$, on peut remarquer que $g(x) = \int_E T_x \mathbb{1}_E d\mu$.
   (b) Remarquer que si $g(x) \neq 0$, $x \in E - E$.
6. (a) Ah bah non ! Vous l'avez déjà eue, l'indication.
   (b) Réduire A dans une base orthonormale.
7. (a) On pourra remarquer que la transformée de Fourier est injective dans $L^1(\mathbb{R}^n)$)
   (b) Utiliser la transformation de Fourier et une fonction bien choisie.
8. On pourra utiliser la formule d'inversion.
9. On pourra utiliser la formule d'inversion.

B.9 Exercices sur la convergence presque sûre

1. Appliquer la loi des grands nombres
2. Pour tout $\varepsilon > 0$, appliquer le lemme de Borel-Cantelli aux événements $\{\frac{X_n}{(\ln n)^{\frac{3}{2}}} > \varepsilon\}$.
3. Discuter en fonction de A la nature de la série de terme général $\mathbb{P}(\frac{X_n}{\ln n} > A)$.
4. S'inspirer de l'exercice précédent en utilisant un équivalent pour la queue de la gaussienne.
5. Pour une suite à veleurs entières, les valeurs d'adhérences sont les valeurs qui sont prises une infinité de fois.
6. On a $T_n = \sum_{k=2}^n \mathbb{1}_{\{U_{n-1}<p\}\Delta\{U_n<p\}}$. On pourra découper T_n en deux sommes de variables aléatoires indépendantes.
7. (a) Passer au logarithme.
   (b) Écrire M_n sous la forme $M_n = PD_nP^{-1}$ et introduire la norme $\|x\|_* = \|P^{-1}x\|_\infty$.
8. On pourra remarquer que pour tout entier n, l'événement "$(Y_n)_{n\geq 1}$ est nulle à partir d'un certain rang" contient l'événement $\{X_n = 0\}$.
9. On pourra montrer que $X_n \in \{0, 1\}$ à partir d'un certain rang.
10. On pourra montrer que $\mathbb{P}(Y_n \neq 0) \leq \exp(-\sum_{k=1}^{n} \exp(-\lambda_k))$.
11. On pourra montrer que $(X_n)_{n\geq 2}$ ne prend la valeur 1 qu'un nombre fini de fois. Enfin, on montrera que

$$1 - p = \prod_{n=2}^{+\infty} (1 - \frac{1}{n^2}).$$

12. (a) Utiliser le lemme de Borel-Cantelli et le lien série-intégrale.

    (b) On pourra commencer par montrer que $\sup_{n\geq 1} \frac{|X_n|}{n} = +\infty$ p.s.

13. *Un raffinement de Erdös-Renyi*

    (a) S'inspirer de la preuve d'Erdös-Renyi.

    (b) On commencera par montrer que

$$\frac{\operatorname{Var} S_{t_N}}{(\mathbb{E}S_{t_N})^2} = \frac{1}{N^2}\sum_{k=1}^{N}(n_k - 1),$$

    où $t_N = n_1 + \cdots + n_N$.

14. *Lemme de Kochen-Stone*

    (a) Si on note $A = \{X > \lambda\mathbb{E}[X]\}$, on pourra noter que

$$\mathbb{E}[X] = \mathbb{E}[X\mathbb{1}_A] + \mathbb{E}[X\mathbb{1}_{A^c}] \leq (\mathbb{E}[X^2]\mathbb{P}(A))^{1/2} + \lambda\mathbb{E}[X].$$

    (b) Utiliser le lemme de Fatou.

    (c) On pourra remarquer que pour tout $\lambda \in]0,1[$,

$$\{\varlimsup_{n\to+\infty} A_n\} \supset \{\varlimsup_{n\to+\infty} \{\mathbb{1}_{A_n} > \lambda\mathbb{P}(A_n)\}\}.$$

    (d) Avec les notations de l'exercice précédent, on trouvera finalement $\mathbb{P}(\varlimsup_{n\to+\infty} A_n) \geq \frac{1}{1+\theta} \geq 1-\theta$. Ainsi Kochen-Stone est toujours plus fin que Erdös-Renyi. C'est surtout plus intéressant lorsque $\theta \in [1,+\infty[$, ce qui est possible d'après la dernière question de l'exercice précédent.

15. Il faut vérifier que l'événement est dans $\mathcal{Q}$, puis montrer grâce à Kochen-Stone qu'il a une probabilité strictement positive.

    Il sera intéressant de remarquer que pour $k \leq \ell$, on a

$$\mathbb{P}(S_{2\ell^2} = 0, S_{2k^2} = 0) = \mathbb{P}(S_{2k^2=0})\mathbb{P}(S_{2(\ell^2-k^2)} = 0).$$

Annexe C

Problèmes

C.1 Densité naturelle des couples premiers entre eux

Notations et résultats admis

- Si E est un ensemble fini, $\mathcal{P}(E)$ désigne l'ensemble des parties de E et $|E|$ le cardinal de E. Si E est non-vide, alors il existe une unique mesure P sur $(E, \mathcal{P}(E))$ vérifiant $\forall x \in E \quad P(\{x\}) = \frac{1}{|E|}$. Cette mesure est une mesure de probabilité, appelée (mesure de) probabilité uniforme sur E. Elle vérifie la propriété suivante :

$$\forall A \in \mathcal{P}(E) \quad P(A) = \frac{|A|}{|E|}.$$

- Pour $x \in \mathbb{R}$, on appelle partie entière de x et on note $\mathrm{Ent}(x)$ l'unique entier n vérifiant $n \leq x < n+1$.
- $\displaystyle\sum_{n=1}^{+\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$.
- Lorsque n et k sont des entiers, l'écriture $k|n$ signifie "k divise n".
  Si les séries de terme général (u_n) et (v_n) sont absolument convergentes, alors on a l'identité

$$\Big(\sum_{n\geq 1} u_n\Big)\Big(\sum_{n\geq 1} v_n\Big) = \sum_{n\geq 1} (u * v)_n,$$

  où l'on a posé

$$(u * v)_n = \sum_{d|n} u_d v_{n/d}.$$

Par exemple

$$(u * v)_6 = u_1 v_6 + u_2 v_3 + u_3 v_2 + u_6 v_1.$$

– Formule de Poincaré (aussi appelée formule du crible) Pour tous événemements $A_1, A_2, ..., A_n$ sous la probabilité P

$$\begin{aligned} P\Big(\bigcup_{i=1}^{n} A_i\Big) &= \sum_{B\in\mathcal{P}(\{1,\ldots,n\})\setminus\varnothing} (-1)^{1+\mathrm{Card}(B)} P\big(\bigcap_{j\in B} A_j\big) \\ &= \sum_{i=1}^{n} P(A_i) - \sum_{1\le i_1<i_2\le n} P(A_{i_1}\cap A_{i_2}) + \cdots \\ &\quad \cdots + (-1)^{k+1} \sum_{1\le i_1<i_2<\ldots<i_k\le n} P(A_{i_1}\cap\cdots\cap A_{i_k}) + \cdots \\ &\quad \cdots + (-1)^{n+1} P(A_1\cap\cdots\cap A_n). \end{aligned}$$

Question préliminaire

Soit donnée une famille de nombres réels $a(k,n)$ pour $k \geq 1, n \geq 1$ entiers. On suppose qu'il existe une suite de nombres réels positifs $(c_k)_{k\geq 1}$ avec les propriétés :

$$\forall (n,k) \in \mathbb{N}^* \times \mathbb{N}^* \quad |a(k,n)| \leq c_k, \qquad \sum_{k=1}^{+\infty} c_k < +\infty.$$

On suppose que pour tout $k \geq 1$, la limite suivante existe :

$$\lim_{n\to+\infty} a(k,n) = a(k,\infty).$$

Montrer que les deux séries $s_n = \sum_{k=1}^{+\infty} a(k,n)$ et $s = \sum_{k=1}^{+\infty} a(k,\infty)$ convergent absolument et que l'on a $\lim_{n\to+\infty} s_n = s$, soit

$$\lim_{n\to+\infty} \sum_{k=1}^{+\infty} a(k,n) = \sum_{k=1}^{+\infty} a(k,\infty)$$

I

Soit $k \in \mathbb{N}^*$. On pose $E_k = \{0,1\}^k$.

Soit $A_k = \{x \in E_k;\ \sum_{i=1}^{k} x_i \text{ est pair}\}$ et $B_k = \{x \in E_k;\ \sum_{i=1}^{k} x_i \text{ est impair}\}$.

1. Montrer que l'application

$$\begin{array}{rcl}\Psi : E_k & \to & E_k \\ (x_1, x_2, \ldots, x_k) & \mapsto & (1 - x_1, x_2, \ldots, x_k)\end{array}$$

réalise une bijection de A_k dans B_k.

2. Montrer que pour tout entier naturel k non nul

$$|A_k| = |B_k| = 2^{k-1}.$$

3. k étudiants assistent à un cours de mathématiques. Quelle est la probabilité pour que le nombre d'étudiants qui s'ennuyent soit pair ? Préciser les hypothèses simplificatrices que vous choisirez de prendre.
   Indication : Repérer l'état de lassitude des étudiants en associant à chaque étudiant un élément de $\{0, 1\}$ suivant l'intérêt qu'il porte au cours.

II

Soit $n \in \mathbb{N}, n \geq 2$. On décompose n en produits de facteurs premiers :

$$n = \prod_{i=1}^{k} p_i^{\alpha_i},$$

où pour tout i dans $\{1, ..., k\}$, p_i est un nombre premier et α_i un entier strictement positif, les p_i étant deux à deux distincts. On pose

$$\mu(n) = \begin{cases} 0 & \text{si il existe } i \text{ tel que } \alpha_i > 1, \\ (-1)^k & \text{sinon.} \end{cases}$$

On pose également $\mu(1) = 1$.
On a ainsi défini une fonction $\mu : \mathbb{N}^* \to \{-1, 0, 1\}$.
Cette fonction est appelée fonction de Moëbius.
On pose

$$\Omega_n = \{0, \ldots, \alpha_1\} \times \{0, \ldots, \alpha_2\} \times \cdots \times \{0, \ldots, \alpha_k\}$$

et on note Ω_n' l'ensemble des diviseurs positifs de n. On note Q la mesure de probabilité uniforme sur Ω_n et Q' la mesure de probabilité uniforme sur Ω_n'.

1. Montrer que $|\Omega_n| = \prod_{i=1}^{k} (\alpha_i + 1)$.

2. Soit f l'application :

$$\begin{array}{rcl} \Omega_n & \to & \mathbb{N} \\ (\nu_1, \ldots, \nu_k) & \mapsto & \prod_{i=1}^{k} p_i^{\nu_i} \end{array}$$

Montrer que Q' est la mesure image de Q par f.

3. On pose $Z_k = \Omega_n \backslash E_k$. Montrer que pour toute fonction g de Ω_n dans $\mathbb{R}$, on a

$$\int_{\Omega_n} g(\omega)\ dQ(\omega) = \int_{A_k} g(\omega)\ dQ(\omega) + \int_{B_k} g(\omega)\ dQ(\omega) \int_{Z_k} g(\omega)\ dQ(\omega)$$

4. Montrer $\forall \nu = (\nu_1, \ldots, \nu_k) \in \Omega_n$

$$(\mu \circ f)(\nu) = \begin{cases} (-1)^{\left(\sum_{i=1}^{k} \nu_i\right)} & \text{si } \nu \in E_k, \\ 0 & \text{sinon.} \end{cases}$$

5. En utilisant le théorème d'intégration par rapport à une mesure image – appelé aussi théorème de transfert – , montrer que

$$\int_{\Omega'_n} \mu(x)\ dQ'(x) = Q(A_k) - Q(B_k).$$

6. Montrer $\forall n \in \mathbb{N}^*$

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{si } n = 1, \\ 0 & \text{sinon.} \end{cases}$$

7. On choisit au hasard - avec équiprobabilité - un diviseur de 2000. Quelle est la probabilité pour que ce nombre soit divisible par un carré autre que 1 ?

III

Pour n fixé, soit Ω l'ensemble des couples d'entiers $(u, v) \in \{1, ..., n\}^2$. On suppose que tous les couples sont équiprobables. On note donc P_n la probabilité définie sur $\mathcal{P}(\Omega)$ par

$$P_n(A) = \frac{1}{n^2}|A|.$$

On note p_n la probabilité pour que deux entiers pris "au hasard" entre 1 et n soient premiers entre eux.
Pour $k \in \mathbb{N}^*$, on pose $F_k = \{(u, v) \in \Omega;\ k|u \text{ et } k|v\}$.

1. Montrer $P_n(F_k) = (\mathrm{Ent}(n/k))^2/n^2$.
2. Montrer que si k et l sont premiers entre eux, alors
$$P_n(F_k \cap F_l) = P_n(F_{kl}).$$
Généraliser au cas de m entiers deux à deux premiers entre eux.
3. Soient $\beta_1, ..., \beta_k$ les nombres premiers compris entre 2 et n.
On pose $F = \bigcup_{i=1}^{k} F_{\beta_i}$.
Montrer $p_n = 1 - P_n(F)$.
4. En utilisant la formule de Poincaré, montrer
$$p_n = \sum_{d=1}^{n} \mu(d) \frac{1}{n^2} (\mathrm{Ent}(\frac{n}{d}))^2.$$
Indication : il pourra être utile de remarquer que pour tout entier d compris entre 1 et n, $\mu(d)$ est non-nul si et seulement si d peut s'écrire comme produit d'éléments deux à deux distincts pris dans l'ensemble $\{\beta_1, \dots, \beta_k\}$.
5. (a) On pose
$$a(k,n) = \begin{cases} \mu(k)\frac{1}{n^2}(\mathrm{Ent}(\frac{n}{k}))^2 & \text{si } 1 \leq k \leq n, \\ 0 & \text{sinon.} \end{cases}$$
Montrer
$$\forall (n,k) \in (\mathbb{N}^*)^2 \quad |a(k,n)| \leq \frac{1}{k^2}$$
ainsi que
$$\forall k \in \mathbb{N}^* \quad \lim_{n \to \infty} a(k,n) = \frac{\mu(k)}{k^2}.$$
(b) Établir que la suite $(p_n)_{n \geq 2}$ converge vers
$$p = \sum_{d=1}^{+\infty} \frac{\mu(d)}{d^2}.$$
6. Calculer p.
Indication : on pourra appliquer le résultat sur les séries absolument convergentes admis en introduction avec $u_n = \frac{\mu(n)}{n^2}$ et $v_n = \frac{1}{n^2}$, puis utiliser II.6.
7. Que pensez vous de l'affirmation suivante :
« La probabilité pour que deux entiers pris au hasard soient premiers entre eux est $6/\pi^2$. » ?

C.2 Preuve de la loi forte des grands nombres

Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires positives de même loi deux à deux indépendantes, admettant un moment d'ordre un. On pose

$$S_n = \sum_{i=1}^{n} X_i \text{ et } Q_n = \frac{1}{n} S_n.$$

On considère également les variables aléatoires tronquées : $X_i^* = X_i 1\!\!1_{X_i \leq i}$ et les sommes et quotients associés : $S_n^* = \sum_{i=1}^n X_i$ et $Q_n^* = S_n^*/n$.

1. Montrer
$$\operatorname{Var} S_n^* \leq \sum_{i=1}^{n} \mathbb{E}(X_i^*)^2.$$
En déduire
$$\operatorname{Var} S_n^* \leq n \mathbb{E} X_1^2 1\!\!1_{X_1 \leq n}.$$
2. Soit $\beta > 1$. On note u_n l'entier le plus proche de β^n. Montrer $u_n \sim \beta^n$. En déduire
$$\sum_{n=N}^{+\infty} \frac{1}{u_n} = O(\frac{1}{u_N}).$$
3. Montrer qu'il existe une constante $C > 0$ telle que
$$\forall N \geq 1 \quad \sum_{n=N}^{+\infty} \frac{1}{u_n} \leq C \frac{1}{u_N}.$$
4. Montrer
$$\sum_{n=1}^{+\infty} \operatorname{Var} Q_{u_n}^* \leq \mathbb{E}\left(X_1^2 \sum_{n=1}^{+\infty} \frac{1}{u_n} 1\!\!1_{X_1 \leq u_n} \right),$$
puis
$$\sum_{n=1}^{+\infty} \operatorname{Var} Q_{u_n}^* \leq \mathbb{E}\left(X_1^2 \sum_{n: u_n \geq X_1} \frac{1}{u_n} \right).$$
5. En déduire que
$$\sum_{n=1}^{+\infty} \operatorname{Var} Q_{u_n}^* < +\infty.$$
6. Montrer que
$$Q_{u_n}^* - \mathbb{E} Q_{u_n}^* \to 0 \text{ p.s.}$$

7. Montrer que
$$\lim_{n\to+\infty} \mathbb{E}Y_n = \mathbb{E}X_1.$$
8. Montrer que $\lim \mathbb{E}Q_n^* = \mathbb{E}X_1$. (Indication : on pourra utiliser le théorème de Césaro)
9. En déduire que
$$Q_{u_n}^* \to \mathbb{E}X_1 \text{ p.s.}$$
10. Montrer que la série de terme général $P(X_n \neq X_n^*)$ converge.
11. A l'aide du lemme de Borel-Cantelli, montrer que pour presque tout ω, il existe un $n_0(\omega)$ tel que les suites $(X_n(\omega))$ et $(X_n^*(\omega))$ coincident à partir du rang $n_0(\omega)$.
12. Montrer que
$$Q_{u_n} \to \mathbb{E}X_1 \text{ p.s.}$$
13. Si $u_n \leq k \leq u_{n+1}$, montrer que
$$\frac{u_n}{u_{n+1}} Q_{u_n} \leq Q_k \leq \frac{u_{n+1}}{u_n} Q_{u_{n+1}}.$$
14. En déduire que
$$\frac{1}{\beta}\mathbb{E}X_1 \leq \underline{\lim}_{k\to+\infty} Q_k \leq \overline{\lim}_{k\to+\infty} Q_k \leq \beta\mathbb{E}X_1 \text{ p.s.}$$
15. On note
$$\Omega_\beta = \{\omega \in \Omega; \frac{1}{\beta}\mathbb{E}X_1 \leq \underline{\lim}_{k\to+\infty} Q_k(\omega) \leq \overline{\lim}_{k\to+\infty} Q_k(\omega) \leq \beta\mathbb{E}X_1\}.$$
Montrer que
$$P(\bigcap_{n=1}^{+\infty} \Omega_{1+1/n}) = 1.$$
16. Montrer que
$$\lim_{k\to+\infty} Q_k = \mathbb{E}X_1 \text{ p.s..}$$
17. Montrer que le résultat de la dernière question demeure vrai si l'on ne suppose plus que les (X_n) sont des variables positives.