Année universitaire 2021-2022

UNIVERSITÉ DE LORRAINE

Olivier GARET

Intégration et Probabilités

Introduction

Le cours contenu dans le présent polycopié reproduit pour l'essentiel (dans les sept premiers chapitres) le contenu de divers enseignements de Licence que j'ai donnés à Orléans, puis à Nancy.

Le cours de ce polycopié a été un des ingrédients de base de l'ouvrage « De l'Intégration aux Probabilités » [6], que j'ai écrit avec Aline Kurtzmann et que nous avons publié aux éditions Ellipses. Vous êtes invités à vous y reporter pour compléter votre culture.

À la fin de chaque chapitre, le présent polycopié contient des exercices qui serviront de base aux travaux dirigés du cours. À la fin du polycopié, on trouve des indications pour chaque exercice. Il est recommandé de ne s'y reporter qu'après avoir un peu cherché.

Les exercices de la première série sont, pour la plupart, ceux dont une correction est proposée dans Garet-Kurtzmann. Cela ne veut pas dire que les autres exercices ne méritent pas votre attention !

Table des matières

Notations

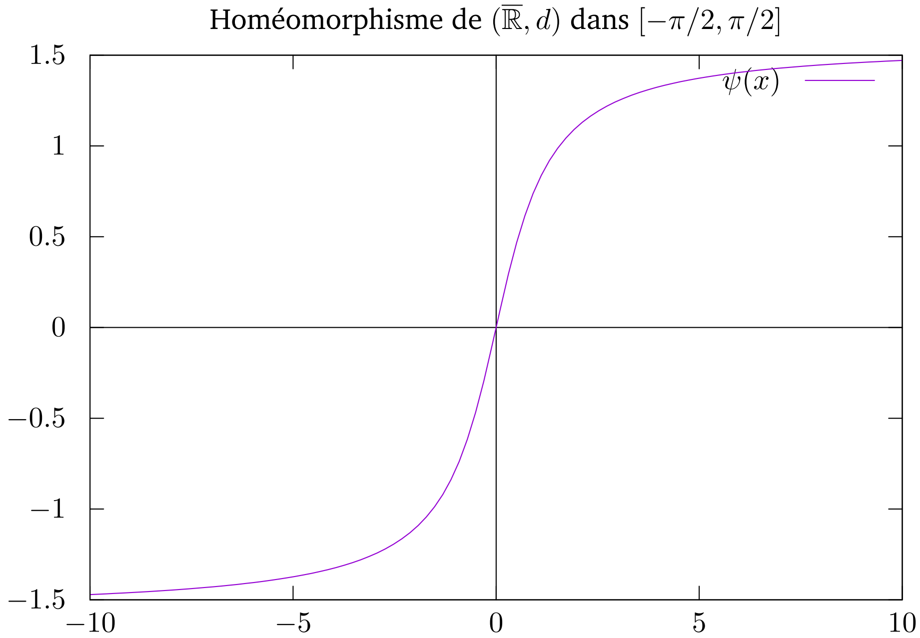
Card(A) ou $|A|$: cardinal de l'ensemble A
$\mathfrak{S}(A)$: ensemble des permutations de A
$\mathfrak{S}_n$: ensemble des permutations de $\{1, \ldots, n\}$
$\mathcal{B}_p(A)$: ensembles des parties de A avec p éléments
$\mathcal{P}(A)$: ensemble des parties de A
M^* : matrice transconjuguée de M
$M_n(\mathbb{K})$: ensemble des matrices $n \times n$ sur le corps $\mathbb{K}$
$\lfloor x \rfloor$: partie entière inférieure de x ($\lfloor \pi \rfloor = \lfloor 3 \rfloor = 3$)
$\lceil x \rceil$: partie entière supérieure de x ($\lceil \pi \rceil = \lceil 4 \rceil = 4$)
$\{x\}$: partie fractionnaire de x : $\{x\} = x - \lfloor x \rfloor$
$n \wedge p$: plus grand commun diviseur (p.g.c.d.) des entiers n et p
$x \wedge y$: minimum des réels x et y
$x \vee y$: maximum des réels x et y
$\langle \cdot, \cdot \rangle$: produit scalaire
$\mathcal{B}(X)$: tribu borélienne de X
$\mathcal{V}(A, \mathcal{A})$: les applications mesurables de $(A, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$
$\overline{\mathcal{V}}(A, \mathcal{A})$: les applications mesurables de $(A, \mathcal{A})$ dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$
$\overline{\mathcal{V}}_+(A, \mathcal{A})$: les applications mesurables de $(A, \mathcal{A})$ dans $(\overline{\mathbb{R}}_+, \mathcal{B}(\overline{\mathbb{R}}_+))$
δ_x : mesure de Dirac au point x
$\mathrm{Ber}(p)$: loi de Bernoulli de paramètre p
$\mathcal{B}(n, p)$: loi binomiale de paramètres n et p
$\mathcal{P}(\lambda)$: loi de Poisson de paramètre λ
$\mathcal{E}(\lambda)$: loi exponentielle de paramètre λ
$\mathcal{N}(m, \sigma^2)$: loi normale de moyenne m et de variance σ^2
$\mathcal{G}(p)$: loi géométrique de paramètre p
$\Gamma(a, \gamma)$: loi Gamma de paramètre de forme a, de paramètre d'échelle γ
$U([a, b])$: loi uniforme sur le segment $[a, b]$
$U(\{a, \ldots, b\})$: loi uniforme sur l'ensemble fini $\{a, \ldots, b\}$
$\mathcal{C}(a, b)$: loi de Cauchy de paramètres a et b
$X_n \Longrightarrow X$: (X_n) converge en loi vers X
$X_n \xrightarrow{\mathbb{P}} X$: (X_n) converge en probabilité vers X
$X_n \xrightarrow{\text{p.s.}} X$: (X_n) converge presque sûrement vers X
i.s. : infiniment souvent ; pour une infinité de valeurs
p.s. : presque sûrement (avec probabilité 1)
p.p. : presque partout (sauf sur un ensemble de mesure nulle)

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 $\psi(x) = \arctan x$ pour x réel $\psi(+\infty) = \pi/2$ et $\psi(-\infty) = -\pi/2$.



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.

Définition. *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.*

Remarque 1.1. *ψ réalise un homéomorphisme croissant de $(\overline{\mathbb{R}}, d)$ dans $[-\frac{\pi}{2}, \frac{\pi}{2}]$ (l'homéomorphisme réciproque est bien sûr le prolongement de la fonction tangente).*

Corollaire 1.2. *De toute suite (a_n) à valeurs dans $(\overline{\mathbb{R}}, d)$, on peut extraire une sous-suite convergente.*

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

Soit une suite $(x_n)_{n\geq 1}$ à valeurs dans $\mathbb{R}$. On peut vérifier que

1. $(x_n)_{n\geq 1}$ converge vers le réel ℓ dans $\overline{\mathbb{R}}$ si et seulement si elle converge vers ℓ dans $\mathbb{R}$,
2. $(x_n)_{n\geq 1}$ converge vers $+\infty$ dans $\overline{\mathbb{R}}$ si et seulement si elle tend vers $+\infty$ lorsque n tend vers l'infini,
3. $(x_n)_{n\geq 1}$ 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 "$-\infty \leq \ell$" et "$\ell \leq +\infty$" pour tout $\ell \in \mathbb{R}$ ainsi que "$-\infty \leq +\infty$". On peut alors énoncer le théorème suivant.

Théorème 1.3. *Toute suite monotone de $(\overline{\mathbb{R}}, d)$ 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 dans $]-\infty, +\infty]$. 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 et majorée, alors elle converge dans $\mathbb{R}$, alors que si elle est croissante non majorée, elle converge vers $+\infty$. $\square$

1.2 Limite supérieure

1.2.1 Limites supérieures, inférieures d'une suite

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

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

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

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

Cette limite existe bien (dans $\overline{\mathbb{R}}$) car la suite (w_n) définie par $w_n = \inf\limits_{k\geq n} a_k$ est croissante.

Exemple: 1) Considérons $a_n = (-1)^n$. On voit que pour n pair, $n = 2p$, $a_{2p} = 1$ tandis que pour n impair, $a_{2p+1} = -1$. Ainsi, pour tout $n \geq 0$, on obtient $\sup\limits_{k\geq n} a_k = 1$ et donc $\overline{\lim\limits_{n\to+\infty}}\ a_n = 1$. De même, pour tout $n \geq 0$, on a $\inf\limits_{k\geq n} a_k = -1$ et donc $\underline{\lim\limits_{n\to+\infty}}\ a_n = -1$.
2) Supposons que $a_n = \frac{2n(-1)^n+1}{n+1}$. Il est aisé de voir que $\lim a_{2n} = 2$ tandis que $\lim a_{2n+1} = -2$. Ainsi, $\overline{\lim\limits_{n\to+\infty}}\ a_n = 2$ tandis que $\underline{\lim\limits_{n\to+\infty}}\ a_n = -2$. On peut remarquer qu'ici, contrairement à ce qu'on observait dans l'exemple précédent, les limites inférieure et supérieure ne sont pas des valeurs prises par la suite. En effet, on a pour tout n : $-2 < a_n < 2$.

Lemme 1.4. *Soient $(x_n)_{n\geq 1}$ une suite à valeurs dans $\overline{\mathbb{R}}$, et f une fonction croissante continue de $(\overline{\mathbb{R}}, d)$ dans $(\overline{\mathbb{R}}, d)$. Alors,*

$$\sup\{f(x_i); i \geq 1\} = f(\sup\{x_i; i \geq 1\}).$$

Démonstration. La suite $(\max\{x_i; 1 \leq i \leq k\})_k$ converge vers $\sup\{x_i; i \geq 1\}$ lorsque k tend vers l'infini. Donc par continuité de f, la suite $(f(\max\{x_i; 1 \leq i \leq k\}))_k$ 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, on obtient l'égalité voulue : $\sup\{f(x_i); i \geq 1\} = f(\sup\{x_i; i \geq 1\})$. □

Définition. *On dit que $a \in \mathbb{R}$ est une valeur d'adhérence pour la suite $(a_n)_n$ s'il existe une sous-suite $(a_{\varphi(n)})_n$ telle que $\lim\limits_{n\to+\infty} a_{\varphi(n)} = a$.*

Théorème 1.5. 1. *$\overline{\lim\limits_{n\to+\infty}}\, a_n$ est la plus grande valeur d'adhérence de la suite (a_n) dans $\overline{\mathbb{R}}$.*

2. *$\underline{\lim\limits_{n\to+\infty}}\, a_n$ est la plus petite valeur d'adhérence de (a_n) dans $\overline{\mathbb{R}}$.*

Démonstration. Nous faisons la démonstration pour la limite supérieure et laissons au lecteur le soin d'adapter cette démonstration pour la limite inférieure. Posons $\overline{\ell} = \overline{\lim\limits_{n\to+\infty}}\, a_n$, et, comme précédemment $v_n = \sup\limits_{k\geq n} a_k$. Montrons d'abord que toute valeur d'adhérence a de (a_n) vérifie $a \leq \overline{\ell}$.

Soit $a = \lim\limits_{n\to+\infty} a_{\phi(n)}$ une valeur d'adhérence. Si $a = -\infty$ ou $\overline{\ell} = +\infty$, il n'y a rien à montrer. Sinon, prenons $\varepsilon > 0$. Il existe N tel que $n \geq N$ entraîne $v_n \leq \overline{\ell} + \varepsilon$, et donc $a_k \leq \overline{\ell} + \varepsilon$ pour $k \geq N$. Comme $\phi(n)$ tend vers l'infini, il existe une constante M telle que $n \geq M$ entraîne $\phi(n) \geq N$. Finalement on a $a_{\phi(n)} \leq \overline{\ell} + \varepsilon$ pour $n \geq M$, d'où $a \leq \overline{\ell} + \varepsilon$. Comme ε est quelconque, on a $a \leq \overline{\ell}$. Il reste à montrer que $\overline{\ell}$ est valeur d'adhérence.

On pose $\phi(1) = 1$. Pour $\psi(x) = \arctan(x)$, on définit par récurrence

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

qui est bien défini, car $\sup\limits_{n\geq\phi(k)+1} \psi(a_n) = \psi\left(\sup\limits_{n\geq\phi(k)+1} a_n\right)$, d'après le lemme 1.19 (ψ est un homéomorphisme, donc continu). Pour $k \geq 1$, on a

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

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

Théorème 1.6.

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

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

Démonstration. Cette fois encore, nous ne prouvons le résultat que pour la limite supérieure. Supposons que x soit tel que $\{n \geq 1; a_n \geq x\}$ est infini. On peut extraire de cet ensemble une suite $(\phi(n))$ strictement croissante d'entiers telle que $(a_{\phi(n)})$ converge vers $z \in \overline{\mathbb{R}}$ et $a_{\phi(n)} \geq x$ pour tout n. Comme $\overline{\lim\limits_{n\to+\infty}}\ a_n$ est plus grande que toutes les valeurs d'adhérence, on a donc

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

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

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

Raisonnons maintenant par l'absurde et supposons que

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

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

Théorème 1.7. *Une suite (a_n) à valeurs dans $\overline{\mathbb{R}}$ converge si et seulement si $\underline{\lim\limits_{n\to+\infty}}\ a_n = \overline{\lim\limits_{n\to+\infty}}\ a_n$. On a alors $\underline{\lim\limits_{n\to+\infty}}\ a_n = \overline{\lim\limits_{n\to+\infty}}\ a_n = \lim\limits_{n\to+\infty} a_n$.*

Démonstration. Si $\underline{\lim\limits_{n\to+\infty}}\ a_n = \overline{\lim\limits_{n\to+\infty}}\ a_n = +\infty$, 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 $\underline{\lim\limits_{n\to+\infty}}\ a_n = \overline{\lim\limits_{n\to+\infty}}\ a_n = -\infty$, 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 $\underline{\lim\limits_{n\to+\infty}}\ a_n = \overline{\lim\limits_{n\to+\infty}}\ a_n = \ell \in \mathbb{R}$. Soit $\varepsilon > 0$. Comme

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

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

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

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

Réciproquement, si (a_n) converge vers $\ell \in \overline{\mathbb{R}}$, $\underline{\lim\limits_{n\to+\infty}}\ a_n$ et $\overline{\lim\limits_{n\to+\infty}}\ a_n$ sont égales à ℓ puisque ce sont des valeurs d'adhérence de $(a_n)_n$ et que ℓ est la seule valeur d'adhérence de la suite $(a_n)_n$. □

Théorème 1.8. *Soient (u_n) et (u'_n) deux suites vérifiant $u_n \leq u'_n$ pour tout n. On a $\overline{\lim\limits_{n\to+\infty}}\ u_n \leq \overline{\lim\limits_{n\to+\infty}}\ u'_n$ et $\underline{\lim\limits_{n\to+\infty}}\ u_n \leq \underline{\lim\limits_{n\to+\infty}}\ u'_n$.*

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

Corollaire 1.9. *Soit $\ell \in \mathbb{R}$. On suppose que pour tout $\varepsilon > 0$,*

$$\underline{\lim}_{n\to+\infty} u_n \geq \ell - \varepsilon \quad \textit{et} \quad \overline{\lim}_{n\to+\infty} u_n \leq \ell + \varepsilon.$$

Alors, (u_n) converge vers ℓ.

Démonstration. On a

$$\ell - \varepsilon \leq \underline{\lim}_{n\to+\infty} u_n \leq \overline{\lim}_{n\to+\infty} u_n \leq \ell + \varepsilon.$$

Comme cela est vrai pour tout $\varepsilon > 0$, on fait tendre ε vers 0 et on obtient

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

Le résultat découle du théorème 1.22. □

1.2.2 Limites supérieures, inférieures d'ensembles

Si $(A_n)_{n\geq 0}$ est une suite d'ensembles, on définit

$$\overline{\lim}_{n\to+\infty} A_n := \bigcap_{n=1}^{+\infty} \bigcup_{k=n}^{+\infty} A_k$$

et

$$\underline{\lim}_{n\to+\infty} A_n := \bigcup_{n=1}^{+\infty} \bigcap_{k=n}^{+\infty} A_k.$$

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 appartiennent à tous ces ensembles à partir d'un certain rang. Nous laissons le soin au lecteur de prouver ce résultat en exercice. Notons l'analogie avec les limites supérieure et inférieure de suites réelles : on remplace le supremum par la réunion et l'infimum par l'intersection.

En probabilité, nous étudierons des suites de fonctions à valeurs réelles, appelées "variables aléatoires", qui sont définies sur un espace Ω, appelé "espace d'état". Nous adoptons donc dès maintenant ces notations et considérons des fonctions $X_n : \Omega \to \mathbb{R}$. Ainsi, dire que $\{n : X_n(\omega) \geq M - \varepsilon\}$ est infini revient à dire que $\omega \in \overline{\lim}_{n\to+\infty} \{\omega : X_n(\omega) \geq M - \varepsilon\}$. On en déduit que

$$\left\{\omega : \overline{\lim}_{n\to+\infty} X_n(\omega) \geq M\right\} = \bigcap_{\varepsilon>0} \overline{\lim}_{n\to+\infty} \{\omega : X_n(\omega) \geq M - \varepsilon\}. \tag{1.1}$$

Par ailleurs, dire que "$\{n : X_n(\omega) \geq M + \varepsilon\}$ est fini" revient à dire qu'à partir d'un certain rang, on a $X_n(\omega) < M + \varepsilon$. Donc si ω est tel que l'ensemble $\{n : X_n(\omega) \geq M + \varepsilon\}$ est fini, alors $\omega \in \underline{\lim}_{n\to+\infty} \{\omega : X_n(\omega) < M + \varepsilon\}$. On en déduit que

$$\left\{\omega : \overline{\lim}_{n\to+\infty} X_n(\omega) \leq M\right\} = \bigcap_{\varepsilon>0} \underline{\lim}_{n\to+\infty} \{\omega : X_n(\omega) < M + \varepsilon\}. \tag{1.2}$$

Si on remplace X_n par $-X_n$ et M par $-M$ dans (1.9), on obtient :

$$\left\{\omega : \overline{\lim_{n\to+\infty}} -X_n(\omega) \leq -M\right\} = \bigcap_{\varepsilon>0} \underline{\lim_{n\to+\infty}} \{\omega : -X_n(\omega) < -M+\varepsilon\},$$

soit

$$\left\{\omega : \underline{\lim_{n\to+\infty}} X_n(\omega) \geq M\right\} = \bigcap_{\varepsilon>0} \underline{\lim_{n\to+\infty}} \{\omega : X_n(\omega) > M-\varepsilon\}. \tag{1.3}$$

Et en passant aux complémentaires dans (1.9) et (1.10), on a

$$\left\{\omega : \overline{\lim_{n\to+\infty}} X_n(\omega) > M\right\} = \bigcup_{\varepsilon>0} \overline{\lim_{n\to+\infty}} \{\omega : X_n(\omega) \geq M+\varepsilon\} \tag{1.4}$$

et

$$\left\{\omega : \underline{\lim_{n\to+\infty}} X_n(\omega) < M\right\} = \bigcup_{\varepsilon>0} \overline{\lim_{n\to+\infty}} \{\omega : X_n(\omega) \leq M-\varepsilon\}. \tag{1.5}$$

Si on fait maintenant subir à la formule (1.8) les mêmes transformations qu'à (1.9), on obtient trois nouvelles formules (à écrire en exercice). Remarquons qu'on peut remplacer dans toutes les formules précédentes $\varepsilon > 0$ par $\frac{1}{p}$ avec $p \in \mathbb{N}^*$.

1.3 Exercices d'analyse

1.3.1 Exercices de la série 1

Exercice 1. Étudier la convergence de chacune des séries de terme général u_n suivant :

a) $u_n = \frac{(n^2+1)2^n}{(2n+1)!}$,

b) $u_n = \left(\frac{n^2-5n+1}{n^2-4n+2}\right)^{n^2}$,

c) $u_n = \left(1+\frac{1}{n}\right)^{n^2} e^{-n}$,

d) $u_n = \left(1+\frac{1}{n}\right)^{-n^2}$.

e) $u_n = \frac{2n+1}{n^2(n+1)^2}$.

Exercice 2. Étudier la nature et calculer éventuellement la somme des séries suivantes :

a) $\sum\limits_{n=2}^{+\infty} \frac{1}{n(n-1)}$. En déduire la convergence de la série $\sum\limits_{n=2}^{+\infty} \frac{1}{n^2}$.

b) $\sum\limits_{n=1}^{+\infty} \frac{1}{(n+\alpha)(n+\alpha+1)}$, avec $\alpha \in \mathbb{R}$, $\alpha \neq 0, -1, \ldots$.

c) $\sum\limits_{n=3}^{+\infty} \frac{14n-18}{n^3-7n+6}$.

d) $\sum\limits_{n=10}^{+\infty} 100\left(\frac{8}{9}\right)^n$ et calculer cette somme.

e) $\sum\limits_{n=1}^{+\infty} \arctan \frac{1}{n^2+n+1}$. (Indication : pour $a, b \in \mathbb{R}$ avec $ab > -1$, on a $\arctan a - \arctan b = \arctan \frac{a-b}{1+ab}$).

Exercice 3. On suppose que la suite $(u_n)_{n\geq 0}$ est telle que $u_n \to 0$ lorsque $n \to +\infty$ et on pose, pour $n \geq 0$: $v_n = u_{2n} + u_{2n+1}$. Montrer que les séries de terme général u_n et de terme général v_n sont de même nature et ont même somme.
Application : montrer la convergence de la série des $\left(\log\left(1+\frac{(-1)^n}{n+2}\right)\right)_{n\geq 0}$.

Exercice 4. Soit $(a_{n,k})_{n\geq 1, k\geq 1}$ une suite double d'éléments de $\overline{\mathbb{R}}_+$. Montrer que $\sum\limits_{n=1}^{+\infty} \sum\limits_{k=1}^{+\infty} a_{n,k} = \sum\limits_{k=1}^{+\infty} \sum\limits_{n=1}^{+\infty} a_{n,k}$.

Exercice 5. Soient $\gamma > -1$, α et β strictement positifs. Étudier la convergence de l'intégrale suivante à l'aide de l'étude de la série associée :

$$\int_0^{+\infty} \frac{x^\gamma}{1 + x^\alpha |\sin x|^\beta}\, dx.$$

Indication : on rappelle que pour $x \in [-\pi/2, \pi/2]$, on a $\frac{2}{\pi}|x| \leq |\sin x| \leq |x|$. Montrer de plus que la série de terme général $\frac{\log n}{n^a}$ converge si et seulement si $a > 1$.

1.3.2 Exercices de la série 2

Exercice 6. Dans tout l'exercice, α désigne un réel quelconque. Discuter la convergence des intégrales suivantes :

1. $\int_0^1 \frac{\sin(x)}{x}\, dx$;
2. $\int_0^1 \frac{\sin(x)}{x^\alpha}\, dx$;
3. $\int_0^1 \frac{\cos(x)}{x^\alpha}\, dx$;
4. $\int_0^1 \frac{x-\sin(x)}{x^\alpha}\, dx$;
5. $\int_0^1 \frac{x-\cos(x)}{x^\alpha}\, dx$;
6. $\int_0^1 \frac{x^\alpha}{1-e^{-x}}\, dx$;
7. $\int_1^{+\infty} \frac{\log(x)^\alpha}{x^2}\, dx$;
8. $\int_1^{+\infty} \frac{\log(1+x)}{x^2}\, dx$;
9. $\int_0^{+\infty} \frac{\log(1+x^2)}{x^2}\, dx$;
10. $\int_0^{+\infty} \frac{\log(1+x)}{x^2}\, dx$;
11. $\int_0^{+\infty} e^{-x}(\log x)\, dx$;
12. $\int_1^{+\infty} e^{-x}x^{\alpha-1}\, dx$.

Exercice 7. Soit (x_n) une suite à valeurs dans $\{0;1\}$. Quelles sont les valeurs possibles pour la limite supérieure et la limite inférieure de la suite (x_n) ? Caractériser par une phrase en français les différentes issues.

Exercice 8.
1. On pose $a_n = (-1)^n(1 + \frac{1}{n})$.

   Déterminer $\underline{\lim}_{n\to+\infty} a_n$ et $\overline{\lim}_{n\to+\infty} a_n$.
2. Soit f une fonction continue croissance de $\mathbb{R}$ dans $\mathbb{R}$. Donner une expression simple de $\underline{\lim}_{n\to+\infty} f(a_n)$ et $\overline{\lim}_{n\to+\infty} f(a_n)$.
3. Même question lorsque (a_n) est remplacée par une suite dont la limite supérieure est 1 et la limite supérieure est -1

Exercice 9. Soit $(a_i,\ i \in I)$ une famille non vide d'éléments de $\overline{\mathbb{R}}$.
1. Démontrer que $\inf(a_i,\ i \in I) = -\sup(-a_i,\ i \in I)$.
2. 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

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

Exercice 10. 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 $\overline{\mathbb{R}}$.

Exercice 11. Soit $(a_n)_{n\geq 1}$ et $(b_n)_{n\geq 1}$ deux suites dans $\mathbb{R}$.
1. Démontrer que

$$\overline{\lim}_{n\to+\infty}(a_n + b_n) \leq \overline{\lim}_{n\to+\infty} a_n + \overline{\lim}_{n\to+\infty} b_n.$$

   Montrer que l'inégalité peut être stricte.

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

$$\varliminf_{n\to+\infty} (a_n + b_n) = \lim_{n\to+\infty} a_n + \varliminf_{n\to+\infty} b_n.$$

Exercice 12. 1. Pour $n \geq 1$, on pose $H_n = \sum_{k=1}^{n} \frac{1}{k}$. Par la méthode de votre choix, montrer que $H_n \sim \log n$.

2. Soit f une fonction continue sur $[0,1]$.
   Pour $n \geq 1$, on pose $S_n = \sum_{k=1}^{n} \frac{1}{k} f(\frac{k}{n})$. Après avoir justifié l'existence d'un $\alpha > 0$ tel que $|f(0) - f(k/n)| \leq \varepsilon$ pour $0 \leq k \leq \alpha n$, montrer que

$$\varlimsup_{n\to+\infty} \frac{|S_n - f(0)H_n|}{H_n} \leq \varepsilon.$$

   Conclure.

3. Donner un équivalent de $\displaystyle\sum_{k=1}^{n} \frac{1}{\sin \frac{k}{n}}$ lorsque n tend vers l'infini.

Exercice 13. *Série de Dirichlet.*
Soient $(u_n)_{n\geq 1}$ une suite de réels, x, y des réels avec $y > x$. Montrer que si la série de terme général $\frac{u_n}{n^x}$ converge, alors la série de terme général $\frac{u_n}{n^y}$ converge.

Exercice 14. *Suites sous-additives (lemme de Fekete)* (voir [4]).
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 $\inf_{n\geq 1} \frac{u_n}{n}$.

1. 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 $\varlimsup_{n\to+\infty} \frac{u_{kn+r}}{nk+r} \leq \frac{u_k}{k}$.
2. Montrer que $\varlimsup_{n\to+\infty} \frac{u_n}{n} \leq \inf_{k\geq 1} \frac{u_k}{k}$.
3. Conclure.
4. Application 1. Pour $A \in M_n(\mathbb{R})$, on pose $|||A||| = \sup_{y\in\mathbb{R}^n\setminus\{0\}} \frac{\|Ay\|_\infty}{\|y\|_\infty}$. Montrer que la suite $|||A^n|||^{1/n}$ converge vers un réel positif.
5. Application 2. Soit E une partie finie de $\mathbb{R}^d$. On note A_n l'ensemble des suites $(u_1, \ldots, u_n)$ qui vérifient
   — $u_1 \in E$
   — $u_{i+1} - u_i \in E$ pour tout $i \in \{1, \ldots, n-1\}$
   — $i \mapsto u_i$ est injective
   Montrer que la suite $|A_n|^{1/n}$ converge vers un réel positif.

Exercice 15. 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 $\lim_{n\to+\infty} (a_n)^n = a > 0$, $\lim_{n\to+\infty} (b_n)^n = b > 0$. Soient $p, q > 0$ avec $p + q = 1$. Déterminer $\lim_{n\to+\infty} (pa_n + qb_n)^n$.

Chapitre 2

Un peu de théorie de la mesure

La théorie des probabilités décrit les événements comme des parties 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*. Afin d'étudier les probabilités, ce qui est notre but, on a besoin des fondements de la théorie de la mesure. Ceci est l'objet de ce chapitre.

Soit 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

On rappelle que $\mathcal{P}(\Omega)$ désigne l'ensemble des parties de Ω. De plus, dire que $A \subset \mathcal{P}(\Omega)$ est équivalent à dire que $A \in \mathcal{P}(\mathcal{P}(\Omega))$.

2.1.1 Axiomes de base

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

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

Remarquons qu'une tribu correspond à toute l'information disponible.

2.1.2 Propriétés

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

— $\Omega \in \mathcal{A}$.

— Pour toute suite $(A_i)_{i\geq 1}$ d'éléments de $\mathcal{A}$, on a $\bigcap\limits_{i=1}^{+\infty} A_i \in \mathcal{A}$.

— Soit $n \geq 1$. Pour toute suite $(A_i)_{i\geq 1}$ d'éléments de $\mathcal{A}$, on a $\bigcup\limits_{i=1}^{n} A_i \in \mathcal{A}$.

— Soit $n \geq 1$. Pour toute suite $(A_i)_{i\geq 1}$ d'éléments de $\mathcal{A}$, on a $\bigcap\limits_{i=1}^{n} A_i \in \mathcal{A}$.

Remarque 2.1. *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 $(A_i)_{i \geq 1}$ *d'éléments de* $\mathcal{A}$ *deux à deux disjoints, on a* $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{A}$.

Nous laissons la démonstration de ce résultat en exercice au lecteur.

2.1.3 Sous-tribus

Si $\mathcal{A}$ est une tribu et si 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

Proposition 2.2. *Soient* Ω *un ensemble et* T *un ensemble de tribus sur* Ω [2]. *On suppose que* T *est non vide. Il peut être fini ou infini, y compris infini non dénombrable. Alors* $\mathcal{B} = \bigcap_{\mathcal{A} \in T} \mathcal{A}$ *est une tribu.*

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

— $\forall \mathcal{A} \in T$, on a $\varnothing \in \mathcal{A}$. Donc $\varnothing \in \bigcap_{\mathcal{A} \in T} \mathcal{A} = \mathcal{B}$.

— Soit $A \in \mathcal{B}$. On doit montrer que $A^c \in \mathcal{B}$. Soit $\mathcal{A} \in T$. Comme on sait que $A \in \mathcal{A}$ et que $\mathcal{A}$ est une tribu, donc $A^c \in \mathcal{A}$. Comme ceci est vrai pour tout $\mathcal{A} \in T$, on a $A^c \in \bigcap_{\mathcal{A} \in T} \mathcal{A} = \mathcal{B}$.

— Soit $(A_i)_{i \in I}$ une famille dénombrable d'éléments de $\mathcal{B}$. On doit montrer que $\bigcup_{i \in I} A_i \in \mathcal{B}$. Soit $\mathcal{A} \in T$. Comme les A_i sont dans $\mathcal{A}$ et $\mathcal{A}$ est une tribu, on a $\bigcup_{i \in I} A_i \in \mathcal{A}$.

Ceci étant vrai pour tout $\mathcal{A} \in T$, on obtient que $\bigcup_{i \in I} A_i \in \bigcap_{\mathcal{A} \in T} \mathcal{A} = \mathcal{B}$.

□

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 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 (au sens de l'inclusion) contenant toutes les tribus $\mathcal{A}_i$. On la note

$$\sigma(\mathcal{A}_i; i \in I).$$

Notons que la tribu engendrée par une tribu est la tribu de départ : $\sigma(\mathcal{A}) = \mathcal{A}$.

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 tous les ensembles A_i est

$$\sigma(\mathcal{A}_i; i \in I).$$

1. Une erreur classique à ne pas commettre : si $\mathcal{B}$ est une sous-tribu de $\mathcal{A}$, et $B \subset A$ 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 inclus dans Ω.

On note cette tribu $\sigma(A_i; i \in I)$. On peut remarquer que
$\sigma(\mathcal{A}_i; i \in I) = \sigma\left(\bigcup_{i\in I} \mathcal{A}_i\right)$.

2.1.5 Tribu borélienne, fonctions mesurables

Soient $(A, \mathcal{A})$ et $(B, \mathcal{B})$ deux espaces mesurables, c'est-à-dire des espaces munis d'une tribu. On dit qu'une application f de A dans B est mesurable de $(A, \mathcal{A})$ dans $(B, \mathcal{B})$ si quel que soit $X \in \mathcal{B}$, son image réciproque $f^{-1}(X)$ est dans $\mathcal{A}$.

Commençons par une remarque simple. Si $f : A \to B$ est une fonction $(A, \mathcal{A}) - (B, \mathcal{B})$ mesurable (c'est-à-dire f est mesurable de $(A, \mathcal{A})$ dans $(B, \mathcal{B})$) et si $g : B \to C$ est $(B, \mathcal{B}) - (C, \mathcal{C})$ mesurable, alors $g \circ f$ est $(A, \mathcal{A}) - (C, \mathcal{C})$ mesurable.

Théorème 2.3 (Théorème fondamental de la mesurabilité)**.** *Soit f une application quelconque d'un ensemble Ω dans un ensemble Ω'. Alors*

1. *Pour toute tribu $\mathcal{T}$ sur Ω', $f^{-1}(\mathcal{T})$ est une tribu sur Ω, où $f^{-1}(\mathcal{T}) = \{f^{-1}(T) : T \in \mathcal{T}\}$.*
2. *Pour tout $\mathcal{A} \in \mathcal{P}(\mathcal{P}(\Omega'))$, on a $\sigma(f^{-1}(\mathcal{A})) = f^{-1}(\sigma(\mathcal{A}))$.*

On appelle $\sigma(f^{-1}(\mathcal{A})) = f^{-1}(\sigma(\mathcal{A}))$ la tribu image de $\mathcal{A}$ par f.

Démonstration. 1. Vérifions que $f^{-1}(\mathcal{T})$ satisfait les axiomes d'une tribu.

— $\varnothing \in f^{-1}(\mathcal{T})$ car $\varnothing = f^{-1}(\varnothing)$ et $\varnothing \in \mathcal{T}$.

— Soit $A \in f^{-1}(\mathcal{T})$. Il existe $B \in \mathcal{T}$ avec $A = f^{-1}(B)$.
On a alors $A^c = (f^{-1}(B))^c = f^{-1}(B^c)$. Comme $B^c \in \mathcal{T}$, on trouve donc $A^c \in f^{-1}(\mathcal{T})$.

— Soit $(A_i)_{i\geq 1}$ une suite d'éléments de $f^{-1}(\mathcal{T})$.
Pour tout i, il existe $B_i \in \mathcal{T}$ avec $A_i = f^{-1}(B_i)$. On a

$$\bigcup_{i=1}^{+\infty} A_i = \bigcup_{i=1}^{+\infty} (f^{-1}(B_i)) = f^{-1}\left(\bigcup_{i=1}^{+\infty} B_i\right).$$

Or on sait que $\bigcup_{i=1}^{+\infty} B_i \in \mathcal{T}$ donc $\bigcup_{i=1}^{+\infty} A_i \in f^{-1}(\mathcal{T})$.

2. Comme $\mathcal{A} \subset \sigma(\mathcal{A})$, on 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 égale à $\sigma(\mathcal{A})$ tout entier, ce qui montre l'inclusion voulue. □

S'il n'y a pas d'ambiguïté 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 Ω telle que f soit une application mesurable de $(\Omega, \mathcal{A})$ dans $(\Omega', \mathcal{B})$. On dit que c'est la tribu engendrée par l'application f. On peut noter que $f : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{B})$ est mesurable si et seulement si la tribu $f^{-1}(\mathcal{B})$ est incluse dans $\mathcal{A}$.

Corollaire 2.4. *Soient $(A, \mathcal{A})$ et $(B, \mathcal{B})$ deux espaces mesurables et $\mathcal{C} \subset \mathcal{B}$ tel 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 $\mathcal{A}$.*

Définition. *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.*

Remarque 2.5. *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.*

Corollaire 2.6. *Soient* 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 peut même énoncer un résultat un peu plus fort.

Corollaire 2.7. *Soient* A *et* B *deux espaces topologiques. On suppose que* $(A_i)_{i\in I}$ *est une partition dénombrable de* A *par des ensembles de* $\mathcal{B}(A)$*. On suppose que* f *est une application de* A *dans* B *telle que pour tout* i*, la restriction* f_i *de* f *à* A_i *est une application continue. Alors,* f *est mesurable de* $(A, \mathcal{B}(A))$ *dans* $(B, \mathcal{B}(B))$.

En prenant $A = \mathbb{R}$*, on a en particulier que les applications continues par morceaux sont mesurables.*

Démonstration. Soit O un ouvert de B. On a $f^{-1}(O) = \underset{i\in I}{\cup} f_i^{-1}(O)$. Or, pour tout $i \in I$, $f_i^{-1}(O)$ est un ouvert de A_i, c'est-à-dire l'intersection de A_i et d'un ouvert de A : c'est donc un borélien de A. Par réunion dénombrable, $f^{-1}(O)$ est encore un borélien de A. Comme les ouverts de B engendrent la tribu borélienne de B, le corollaire 2.4 permet de conclure. □

Théorème 2.8. *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 non-vide de $\mathbb{R}^d$. Soit $x \in O$: 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

$$O = \underset{x\in O}{\cup} U(x).$$

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

$$O = \underset{x\in O/\sim}{\cup} U(x).$$

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

Remarque 2.9. *Pour le lecteur n'appréciant pas le passage au quotient, voici une autre démonstration du résultat précédent (pour montrer que* O *est dans* $\mathcal{T}$*).*
Notons P *l'ensemble des pavés ouverts de* $\mathbb{R}^d$ *à extrémités rationnelles. On a alors*

$$O = \underset{U\in\{U(x):\ x\in O\}}{\cup} U.$$

Comme $\{U(x) :\ x \in O\} \subset P$ *qui est dénombrable, cela prouve bien que* 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 $\mathbb{R}$ (ou $\mathbb{Q}$).

Corollaire 2.10. *Soient $(A, \mathcal{A})$ un espace mesurable, 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 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*

On note dans la suite $\mathcal{V}(A, \mathcal{A})$ 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

Définition. *Soient $(\Omega, \mathcal{A})$ et $(\Omega', \mathcal{B})$ deux espaces mesurables. 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.*

Remarque 2.11. *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 2.12. *On suppose que $\mathcal{A} = \sigma((A_i)_{i \in I})$ et $\mathcal{B} = \sigma((B_j)_{j \in J})$. On suppose en outre qu'il existe I' et J' dénombrables avec $I' \subset I$, $J' \subset J$ et tels que $\Omega = \underset{i \in I'}{\cup} A_i$ et $\Omega' = \underset{j \in J'}{\cup} B_j$. 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 $\mathcal{C}_A = \{B \in \mathcal{B} : A \times B \in \mathcal{O}\}$. Montrons que pour tout $i \in I$, $\mathcal{C}_{A_i} = \mathcal{B}$. Comme $\mathcal{C}_{A_i}$ contient les B_j qui engendrent $\mathcal{B}$, il suffit de voir que $\mathcal{C}_{A_i}$ est une tribu. On a $A_i \times \varnothing = \varnothing \in \mathcal{O}$. De même, il est facile de voir que $\mathcal{C}_{A_i}$ est stable par réunion dénombrable (laissé au lecteur). On en déduit que $\Omega' = \cup_{j \in J'} B_j \in \mathcal{C}_{A_i}$. Pour $B \in \mathcal{C}_{A_i}$, on a $A_i \times (\Omega' \backslash B) = (A_i \times \Omega') \backslash (A_i \times B) \in \mathcal{O}$, d'où $B^c \in \mathcal{C}_{A_i}$. $\mathcal{C}_{A_i}$ est donc bien une sous-tribu de $\mathcal{B}$: elle contient les B_j qui engendrent $\mathcal{B}$: c'est $\mathcal{B}$. Notons $\mathcal{D} = \{A \in \mathcal{A} : \mathcal{C}_A = \mathcal{B}\}$. En procédant comme précédemment, le lecteur montre (laissé en exercice) que $\mathcal{D}$ est une sous-tribu de $\mathcal{A}$. Mais $\mathcal{D}$ contient les A_i. Comme les A_i engendrent $\mathcal{A}$, on a $\mathcal{D} = \mathcal{A}$, 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 engendrées, on a $\mathcal{A} \otimes \mathcal{B} \subset \mathcal{O}$. L'inclusion réciproque est évidente. □

Théorème 2.13. *Soient f une application de Ω_3 dans Ω_1, g une application de Ω_3 dans Ω_2. On définit une application $F : \Omega_3 \to \Omega_1 \times \Omega_2$ par $F(x) = (f(x), g(x))$. L'application F est $(\Omega_3, \mathcal{C}) - (\Omega_1 \times \Omega_2, \mathcal{A} \otimes \mathcal{B})$ mesurable si et seulement si f est $(\Omega_3, \mathcal{C}) - (\Omega_1, \mathcal{A})$ mesurable et g est $(\Omega_3, \mathcal{C}) - (\Omega_2, \mathcal{B})$ mesurable.*

Démonstration. La condition est nécessaire car $f = \pi_1 \circ F$ et $g = \pi_2 \circ F$.
Ainsi lorsque F est $(\Omega_3, \mathcal{C}) - (\Omega_1 \times \Omega_2, \mathcal{A} \otimes \mathcal{B})$ mesurable, comme π_1 est
$(\Omega_1 \times \Omega_2, \mathcal{A} \otimes \mathcal{B}) - (\Omega_1, \mathcal{A})$ mesurable, f est mesurable en tant que composée d'applications mesurables. Pour les mêmes raisons, g est mesurable. Supposons maintenant que f est $(\Omega_3, \mathcal{C}) - (\Omega_1, \mathcal{A})$ mesurable et g est $(\Omega_3, \mathcal{C}) - (\Omega_2, \mathcal{B})$ mesurable. Intéressons-nous maintenant à F.
Soit $A \times B \in \mathcal{A} \times \mathcal{B}$. On a $F^{-1}(A \times B) = f^{-1}(A) \cap g^{-1}(B)$. Comme f et g sont mesurables, $f^{-1}(A)$ et $g^{-1}(B)$ sont dans $\mathcal{C}$, donc leur intersection aussi. Ainsi pour tout $A \times B \in \mathcal{A} \times \mathcal{B}$, on a $F^{-1}(A \times B) \in \mathcal{C}$. Mais les ensembles $A \times B \in \mathcal{A} \times \mathcal{B}$ engendrent $\mathcal{A} \otimes \mathcal{B}$, donc F est bien $(\Omega_3, \mathcal{C}) - (\Omega_1 \times \Omega_2, \mathcal{A} \otimes \mathcal{B})$ mesurable. □

Théorème 2.14. *Soient* $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2), (\Omega_3, \mathcal{F}_3)$ *trois espaces mesurables. 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 2.12, 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 2.4 permet alors de conclure. □

L'extension du théorème 2.14 au produit d'un nombre fini quelconque d'espaces mesurés se fait alors aisément par récurrence.

Théorème 2.15. *Pour tout entier* $d \geq 2$*, on a*

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

Démonstration. Il nous suffit de montrer que pour tout $d \geq 1$, on a l'identité $\mathcal{B}(\mathbb{R}^{d+1}) = \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R})$, puis de conclure par récurrence. Or, d'après le théorème 2.8, 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 2.12, les produits $A \times B$ engendrent la tribu $\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R})$. Or 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 2.8, engendrent la tribu $\mathcal{B}(\mathbb{R}^{d+1})$. □

Théorème 2.16. *Soit* f, g *des applications mesurables de* $(\Omega, \mathcal{A})$ *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* $(\Omega, \mathcal{A})$ *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* $(\Omega, \mathcal{A})$ *dans* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Démonstration. Avec les notations du théorème 2.13, $H = G \circ F$ car on sait que $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$. Pour les cas particuliers de fonctions G données dans l'énoncé, notons que G étant 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

Définition. *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. $\forall A \in \mathcal{A}$*, on a* $A^c \in \mathcal{A}$.
3. *Pour tous* A *et* B *dans* $\mathcal{A}$*, on a* $A \cup B \in \mathcal{A}$.

Remarque 2.17. *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 appelées σ-algèbres, la lettre σ étant traditionnellement attachée aux propriétés liées à des familles dénombrables.

Remarque 2.18. *En anglais,*
— *algèbre se dit "field", ou "algebra".*
— *tribu (σ-algèbre) se dit "σ-field", ou "σ-algebra".*

2.2.2 Espace mesuré

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

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

vérifiant les propriétés suivantes :

1. $\mu(\varnothing) = 0$.
2. Pour toute suite $(A_i)_{i\geq 1}$ d'éléments de $\mathcal{A}$ deux à deux disjoints et telle que $\overset{+\infty}{\underset{i=1}{\cup}} A_i \in \mathcal{A}$, on a
$$\mu\left(\overset{+\infty}{\underset{i=1}{\cup}} A_i\right) = \sum_{i=1}^{+\infty} \mu(A_i).$$

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

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

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

et $\mu(A) = 0$.

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

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

Proposition 2.19. *Soient $\mathcal{A}$ une algèbre et μ une mesure sur $(\Omega, \mathcal{A})$. Les propriétés suivantes sont alors des conséquences relativement immédiates des définitions :*

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

10. *Si $(A_i)_{i\geq 1}$ est une suite croissante d'éléments de $\mathcal{A}$ (c'est-à-dire que $\forall i \in \mathbb{N}^*, \quad A_i \subset A_{i+1}$) telle que $A = \bigcup_{i=1}^{+\infty} A_i \in \mathcal{A}$, alors la suite $(\mu(A_i))_{i\in\mathbb{N}}$ est monotone, croissante, et converge vers $\mu(A)$.*
11. *Si $(A_i)_{i\geq 1}$ est une suite décroissante d'éléments de $\mathcal{A}$ (c'est-à-dire que $\forall i \in \mathbb{N}^*, \quad A_{i+1} \subset A_i$), avec $\mu(A_1) < +\infty$ et si $A = \bigcap_{i=1}^{+\infty} A_i \in \mathcal{A}$, alors la suite $(\mu(A_i))_{i\in\mathbb{N}}$ 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 $\mu(A) + \mu(B) = \mu(A \cup B) = \mu(E)$.
4. Les ensembles $A\backslash B$, $B\backslash A$ et $A \cap B$ sont disjoints et leur réunion est $A \cup B$, donc d'après la propriété 1, on a

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

car $A\backslash B$ et $A \cap B$ sont disjoints, de réunion A, tandis que $B\backslash A$ et $A \cap B$ sont disjoints, de réunion B.

5. Si $A \subset B$, alors B est la réunion disjointe de A et de $B\backslash A$.
Donc $\mu(B) = \mu(A) + \mu(B\backslash A) \geq \mu(A)$.
6. Si $\mu(A) = \mu(B) = +\infty$, alors le résultat est trivial. Supposons donc $\mu(A) < +\infty$. Dans ce cas, comme $A\cap B \subset A$, on a par le point précédent que $\mu(A\cap B) \leq \mu(A) < +\infty$. Il suffit ensuite d'appliquer la relation 4 en remarquant que $\mu(A\cap B) \geq 0$.
7. $(A\cap B) \subset A$, donc d'après la propriété 5, $\mu(A\cap B) \leq \mu(A)$. 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 5, $\mu(A) \leq \mu(A\cup B)$. 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 $n \geq 2$, $B_n = A_n \backslash \left(\bigcup_{i=1}^{n-1} B_i \right)$. Par construction, les $(B_n)_{n\geq 1}$ 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

$$\mu\left(\bigcup_{i=1}^{+\infty} A_i\right) = \mu\left(\bigcup_{i=1}^{+\infty} B_i\right) = \sum_{i=1}^{+\infty} \mu(B_i) \leq \sum_{i=1}^{+\infty} \mu(A_i).$$

10. Comme $A_i \subset A_{i+1}$, on a $\mu(A_i) \leq \mu(A_{i+1})$, donc la suite est croissante. Comme on a pour tout i, $A_i \subset A$, la suite $(\mu(A_i))_{i\geq 1}$ 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,

$$\begin{aligned} \mu(A) &= \mu\left(\bigcup_{i=1}^{+\infty} B_i\right) = \sum_{i=1}^{+\infty} \mu(B_i) = \lim_{n\to+\infty} \sum_{i=1}^{n} \mu(B_i) \\ &= \lim_{n\to+\infty} \mu\left(\bigcup_{i=1}^{n} B_i\right) = \lim_{n\to+\infty} \mu(A_n). \end{aligned}$$

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 muni d'une tribu. Soit $x \in \Omega$. On appelle mesure de Dirac en x et on note δ_x la mesure définie par

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

Vérifions brièvement que δ_x est une mesure. Il est évident que δ_x est à valeurs dans $[0, +\infty]$. 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

$$\delta_x\left(\bigcup_{i=1}^{+\infty} A_i\right) = 0 = \sum_{i=1}^{+\infty} \delta_x(A_i).$$

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

$$\sum_{i=1}^{+\infty} \delta_x(A_i) = 1 = \delta_x\left(\bigcup_{i=1}^{+\infty} A_i\right).$$

2.2.4 Mesure de comptage

Soit $(\Omega, \mathcal{F})$ un espace mesurable. 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

$$\sum_{i=1}^{+\infty} C(A_i) = +\infty = C\left(\bigcup_{i=1}^{+\infty} A_i\right).$$

De même s'il y a une infinité de A_i non vides, $\bigcup_{i=1}^{+\infty} A_i$ est infini et la somme $\sum_{i=1}^{+\infty} C(A_i)$ a une infinité de termes qui dépassent 1 donc encore une fois $\sum_{i=1}^{+\infty} C(A_i) = +\infty = C\left(\bigcup_{i=1}^{+\infty} A_i\right)$. 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'ensembles finis et alors la formule recherchée est bien connue.

2.2.5 Opérations 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 Mesure image

Soient $(\Omega, \mathcal{F}, \mu)$ un espace mesuré et f une application de Ω dans Ω'. On considère la tribu

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

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

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

Vérifions que μ_f est bien une mesure. Il est immédiat que μ_f prend des valeurs positives, et $\mu_f(\varnothing) = \mu(f^{-1}(\varnothing)) = \mu(\varnothing) = 0$. Si les $(A_n)_{n\geq 1}$ sont des éléments de $\tilde{\sigma}(\mathcal{F}, f)$ deux à deux disjoints, les $f^{-1}(A_n)$ sont des éléments de $\mathcal{F}$ deux à deux disjoints, et on a

$$\begin{aligned} \mu_f\left(\bigcup_{n=1}^{+\infty} A_n\right) &= \mu\left(f^{-1}(\bigcup_{n=1}^{+\infty} A_n)\right) = \mu\left(\bigcup_{n=1}^{+\infty} f^{-1}(A_n)\right) \\ &= \sum_{n=1}^{+\infty} \mu(f^{-1}(A_n)) = \sum_{n=1}^{+\infty} \mu_f(A_n), \end{aligned}$$

donc μ_f est bien une mesure.

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 $\tilde{\sigma}(\mathcal{F}, f)$.

2.2.7 Extension d'une mesure – mesure de Lebesgue

Définition. *Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré. Lorsqu'il existe un recouvrement dénombrable de Ω par des sous-ensembles de mesure finie, c'est-à-dire lorsqu'il existe une suite $(E_n)_{n\in\mathbb{N}}$ d'éléments de la tribu, tous de mesure finie, avec*

$$\Omega = \bigcup_{n=0}^{+\infty} E_n$$

on dit que μ est une mesure $\sigma-$finie. Quitte à remplacer chaque E_n par $\bigcup_{k=0}^{n} E_k$, on peut supposer que la suite de sous-ensembles (E_n) est croissante.

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

Théorème 2.20 (prolongement de Hahn)**.** *Étant données 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*

$$\bar{\mu}(A) = \inf\left(\{\sum_{n=1}^{+\infty} \mu(A_n);\ \textit{pour tout } n, A_n \in \mathcal{F} \textit{ et } A \subset \bigcup_{n=1}^{\infty} A_n\}\right)$$

est une mesure sur $\sigma(\mathcal{F})$ qui prolonge μ. Ce prolongement est unique si μ est une mesure 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 2.21. *Si μ est une mesure finie sur la tribu $\mathcal{F}$ 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 $\mu(A\Delta A') \leq \varepsilon$, où $A\Delta A' = (A \cup A') \setminus (A \cap A')$.*

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 $A = \bigcup_{n\geq 1} A_n$ 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 $\mu\left(A\setminus \bigcup_{k=1}^{n} A_k\right) \leq \varepsilon/2$. On a

$$\begin{aligned}
\mu\left(A\Delta \bigcup_{k=1}^{n} A'_k\right) &\leq \mu\left(A\Delta \bigcup_{k=1}^{n} A_k\right) + \mu\left(\bigcup_{k=1}^{n} A'_k\Delta \bigcup_{k=1}^{n} A_k\right) \\
&\leq \frac{\varepsilon}{2} + \sum_{k=1}^{n} \frac{\varepsilon}{2^{k+1}} \leq \varepsilon.
\end{aligned}$$

□

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

Théorème 2.22. *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 $\mathbb{R}$.

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

Remarque 2.23. *λ est l'unique mesure sur la tribu borélienne telle que la mesure d'un segment est la longueur dudit segment. En effet, on peut retrouver la mesure de tous les intervalles à partir de la longueur des segments avec le théorème de continuité séquentielle croissante.*

Corollaire 2.24. *La mesure de Lebesgue est invariante par les translations : pour tout borélien A et tout réel t, on a $\lambda(A) = \lambda(\tau_t^{-1}(A))$, où $\tau_t(x) = x + t$.*

Démonstration. Notons μ la mesure image de λ par τ_t. Pour des réels a et b avec $a \leq b$, $\mu([a, b]) = \lambda([a - t, b - t]) = b - a$, donc μ est la mesure de Lebesgue d'après le résultat d'unicité. □

Comme un ensemble dénombrable est réunion dénombrable de singletons, notons que tout ensemble dénombrable est de mesure de Lebesgue nulle. Par exemple, l'ensemble des rationnels est de mesure nulle.

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 $\phi(x) = \tan x$ si $x \in]-\pi/2, \pi/2[$, $\phi(-\pi/2) = -\infty$ et $\phi(+\pi/2) = +\infty$.

On définit une métrique sur $\overline{\mathbb{R}}$ par $d(x, y) = |\phi^{-1}(x) - \phi^{-1}(y)|$. 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 $\mathbb{R}$ (et $\overline{\mathbb{R}}$)

$\mathbb{R}$ 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 œuvre une technique classique à connaître.

Théorème 2.25. *Soient f et 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. On a

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

Comme f est mesurable de $(\Omega, \mathcal{F})$ dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ et que $[q, +\infty] \in \mathcal{B}(\overline{\mathbb{R}})$, on a $f^{-1}([q, +\infty]) \in \mathcal{F}$. 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 2.26. *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 ensembles 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

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

ou, en adoptant le formalisme probabiliste :

$$\{f > x\} = \bigcup_{n=1}^{+\infty} \{f_n > x\}.$$

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 $\left\{\underline{\lim}_{n\to+\infty} f_n < \overline{\lim}_{n\to+\infty} f_n\right\}$ qui est dans $\mathcal{F}$ 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

2.4.1 Exercices de la série 1

Exercice 16. Montrer que $\mathcal{A} = \{A \subset \mathbb{N}; A$ ou A^c est fini$\}$ est une algèbre sur $\mathbb{N}$. Est-ce une tribu ? Mêmes questions si on remplace $\mathbb{N}$ par $\mathbb{R}$.

Exercice 17. 1. On sait que la tribu borélienne de $\mathbb{R}$ est engendrée par les ensembles de la forme $]a,b[$, avec a et b dans $\mathbb{Q}$. Montrer que la tribu borélienne est la tribu engendrée par
   (a) les intervalles fermés $[a,b]$, où a et b sont dans $\mathbb{Q}$,
   (b) les intervalles semi-ouverts $[a,b[$, où a et b sont dans $\mathbb{Q}$,
   (c) les demi-droites $[a,+\infty[$, où $a \in \mathbb{Q}$.

2. On rappelle que la tribu borélienne de $\mathbb{R}^2$ est engendrée par les pavés rectangulaires à côtés rationnels. Montrer que

$$\mathcal{B}(\mathbb{R}^2) = \sigma([a,b]\times[c,d]) = \sigma(B(x,r); x \in \mathbb{R}^2, r > 0)$$

où $B(x,r)$ est la boule ouverte de centre x et de rayon r (c'est-à-dire l'ensemble des points qui sont à une distance de x strictement plus petite que r).

Exercice 18. Soient $\mathcal{A}$ une σ-algèbre (ou tribu) et $(A_n)_{n\geq 1}$ une suite d'événements de $\mathcal{A}$.

1. Montrer que

$$\underline{\lim}_{n\to+\infty} A_n \in \mathcal{A}, \quad \overline{\lim}_{n\to+\infty} A_n \in \mathcal{A}, \text{ et } \underline{\lim}_{n\to+\infty} A_n \subset \overline{\lim}_{n\to+\infty} A_n.$$

2. Montrer que

$$\overline{\lim}_{n\to+\infty} \mathbb{1}_{A_n} - \underline{\lim}_{n\to+\infty} \mathbb{1}_{A_n} = \mathbb{1}_{\{\overline{\lim}_{n\to+\infty} A_n \setminus \underline{\lim}_{n\to+\infty} A_n\}}.$$

Exercice 19. Soient $(E,\mathcal{T})$ un espace mesurable et $(\mu_n)_n$ une suite de mesures sur $(E,\mathcal{T})$. Pour tout $A \in \mathcal{T}$, on pose $\mu(A) := \sum_{n=0}^{+\infty} \mu_n(A)$.

1. Montrer que μ est une mesure sur $(E,\mathcal{T})$.
2. On suppose que les mesures μ_n sont des probabilités (c'est-à-dire que pour tout n, on suppose que $\mu_n(E) = 1$), et on considère une suite $(p_n)_n$ de réels positifs tels que $\sum_{n=0}^{+\infty} p_n = 1$. Pour tout $A \in \mathcal{T}$, on pose $\mu(A) = \sum_{n=0}^{+\infty} p_n\mu_n(A)$. Vérifier que μ est une probabilité sur $(E,\mathcal{T})$.
3. Application. On considère les mesures $\mu_1 = \sum_{p=1}^{+\infty} \delta_p$ et $\mu_2 = \sum_{p=1}^{+\infty} p\delta_p$. Pour chacune de ces mesures, calculer la mesure des ensembles (mesurables) suivants :
   (a) $A_n := [n, n+1+\frac{1}{n^2}]$, où $n \in \mathbb{N}^*$,
   (b) $B_n := \bigcup_{p=1}^{n} A_p$ et $B = \bigcup_{n\geq 1} A_n$,
   (c) $C_n := \bigcap_{p=1}^{n} A_p$ et $C = \bigcap_{n\geq 1} A_n$.

Exercice 20. 1. Montrer qu'une tribu engendrée par une famille finie de parties est engendrée par une partition finie.

2. Soient Ω un ensemble, $(\Omega_i)_{i\in I}$ une partition de Ω indexée par un ensemble I quelconque. On pose $\mathcal{G} = \sigma(\Omega_i, i \in I)$ et

$$\mathcal{T} = \{A \in \mathcal{P}(\Omega); \forall i \in I \quad A \cap \Omega_i = \varnothing \text{ ou } A \cap \Omega_i = \Omega_i\}.$$

   (a) Montrer que $\mathcal{T}$ est une tribu.
   (b) Décrire $\mathcal{G}$, lorsque I est fini ou dénombrable.
   (c) Soient $x \in \Omega$ et i_0 l'unique $i \in I$ tel que $x \in \Omega_i$. Montrer que Ω_{i_0} est le plus petit des $A \in \mathcal{G}$ tels que $x \in A$.
3. Montrer que le résultat de la question 1. est faux si on remplace fini par dénombrable. (Conseil : utiliser la tribu borélienne de $\mathbb{R}$.)

Exercice 21. Soit $f_n : X \to \mathbb{R}$ une suite de fonctions mesurables. Montrer que l'ensemble suivant est mesurable :

$$A := \{x \in X; f_n(x) \text{ n'est pas de Cauchy}\}.$$

Exercice 22. *Exemple de partie de $\mathbb{R}$ qui n'est pas un borélien.*
Comme $\mathbb{R}$ est un $\mathbb{Q}$-espace vectoriel, le sous $\mathbb{Q}$-espace vectoriel de $\mathbb{R}$ engendré par 1 admet un supplémentaire S, c'est-à-dire que S est un $\mathbb{Q}$-espace vectoriel inclus dans $\mathbb{R}$ et que tout $x \in \mathbb{R}$ s'écrit sous la forme $x = q + s$ pour un unique couple $(q,s) \in \mathbb{Q} \times S$.

1. En déduire qu'il existe $S' \subset [0,1[$ tel que tout $x \in \mathbb{R}$ s'écrit sous la forme $x = q' + s'$ pour un unique couple $(q', s') \in \mathbb{Q} \times S'$.
2. Montrer qu'il n'existe pas de mesure μ non nulle sur $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ qui soit invariante par translation et assigne une masse finie à l'intervalle $[0,1]$.
3. Montrer que $\mathcal{B}(\mathbb{R}) \neq \mathcal{P}(\mathbb{R})$.

Exercice 23. On se place sur $(\mathbb{N}^*, \mathcal{P}(\mathbb{N}^*))$.

1. La fonction définie par $\mu(A) = \sum\limits_{n \in A} \frac{1}{n^2}$ si A est fini et $\mu(A) = +\infty$ si A est infini est-elle une mesure ?
2. Montrer que la fonction définie par $\mu(A) = \sum\limits_{n \in A} \frac{1}{n^2}$ si l'ensemble A est non vide et $\mu(A) = 0$ sinon est une mesure.

Exercice 24. Soit μ la mesure du dénombrement sur $\mathbb{N}$ (c'est-à-dire pour tout $A \subset \mathbb{N}$, $\mu(A) = \text{card}(A)$). Pour tout $n \in \mathbb{N}$, posons $A_n = \{p; p \geq n\}$.
A-t-on $\mu\left(\bigcap\limits_{n=1}^{+\infty} A_n\right) = \lim\limits_{n \to +\infty} \mu(A_n)$? Expliquer.

Exercice 25. Soit $(X, \mathcal{T}, \mu)$ un ensemble mesuré. Soient $A, B \in \mathcal{T}$ avec $A \subset B$ et $\mu(A) = \mu(B) < +\infty$.
Montrer que pour tout $C \in \mathcal{T}$, on a $\mu(A \cap C) = \mu(B \cap C)$.

Exercice 26. Soit $\mathcal{T} := \{A \in \mathcal{B}(\mathbb{R}); A = -A\}$ où $-A = \{-x; x \in A\}$.

1. Montrer que $\mathcal{T}$ est une tribu.
2. Les applications $f : x \mapsto e^x$, $g : x \mapsto x^3$ et $h : x \mapsto \cos(x)$ sont-elles $(\mathcal{B}(\mathbb{R}), \mathcal{T})$-mesurables ? $(\mathcal{T}, \mathcal{B}(\mathbb{R}))$-mesurables ? $(\mathcal{T}, \mathcal{T})$-mesurables ?
3. Caractériser les applications de $\mathbb{R}$ dans $\mathbb{R}$ qui sont $(\mathcal{T}, \mathcal{B}(\mathbb{R}))$-mesurables.

2.4.2 Exercices de la série 2

Exercice 27. Soit $\Omega = \{1, 2, 3, 4, 5, 6\}$. Pour $x \in \Omega$, on pose $f(x) = \mathbb{1}_{\{x>3\}}$ et $g(x) = \mathbb{1}_{\{3|x\}}$.
Décrire $\sigma(f)$, $\sigma(g)$, puis $\sigma(f, g)$.

Exercice 28. En vous inspirant de l'exercice précédent, montrer que la tribu engendrée par une fonction d'un ensemble E vers un ensemble fini F est en bijection avec la tribu de toutes les parties d'un (éventuellement autre) ensemble fini.

Exercice 29. Soit μ une mesure sur $(\Omega, \mathcal{F})$. Soit $B \in \mathcal{F}$ fixé. Montrer que $A \mapsto \mu(A \cap B)$ définit une mesure sur $(\Omega, \mathcal{F})$.

Exercice 30.

1. 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 ?
2. 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, on a $\{f < x\} = \{g < x\}$ μ-presque partout, alors $f = g$ μ-presque partout.

Exercice 31. On pose
$$\mu_n = \sum_{k=1}^{2n} \frac{1}{k}\delta_k.$$
Expliquez brièvement pourquoi μ_n est une mesure et calculer $\mu_5([0,\pi])$. Étudier le comportement asymptotique de $\mu_n([n,+\infty[)$.

Exercice 32. Soit a un réel. On définit la translation τ_a sur $\mathbb{R}$ 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.

Exercice 33. 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 $\mathcal{A} \times \mathcal{B}$.

Exercice 34. 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

$\mathcal{C} = \{]-\infty, a[; a \in \mathbb{Q}\}$ $\quad$ $\mathcal{F} = \{[a,b]; a,b \in \mathbb{Q}\}$
$\mathcal{D} = \{]-\infty, a]; a \in \mathbb{Q}\}$ $\quad$ $\mathcal{G} = \{]a,b]; a,b \in \mathbb{Q}\}$
$\mathcal{E} = \{[a,b[; a,b \in \mathbb{Q}\}$ $\quad$ $\mathcal{I} = \{[a,+\infty[; a \in \mathbb{Q}\}$.

Exercice 35. Pour tout entier n 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 $(A_p)_{p\in\mathcal{P}}$.

1. Montrer que l'ensemble C des entiers qui sont premiers avec 2000 est $\mathcal{T}$-mesurable.
2. Montrer que l'ensemble $B = \{2^k; k \in \mathbb{N}^*\}$ des puissances de deux est $\mathcal{T}$-mesurable.

Exercice 36. 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).

Exercice 37. *Support d'une mesure sur* $\mathbb{R}^d$.

1. 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.
2. 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 $r_n \in \mathbb{Q}_*^+$ telles que $O = \bigcup_{n\in D} B(x_n, r_n)$.
3. 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 l'espace $(\Omega, \mathcal{F})$.

3.1 Espace probabilisé

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

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

vérifiant les propriétés suivantes :

1. $\mathbb{P}(\varnothing) = 0$, $\mathbb{P}(\Omega) = 1$.
2. Pour toute suite $(A_i)_{i\geq 1}$ d'éléments de $\mathcal{F}$ deux à deux disjoints, on a $\mathbb{P}\left(\bigcup\limits_{i=1}^{+\infty} A_i\right) = \sum\limits_{i=1}^{+\infty} \mathbb{P}(A_i)$.

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

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

Dans le contexte des probabilités, on appelle *événement* [1] tout élément de la tribu $\mathcal{F}$.

Remarque 3.1 (sur le vocabulaire)**.** *L'image $\mathbb{P}(A)$ d'un événement A par l'application $\mathbb{P}$ est appelée probabilité de cet événement. Ainsi, le mot « probabilité » peut désigner à la fois une application et la valeur de cette application en un point. Le contexte doit permettre de lever toute ambiguïté.*

Proposition 3.2. *Les propriétés suivantes sont alors des conséquences relativement immédiates des définitions :*

1. *Pour toute suite $(A_i)_{1\leq i\leq n}$ d'éléments de $\mathcal{F}$ deux à deux disjoints, on a $\mathbb{P}\left(\bigcup\limits_{i=1}^{n} A_i\right) = \sum\limits_{i=1}^{n} \mathbb{P}(A_i)$.*
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. $\forall A, B \in \mathcal{F}, \quad (A \subset B) \Longrightarrow (\mathbb{P}(A) \leq \mathbb{P}(B))$.

1. En 1990, l'Académie française adopte la graphie évènement, mais note que « la graphie ancienne événement n'est cependant pas considérée comme fautive, encore que rien ne la justifie plus ».

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* $\mathcal{F}$, $\mathbb{P}\left(\bigcup_{i=1}^{+\infty} A_i\right) \leq \sum_{i=1}^{+\infty} \mathbb{P}(A_i)$.
10. *Si* $(A_i)_{i \geq 1}$ *est une suite croissante d'événements (c'est-à-dire que pour tout* $i \geq 1, \quad A_i \subset A_{i+1}$*) et si l'on pose* $A = \bigcup_{i=1}^{+\infty} A_i$*, alors la suite* $(\mathbb{P}(A_i))_{i \geq 1}$ *est monotone, croissante, et converge vers* $\mathbb{P}(A)$ *quand* i *tend vers* $+\infty$.
11. *Si* $(A_i)_{i \geq 1}$ *est une suite décroissante d'événements (c'est-à-dire que pour tout* $i \geq 1 \quad A_{i+1} \subset A_i$*) et si l'on pose* $A = \bigcap_{i=1}^{+\infty} A_i$*, alors la suite* $(\mathbb{P}(A_i))_{i \geq 1}$ *est monotone, décroissante, et converge vers* $\mathbb{P}(A)$ *quand* i *tend vers* $+\infty$.

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

Théorème 3.3. *Soient* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace probabilisé et* $(A_t)_{t \in T}$ *des éléments de* $\mathcal{F}$ *deux à deux disjoints. Alors* $J = \{t \in T; \mathbb{P}(A_t) > 0\}$ *est fini ou dénombrable.*

Démonstration. On a $J = \bigcup_{n=1}^{+\infty} J_n$, avec $J_n = \{t \in T; \mathbb{P}(A_t) \geq 1/n\}$. On va montrer que J_n est fini, ce qui montrera que J est fini ou dénombrable, comme réunion dénombrable d'ensembles finis. Soient $n \geq 1$ et $I \subset J_n$ avec I fini. On a

$$\frac{|I|}{n} \leq \sum_{t \in I} \mathbb{P}(A_t) = \mathbb{P}\left(\bigcup_{t \in I} A_t\right) \leq \mathbb{P}(\Omega) = 1,$$

donc $|I| \leq n$. On en déduit que $|J_n| \leq n$, ce qui donne le résultat voulu. □

3.2 Partitions et probabilités

Tout d'abord, nous donnons la définition d'une partition, qui sera à nouveau utilisée plus tard, lors des rappels de dénombrement A.3.3.

Définition. *On appelle partition de l'ensemble* Ω *une famille* $(A_i)_{i \in I}$ *de parties de* Ω *deux à deux disjointes telles que* $\Omega = \bigcup_{i \in I} A_i$.

Remarquons que si Ω est un ensemble fini et les $(A_i)_{i \in I}$ forment une partition de Ω, alors $|\Omega| = \sum_{i \in I} |Ai|$.

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 3.4. *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace probabilisé. Soient* I *un ensemble d'indices fini ou dénombrable et* $(\Omega_i)_{i \in I}$ *une partition* $\mathcal{F}-$*mesurable 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 $(\Omega_i)_{i \in I}$ est une partition de Ω, la famille $(A \cap \Omega_i)_{i \in I}$ est une partition de A. A est donc réunion disjointe des $(A \cap \Omega_i)_{i \in I}$, et donc $\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \cap \Omega_i)$. □

3.3 Probabilité conditionnelle

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

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

$\mathbb{P}(A|B)$ se lit « probabilité de A sachant B »et se note aussi parfois $\mathbb{P}_B(A)$. On a évidemment

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

Remarque 3.5. *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 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 3.6. *Soient* $n \geq 2$ *et* $E_1, \dots, E_n$ *des événements vérifiant*

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

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

$$\begin{aligned} \mathbb{P}(E_n) &= \mathbb{P}(E_n|E_{n-1})\mathbb{P}(E_{n-1}|E_{n-2}) \dots \mathbb{P}(E_2|E_1)\mathbb{P}(E_1) \\ &= \left(\prod_{k=2}^{n} \mathbb{P}(E_k|E_{k-1}) \right) \mathbb{P}(E_1). \end{aligned}$$

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 : $0.8 \times 0.6 \times 0.1 = 0.048$. Le théorème ci-dessus est parfois énoncé sous la forme plus compliquée – mais équivalente – suivante.

Théorème 3.7. *Soit* $n \geq 2$. *On suppose que* $A_1, \dots, A_n$ *sont des événements tels que* $\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) > 0$. *Alors*

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

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. □

3.3.2 Conditionnement par tous les cas possibles

Le résultat suivant est l'expression en termes de probabilités conditionnelles du principe de partition.

Théorème 3.8. *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace probabilisé. Soient* I *un ensemble d'indices fini ou dénombrable et* $(\Omega_i)_{i\in I}$ *une partition* $\mathcal{F}-$*mesurable 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 3.4, 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 3.9. *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace probabilisé. Soient* I *un ensemble d'indices fini ou dénombrable et* $(\Omega_i)_{i\in I}$ *une partition mesurable de* Ω *telle que pour tout* $i \in I$, $\Omega_i \in \mathcal{F}$ *avec* $\mathbb{P}(\Omega_i) > 0$. *Soit* A *un événement 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. On écrit

$$\mathbb{P}(\Omega_j|A) = \frac{\mathbb{P}(A\cap\Omega_j)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|\Omega_j)\mathbb{P}(\Omega_j)}{\mathbb{P}(A)}$$

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 au 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 obtient 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 A et B sont indépendants sous $\mathbb{P}$ si on a

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

Si il n'y a pas d'ambiguité quant à la mesure de probabilité considérée, la précision "sous $\mathbb{P}$" est souvent sous-entendue.

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

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

3.4.2 Tribus indépendantes

Soient $(\Omega, \mathcal{F}, \mathbb{P})$ 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 $\mathbb{P}$ 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 $\mathbb{P}$ si pour tout sous-ensemble fini $J \subset I$, on a

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

Remarque 3.10. — *On peut noter qu'une famille de tribus est indépendante si toute sous-famille finie est indépendante. En particulier, si $I = \mathbb{N}$, pour montrer que les tribus $(\mathcal{A}_i)_{i\in\mathbb{N}}$ sont indépendantes, comme toute partie finie de $\mathbb{N}$ est inclus dans un ensemble $\{0, \ldots, n\}$, il suffit de montrer que pour tout $n \geq 1$, les tribus $\mathcal{A}_0, \ldots, \mathcal{A}_n$ sont indépendantes.*

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

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

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. En particulier, si $I = \{0, \ldots, n\}$, pour montrer que les tribus $(\mathcal{A}_i)_{i\in I}$ sont indépendantes, il suffit de montrer que pour tout $k \in \{1, \ldots, n\}$, la tribu $\mathcal{A}_k$ est indépendante de la tribu engendrée par $\mathcal{A}_0, \ldots, \mathcal{A}_{k-1}$: on prouve alors par récurrence sur $k \leq n$:

$$\forall (A_i)_{0\leq i\leq k} \in \prod_{i=0}^{k} \mathcal{A}_i, \quad \mathbb{P}\left(\bigcap_{i=0}^{k} A_i\right) = \prod_{i=0}^{k} \mathbb{P}(A_i).$$

Ceci peut être utilisé en conjonction avec le premier point pour montrer qu'une suite de tribus est indépendante.

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 3.11 (utile)**.** *Si les tribus $\mathcal{A}$ et $\mathcal{B}$ sont indépendantes sous $\mathbb{P}$, si $\mathcal{A}'$ 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 $\mathbb{P}$.*

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 reportée à la fin du chapitre par souci de lisibilité.

Théorème 3.12 (Critère d'identification d'une mesure σ-finie)**.** *Soient $\mathbb{P}$ et $\mathbb{Q}$ deux mesures sur $(\Omega, \mathcal{F})$. On suppose qu'il existe un π-système $\mathcal{C}$ qui engendre $\mathcal{F}$ ($\sigma(\mathcal{C}) = \mathcal{F}$) et sur lequel $\mathbb{P}$ et $\mathbb{Q}$ coïncident, c'est-à-dire que*

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

et qu'il existe une famille croissante Ω_n d'éléments de $\mathcal{C}$ avec $\Omega = \bigcup_{n=1}^{+\infty} \Omega_n$ et $\mathbb{P}(\Omega_n) < +\infty$ pour tout n. Alors $\mathbb{P} = \mathbb{Q}$.

Corollaire 3.13 (Critère d'identification d'une probabilité)**.** *Soient $\mathbb{P}$ et $\mathbb{Q}$ deux mesures de probabilité sur $(\Omega, \mathcal{F})$. On suppose qu'il existe un π-système $\mathcal{C}$ qui engendre $\mathcal{F}$ ($\sigma(\mathcal{C}) = \mathcal{F}$) et sur lequel $\mathbb{P}$ et $\mathbb{Q}$ coïncident, c'est-à-dire que*

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

Alors $\mathbb{P} = \mathbb{Q}$.

Démonstration. Remarquer que la deuxième condition du théorème est toujours vérifiée si $\mathbb{P}$ est une mesure de probabilité. $\square$

Théorème 3.14. *Soient $\mathcal{C}$ 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 couple $(A, B) \in \mathcal{C} \times \mathcal{D}$, 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 $\mathcal{T}_A = \{B \in \mathcal{B}, \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\}$. Considérons d'abord le cas où $A \in \mathcal{C}$. Si $\mathbb{P}(A) = 0$, alors A est indépendant de tout événement, donc $\mathcal{T}_A = \mathcal{B}$. Si $\mathbb{P}(A) \neq 0$, on peut définir sur $\mathcal{B}$ la probabilité conditionnelle $\mathbb{P}_A$ par

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

Les probabilités $\mathbb{P}$ et $\mathbb{P}_A$ coïncident sur $\mathcal{D}$. Comme $\mathcal{D}$ est un π-système qui engendre $\mathcal{B}$, $\mathbb{P}$ et $\mathbb{P}_A$ 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

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

Or $\mathcal{B} = \sigma(\mathcal{D})$ 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 ce que nous voulions, soit :

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

$\square$

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

Cette section peut être omise en première lecture.

Définition. *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 $\mathcal{L}$, $\underset{n\geq 1}{\cup} A_n \in \mathcal{L}$.*

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 $A'_n = \overset{n}{\underset{k=1}{\cup}} A_k = \Omega \backslash \left(\overset{n}{\underset{k=1}{\cap}} A_k^c \right)$, 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, on a $A = \underset{n\geq 1}{\cup} A_n = \underset{n\geq 1}{\cup} A'_n \in \mathcal{A}$, d'après le troisième axiome d'un λ-système.

Théorème 3.15 (Théorème $\lambda - \pi$)**.** *Si $\mathcal{P}$ est un π-système et si $\mathcal{L}$ est un λ-système, alors $\mathcal{P} \subset \mathcal{L}$ entraîne $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

Démonstration. Admis, voir par exemple Garet-Kurtzmann. □

Démonstration du théorème 3.12. Soit $n \geq 1$. Considérons l'ensemble $\mathcal{L}$ des éléments A de la tribu $\mathcal{F}$ engendrée par le π-système tels que

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

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 $\lambda - \pi$ de Dynkin, pour voir que pour tout A, on a l'identité $\mathbb{P}(A \cap \Omega_n) = \mathbb{Q}(A \cap \Omega_n)$. Le théorème de continuité séquentielle croissante permet alors de conclure. □

3.6 Exercices sur le formalisme probabiliste

3.6.1 Exercices de la série 1

Exercice 38. Soit A un événement d'un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. La fonction indicatrice de A, notée $\mathbb{1}_A$, est donnée par :

$$\mathbb{1}_A : \Omega \to \{0, 1\}, \;\; \omega \mapsto 1 \text{ si } \omega \in A, \; 0 \text{ sinon.}$$

Calculer $\mathbb{1}_\Omega, \mathbb{1}_\varnothing$.
Pour $A, B \subset \Omega$, exprimer $\mathbb{1}_{A\cap B}, \mathbb{1}_{A^c}, \mathbb{1}_{A\backslash B}, \mathbb{1}_{A\Delta B}$ à l'aide de $\mathbb{1}_A, \mathbb{1}_B$ par des formules simples.
(*Rappel* : $A\Delta B = (A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$ s'appelle la différence symétrique entre A et B.)
Montrer que $\mathbb{1}_{A\cup B} = 1 - (1 - \mathbb{1}_A)(1 - \mathbb{1}_B)$.

Exercice 39. Soient I un ensemble fini et $(A_i)_{i\in I}$ des événements indépendants. Montrer que les événements $(A_i^c)_{i\in I}$ sont indépendants.

Exercice 40. On note $\Omega = \mathbb{N}^*$, que l'on munit de la tribu $\mathcal{F} = \mathcal{P}(\Omega)$ et de la mesure de comptage C. Pour $s > 1$, on pose

$$\zeta(s) = \sum_{k=1}^{+\infty} \frac{1}{k^s} < +\infty.$$

1. Montrer que $\lim\limits_{s\to 1^+} \zeta(s) = +\infty$.

2. Soit $s > 1$. On note μ_s la mesure sur $(\Omega, \mathcal{F})$ telle que

$$\forall i \in \mathbb{N}^* \quad \mu_s(\{i\}) = \frac{1}{\zeta(s)} \frac{1}{i^s}.$$

Vérifier que μ_s est une mesure de probabilité.

Soit p un entier naturel non-nul. Montrer que $\mu_s(p\mathbb{N}^*) = \frac{1}{p^s}$.

3. On note $(p_k)_{k\geq 1}$ la suite des nombres premiers. On pose $A_k = p_k\mathbb{N}^*$. Montrer que

$$\{1\} = \bigcap_{k=1}^{+\infty} A_k^c.$$

En déduire soigneusement que

$$\frac{1}{\zeta(s)} = \lim_{n\to+\infty} \mu_s\left(\bigcap_{k=1}^{n} A_k^c\right).$$

4. Montrer que les événements $(A_k)_{k\geq 1}$ sont indépendants sous μ_s.
5. Donner une preuve probabiliste des identités

$$\frac{1}{\zeta(s)} = \prod_{k=1}^{+\infty} (1 - p_k^{-s})$$

(développement de Zêta en produit eulérien) et

$$\forall s > 1 \quad \log \zeta(s) = \lim_{n\to+\infty} \sum_{k=1}^{n} -\log(1 - p_k^{-s}).$$

6. Montrer que les séries de termes généraux respectifs $\left(-\log(1-\frac{1}{p_k})\right)_{k\geq 1}$ et $\left(\frac{1}{p_k}\right)_{k\geq 1}$ sont divergentes.
7. Dans cette question, on s'attache à montrer le résultat suivant : il n'existe pas de mesure de probabilité μ sur $\mathbb{N}$ telle que pour tout $n \geq 1$, on ait $\mu(n\mathbb{N}) = \frac{1}{n}$ [2].
On raisonne donc par l'absurde et on suppose qu'une telle mesure de probabilité existe.
   (a) Soient n et ℓ des entiers avec $\ell > n \geq 1$. Établir l'inégalité

   $$\mu(\{n\}) \leq \prod_{i=n}^{\ell} \left(1 - \frac{1}{p_i}\right).$$

   Indication : on pourra remarquer que les $(p_i\mathbb{N})_{i\geq 1}$ sont indépendants sous μ.
   (b) Conclure.

Exercice 41. 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.

2. Si vous savez déjà ce qu'est une variable aléatoire, on peut présenter le résultat sous la forme plus agréable suivante : 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

$$\forall n \geq 1 \quad \mathbb{P}(n \text{ divise } X) = \frac{1}{n}.$$

En effet, la mesure de probabilité $\mu = \mathbb{P}_X$ contredirait le résultat de l'exercice.

1. Proposer 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 $\mathfrak{S}_n$ sans point fixe :
$$d_n = \text{Card}(\{\sigma \in \mathfrak{S}_n; \forall 1 \leq i \leq n \quad \sigma(i) \neq i\}).$$
(On pose $d_0 = 1$.)
2. Pour $0 \leq k \leq n$, on note A_k^n l'ensemble des permutations de $\mathfrak{S}_n$ ayant exactement k points fixes :
$$A_k^n = \{\sigma \in \mathfrak{S}_n; \text{Card}(\{i \in 1, \dots, d \quad | \quad \sigma(i) = i\}) = k\}.$$
Montrer $\text{Card}(A_k^n) = \binom{n}{k} d_{n-k}$. En déduire
$$\sum_{k=0}^{n} \binom{n}{k} d_k = n!$$
3. Soit Φ l'endomorphisme de $\mathbb{R}_n[X]$ défini par $\Phi(P) = P(X+1)$ (où 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}.
4. Montrer que $(d_0, d_1, \dots, d_n).M = (0!, 1!, \dots, n!)$. En déduire que
$$p_n = \frac{d_n}{n!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$
5. Montrer $\lim\limits_{n \to +\infty} p_n = \frac{1}{e}$. Montrer que pour $n \geq 2$, d_n est l'entier le plus proche de $\frac{n!}{e}$.

Exercice 42. *Calcul probabiliste de la formule de l'indicatrice d'Euler.*
On note Ω_n l'ensemble des entiers de 1 à n. On note $n = \prod_{i \in I} p_i^{\alpha_i}$ la décomposition de n en produits de facteurs premiers. Le but de cet exercice est de déterminer le nombre $\phi(n)$ qui est le cardinal de l'ensemble G_n des entiers compris entre 1 et n qui sont premiers avec n. On note $\mathbb{P}$ la loi uniforme sur Ω_n.

1. Pour d divisant n, on note $A_d = \{k \in \Omega_n; d | k\}$. Calculer $\mathbb{P}(A_d)$.
2. Soit $d_1, \dots, d_r$ des diviseurs de n premiers entre eux.
Calculer $\mathbb{P}\left(\bigcap\limits_{i=1}^{r} A_{d_i}\right)$. Qu'en déduire pour les événements $(A_{p_i})_{i \in I}$?
3. Montrer que $\mathbb{P}(G_n) = 1 - \mathbb{P}(\bigcup\limits_{i} A_{p_i})$.
4. Montrer enfin que $\frac{\phi(n)}{n} = \prod\limits_{i \in I} (1 - 1/p_i)$.

3.6.2 Exercices de la série 2

Exercice 43. Déterminer toutes les mesures de probabilité μ sur $\mathcal{B}(\mathbb{R})$ vérifiant $\mu(\mathbb{N}^*) = 1$ et pour tout $n \in \mathbb{N}^*$, $\mu(\{n\}) = \mu([n+1, +\infty[)$.

Exercice 44. Soit $(A_n)_{n \geq 1}$ une suite d'événements indépendants, tous de probabilité non nulle. On pose
$$A = \bigcap_{n=1}^{+\infty} A_n$$
et
$$B = \varlimsup_{n \to +\infty} A_n = \bigcup_{n=1} \bigcap_{k=n}^{+\infty} A_k.$$
Montrer que $\mathbb{P}(A) = 0$ si et seulement si $\mathbb{P}(B) = 0$

Exercice 45. *Densité naturelle, densité de Dirichlet. Ensemble des entiers sans carré.*
Un entier est dit sans carré s'il n'est divisible par le carré d'aucun entier qui n'est pas une unité, ou de manière équivalente, par le carré d'aucun nombre premier.

On dit qu'une partie E de $\mathbb{N}^*$ admet une densité naturelle, et que cette densité naturelle vaut ℓ si

$$\lim_{n\to+\infty} \mathbb{P}_n(E) = \ell, \text{ où } \mathbb{P}_n(E) = \frac{|\{1,\ldots,n\}\cap E|}{n}.$$

On dit également qu'une partie E de $\mathbb{N}^*$ admet une densité de Dirichlet, et que cette densité de Dirichlet vaut ℓ si

$$\lim_{s\to 1^+} \mu_s(E) = \ell,$$

où μ_s est la mesure de probabilité introduite à l'exercice corrigé 41 sous le nom de loi Zêta de paramètre s.

Soit A l'ensemble des entiers naturels non nuls sans carré. On va montrer que A admet une densité naturelle (resp. de Dirichlet) et la calculer.

Dans la suite de l'exercice, on note $(p_i)_{i\geq 1}$ la suite des nombres premiers.

1. Soient n et N des entiers tels que p_N est le plus grand nombre premier inférieur ou égal à n. Montrer que

$$\mathbb{P}_n(A) = \frac{1}{n} \sum_{B\subset\{1,\ldots,N\}} (-1)^{|B|} \lfloor \frac{n}{\prod_{i\in B} p_i^2} \rfloor = \frac{1}{n}\sum_{k\in A} \mu(k) \lfloor \frac{n}{k^2} \rfloor,$$

   où μ est la fonction de Möbius : $\mu(n) = 0$ si $n \notin A$ et $\mu(n) = (-1)^i$ si $n \in A$ s'écrit comme produit de i facteurs premiers.

   Pour la deuxième égalité, remarquer que $\Phi : B \mapsto \prod\limits_{i\in B} p_i$ réalise une bijection de $\mathcal{P}_f(\mathbb{N}^*)$ dans A (où $\mathcal{P}_f(\mathbb{N}^*)$ est l'ensemble des parties finies de $\mathbb{N}^*$).
2. En déduire que A admet une densité naturelle, qui vaut

$$\lim_{n\to+\infty} \mathbb{P}_n(A) = \sum_{k\in A} \frac{\mu(k)}{k^2}.$$

3. Soit $s > 1$. Montrer que $\mu_s(A) = \prod\limits_{i=1}^{+\infty} \left(1 - \frac{1}{p_i^{2s}}\right) = \frac{1}{\zeta(2s)}$. En déduire que A admet une densité de Dirichlet et la calculer.
4. Soit E une partie de $\mathbb{N}^*$ admettant une densité naturelle ℓ. Montrer que ℓ est aussi la densité de Dirichlet de E.
   Indication : on a $\mu_s(E) = \frac{1}{\zeta(s)} \sum\limits_{k=1}^{+\infty} \frac{\mathbb{1}_E(k)}{k^s}$. Écrire $\mathbb{1}_E(k)$ sous la forme $N_k(E) - N_{k-1}(E)$, puis faire une transformation d'Abel.
5. On rappelle que $\zeta(2) = \frac{\pi^2}{6}$. Montrer que la densité naturelle de A est $\frac{6}{\pi^2}$.
6. Pour $n \geq 1$, calculer les densités de $n\mathbb{N}^*$. Comparer avec le résultat de l'exercice 41.

Exercice 46. 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.

Exercice 47. Donner un exemple de trois événements A_1, A_2, A_3 qui ne sont pas indépendants et pour lesquels

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

Chapitre 4

Intégrales

Jusqu'ici, nous n'avons parlé que de mesures et nullement d'intégrales. Le présent chapitre va combler cette lacune.

Nous commençons par donner la définition de l'intégrale dite « de Lebesgue »[1] et en énoncer les propriétés fondamentales. Afin de se concentrer sur la pratique, certains résultats seront admis dans un premier temps et démontrés à la fin du chapitre.

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 mesurable positive f, on définit l'intégrale de f, notée $\int f \, d\mu$ ou encore $\int f(x) \, d\mu(x)$ par

$$\int f \, d\mu = \sup \sum_i \inf\{f(\omega); \omega \in \Omega_i\} \mu(\Omega_i), \tag{4.1}$$

où le supremum porte sur toutes les partitions mesurables finies de Ω, c'est-à-dire les partitions $(\Omega_i)_{i \in I}$ que l'on peut indexer par un ensemble fini I et telles que pour tout $i \in I$, $\Omega_i \in \mathcal{F}$. Observons que cette quantité peut être infinie.

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

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

Définition. *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'une finie, l'autre infinie, 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 s'il existe E mesurable avec $\mu(E) = 0$ tel que pour tout $x \in \Omega \backslash E$, $\mathcal{P}(x)$ est vérifiée. Au lieu de μ-presque partout, nous écrirons parfois μ-p.p.*

1. Henri Lebesgue (1875-1941) est un mathématicien français. Il a soutenu sa thèse *Intégrale, longueur, aire* en 1902, pendant qu'il était professeur de lycée à Nancy.

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 $f \leq g$ μ-presque partout, alors on a $\int f \, d\mu \leq \int g \, d\mu$, avec égalité si et seulement si $f = g$ μ-presque partout. En particulier, si $f \geq 0$ 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

$$\int \alpha f + \beta g \, d\mu = \alpha \int f \, d\mu + \beta \int g \, d\mu.$$

— **Convergence monotone (ou théorème de Beppo Levi[2])** :
Si $(f_n)_{n \geq 1}$ est une suite croissante de fonctions mesurables 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 intégrale. 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 reportons les preuves à une section ultérieure qui viendra en fin de chapitre.

Remarque 4.1. *Pour définir l'intégrale* (4.1)*, la mesurabilité de f n'est nullement requise, ce qui peut laisser perplexe. Cependant, pour que l'intégrale jouisse des propriétés intéressantes que nous venons d'énoncer, l'hypothèse de mesurabilité va servir, comme le constateront les courageux lecteurs de la dernière section. De plus, dans la pratique, nous n'intégrons que des fonctions mesurables.*

4.1.3 Les grands théorèmes : lemme de Fatou et convergence dominée

Théorème 4.2 (Lemme de Fatou[3])**.** *Pour toute suite $(f_n)_{n \geq 1}$ de fonctions mesurables positives, on a*

$$\int \varliminf_{n \to +\infty} f_n \, d\mu \leq \varliminf_{n \to +\infty} \int f_n \, d\mu.$$

Démonstration. Il suffit de poser $g_n = \inf_{k \geq n} f_k$. La suite $(g_n)_{n \geq 1}$ est une suite croissante, dont la limite est, par définition, $g = \lim_{n \to +\infty} g_n = \varliminf_{n \to +\infty} f_n$. On a pour tout n

$$f_n \geq g_n \quad \text{et donc} \quad \int f_n \, d\mu \geq \int g_n \, d\mu.$$

D'où

$$\varliminf_{n \to +\infty} \int f_n \, d\mu \quad \geq \quad \varliminf_{n \to +\infty} \int g_n \, d\mu.$$

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. Dans cette inégalité, les termes peuvent valoir $+\infty$. $\square$

Théorème 4.3 (Convergence dominée)**.** *Si $(f_n)_{n \geq 1}$ est une suite de fonctions mesurables 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 $\int f \, d\mu$.*

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

3. Mathématicien français (1878-1929). Avec sa thèse, *Séries trigonométriques et séries de Taylor*, c'est un des premiers utilisateurs de la théorie de l'intégration de Lebesgue.

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, ce qui donne

$$\int \varliminf_{n\to+\infty} (g + f_n)\, d\mu \le \varliminf_{n\to+\infty} \int (g + f_n)\, d\mu$$

et

$$\int \varliminf_{n\to+\infty} (g - f_n)\, d\mu \le \varliminf_{n\to+\infty} \int (g - f_n)\, d\mu$$

soit

$$\int g\, d\mu + \int f\, d\mu \le \int g\, d\mu + \varliminf_{n\to+\infty} \int f_n\, d\mu$$

et

$$\int g\, d\mu - \int f\, d\mu \le \int g\, d\mu - \varlimsup_{n\to+\infty} \int f_n\, d\mu.$$

Comme $|\int g\, d\mu| < +\infty$, on peut simplifier et on obtient

$$\int f\, d\mu \le \varliminf_{n\to+\infty} \int f_n\, d\mu$$

et

$$\varlimsup_{n\to+\infty} \int f_n\, d\mu \le \int f\, d\mu$$

ce qui montre le résultat voulu. □

4.2 Intégration sur un ensemble

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

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

Théorème 4.4. *Si f est intégrable et si $(A_n)_{n\ge1}$ est une partition dénombrable mesurable de Ω, alors*

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

Démonstration. On pose $f_n = f \times \mathbb{1}_{\bigcup_{k=1}^{n} A_k} = f \times (\sum_{k=1}^{n} \mathbb{1}_{A_k})$ 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 4.5. *Soit $(\Omega, \mathcal{F})$ un espace mesurable. 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 $\inf_{\Omega_i} f \leq f(x)$ est une 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 supremum 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 4.6. *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*

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

où cette somme peut valoir $+\infty$. 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. Soient f positive et $(\Omega_i)_{i\in I}$ une partition mesurable finie de Ω. On a

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

D'où en sommant sur I

$$\sum_{i\in I} C(\Omega_i) \inf_{\Omega_i} f \leq \sum_{i\in I} \sum_{k\in\Omega} \left(\mathbb{1}_{\Omega_i}(k) f(k) \right).$$

Cependant, par le théorème de Fubini, on a

$$\sum_{i\in I} \sum_{k\in\Omega} \left(\mathbb{1}_{\Omega_i}(k) f(k) \right) = \sum_{k\in\Omega} \left(\sum_{i\in I} \mathbb{1}_{\Omega_i}(k) f(k) \right) = \sum_{k\in\Omega} f(k)$$

d'où en passant au supremum

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

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

$$\sum_{k\in F} f(k) + |F^c| \inf_{F^c} f \geq \sum_{k\in F} f(k).$$

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)

Définition. *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 4.7. *Toute fonction mesurable positive f (éventuellement infinie) peut s'écrire comme limite simple d'une suite croissante de fonctions simples (f_n).*

Démonstration. Il est naturel de chercher f_n sous la forme $f_n = \phi_n \circ f$, où ϕ_n est une suite croissante de fonctions qui convergent ponctuellement vers l'identité, chaque fonction f_n ne prenant qu'un nombre fini de valeurs. On définit sur $[0,+\infty]$ une fonction ϕ_n par

$$\phi_n(x) = 2^{-n}\lfloor 2^n x\rfloor \mathbb{1}_{[0,n]}(x) \text{ pour } x < +\infty \text{ et } \phi_n(+\infty) = n.$$

La prise de la partie entière d'un multiple, suivie d'une division, est classique dans l'approximation par des rationnels. Plus originale, l'intervention des 2^n est, on le verra, destinée à assurer la monotonie. Évidemment la suite $(\phi_n(\infty))_{n\geq 1}$ tend en croissant vers $+\infty$. Soit $x \geq 0$. Il est immédiat que $\mathbb{1}_{[0,n+1]}(x) \geq \mathbb{1}_{[0,n]}(x)$. Posons $y = 2^n x$. On a $y \geq \lfloor y\rfloor$, donc $2y \geq 2\lfloor y\rfloor$. Comme $2\lfloor y\rfloor$ est entier, on a aussi $\lfloor 2y\rfloor \geq 2\lfloor y\rfloor$, ce qui nous donne finalement $\phi_{n+1}(x) \geq \phi_n(x)$. D'autre part pour $n \geq x$, on a $x - 2^{-n} \leq \phi_n(x) \leq x$, donc $\phi_n(x)$ tend vers x. Il suffit alors de poser $f_n(x) = \phi_n(f(x))$. □

Remarque 4.8. *On peut montrer que la convergence précédente est uniforme lorsque f est bornée.*

4.3.4 Intégration par rapport à une somme de deux mesures

Théorème 4.9. *Soient $(\Omega, \mathcal{F})$ un espace mesurable, 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*

$$\int_\Omega f \, d(\mu + \nu) = \int_\Omega f \, d\mu + \int_\Omega f \, d\nu. \tag{4.2}$$

Dans le cas où f est de signe quelconque, si $\int_\Omega |f| \, d\mu$ et $\int_\Omega |f| \, d\nu$ sont finies, alors f est intégrable par rapport à $\mu + \nu$ et on a encore (4.2).

Démonstration. Dans le cas où f est l'indicatrice d'un élément de $\mathcal{F}$, l'identité (4.2) 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 (4.2) 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 4.7 et le théorème de convergence monotone, il s'ensuit que l'identité (4.2) 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

On a ici $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$.

Théorème 4.10. *Soit f une fonction continue par morceaux sur un intervalle compact $[a, b]$. Alors*

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

Ainsi, on a identité entre les intégrales de Lebesgue (à gauche) et Riemann (à droite) pour les fonctions continues par morceaux sur un intervalle compact.

Démonstration. Grâce à 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

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

Comme f est continue sur $[a, b]$, son module est borné par une constante M. Comme f est continue sur $[a, b]$, $f_n(x)$ converge vers $f(x)$ pour tout $x \in [a, b]$. 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 reconnaît 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 4.11. *Soit f une fonction positive continue par morceaux sur un intervalle semi-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. La suite $(f\mathbb{1}_{[a,b_n]})$ 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 $\int_{[a,b_n]} f(x)\ d\lambda(x)$. 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 4.12. *Soit f une fonction continue par morceaux sur un intervalle semi-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 revient à 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. La suite $(f\mathbb{1}_{[a,b_n]})$ converge vers $f\mathbb{1}_{[a,b[}$ et $|f\mathbb{1}_{[a,b_n]}| \leq |f|$, donc d'après le théorème de convergence dominée $\int_{[a,b[} f(x)\ d\lambda(x)$ 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. □

Remarque 4.13. *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.

À titre culturel, on peut noter le résultat suivant dû à Lebesgue :

Théorème 4.14. *Soit f une fonction mesurable et bornée sur un intervalle $[a, b]$. Alors f est Riemann-intégrable sur $[a, b]$ si et seulement si l'ensemble des points de discontinuité de f est de mesure de Lebesgue nulle.*

Démonstration. On va se contenter de démontrer que si l'ensemble des points de discontinuité de f est de mesure de Lebesgue nulle, alors f est Riemann-intégrable sur $[a, b]$, c'est-à-dire que si on a

$$a = a_0^n \leq x_0^n \leq a_1^n \leq x_1^n \leq \dots x_{n-1}^n \leq a_n^n = b,$$

et que la suite $e_n = \max\limits_{1\leq i\leq n} (a_i^n - a_{i-1}^n)$ est de limite nulle, alors la suite $S_n = \sum\limits_{k=1}^{n} (a_k - a_{k-1})f(x_{k-1})$ converge toujours vers une même limite ne dépendant pas des subdivisions choisies. On procède comme dans la preuve du théorème 4.10 : soit M une constante bornant f sur $[a, b]$.
Si l'on pose $f_n = \sum\limits_{i=1}^{n} f(x_{i-1}^n)\mathbb{1}_{]a_{i-1}^n, a_i^n[}$, alors f_n converge vers f en les points de continuité de f, donc presque partout. Comme $|f_n| \leq M$, le théorème de convergence dominée s'applique et $S_n = \int_{[a,b]} f_n\ d\lambda$ converge vers $\int_{[a,b]} f\ d\lambda$. □

Ce résultat illustre bien la supériorité de l'intégrale de Lebesgue sur l'intégrale de Riemann. Pour vous en convaincre, essayez donc de montrer, à l'aide de la seule définition de la Riemann-intégrabilité, la convergence des sommes de Riemann associées à la fonction $x \mapsto \{\frac{1}{x}\}$ sur $]0, 1]$. (La valeur de l'intégrale sera calculée en exercice.)

Terminons par quelques remarques élémentaires : l'intérêt de l'intégrale de Riemann, c'est qu'on a appris (dans certains cas) à la calculer ! Plus précisément, le théorème fondamental de l'analyse nous enseigne que 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 $F \circ \phi$ est $\phi'.(f \circ \phi)$, ce qui nous dit que $F \circ \phi$ est une primitive de $\phi'.(f \circ \phi)$, et donc

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

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

4.5 Intégrale d'une fonction à valeurs complexes

Définition. *Soient $(\Omega, \mathcal{F}, \mu)$ un espace mesuré et f une fonction de Ω dans $\mathbb{C}$. On dit que f est mesurable sur $(\Omega, \mathcal{F}, \mu)$ si ses parties réelle et imaginaire le sont. On dit que f est intégrable sur $(\Omega, \mathcal{F}, \mu)$ si ses parties réelle et imaginaire le sont. Ainsi si f s'écrit $f = f_1 + if_2$ où f_1 et f_2 sont des fonctions à valeurs réelles mesurables, on peut définir $\int_\Omega f\, d\mu$ par*

$$\int_\Omega f\, d\mu = \left(\int_\Omega f_1\, d\mu\right) + i\left(\int_\Omega f_2\, d\mu\right).$$

Il n'est alors pas très difficile de voir que les fonctions intégrables sur $(\Omega, \mathcal{F}, \mu)$ à valeurs dans $\mathbb{C}$ forment un $\mathbb{C}$-espace vectoriel et que l'intégrale ainsi définie est $\mathbb{C}$-linéaire. Quelles que soient les fonctions complexes intégrables sur $(\Omega, \mathcal{F}, \mu)$ et quels que soient $\alpha, \beta \in \mathbb{C}$, on a

$$\int_\Omega (\alpha f + \beta g)\, d\mu = \alpha \int_\Omega f\, d\mu + \beta \int_\Omega g\, d\mu.$$

Si l'on sait que f est mesurable (c'est-à-dire que la partie réelle f_1 et la partie imaginaire f_2 de f le sont), alors comme pour tout $i \in \{1,2\}$,

$$|f_i| \leq |f| \leq |f_1| + |f_2|,$$

f sera intégrable si et seulement si $|f|$ l'est.

Enfin, il sera souvent utile de connaître le résultat suivant : si a et b sont des nombres réels avec $a < b$ et z un nombre complexe non nul, on a

$$\int_{[a,b]} e^{zx}\, d\lambda(x) = \frac{e^{bz} - e^{az}}{z}.$$

Dans la suite, la plupart des théorèmes seront énoncés pour des fonctions à valeurs réelles, mais dans le cas de fonctions à valeurs complexes, on pourra souvent démontrer un résultat analogue en considérant séparément les parties réelle et imaginaire.

Par exemple, on peut énoncer :

Théorème 4.15 (Convergence dominée pour des fonctions complexes). *Si $(f_n)_{n\geq 1}$ est une suite de fonctions mesurables complexes 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 $\left(\int f_n\, d\mu\right)_{n\geq 1}$ converge vers $\int f\, d\mu$.*

Démonstration. Comme $|\text{Re } f_n| \leq |f_n| \leq g$, on peut appliquer le théorème de convergence dominée à $(\text{Re } f_n)_{n\geq 1}$. Idem pour $(\text{Im } f_n)_{n\geq 1}$. □

Le théorème "évident" suivant mérite tout de même une démonstration :

Théorème 4.16. *si f est une fonction complexe intégrable sur $(\Omega, \mathcal{F}, \mu)$, alors*

$$|\int_\Omega f\ d\mu| \leq \int_\Omega |f|\ d\mu.$$

Démonstration. Soit $a \in \mathbb{C}$

$$\begin{aligned}\text{Re } (a \int_\Omega f\ d\mu) = \text{Re } (\int_\Omega af\ d\mu) &= \int_\Omega \text{Re } af\ d\mu \\ \leq \int_\Omega |af|\ d\mu &= |a| \int_\Omega |f|\ d\mu\end{aligned}$$

Si on prend $a = \overline{\int_\Omega f\ d\mu}$, on obtient

$$|\int_\Omega f\ d\mu|^2 \leq |\int_\Omega f\ d\mu| \int_\Omega |f|\ d\mu,$$

ce qui donne le résultat voulu. □

4.6 Identifier des mesures par leurs intégrales

Les intégrales caractérisent des mesures.

Théorème 4.17. *Soit μ et ν deux mesures sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ qui donnent chacune une masse finie aux compacts de $\mathbb{R}^d$. On suppose que pour toute fonction continue à support compact f, on a $\int_{\mathbb{R}^d} f\ d\mu = \int_{\mathbb{R}^d} f\ d\nu$. Alors $\mu = \nu$.*

Démonstration. Les compacts de $\mathbb{R}^d$ forment un π-système qui engendre la tribu borélienne de $\mathbb{R}^d$ (par exemple car les pavés ouverts s'écrivent comme réunion dénombrable de pavés compacts), donc il suffit de montrer que μ et ν coïncident sur les compacts. Soit f_n la fonction de $\mathbb{R}_+$ dans $\mathbb{R}_+$ définie par $f_n(x) = (1 - nx)^+$. f est continue, vaut 1 en 0 et est nulle sur $[1/n, +\infty[$. Soit K un compact de $\mathbb{R}^d$, et posons $g_n(x) = f_n(d(x, K))$, où $d(x, K) = \inf d(x, y);\ y \in K$. g_n est continue, comme composition d'applications continues, et converge simplement vers l'indicatrice de K. Comme $|g_n| \leq \mathbb{1}_{K+\overline{B}(0,1)}$ qui est intégrable par rapport à μ et ν, le théorème de convergence dominée dit que $\int_{\mathbb{R}^d} g_n\ d\mu$ converge vers $\int_{\mathbb{R}^d} \mathbb{1}_K\ d\mu = \mu(K)$ et $\int_{\mathbb{R}^d} g_n\ d\nu$ vers $\int_{\mathbb{R}^d} \mathbb{1}_K\ d\nu = \nu(K)$. Vu l'hypothèse faite, $\int_{\mathbb{R}^d} g_n\ d\mu = \int_{\mathbb{R}^d} g_n\ d\nu$ pour tout n, donc $\mu(K) = \nu(K)$. Comme μ et ν coïncident sur les compacts, on a donc bien $\mu = \nu$. □

4.7 Applications aux intégrales à paramètre

4.7.1 Continuité d'une intégrale dépendant d'un paramètre

Théorème 4.18. *Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré. Soit $f(x, t)$ une fonction (à valeurs réelles) 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 $t \in T$. Comme T est un espace métrique, la continuité est caractérisée par le comportement des suites : 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 convergeant vers 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 $|f_n| \leq g$. $\square$

4.7.2 Dérivabilité d'une intégrale dépendant d'un paramètre

Théorème 4.19. *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 $t \in T$:*

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

Alors

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

définit une fonction dérivable 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 $f_n(x) = \frac{f(x,t_n) - f(x,t)}{t_n - t}$.
La suite de fonctions mesurables $(f_n)_{n\geq 1}$ converge μ-presque partout vers $\frac{\partial f}{\partial t}(x,t)$, ce qui assure la mesurabilité de $x \mapsto \frac{\partial f}{\partial t}(x,t)$. De plus, d'après l'inégalité des accroissements finis, on a $|f_n(x)| \leq g(x)$ $\mu-$presque partout. Ainsi, d'après le théorème de convergence dominée, $\int_\Omega f_n\ d\mu$ converge vers l'intégrale $\int_\Omega \frac{\partial f}{\partial t}(x,t)\ d\mu(x)$. Cependant $\int_\Omega f_n\ d\mu = \frac{F(t_n)-F(t)}{t_n - t}$, ce qui donne donc le résultat voulu. $\square$

Remarque 4.20 (importante)**.** *Lorsque l'on veut démontrer la continuité (ou la dérivabilité) de $F(t)$ définie 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 $\mathcal{V}$ de t tel que l'on ait une majoration uniforme pour les $t \in \mathcal{V}$.*

4.7.3 Exercice : la fonction Gamma

On note Γ la fonction définie sur $]0, +\infty[$ par la relation

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

Pour n entier naturel non nul, on pose

$$H_n = \sum_{k=1}^{n} \frac{1}{k} = \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n}.$$

1. Vérifier que Γ est bien définie.
2. Démontrer que la fonction Γ est dérivable sur $]0, +\infty[$, avec
$$\forall x > 0 \quad \Gamma'(x) = \int_0^{+\infty} e^{-t} t^{x-1} \log t \, dt.$$
3. Montrer que la suite $(u_n)_{n\geq 1}$ définie par $u_n = H_n - \log n$ admet une limite lorsque n tend vers l'infini. On notera γ cette limite, appelée constante d'Euler.
4. Montrer que pour tout entier $n \geq 1$, $H_n = \int_0^1 \frac{1-(1-v)^n}{v} \, dv$.
5. En déduire que, pour tout entier naturel $n \geq 1$,
$$\frac{H_{n+1}}{n+1} = -\int_0^1 (1-v)^n \log v \, dv.$$
6. Établir que pour tout $t \geq 0$, $1 - t \leq e^{-t}$.
7. On pose $I_n = \int_0^n (1 - \frac{t}{n})^n \log t \, dt$. Montrer que
$$\lim_{n\to+\infty} I_n = \int_0^{+\infty} e^{-t} \log t \, dt.$$
8. Montrer que $\gamma = -\Gamma'(1)$.
Indication : on pourra montrer que $I_n = \frac{n}{n+1}(\log n - H_{n+1})$.

Solution

1. La fonction $t \mapsto e^{-t}t^{x-1}$ est continue sur $]0, +\infty[$, donc localement intégrable sur $]0, +\infty[$. Reste à étudier l'intégrabilité en 0 et en l'infini. En 0, $e^{-t}t^{x-1} \sim t^{x-1}$, et comme $x - 1 > -1$ et que t^{x-1} est de signe constant au voisinage de 0, l'intégrabilité en 0 découle du critère d'équivalence avec une intégrale "de type Riemann" classique. En l'infini, $e^{-t}t^{x-1} = o(e^{-t/2})$, ce qui donne la convergence en l'infini.
2. Soient a, b réels avec $0 < a < b < +\infty$. Pour tout $t > 0$, on a
$$\frac{\partial}{\partial x}\left(e^{-t}t^{x-1}\right) = e^{-t}t^{x-1} \log t.$$
Comme pour tout $t > 0$ et tout $x \in]a, b[$, on a
$$\begin{aligned} |e^{-t}t^{x-1} \log t| &= (-\log t)e^{-t}t^{x-1} \, \mathbb{1}_{]0,1[}(t) + e^{-t}t^{x-1} \log t \, \mathbb{1}_{[1,+\infty[}(t) \\ &\leq (-\log t)e^{-t}t^{a-1} \, \mathbb{1}_{]0,1[}(t) + e^{-t}t^{b-1} \log t \, \mathbb{1}_{[1,+\infty[}(t), \end{aligned}$$
on pourra appliquer le théorème de dérivation sous le signe intégrale dès qu'il sera acquis que
$$t \mapsto (-\log t)e^{-t}t^{a-1} \, \mathbb{1}_{]0,1[}(t) + e^{-t}t^{b-1} \log t \, \mathbb{1}_{[1,+\infty[}(t)$$
est intégrable sur $]0, +\infty[$.
— Cette fonction est continue, donc localement intégrable.
— En 0, on a $\log t = o(t^{-a/2})$ et $e^{-t} \sim 1$, d'où la relation de comparaison $(-\log t)e^{-t}t^{a-1} = o(t^{a/2-1})$, ce qui donne l'intégrabilité en 0.
— En $+\infty$, comme $\log t = o(t)$, on a $e^{-t}t^{b-1} \log t = o(e^{-t}t^b)$, mais $t^b = o(e^{t/2})$, donc finalement $e^{-t}t^{b-1} \log t = o(e^{-t/2})$, ce qui donne l'intégrabilité en l'infini.

Ainsi, la fonction Γ est dérivable sur $]a, b[$, avec
$$\forall x \in]a, b[\quad \Gamma'(x) = \int_0^{+\infty} e^{-t}t^{x-1} \log t \, d\lambda(t).$$
Comme la dérivabilité est une propriété locale et que tout point de $]0, +\infty[$ admet un voisinage de la forme $]a, b[$, avec a, b réels vérifiant $0 < a < b < +\infty$, le résultat s'ensuit.

3. On a le comportement suivant

$$\begin{aligned}|u_n - u_{n-1}| &= |(H_n - H_{n-1}) - (\log n - \log(n-1))| \\ &= \left|\frac{1}{n} + \log(1 - 1/n)\right| = \left|\frac{1}{n} - \frac{1}{n} + O\left(\frac{1}{n^2}\right)\right| = O\left(\frac{1}{n^2}\right).\end{aligned}$$

Ainsi la série de terme général $u_n - u_{n-1}$ converge, mais on a la relation télescopique $\sum\limits_{k=2}^{n} (u_k - u_{k-1}) = u_n - u_1$, donc la suite $(u_n)_{n\geq 1}$ converge.

4. Pour tout $v \in]0, 1[$, on a en posant $u = 1 - v$

$$\begin{aligned}\int_0^1 \frac{1-(1-v)^n}{v}\,dv &= \int_0^1 \frac{1-u^n}{1-u}\,du = \int_0^1 \left(\sum_{k=0}^{n-1} u^k\right) du \\ &= \sum_{k=0}^{n-1} \int_0^1 u^k\,du = \sum_{k=0}^{n-1} \frac{1}{k+1} = H_n.\end{aligned}$$

5. On fait une intégration par parties : pour tout $\varepsilon > 0$, on a

$$\begin{aligned}&\int_\varepsilon^1 (1-v)^n \log v\,dv \\ &= [\frac{1-(1-v)^{n+1}}{n+1}\log v]_\varepsilon^1 - \frac{1}{n+1}\int_\varepsilon^1 \frac{1-(1-v)^{n+1}}{v}\,dv \\ &= -\frac{1-(1-\varepsilon)^{n+1}}{n+1}\log\varepsilon - \frac{1}{n+1}\int_\varepsilon^1 \frac{1-(1-v)^{n+1}}{v}\,dv\end{aligned}$$

Cependant, on a l'équivalent en 0 : $-\frac{1-(1-\varepsilon)^{n+1}}{n+1}\log\varepsilon \sim -\varepsilon\log\varepsilon$, d'où en faisant tendre ε vers 0 :

$$\int_0^1 (1-v)^n \log v\,dv = -\frac{1}{n+1}\int_0^1 \frac{1-(1-v)^{n+1}}{v}\,dv = -\frac{H_{n+1}}{n+1}$$

grâce à la question précédente.

6. La fonction $t \mapsto f(t) = 1 - e^{-t}$ a comme dérivée e^{-t}, majorée par 1 sur $\mathbb{R}_+$. D'après l'inégalité des accroissements finis, on a $f(t) - f(0) \leq t$, d'où l'inégalité voulue.

7. Posons, pour $t > 0$ et $n \geq 1$, $f_n(t) = (1 - \frac{t}{n})^n(\log t)\,\mathbb{1}_{[0,n]}(t)$. Comme f_n est continue par morceaux, on a

$$\int_{]0,+\infty[} f_n(x)\,d\lambda(x) = \int_{]0,n]} f_n(x)\,d\lambda(x) = \int_0^n f_n(t)\,dt = I_n.$$

Pour $n \geq t$, on a $f_n(t) = (1 - \frac{t}{n})^n \log t$.
Mais $(1-\frac{t}{n})^n = \exp(\log(1-\frac{t}{n})^n) = \exp(n\log(1-t/n))$: lorsque n tend vers l'infini $\log(1-t/n) \sim -t/n$, d'où $n\log(1-t/n) \sim -t$. Ainsi, $\lim\limits_{n\to+\infty} n\log(1-t/n) = -t$, d'où $\lim\limits_{n\to+\infty} (1-t/n)^n = e^{-t}$. Finalement, pour tout $t > 0$, $\lim\limits_{n\to+\infty} f_n(t) = e^{-t}\log t$.
On a pour t compris entre 0 et n

$$\left|\left(1-\frac{t}{n}\right)^n \log t\right| \leq (e^{-t/n})^n|\log t| = |\log t|e^{-t},$$

d'où $|f_n(t)| \leq |\log t|e^{-t}$. La dernière inégalité est encore vérifiée pour $t > n$: les termes sont tous nuls. Ainsi, on a sur $]0, +\infty[$ l'inégalité $|f_n(t)| \leq |\log t|e^{-t}$. Comme, on l'a vu au 2, cette fonction est intégrable, on peut alors appliquer le théorème de convergence dominée et on a

$$\lim_{n\to+\infty} I_n = \int_0^{+\infty} e^{-t}\log t\,dt.$$

8. Un simple changement de variable affine donne

$$\begin{aligned}\int_0^n f_n(t)\ dt &= n\int_0^1 f_n(ny)\ dy = n\int_0^1 (1-y)^n \log(ny)\ dy.\\ &= n\log n\int_0^1 (1-y)^n\ dy + n\int_0^1 (1-y)^n \log y\ dy.\\ &= \frac{n\log n}{n+1} - n\frac{H_{n+1}}{n+1},\end{aligned}$$

soit $I_n = \frac{n}{n+1}(\log n - H_{n+1}) = \frac{n}{n+1}(\log n - H_n - \frac{1}{n+1})$.
Comme $\lim\limits_{n\to+\infty} (H_n - \log n) = \gamma$, cela nous donne $\lim\limits_{n\to+\infty} I_n = -\gamma$. Or, d'après la question précédente, $\lim\limits_{n\to+\infty} I_n = \int_0^{+\infty} e^{-t}\log t\ dt$, qui d'après la question 2, est égale à $\Gamma'(1)$.
On obtient donc l'identité $\gamma = -\Gamma'(1)$.

4.8 Mesures à densité

4.8.1 Définition et premières propriétés

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 sur $(\Omega, \mathcal{F})$ (exercice laissé au lecteur). On note parfois $f.\mu$ la mesure ainsi définie.

Définition. *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 un abus de langage : en effet, une même mesure ne peut-elle admettre plusieurs densités par rapport à μ ?

Proposition 4.21. *Soit ν une mesure σ-finie. Soient f et g deux fonctions mesurables étant toutes deux des densités de ν par rapport à μ. Alors $f = g$ μ-presque partout.*

Démonstration. Supposons d'abord ν finie. Posons $A_+ = \{\omega : f(\omega) > g(\omega)\}$.
On a $0 = \nu(A_+) - \nu(A_+) = \int_{A_+} f\ d\mu - \int_{A_+} g\ d\mu = \int_{A_+} (f-g)\ d\mu$.
De même si l'on pose $A_- = \{\omega : f(\omega) < g(\omega)\}$, on a encore la relation $0 = \nu(A_-) - \nu(A_-) = \int_{A_-} f\ d\mu - \int_{A_-} g\ d\mu = \int_{A_-} (f-g)\ d\mu$. Cependant $|f-g| = (f-g)\mathbb{1}_{A_+} - (f-g)\mathbb{1}_{A_-}$, donc

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

Ce qui implique que $f = g$ μ-presque partout.
Cas général : on pose $\nu_n(A) = \nu(A \cap \Omega_n)$, où (Ω_n) est une suite croissante d'ensembles de mesure finie de réunion Ω. ν_n est une mesure finie et admet les densités $f\mathbb{1}_{\Omega_n}$ et $g\mathbb{1}_{\Omega_n}$ qui coïncident donc μ-presque partout :
on a $f\mathbb{1}_{\Omega_n} = g\mathbb{1}_{\Omega_n}$ μ-presque partout, et à la limite $f = g$ μ-p.p. □

Théorème 4.22. *On suppose que ν est une mesure admettant f comme densité par rapport à μ. Alors, pour toute fonction mesurable g*

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

Si cette quantité est finie, on a alors

$$\int g \, d\nu = \int g f \, d\mu. \tag{4.4}$$

Démonstration. Si $g = \mathbb{1}_A$ avec $A \in \mathcal{F}$, (4.4) est immédiat. Par linéarité, (4.4) est également vérifiée lorsque g est une fonction simple positive. En utilisant le lemme 4.7 et le théorème de convergence monotone, il s'ensuit que (4.4) est vraie pour toute fonction mesurable positive, en particulier (4.3) est vraie pour toute fonction mesurable g. Supposons 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 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 des deux, on obtient donc l'égalité $\int (g^+ - g^-) \, d\nu = \int (g^+ - g^-) f \, d\mu$, soit (4.4). □

4.8.2 Décomposition de Lebesgue

Définition. *On dit que la mesure μ est une mesure absolument continue par rapport à λ, ce qui est noté $\mu \ll \lambda$, si pour tout borélien $A \in \mathcal{B}(\mathbb{R})$, on a : $\lambda(A) = 0$ implique $\mu(A) = 0$. On dit que la mesure ν est une mesure singulière par rapport à λ, ce que l'on note $\nu \perp \lambda$, s'il existe $N \in \mathcal{B}(\mathbb{R})$ tel que $\lambda(N) = 0$ et $\nu(N^c) = 0$.*

Remarque 4.23. *Un tel borélien N n'est pas nécessairement unique.*

Théorème 4.24. *Toute mesure $\sigma-$finie μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ se décompose de façon unique sous la forme $\mu = \nu_1 + \nu_2$, où ν_1 est absolument continue par rapport à la mesure de Lebesgue λ et ν_2 est singulière par rapport à λ.*

Démonstration. Montrons tout d'abord l'existence d'une telle décomposition. Considérons l'ensemble des négligeables pour la mesure λ :

$$\mathcal{N} = \{A \in \mathcal{B}(\mathbb{R}); \ \lambda(A) = 0\}.$$

Posons $\alpha = \sup\{\mu(A); \ A \in \mathcal{N}\}$. Si $\alpha = 0$, alors $\mu = \nu_1$ et il n'y a rien à démontrer. Supposons donc $\alpha > 0$. Dans ce cas, il existe une suite croissante (A_n) d'éléments de $\mathcal{N}$ telle que $\mu(A_n) \geq \alpha - \frac{1}{n}$ pour tout $n \in \mathbb{N}^*$. Ainsi, l'ensemble $A = \cup_{n \geq 1} A_n \in \mathcal{N}$ vérifie $\mu(A) = \alpha$. Soit maintenant $B \subset A^c$ tel que $\lambda(B) = 0$. On se demande s'il est possible d'avoir $\mu(B) > 0$. On sait que $\mu(A \cup B) = \mu(A) + \mu(B) \geq \alpha$ et $A \cup B \in \mathcal{N}$ donc $\mu(A \cup B) \leq \alpha$. On voit donc que nécessairement $\mu(B) = 0$. Posons maintenant $\nu_1 = \mathbb{1}_{A^c}\mu$ et $\nu_2 = \mathbb{1}_A \mu$. La mesure ν_1 admet la densité $\mathbb{1}_{A^c}$ par rapport à μ. On a

$$\nu_1(B) = \int_B \mathbb{1}_{A^c} \, d\mu = \int \mathbb{1}_{A^c} \mathbb{1}_B \, d\mu = \int \mathbb{1}_{A^c \cap B} \, d\mu = \mu(A^c \cap B).$$

Ainsi, si B est tel que $\lambda(B) = 0$, alors $\nu_1(B) = \mu(A^c \cap B) = 0$ et donc $\nu_1 \ll \lambda$. On remarque de plus que $\lambda(A^c) = 0$ et $\nu_2(A^c) = \mu(A \cap A^c) = 0$. Donc ν_2 est bien singulière par rapport à λ.

Pour montrer l'unicité de la décomposition, supposons que $\mu = \nu_1' + \nu_2'$, avec $\nu_1' \ll \lambda$ et $\nu_2' \perp \lambda$. Choisissons donc des ensembles $A, B \in \mathcal{B}(\mathbb{R})$ tels que $\nu_2(A) = \lambda(A^c) = \nu'{}_2(B) = \lambda(B^c) = 0$. On a alors

$$\nu_2(A \cap B) = \nu_2'(A \cap B) = \nu_1(A^c \cup B^c) = \nu_1'(A^c \cup B^c) = 0.$$

Donc $\nu_1 = \mathbb{1}_{A \cap B} \, \nu_1 = \mathbb{1}_{A \cap B} \, \mu = \mathbb{1}_{A \cap B} \, \nu_1' = \nu_1'$ et $\nu_2 = \mu - \nu_1 = \mu - \nu_1' = \nu_2'$.

□

Remarque 4.25. *En réalité, on peut dire un peu plus : la mesure ν_1 apparaissant dans la décomposition du théorème admet une densité par rapport à la mesure de Lebesgue. Ce résultat constitue le théorème de Radon-Nikodým. Ce résultat, que nous ne démontrerons pas ici, peut être établi à l'aide de techniques hilbertiennes (voir par exemple Rudin [10]).*

4.9 Intégration par rapport à une mesure image : le théorème de transfert

Le théorème qui suit est un résultat très utile, dont la portée n'est malheureusement pas toujours bien comprise. Dans un certain sens, on peut considérer que c'est ce théorème qui légitime l'introduction du concept de mesure image, puisqu'il exprime que dès lors qu'on sait décrire la mesure image μ_T, on saura calculer les intégrales de fonctions de la forme $f \circ T$. C'est en probabilités que son intérêt est le plus évident ; sa maîtrise est un objectif important d'un cours de probabilités de ce niveau. En analyse, ce théorème permet de calculer certaines intégrales avec une redoutable efficacité, voir par exemple la preuve du calcul du volume de la boule unité que nous proposons en fin de ce chapitre.

Théorème 4.26. *Soient $(\Omega, \mathcal{F}, \mu)$ 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*

$$\int_{\Omega'} f(y)\, d\mu_T(y) = \int_{\Omega} (f \circ T)(x)\, d\mu(x). \tag{4.5}$$

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'autre part on a $\mathbb{1}_A \circ T = \mathbb{1}_{T^{-1}(A)}$, donc $\int_\Omega f \circ T\, d\mu = \int_\Omega \mathbb{1}_{T^{-1}(A)}\, d\mu = \mu(T^{-1}(A))$. L'égalité (4.5) est donc vérifiée dans le cas où f est l'indicatrice d'un ensemble $A \in \mathcal{F}'$. Par linéarité, elle est 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 f mesurable de $(\Omega', \mathcal{F}')$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on a $\int_{\Omega'} |f|\, d\mu_T = \int_\Omega |f| \circ T\, d\mu$, ce qui montre bien que 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 $\int_{\Omega'} f^-(y)\, d\mu_T(y) = \int_\Omega f^- \circ T(x)\, d\mu(x)$ de l'identité $\int_{\Omega'} f^+(y)\, d\mu_T(y) = \int_\Omega f^+ \circ T(x)\, d\mu(x)$, on obtient le résultat voulu. □

4.10 Mesure produit

L'introduction de la notion de mesure produit vise plusieurs buts :

— donner un sens mathématique à la notion intuitive d'aire dans $\mathbb{R}^2$, ou de volume dans $\mathbb{R}^3$,

— permettre le calcul d'intégrales de plusieurs variables,

— introduire un cadre mathématique qui permettra, dans un contexte probabiliste, de manier efficacement la notion d'indépendance.

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 $\mathcal{X} \times \mathcal{Y}$.

4.10.1 Construction de la mesure produit

Lemme 4.27. *Pour tout $A \in \mathcal{X} \otimes \mathcal{Y}$, $x \in X$ et $y \in Y$, on note*

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

Alors $A_x \in \mathcal{Y}$ et $A^y \in \mathcal{X}$. De plus, si f est une fonction mesurable de $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$ vers $(C, \mathcal{C})$, 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 assertion. Fixons $x \in X$ et montrons que $f_x^1 : y \mapsto f(x, y)$ est $(Y, \mathcal{Y}) - (C, \mathcal{C})$ mesurable. Notons $\pi_x^1 : Y \to X \times Y$ qui à y associe (x, y). π_x^1 est $(Y, \mathcal{Y}) - (X \times Y, \mathcal{X} \otimes \mathcal{Y})$ mesurable car chacune des composantes est mesurable. Maintenant, l'identité $f_x^1 = f \circ \pi_x^1$ donne la mesurabilité voulue, par composition d'applications mesurables.

Revenons à la première proposition. La section verticale de niveau x : $A_x = \{y \in Y : (x, y) \in A\}$ peut s'écrire comme une image réciproque puisque $A_x = (f_x^1)^{-1}(\{1\})$, avec $f_x^1(y) = \mathbb{1}_A(x, y)$. Or l'application de $X \times Y$ dans $\mathbb{R}$ qui à (x, y) associe $\mathbb{1}_A(x, y)$ est $(X \times Y, \mathcal{X} \otimes \mathcal{Y}) - (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable ; comme d'après ce qui précède, f_x^1 est $(Y, \mathcal{Y}) - (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable, on a donc $A_x \in \mathcal{Y}$.

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

Remarque 4.28. *On dit parfois que A^y et A_x sont des sections de l'ensemble A.*

Théorème 4.29. *Soient $(X, \mathcal{X}, \mu)$ et $(Y, \mathcal{Y}, \nu)$ deux espaces mesurés dont les mesures μ et ν sont σ-finies. Alors, il existe une unique mesure m sur l'ensemble $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$ telle 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 $E \in \mathcal{X} \otimes \mathcal{Y}$, les fonctions $x \mapsto \nu(E_x)$ et $y \mapsto \mu(E^y)$ sont mesurables et l'on a

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

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 $x \mapsto \nu(E_x)$ 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, prenons $E = A \times B$, avec $A \in \mathcal{X}$ et $B \in \mathcal{Y}$: on a alors $E_x = \{y \in B : (x, y) \in A \times B\}$. Ainsi $E_x = B$ si $x \in A$ et $\varnothing$ sinon, et donc $\nu(E_x) = \nu(B)$ si $x \in A$ et 0 sinon. Ainsi, $\nu(E_x) = \nu(B)\mathbb{1}_A(x)$, 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 contenant 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

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

et de même on pourrait définir

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

Prenons à nouveau $E = A \times B$ et $E_x = \{y \in B{:}(x,y) \in A \times B\}$. Ainsi $E_x = B$ si $x \in A$ et $\varnothing$ sinon, et donc $\nu(E_x) = \nu(B)$ 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 $m_2(E) = \int_Y \mathbb{1}_B \mu(A)\, d\nu = \mu(A)\nu(B)$. Donc m_1 et m_2 sont 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 :
$X = \bigcup_{i=1}^{+\infty} A_i$ et $Y = \bigcup_{j=1}^{+\infty} B_j$. Notons $m^{i,j}$ la mesure associée comme précédemment aux mesures traces $\nu_{|A_i}$ et $\mu_{|B_j}$. En d'autres termes

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

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

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

□

Remarque 4.30. *Il n'est pas difficile de voir que si μ, ν sont des mesures σ-finies, a et b des réels positifs, alors $(a\mu) \otimes (b\nu) = (ab)(\mu \otimes \nu)$ (utiliser la partie unicité du théorème).*

Exercice. Soient $(X, \mathcal{X}, \mu)$ un espace mesuré où μ est 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 $\mu \otimes \lambda$.

En effet, 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 B de $\mathbb{R}$, l'ensemble $E = A \times B$ vérifie $m(E) = m(T^{-1}(E))$. On a

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

Comme on l'a déjà vu, $E_x = B$ si $x \in A$, tandis que $E_x = \varnothing$ si $x \notin A$. Ainsi, on a $\lambda(E_x) = \lambda(B)\, \mathbb{1}_A(x)$. Par ailleurs,

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

qui est donc égal à $B - f(x)$ si $x \in A$, zéro sinon. Ainsi,

$$\lambda((T^{-1}(E))_x) = \lambda(B - f(x))\, \mathbb{1}_A(x).$$

Mais 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 pour obtenir l'égalité voulue.

4.10.2 Théorèmes de Fubini et Tonelli

La partie qui précède aura peut-être semblé un peu fastidieuse au lecteur. Mais maintenant le plus dur est fait, et nous allons voir comment, avec les théorèmes de Fubini et Tonelli, on peut concrètement calculer des intégrales de fonctions de plusieurs variables. Les résultats suivants se montrent toujours en trois étapes.
On commence par les prouver pour une fonction indicatrice quelconque, puis pour une fonction simple quelconque et on conclut par un argument d'approximation par des fonctions simples (énoncé dans le lemme 4.7).

Théorème 4.31 (Tonelli)**.** *Soient $(X, \mathcal{X}, \mu)$ et $(Y, \mathcal{Y}, \nu)$ deux espaces mesurés dont les mesures μ et ν sont σ-finies et $f \in \mathcal{V}_+(X \times Y, \mathcal{X} \otimes \mathcal{Y})$. f est donc une fonction positive.*

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 $(\overline{\mathbb{R}}_+, \mathcal{B}(\overline{\mathbb{R}}_+))$ et

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

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

Enfin, on a les égalités

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

Démonstration. La mesurabilité de $y \mapsto f(x,y)$ est une conséquence immédiate du lemme 4.28.

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

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

La deuxième partie du Théorème 4.30 assure la mesurabilité de l'application $x \mapsto \int_Y f(x,y)\ d\nu(y)$ ainsi que

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

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 4.7 et le théorème de convergence monotone. □

Théorème 4.32 (Fubini)**.** *Soient $(X, \mathcal{X}, \mu)$ 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

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

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

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

□

Les théorèmes de Tonelli et Fubini sont intimement liés : très souvent, on utilise d'abord le théorème de Tonelli afin de pouvoir appliquer le théorème de Fubini. **Exercice.** Pour $x > 0$, on pose $F(x) = \int_0^{+\infty} \frac{(\cos t)(1-\cos t)}{t} e^{-xt} \, dt$. Montrer que F est intégrable sur $]0, +\infty[$ et que

$$\int_{]0,+\infty[} F(x) \, d\lambda(x) = \int_{]0,+\infty[} \frac{(\cos t)(1-\cos t)}{t^2} \, d\lambda(t).$$

Posons $f(x,t) = \frac{(\cos t)(1-\cos t)}{t} e^{-xt}$. On a $|f(x,t)| \leq \frac{1-\cos t}{t} e^{-xt}$, d'où pour tout $t > 0$,

$$\int_{]0,+\infty[} |f(x,t)| \, d\lambda(x) \leq \int_{]0,+\infty[} \frac{1-\cos t}{t} e^{-xt} \, d\lambda(x) = \frac{1-\cos t}{t^2}.$$

La fonction $t \mapsto \frac{1-\cos t}{t^2}$ est intégrable sur $]0, +\infty[$: en effet, elle est continue, admet une limite $1/2$ en 0 et est en $O(1/t^2)$ en l'infini. D'après le théorème de Tonelli, f (ou $|f|$) est intégrable sur $]0, +\infty[\times]0, +\infty[$. D'après Tonelli, l'application $x \mapsto \int_{]0,+\infty[} |f(x,t)| \, d\lambda(t)$ est intégrable par rapport à la mesure de Lebesgue sur $]0, +\infty[$. Comme $0 \leq F(x) \leq \int_{]0,+\infty[} |f(x,t)| \, d\lambda(t)$, F l'est aussi. Comme f est intégrable sur $]0, +\infty[\times]0, +\infty[$, le théorème de Fubini nous dit que son intégrale est égale à $\int_{]0,+\infty[} F(x) \, d\lambda(x)$ d'une part (intégration en t, puis en x), et à

$$\int_{]0,+\infty[} \left(\int_{]0,+\infty[} f(x,t) \, d\lambda(x) \right) \, d\lambda(t)$$

d'autre part (intégration en x puis en t), soit $\int_{]0,+\infty[} \frac{(\cos t)(1-\cos t)}{t^2} \, d\lambda(t)$.

4.10.3 Associativité de la mesure produit

Soient $(X,\mathcal{X},\mu),(Y,\mathcal{Y},\nu),(Z,\mathcal{Z},\gamma)$ 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 $(\mu\otimes\nu)\otimes\gamma$ par ϕ et la mesure image m_2 de $\mu\otimes(\nu\otimes\gamma)$ par ψ sont égales : on note simplement cette mesure $\mu\otimes\nu\otimes\gamma$.

Montrons que m_1 et m_2 sont égales. On a :

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

et

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

ce qui montre que les mesures coïncident.

4.10.4 Convolution de mesures

Définition. *Soient 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 $(x,y)\mapsto x+y$.*

Proposition 4.33. *Si μ, ν sont des mesures σ-finies, a et b des réels positifs, alors*
- $(a\mu)*(b\nu)=(ab)(\mu*\nu)$.
- $\mu*0=0*\mu=0$*, où* 0 *désigne la mesure nulle.*

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

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

De plus, on a $\mu*0(A)=(\mu\otimes 0)(f^{-1}(A))=0$, et de même
$0*\mu(A)=(0\otimes\mu)(f^{-1}(A))=0$. □

4.11 Les théorèmes généraux et la mesure de comptage

Un certain nombre de théorèmes généraux donnent des résultats très pratiques lorsqu'ils sont utilisés avec la mesure de comptage. On va juste en énoncer 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) mesure(s) en jeu la mesure de comptage. Bien sûr, il retrouvera parfois des résultats connus.

Théorème 4.34 (Série à paramètre)**.** *Soit une famille de nombres réels $a(k,n)$ pour $k\ge 1,n\ge 1$ entiers. On suppose qu'il existe une suite de nombres réels positifs $(c_k)_{k\ge1}$ avec les propriétés :*

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

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

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

Alors les deux séries $s_n = \sum\limits_{k=1}^{+\infty} a(k,n)$ et $s = \sum\limits_{k=1}^{+\infty} a(k,\infty)$ convergent absolument et on a de plus $\lim\limits_{n\to+\infty} s_n = s$, 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. □

Démontrer le théorème 4.35 par des moyens élémentaires (avec des ε) est également un exercice très instructif que nous vous conseillons vivement.

Théorème 4.35. *Soient $(\Omega, \mathcal{F}, \mu)$ un espace mesuré et $(f_n)_{n\geq 1}$ une suite de fonctions mesurables positives de $(\Omega, \mathcal{F}, \mu)$ dans $(\overline{\mathbb{R}}_+, \mathcal{B}(\overline{\mathbb{R}}_+))$.*
Si on pose $f = \sum\limits_{n=1}^{+\infty} f_n$, alors

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

Les deux quantités peuvent valoir $+\infty$.

Démonstration. On peut, au choix, appliquer le théorème de Tonelli à la fonction $f(n,x) = f_n(x)$ que l'on intègre sur $\mathbb{N}^* \times \Omega$, ou encore appliquer le théorème de convergence monotone aux sommes partielles. □

Remarque 4.36. *Série de fonctions*
Voici une conséquence de ce théorème. Si la série de terme général $\int_\Omega f_n \, d\mu$ converge, alors f est intégrable. En particulier $f(\omega)$ est fini pour μ-presque tout ω. Nous en déduisons que, pour une suite (f_n) de fonctions mesurables de signe quelconque, si la série de terme général $\int_\Omega |f_n| \, d\mu$ converge, alors la série de terme général $f_n(\omega)$ converge (absolument) pour μ-presque tout ω.

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

Définition. *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). En dimension deux, $\lambda^{\otimes 2}(A)$ est l'aire de A ; en dimension trois, $\lambda^{\otimes 3}(A)$ est le volume de A.*

4.12.1 Transformations affines

Théorème 4.37. *Soit $y \in \mathbb{R}^d$. La translation $x \mapsto x + y$ laisse invariante la mesure de Lebesgue sur $\mathbb{R}^d$.*

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

L'invariance de la mesure de Lebesgue sur $\mathbb{R}^d$ par les translations en fait une mesure très particulière. Plus précisément, on a le résultat suivant :

Théorème 4.38. *Soit m une mesure sur $\mathbb{R}^d$ invariante par les translations et telle que $m(B(0,1)) < +\infty$. Alors $m = \frac{m(B(0,1))}{\lambda(B(0,1))}\lambda$.*

Démonstration. $\mathbb{R}^d$ est réunion dénombrable de translatés de la boule unité, donc m est σ-finie, ce qui permet d'appliquer Tonelli. Si $m(B(0,1)) = 0$, $m(\mathbb{R}^d) = 0$ et il n'y a rien à démontrer. Sinon, posons $g = \frac{1}{m(B(0,1))}\mathbb{1}_{B(0,1)}$. Par construction, on a $\int_{\mathbb{R}^d} g\ dm = 1$. On pose $\alpha = \int_{\mathbb{R}^d} g(-x)\ d\lambda(x) = \frac{\lambda(B(0,1))}{m(B(0,1))}$. Soit f une fonction mesurable positive. On a, en appliquant plusieurs fois les invariances de λ et de m ainsi que le théorème de Tonelli :

$$\begin{aligned}\int f\ d\lambda &= \Big(\int f\ d\lambda\Big)\Big(\int g\ dm\Big)\\ &= \int g(y)\Big(\int f(x)\ d\lambda(x)\Big)\ dm(y)\\ &= \int g(y)\Big(\int f(x+y)\ d\lambda(x)\Big)\ dm(y)\\ &= \int\Big(\int g(y)f(x+y)\ d\lambda(x)\Big)\ dm(y)\\ &= \int\Big(\int g(y)f(x+y)\ dm(y)\Big)\ d\lambda(x)\\ &= \int\Big(\int g(y-x)f(y)\ dm(y)\Big)\ d\lambda(x)\\ &= \int\Big(\int g(y-x)\ d\lambda(x)\Big) f(y)\ dm(y)\\ &= \int \alpha f(y)\ dm(y).\end{aligned}$$

Reste à trouver la valeur de α. En prenant $f = \mathbb{1}_A$, on obtient $\lambda(A) = \alpha m(A)$, ce qui est le résultat voulu. □

Revenons aux propriétés d'invariance de la mesure de Lebesgue.

Théorème 4.39. *Soit $M \in SL_d(\mathbb{R})$ (c'est-à-dire que $\det M = 1$). L'application $x \mapsto Mx$ laisse invariante la mesure de Lebesgue sur $\mathbb{R}^d$.*

Démonstration. Par un théorème d'algèbre linéaire, toute matrice de $SL_d(\mathbb{R})$ peut s'écrire comme produit de matrices de transvections, c'est-à-dire de matrices de la forme $I_n + \alpha E_{ij}$ avec $i \neq j$, où E_{ij} est la matrice dont tous les coefficients sont nuls, sauf celui en (i,j) qui vaut 1. Ainsi, il suffit de montrer le résultat pour une matrice de transvection. Mais c'est alors un cas particulier de l'exercice 2 vu précédemment : on identifie $\mathbb{R}^d$ à $\mathbb{R} \times \mathbb{R}^{\{2,\ldots,d\}}$, on prend $\mu = \lambda^{d-1}$ et $f(x) = \alpha\langle x, e_j\rangle e_i$. □

Théorème 4.40. *Soit $M \in M_d(\mathbb{R})$. Pour tout borélien A, on a*

$$\lambda^d(MA) = |\det M|\lambda^d(A).$$

En particulier, pour tout $c \in \mathbb{R}$, $\lambda^d(cA) = |c|^d\lambda^d(A)$.

Démonstration. Dans le cas où $M = \text{diag}(\lambda, 1, \ldots, 1)$, 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 alors écrire $M = \text{diag}(\det M, 1, \ldots, 1)N$, où $N \in SL_d(\mathbb{R})$. Ainsi

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

Reste le cas où M n'est pas 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 Im M est un espace vectoriel de dimension au plus $d-1$, il existe donc une application inversible qui envoie Im M dans $\mathbb{R}^{d-1} \times \{0\}$. Comme $\mathbb{R}^{d-1} \times \{0\}$ est de mesure nulle, Im M aussi. □

Corollaire 4.41. *Si M est inversible, la mesure image de λ^d par $x \mapsto Mx + b$ est $\frac{1}{|\det M|}\lambda^d$.*

Un calcul de volume : le volume d'un cône

Le calcul d'une aire, ou d'un volume repose souvent sur l'utilisation du théorème de Tonelli ou de son ancêtre, le théorème 4.30. Voyons par exemple comment calculer le volume d'un cône. Une surface plane de $\mathbb{R}^3$ est un borélien dont la dimension affine est égale à deux. Si A est une surface plane et x un point de $\mathbb{R}^3$ qui n'est pas dans l'espace affine engendré par A, le cône de sommet x et de base A est

$$C = \{(1-\theta)x + \theta u; (\theta, u) \in [0,1] \times A\}.$$

Si $A = A' \times \{h\}$, avec $h > 0$ et $x = (0,0,0)$, en découpant le volume par tranches horizontales (perpendiculaires à « l'axe des z »), on a

$$\lambda^3(C) = \int_{\mathbb{R}} \lambda^2(C_z)\, d\lambda(z);$$

où $C_z = \{(x,y) \in \mathbb{R}^2; (x,y,z) \in C\}$. On a $C_z = \frac{z}{h}A'$ pour $z \in [0,h]$, $C_z = \varnothing$ sinon. On en déduit

$$\begin{aligned}\lambda^3(C) &= \int_{[0,h]} \lambda^2(\frac{z}{h}A')\, d\lambda(z) = \int_{[0,h]} \lambda^2(A')\frac{z^2}{h^2}\, d\lambda(z) \\ &= \frac{h}{3}\lambda^2(A').\end{aligned}$$

On retrouve donc la formule bien connue : le volume du cône est égal au tiers du produit de la hauteur fois l'aire de la base.

Corollaire 4.42. *Soient M une matrice inversible et $b \in \mathbb{R}^d$.*
On pose $T(x) = Mx + b$. Soit 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 $\mathbb{R}^d$. Notons μ_2 la mesure image de μ_1 par T. D'après le théorème de transfert,

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

ce qui donne le résultat voulu. □

4.12.2 Intégration des fonctions radiales

Théorème 4.43. *Soit $\|.\|$ une norme quelconque sur $\mathbb{R}^n$. On note V le volume de la boule unité pour cette norme. Alors, pour toute fonction ϕ mesurable de $\mathbb{R}_+$ dans $\mathbb{R}$, l'application $x \mapsto \phi(\|x\|)$ est intégrable par rapport à $\lambda^{\otimes n}$ si et seulement si $\int_{\mathbb{R}_+} nt^{n-1}|\phi(t)| \, d\lambda(t) < +\infty$ et alors*

$$\int_{\mathbb{R}^n} \phi(\|x\|) \, d\lambda^{\otimes n}(x) = V \int_{\mathbb{R}_+} \phi(t) nt^{n-1} \, d\lambda(t).$$

Démonstration. D'après le théorème de transfert, $\phi \circ \|.\|$ est intégrable si et seulement si ϕ est intégrable par rapport à la mesure image de $\lambda^{\otimes n}$ par $\|.\|$, et on aura alors

$$\int_{\mathbb{R}^n} \phi(\|x\|) \, d\lambda^{\otimes n}(x) = \int_{\mathbb{R}_+} \phi(t) \, dm(t).$$

Il suffit donc de caractériser m. Soit $a \geq 0$. En utilisant successivement la définition d'une mesure image, l'homogénéité d'une norme, et la propriété d'échelle de la mesure de Lebesgue sur $\mathbb{R}^n$, on a

$$m([0,a]) = \lambda^{\otimes n}(B(0,a)) = \lambda^{\otimes n}(aB(0,1)) = a^n \lambda^{\otimes n}(B(0,1)) = Va^n.$$

Comme les intervalles $[0,a]$ forment un π-système qui engendre la tribu borélienne de $\mathbb{R}_+$, avec $\mathbb{R}_+ = \bigcup_{i=0}^{+\infty} [0,i]$, le théorème 3.12 nous dit que la connaissance de m sur les intervalles $([0,a])_{a \in \mathbb{R}_+}$ permet de l'identifier. Or il est facile de voir que

$$Va^n = \int_{[0,a]} Vnt^{n-1} \, d\lambda(t),$$

donc m est la mesure dont la densité par rapport à la mesure de Lebesgue est $t \mapsto Vnt^{n-1}\mathbb{1}_{[0,+\infty[}(t)$. On en déduit que

$$\int_{\mathbb{R}_+} \phi(t) \, dm(t) = V \int_{\mathbb{R}_+} \phi(t) nt^{n-1} \, d\lambda(t),$$

ce qui est le résultat voulu. □

Corollaire 4.44. Calcul de l'intégrale de Gauss

$$I = \int_{-\infty}^{+\infty} \exp(-\frac{x^2}{2}) = \sqrt{2\pi}.$$

Démonstration. Le théorème de Tonelli donne

$$I^2 = \int_{\mathbb{R}^2} \exp(-\frac{x_1^2 + x_2^2}{2}) \, d(\lambda \otimes \lambda)(x_1, x_2) = \int_{\mathbb{R}^2} \phi(\|x\|) \, d\lambda^2(x),$$

avec $\phi(x) = \exp(-\frac{x^2}{2})$. Si on note V_2 le volume de la boule unité euclidienne de $\mathbb{R}^2$, avec la formule d'intégration d'une fonction radiale, on a donc

$$I^2 = V_2 \int_0^{+\infty} 2r\phi(r) \, dr = 2V_2 \lim_{M \to +\infty} [-\exp(-\frac{r^2}{2})]_0^M = 2\pi,$$

car on sait que $V_2 = \pi$. □

On voit sur cet exemple que même en petite dimension, le théorème d'intégration d'une fonction radiale est d'usage plus simple que le changement de variable polaire.

Corollaire 4.45. *Le volume de la boule unité euclidienne de* $\mathbb{R}^n$ *est*

$$V_n = \frac{2\pi^{n/2}}{n\Gamma(n/2)}.$$

Démonstration. On prend $\phi(x) = \exp(-\frac{x^2}{2})$. Le théorème de Tonelli donne

$$\int_{\mathbb{R}^n} \exp(-\frac{\|x\|_2^2}{2})\ d\lambda^{\otimes n}(x) = \left(\int_{\mathbb{R}} \exp(-\frac{x^2}{2})\ d\lambda(x)\right)^n = (2\pi)^{n/2}.$$

D'autre part le changement de variable $u = t^2/2$ donne

$$\int_{\mathbb{R}_+} \exp(-\frac{t^2}{2}) nt^{n-1}\ d\lambda(t) = \int_{\mathbb{R}_+} \exp(-u) n(2u)^{n/2-1}\ d\lambda(u) = 2^{\frac{n}{2}-1} n\Gamma(\frac{n}{2}).$$

En faisant le quotient et en appliquant le théorème précédent, on obtient le résultat voulu. $\square$

Remarque 4.46. *L'astuce est évidemment de trouver une fonction* ϕ *pour laquelle on sait calculer les deux intégrales. Ce n'est tout de même pas si fréquent. La méthode permet également de calculer le volume de la boule unité de* $\|.\|_p$*, définie par* $\|x\|_p^p = \sum_{i=1}^n |x_i|^p$*, en prenant* $\phi(x) = \exp(-x^p)$ *(exercice laissé au lecteur ; on trouvera comme volume* $\frac{2^n\Gamma(\frac{1}{p}+1)^n}{\Gamma(\frac{n}{p}+1)}$*).*

4.13 Preuve des propriétés de base de l'intégrale

4.13.1 Premiers résultats

L'implication $(f \le g) \implies \int f\ d\mu \le \int g\ d\mu$ 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

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

et

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

Lemme 4.47. *Si une fonction est* μ*-presque partout nulle, alors 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 une fonction f positive est μ-presque partout nulle, alors pour tout i tel que $\inf\{f(\omega); \omega \in \Omega_i\} > 0$, on a $\mu(\Omega_i) \le \mu(f \ge \inf\{f(\omega); \omega \in \Omega_i\}) \le \mu(f > 0) = 0$. Cela entraîne que pour toute partition finie, $I((\Omega_i)_{i\in I}, f) = 0$, d'où la nullité de l'intégrale. $\square$

Lemme 4.48. *Soient* f *mesurable positive,* $\alpha > 0$ *et* A *mesurable. Alors*

$$\int (f + \alpha \mathbb{1}_A)\ d\mu = \int f\ d\mu + \alpha\mu(A).$$

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

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

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

Ainsi

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

d'où en passant au supremum :

$$\int (f + \alpha\mathbb{1}_A)\, d\mu \leq \int f\, d\mu + \alpha\mu(A).$$

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

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

En passant au supremum 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 $\int f\, d\mu \geq \alpha\mu(f \geq \alpha)$[4]. Ainsi, si f positive est d'intégrale nulle, on a pour tout n : $0 = \int f\, d\mu \geq \frac{1}{n}\mu\left(f \geq \frac{1}{n}\right)$, soit $\mu(f \geq 1/n) = 0$, et donc d'après le théorème de continuité séquentielle décroissante, $\mu(f > 0) = 0$.

4.13.2 Démonstration du théorème de Beppo Levi

1. On va d'abord montrer une forme très faible de ce résultat : si f est une fonction simple positive et $(E_n)_{n\geq 1}$ une suite croissante d'ensembles mesurables 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.

4. 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.

2. Passons au cas général. On considère (f_n) une suite croissante de fonctions mesurables positives tendant vers f et on veut montrer que $\int f_n \, d\mu$ tend vers $\int f \, d\mu$. Bien sûr, la suite $(\int f_n \, d\mu)_n$ 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. Comme les fonctions f_n et g sont mesurables, $E_n \in \mathcal{F}$. 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, d'après la forme faible du théorème [5], c'est-à-dire la partie 1. de cette démonstration,

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

On a

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

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 l'inégalité souhaitée

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

4.13.3 Preuve de la linéarité

Soient f, g deux fonctions mesurables positives, $\alpha > 0$. On veut montrer que $\int (f + \alpha g) \, d\mu = \int f \, d\mu + \alpha \int g \, d\mu$. Soient $(f_n), (g_n)$ 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. On a $\int |\alpha g| \, d\mu \leq \int |\alpha| g^+ + |\alpha| g^- \, d\mu < +\infty$ en utilisant la linéarité pour les fonctions positives. Ainsi, on obtient 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 $\alpha = 1$. Comme $|f + g| \leq |f| + |g|$, $f + g$ 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

$$\int f^+ + \int g^+ + \int h^- = \int f^- + \int g^- + \int h^+.$$

5. Noter que si les f_n n'étaient pas mesurables, E_n ne serait pas nécessairement dans $\mathcal{F}$ et on ne pourrait invoquer la forme faible. L'hypothèse de mesurabilité sert donc bien à quelque chose !

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 $\int f\, d\mu \leq \int g\, d\mu$. Écrivons $f = f\, \mathbb{1}_{\{f\leq g\}} + f\, \mathbb{1}_{\{f>g\}}$ et $g = g\, \mathbb{1}_{\{f\leq g\}} + g\, \mathbb{1}_{\{f>g\}}$. Comme $f\, \mathbb{1}_{\{f>g\}}$ et $g\, \mathbb{1}_{\{f>g\}}$ sont nulles et donc d'intégrale 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.

4.14 Exercices sur les intégrales

4.14.1 Exercices de la série 1

Exercice 48. Montrer l'existence de $\int_{]0,1]}\{\frac{1}{x}\}\, d\lambda(x)$, puis montrer que

$$\int_{]0,1]}\{\frac{1}{x}\}\, d\lambda(x) = 1 - \gamma,$$

où γ est la constante d'Euler, qu'on a introduite à la section 4.7.3. [6]

Exercice 49. *Continuité de la transformée de Laplace.*
Soit f une fonction continue de $\mathbb{R}_+$ dans $\mathbb{C}$. On suppose que l'intégrale impropre $\int_0^{+\infty} f(t)\, dt$ existe (c'est-à-dire que $\int_0^T f(t)\, dt$ admet une limite quand T tend vers $+\infty$). Montrer que pour tout $\lambda \geq 0$, l'intégrale impropre $\int_0^{+\infty} e^{-\lambda t} f(t)\, dt$ existe et que la fonction $\lambda \mapsto \int_0^{+\infty} e^{-\lambda t} f(t)\, dt$ est continue sur $\mathbb{R}_+$.

Exercice 50. *Théorème du retour de Poincaré.*
Soient $(\Omega, \mathcal{F}, \mu)$ 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.

1. On pose $N(x) = \sum\limits_{k=0}^{+\infty} \mathbb{1}_A(T^k(x))$ 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.
2. On pose $Z(x) = Y(Tx)$. Montrer que $Y(x) = e^{-\mathbb{1}_A(x)} Z(x)$, puis que $\int Y(x)\, d\mu(x) = \int Z(x)\, d\mu(x)$.
3. Conclure.

Exercice 51.

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

1. $\int_0^{+\infty} \frac{\log(x+n)}{n} e^{-x} \cos x\, dx$.
2. $\int_0^1 \frac{1+nx}{(1+x)^n}\, dx$.
3. $\int_1^{+\infty} \frac{1+nx}{(1+x)^n}\, dx$.

Exercice 52. Les fonctions $f(x, y)$ suivantes sont-elles intégrables sur le pavé $[0, 1] \times [0, 1]$?

1. $(x - y)(x^2 + y^2)^{-3/2}$.
2. $(1 - xy)^p$ avec $p < 0$.

---

6. On pourra trouver d'autres intégrales du même style avec des applications à la théorie des nombres dans l'ouvrage de Pólya et Szegö [9], Partie II, chapitre 1, paragraphe 5.

3. $(x^2 - y^2)/(x^2 + y^2)^2$.
4. $(x - y)/(x + y)^3$.

Exercice 53. *Intégrales de Wallis.*

1. On pose
$$W_n = \int_0^{\pi/2} (\cos\theta)^n \, d\theta.$$
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.
2. Montrer que $W_nW_{n+1} \leq W_n^2 \leq W_nW_{n-1}$. En déduire l'équivalent à l'infini
$$W_n \sim \sqrt{\frac{\pi}{2n}}.$$

Pour d'autres méthodes de calcul de cet équivalent, on peut se reporter à l'exercice 73 ainsi qu'aux indications de l'exercice 91.

Exercice 54. *Calcul de l'intégrale de Gauss.*

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

Exercice 55. *Calcul de* $\Gamma(1/2)$.
Connaissant la valeur de l'intégrale de Gauss, montrer que
$$\Gamma(1/2) = \sqrt{\pi}.$$

Exercice 56. 1. À l'aide d'un développement en série entière, montrer que
$$\sum_{n=0}^{+\infty} \frac{(-1)^n}{3n+1} = \int_0^1 \frac{dx}{x^3+1}.$$

2. On donne
$$\frac{1}{x^3+1} = \frac{1}{3}\left(\frac{1}{1+x} - \frac{x-2}{x^2-x+1}\right).$$
Montrer que
$$\sum_{n=0}^{+\infty} \frac{(-1)^n}{3n+1} = \frac{3\log 2 + \sqrt{3}\pi}{9}.$$

Exercice 57. *Calcul de l'intégrale de Dirichlet avec Fubini.*
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 $I_a = \displaystyle\int_{[0,a]\times[0,+\infty[} f(x, y) \, d\lambda \otimes \lambda(x, y)$. Déterminer la limite de I_a quand a tend vers $+\infty$. En déduire la valeur de $\displaystyle\int_0^{+\infty} \frac{\sin x}{x} \, dx$.

Exercice 58. *Calcul d'intégrales liées aux intégrales de Fresnel.*
Le but de cet exercice est le calcul des intégrales de la forme
$$\int_0^{+\infty} e^{iu} u^{\alpha-1} \, du, \quad 0 < \alpha < 1$$
et d'intégrales liées.

1. On pose, pour $\lambda \geq 0$,
$$\phi(\lambda) = \int_0^{+\infty} e^{-\lambda u}\, e^{iu} u^{\alpha-1}\, du.$$
Montrer que ϕ définit une fonction continue sur $[0, +\infty[$.
2. Montrer que ϕ est dérivable sur $]0, +\infty[$ et satisfait l'équation différentielle $\phi'(\lambda) = \frac{\alpha}{i-\lambda}\phi(\lambda)$. En déduire que pour tout $\lambda > 0$, on a
$$\phi(\lambda) = \phi(0)\frac{1}{(\lambda^2+1)^{\alpha/2}} \exp(-\alpha i \arctan \lambda).$$
3. Montrer que pour tout $\lambda > 0$, on a
$$\phi(0) = (1+\lambda^{-2})^{\alpha/2} \exp(\alpha i \arctan \lambda) \int_0^{+\infty} e^{-x} e^{i\frac{x}{\lambda}} x^{\alpha-1}\, dx.$$
4. En déduire que
$$\int_0^{+\infty} e^{iu} u^{\alpha-1}\, du = \exp\left(i\alpha\frac{\pi}{2}\right)\Gamma(\alpha),$$
et en particulier, comme $\Gamma(1/2) = \sqrt{\pi}$,
$$\int_0^{+\infty} \frac{\cos u}{\sqrt{u}}\, du = \int_0^{+\infty} \frac{\sin u}{\sqrt{u}}\, du = \sqrt{\frac{\pi}{2}}.$$
5. Calculer les intégrales de Fresnel
$$\int_0^{+\infty} \cos(u^2)\, du \text{ et } \int_0^{+\infty} \sin(u^2)\, du.$$

Exercice 59. *Intégrale de Dirichlet : semi-convergence et un équivalent.*
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 $\mathbb{R}_+$. Montrer de plus l'équivalent à l'infini $\int_0^{n\pi} \frac{|\sin t|}{t}\, dt \sim \frac{2}{\pi}\log n$.

Exercice 60. *Calcul de l'intégrale de Dirichlet à l'aide d'une intégrale à paramètre.*

1. On pose, pour $x \geq 0$, $F(x) = \int_0^{+\infty} \frac{\sin t}{t} e^{-xt}\, dt$. Montrer que F est correctement définie et définit une fonction continue sur $\mathbb{R}_+$.
2. Montrer que F est dérivable sur $]0, +\infty[$, avec $F'(x) = -\frac{1}{1+x^2}$.
3. Calculer F et en déduire la valeur de $\int_0^{+\infty} \frac{\sin t}{t}\, dt$.

Exercice 61. Calculer l'intégrale $I = \int_{V(a,b,c)} (x^2+y^2+z^2)\, d\lambda^{\otimes 3}(x,y,z)$, où
$$V(a,b,c) = \left\{(x,y,z) \in \mathbb{R}^3,\ x \geq 0,\ y \geq 0,\ z \geq 0,\ \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leq 1\right\}.$$
On rappelle que le volume de la boule unité de $\mathbb{R}^3$ est $\frac{4}{3}\pi$.
Indication : commencer par traiter le cas où $a = b = c = 1$.

Exercice 62. *Fonction Gamma : formule de récurrence et application au calcul de la transformée de Fourier de la gaussienne.*

1. Montrer que la fonction Γ vérifie, pour tout réel $x > 0$, l'égalité
$\Gamma(x+1) = x \cdot \Gamma(x)$. En particulier, vérifier que pour tout entier $n \geq 0$, on a :
$$\Gamma(n+1) = n! \quad \text{et} \quad \Gamma\left(n+\frac{1}{2}\right) = \Gamma\left(\frac{1}{2}\right) \cdot \frac{(2n)!}{2^{2n} \cdot n!}.$$

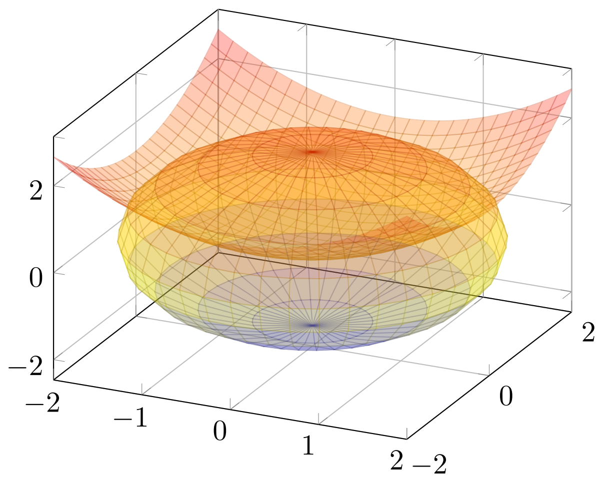
2. On donne $\Gamma(1/2) = \sqrt{\pi}$. Montrer que pour tout entier naturel n, on a
$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} t^{2n} e^{-\frac{t^2}{2}} d\lambda(t) = \frac{(2n)!}{2^n n!}.$$
3. On pose $G(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$. Montrer que pour tout x réel
$$\int_{\mathbb{R}} G(t) e^{itx}\, d\lambda(t) = \int_{\mathbb{R}} G(t) \cos(tx)\, d\lambda(t) = \sqrt{2\pi} G(x).$$
4. En deuxième lecture du livre : interpréter les résultats démontrés en termes probabilistes.

Exercice 63. *Fonction Gamma et fonction Zêta.*
Pour $s > 1$, on pose $\zeta(s) = \sum\limits_{n=1}^{+\infty} \frac{1}{n^s}$. Montrer que l'on a
$$\zeta(s)\Gamma(s) = \int_0^{+\infty} t^{s-1} \frac{e^{-t}}{1 - e^{-t}}\, dt.$$

Exercice 64. *Formule de Stirling.*

1. Montrer que $\int_{2n}^{+\infty} x^n e^{-x}\, dx = o(\Gamma(n+1))$.
2. Montrer l'équivalent $\frac{\Gamma(n+1)}{e^{-n} n^{n+1/2}} \sim \int_{-\sqrt{n}}^{\sqrt{n}} (1 + \frac{u}{\sqrt{n}})^n e^{-u\sqrt{n}}\, du$.
3. Montrer que pour tout $x \in]-1, 1[$, on a $\ln(1+x) - x \leq -\frac{x^2}{6}$.
4. On rappelle que $\int_{-\infty}^{+\infty} e^{-t^2/2}\, dt = \sqrt{2\pi}$. Montrer que on a l'équivalent suivant en $+\infty$:
$$\Gamma(n+1) \sim \sqrt{2\pi} e^{-n} n^{n+1/2}.$$

Exercice 65. On note S la boule de $\mathbb{R}^3$ centrée en l'origine et de rayon 2 et H^- la surface inférieure de frontière paraboloïdique (elliptique) :
$$H^- = \{(x, y, z) \in \mathbb{R}^3 ; 3z \leq x^2 + y^2\}.$$



Pour toute partie A de $\mathbb{R}^3$, on note A_z sa tranche de niveau z :
$$A_z = \{(x, y) \in \mathbb{R}^2 ; (x, y, z) \in A\}.$$

1. Soit $z \in \mathbb{R}$. Montrer que
$$(S \cap H^-)_z = \begin{cases} S_z & \text{si } z < 0 \\ \varnothing & \text{si } z > 1 \\ \{(x, y) \in \mathbb{R}^2 ; 3z \leq x^2 + y^2 \leq 4 - z^2\} & \text{si } z \in [0, 1] \end{cases}$$

2. Montrer que $\lambda^{\otimes 3}(S \cap H^-) = \frac{15}{2}\pi$.
3. Le centre de gravité d'un solide homogène représenté par le borélien borné A est le point
$$\frac{1}{\lambda(A)} \int_A (x, y, z)\ d\lambda^{\otimes 3}(x, y, z).$$
Montrer que le centre de gravité de $S \cap H^-$ est le point de cooordonnées $(0, 0, -\frac{13}{30})$.
4. On pose
$$H^+ = \{(x, y, z) \in \mathbb{R}^3; 3z > x^2 + y^2\}.$$
Montrer que le centre de gravité de $S \cap H^+$ est le point de cooordonnées $(0, 0, \frac{39}{38})$. Indication : on conseille ne pas refaire tous les calculs, de remarquer plutôt que le centre de gravité de S est l'origine.

Exercice 66. *Méthode de Laplace.*
Soient g et h des fonctions mesurables de $\mathbb{R}_+$ dans $\mathbb{R}$. On suppose que :
- h est strictement décroissante,
- on a en 0 : $h(x) = h(0) - cx^\beta + o(x^\beta)$ avec $c > 0$ et $\beta > 0$,
- on a en 0 : $g(x) \sim Ax^\alpha$ avec $\alpha > -1$ et $A \neq 0$,
- $\int_{\mathbb{R}_+} |g(x)|e^{h(x)}\ dx < +\infty$.

Le but de l'exercice est de montrer que lorsque t tend vers l'infini, on a
$$I(t) = \int_0^{+\infty} g(x)e^{th(x)}\ d\lambda(x) \sim Ae^{h(0)t}(ct)^{-\frac{\alpha+1}{\beta}} K_{\alpha,\beta},$$
avec $K_{\alpha,\beta} = \frac{1}{\beta}\Gamma(\frac{\alpha+1}{\beta})$. On vérifie sans peine que le résultat est encore vrai si les intégrales et les fonctions ne sont pas définies sur $[0, +\infty[$, mais sur $[0, M[$ avec $0 < M < +\infty$.

1. Soit $\delta > 0$. On pose
$$I_\delta(t) = \int_{[0,\delta]} g(x)e^{th(x)}\ d\lambda(x) \text{ et } R_\delta(t) = \int_{[\delta,+\infty[} g(x)e^{th(x)}\ d\lambda(x).$$
Montrer que
$$\lim_{t\to+\infty} \frac{R_\delta(t)}{e^{h(0)t}t^{-\frac{\alpha+1}{\beta}}} = 0.$$
2. Exprimer $\int_{[0,+\infty[} u^\alpha e^{-u^\beta}\ d\lambda(u)$ à l'aide de la fonction Γ.
3. Pour $\delta > 0$, on pose $q_\delta(x) = \frac{g(x)}{Ax^\alpha}\mathbb{1}_{[0,\delta]}(x)$ et $r(x) = \frac{h(0)-h(x)}{cx^\beta}$. Vérifier qu'il existe $\delta > 0$ tel que $|q_\delta(x)| \leq 2$ et $r(x) \geq 1/2$ pour $x \in]0, \delta]$.
Montrer que
$$\frac{I_\delta(t)}{Ae^{h(0)t}(ct)^{-\frac{\alpha+1}{\beta}}} = \int_{\mathbb{R}_+} q_\delta\left(\frac{u}{(ct)^{1/\beta}}\right) u^\alpha e^{-r\left(\frac{u}{(ct)^{1/\beta}}\right)u^\beta}\ d\lambda(u).$$
4. Conclure.

Exercice 67. Soient μ et ν deux mesures σ-finies sur $\mathbb{R}^n$. On suppose que ν admet une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$. Montrer que $\mu * \nu$ admet une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$.

4.14.2 Exercices de la série 2

Exercice 68. 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 $\mu = \nu$.

Exercice 69. Soient μ une mesure finie sur $(\Omega, \mathcal{F})$ et f une application finie μ-presque partout. Montrer que les quatre conditions suivantes sont équivalentes :

1. f est intégrable par rapport à μ.
2. $\int_{\{|f|>n\}} |f| \; d\mu$ tend vers 0 quand n tend vers l'infini
3. $\sum_{n\geq 1} n\mu(n < |f| \leq n+1) < +\infty$.
4. $\sum_{n\geq 0} \mu(|f| > n) < +\infty$.

Indication : montrer $a \iff b, a \iff c, d \Longrightarrow c, (c\&b) \Longrightarrow d$.

Exercice 70. Soient 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.

Exercice 71. *Intégration par rapport à une somme de mesures.*
Soit $(\mu_i)_{i\in\mathbb{N}}$ une suite de mesures définies sur un espace $(\Omega, \mathcal{F})$.

1. Montrer que $\mu = \sum_{i\in\mathbb{N}} \mu_i$ est une mesure sur $(\Omega, \mathcal{F})$.
2. Montrer que pour f mesurable positive, puis pour f intégrable, on a

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

Exercice 72. 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.

Exercice 73. 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.

Exercice 74. 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.

Exercice 75.

1. Calculer $I_n = \int_0^1 x^n \log x \; dx$ pour $n \in \mathbb{N}$.
2. En déduire la valeur de $\int_0^1 \frac{\log x}{1-x} \; dx$, sachant que $\sum\limits_{n=1}^{+\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$.
3. En calculant de deux manières différentes $\displaystyle\int_0^1 (1-x)^n \ln(x) \; dx$, montrer que pour $n \geq 0$, on a

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

Exercice 76. Démontrer que

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

Exercice 77. Calculer $\int_0^1 \frac{1}{1+x^2}\, dx$. En déduire $\sum_{n=0}^{+\infty} \frac{(-1)^n}{2n+1}$.

Exercice 78. Démontrer que $\displaystyle\int_0^{+\infty} \frac{\sin ax}{e^x - 1}\, dx = \sum_{n=1}^{+\infty} \frac{a}{n^2+a^2}$.

Exercice 79. Démontrer que $\displaystyle\int_0^1 \frac{(x\log x)^2}{1+x^2}\, dx = 2\sum_{n=0}^{+\infty} \frac{(-1)^n}{(2n+3)^3}$.

Exercice 80. *Calcul de* $\sum_{n\geq 1} \frac{1}{n^2}$ *avec Fubini.*
Calculer de deux façons différentes :

$$J = \int_{-1}^1 \int_{-1}^1 \frac{1}{1+2xy+y^2}\, dx\, dy.$$

Exercice 81. Montrer que pour tout $z \in \mathbb{C}$, $(1+z/n)^n$ tend vers $\exp(z)$ lorsque n tend vers l'infini.

Exercice 82. *Régularité de la fonction Gamma.*
Démontrer que la fonction Γ est de classe $\mathcal{C}^\infty$ sur $]0,+\infty[$ et que pour tout entier $n \geq 1$,

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

Exercice 83. *Intégrale de Wallis de deuxième espèce.*
Trouver par convergence dominée l'équivalent de l'intégrale de Wallis de seconde espèce :

$$\int_0^\pi \left(\frac{\sin x}{x}\right)^n dx \sim \sqrt{\frac{3\pi}{2n}}.$$

Exercice 84. *Gamma et Zêta, la suite.*
Soit $s > 0$. Montrer que

$$\int_0^{+\infty} \frac{t^{s-1}}{1+e^t}\, dt = \Gamma(s)\eta(s),\ \text{avec } \eta(s) = \sum_{k=1}^{+\infty} \frac{(-1)^{k-1}}{k^s}$$

η est la fonction êta de Dirichlet.
Lorsque $s > 1$, montrer que $\eta(s) = (1 - \frac{1}{2^{s-1}})\zeta(s)$.

Exercice 85. Montrer que pour $\alpha > 2$, on a

$$\int_1^{+\infty} \frac{\{x\}}{x^\alpha}\, dx = \frac{1}{\alpha-2} - \frac{\zeta(\alpha-1)}{\alpha-1}.$$

À l'aide du résultat de l'exercice 53, en déduire le développement asymptotique en 0 à droite : $\zeta(1+h) = \frac{1}{h} + \gamma + o(1)$, où γ est la constante d'Euler.

Chapitre 5

Lois des variables et des vecteurs aléatoires

Rappelons que 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. Nous notons λ la mesure de Lebesgue.

5.1 Notions générales

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

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

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

Par définition, $X^{-1}(A) = \{\omega \in \Omega; X(\omega) \in A\}$.

Afin de simplifier les notations, 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 $\mathbb{P}_X(A)$. On utilise souvent $A = \{x\}$ ou $A =]-\infty, x]$, etc. De plus, l'événement $X^{-1}(\{x\})$ est noté $\{X = x\}$, l'événement $X^{-1}(]-\infty, x])$ est noté $\{X \leq x\}$, etc.
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 (puisque continue), 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

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

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 de la théorie 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

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

Remarque 5.1. *On appelle parfois fonction de queue de X la fonction $1 - F_X$.*

Théorème 5.2. *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 $\mathbb{P}_Y$ coïncident sur les ensembles de la forme $]-\infty, t_1]\times]-\infty, t_2]\times\cdots\times]-\infty, t_d]$. Or ces ensembles forment un π-système qui engendre $\mathcal{B}(\mathbb{R}^d)$, donc $\mathbb{P}_X$ et $\mathbb{P}_Y$ sont égales. □

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

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

Théorème 5.3. *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 $\mathbb{R}$.*
— $\lim\limits_{t\to-\infty} F_X(t) = 0$ *et* $\lim\limits_{t\to+\infty} F_X(t) = 1$.
— *En tout point $x \in \mathbb{R}$, F_X est continue à droite.*
— *En tout point $x \in \mathbb{R}$, F_X admet une limite à gauche qui est égale à $F(x)$ si et seulement si $\mathbb{P}(X = x) = 0$.*
— *L'ensemble des points de discontinuité de F est fini ou dénombrable.*

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

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

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

— Posons pour $n \geq 1$, $A_n =]-\infty, -n]$. On a $A_{n+1} \subset A_n$ et $\bigcap\limits_{n=1}^{+\infty} A_n = \varnothing$, d'où $\lim\limits_{n\to+\infty} \mathbb{P}_X(A_n) = \mathbb{P}_X(\varnothing) = 0$. 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

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

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

Posons pour $n \geq 1$, $A_n =]-\infty, n]$. On a $A_n \subset A_{n+1}$ et $\bigcup_{n=1}^{+\infty} A_n = \mathbb{R}$, 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 $\mathbb{P}_X(A_n) \geq 1 - \varepsilon$. Comme F_X est croissante et majorée par 1, on a

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

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

— Soient $t \in \mathbb{R}$, $(t_n)_{n \geq 1}$ une suite décroissante convergeant vers t. Posons, pour $n \geq 1$, $A_n =]-\infty, t_n]$. On a $A_{n+1} \subset A_n$ et $\bigcap_{n=1}^{+\infty} A_n =]-\infty, t]$, 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 son ensemble de définition. On a

$$\begin{aligned} F_X(x) - \lim_{y \to x^-} F_X(y) &= \lim_{n \to +\infty} F_X(x) - F_X(x - 1/n) \\ &= \lim_{n \to +\infty} \mathbb{P}(X \in]x - 1/n, x]) = \mathbb{P}(X = x) \end{aligned}$$

d'après le théorème de continuité séquentielle décroissante.

— D'après ce qui précède, les points de discontinuité de F_X sont les points x tels que $\mathbb{P}(X = x) > 0$: il y en a au plus un ensemble dénombrable d'après le théorème 3.3.

□

Remarque 5.4. *On parle de la fonction de répartition d'une variable aléatoire, mais il s'agit en fait de la fonction de répartition de la loi de la variable aléatoire.*

Retrouver la loi lorsque la fonction de répartition n'est pas continue

Théorème 5.5. *Soit F la fonction de répartition associée à la loi μ. On suppose que F est de classe C^1 par morceaux, avec les points de discontinuité $a_1, \dots, a_n$. Alors μ se décompose en la somme d'une partie à densité, f qui est la dérivée de F là où F est dérivable, et d'une partie discrète qui est $\nu = \sum_{i=1}^{n} \mu(a_i)\delta_{a_i}$.*

Démonstration. On doit montrer que pour tout $t \in \mathbb{R}$,

$$F(t) = \int_{]-\infty, t]} f(x)\, d\lambda(x) + \nu(]-\infty, t]).$$

Soient $T < t < a_1$. D'après le théorème fondamental de l'analyse,

$$F(t) - F(T) = \int_t^T f(x)\, dx = \int_{]T,t]} f(x)\, d\lambda(x).$$

En faisant tendre T vers $-\infty$, on obtient à gauche $F(t)$ et à droite $\int_{]-\infty,t]} f(x)\, d\lambda(x)$ avec le théorème de convergence monotone. (H_0) est donc vraie, où l'on note

$$(H_i) \quad \forall t < a_i, \quad F(t) = \int_{]-\infty,t]} f(x)\, d\lambda(x) + \sum_{j<i} \mu(a_j).$$

Il suffit alors de montrer que $(H_i) \implies (H_{i+1})$ pour conclure. On a

$$F(a_i) = \mu(a_i) + \lim_{t \to a_i^-} F(t) = \int_{]-\infty, a_i]} f(x)\, d\lambda(x) + \sum_{j<i} \mu(a_j) + \mu(a_i)$$

Soient $a_i < T < t < a_{i+1}$. D'après le théorème fondamental de l'analyse,

$$F(t) - F(T) = \int_T^t f(x)\, dx = \int_{]T,t]} f(x)\, d\lambda(x).$$

En faisant tendre T vers a_i, on obtient à gauche $F(t) - F(a_i)$ et à droite $\int_{]a_i,t]} f(x)\, d\lambda(x)$ avec le théorème de convergence monotone. En ajoutant les deux égalités on a

$$F(t) = \int_{]-\infty,t]} f(x)\, d\lambda(x) + \sum_{j<i+1} \mu(a_j).$$

□

On va terminer cette sous-section consacrée aux fonctions de répartition par un théorème général sur la convergence des fonctions de répartition qui sera utile plus tard quand on s'intéressera à la convergence des lois.

5.1.2 Tribu engendrée par une ou plusieurs variables aléatoires

Définition. *Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, X une variable aléatoire (ou un vecteur aléatoire) sur cet 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.

Pareillement, la tribu engendrée par une famille de variables $(X_i)_{i \in I}$, que l'on note $\sigma((X_i)_{i \in I})$, est la tribu

$$\sigma(X_i : i \in I) = \sigma\left(\underset{i \in I}{\cup} \sigma(X_i) \right).$$

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

En effet, on a

$$\{X = Y\} = \underset{k \in \mathbb{N}}{\cup} \{X = Y\} \cap \{X = k\} = \underset{k \in \mathbb{N}}{\cup} \{X = k\} \cap \{Y = k\}.$$

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.

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

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

Remarque 5.6. *De nombreux théorèmes et exercices de ce cours commenceront ainsi : "soit $(X_n)_{n\ge 1}$ une suite de variables (ou vecteurs) aléatoires indépendant(e)s avec X_i qui suit la loi μ_i". L'existence d'un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ sur lequel peuvent vivre de telles variables n'est pas évidente ; elle peut être obtenue comme conséquence d'un théorème profond de théorie de la mesure, appelé théorème de prolongement de Kolmogorov. Ce théorème ne sera pas démontré ici, nous renvoyons à Billingsley [1] pour une preuve. Dans le cas d'une suite de variables (et non de vecteurs) aléatoires, le résultat sera vu en exercice dans un chapitre ultérieur.*

Théorème 5.7. *Soit $(X_i)_{i\in I}$ une collection de vecteurs aléatoires indépendants. On suppose que X_i est à valeurs dans $\mathbb{R}^{n_i}$. Soit $(f_i)_{i\in I}$ une famille d'applications telle que pour tout i, f_i est 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 équivalente à 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, leurs sous-tribus $(\sigma(Y_i))_{i\in I}$ le sont aussi. □

Exemple: Si X, Y et Z sont indépendantes, alors $\mathrm{ch}(X), Y^2$ et Z^3 sont indépendantes. En fait, nous voudrions pouvoir dire aussi que $\mathrm{ch}(X) + Y^2$ est indépendante de Z^3. Pour cela, il faudrait savoir que (X, Y) est indépendant de Z, auquel cas nous pourrions utiliser les fonctions $f(x, y) = \mathrm{ch}\, x + y^2$ et $g(z) = z^3$. Ceci est vrai. En effet, on a le résultat suivant.

5.2.1 Retour sur l'indépendance des tribus

Théorème 5.8. *Soit $(\mathcal{A}_i)_{i\in I}$ 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 $\sigma(\mathcal{A}_j; j \in J)$ et $\sigma(\mathcal{A}_k; k \in K)$ sont indépendantes.

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

$$\mathcal{C} = \bigcup_{E\subset J,\ E \text{ fini}} \{ \bigcap_{x\in E} A_x ; \forall x \in E \quad A_x \in \mathcal{A}_x\},$$

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

$$\mathcal{D} = \bigcup_{E\subset K,\ E \text{ fini}} \{ \bigcap_{x\in E} A_x ; \forall x \in E \quad A_x \in \mathcal{A}_x\}.$$

Si $B_1 \in \mathcal{C}$, B_1 peut s'écrire sous la forme $B_1 = \bigcap_{x\in E} A_x$, où $E \subset J$ et E fini et où $\forall x \in E$, $A_x \in \mathcal{A}_x$. De même, si $B_2 \in \mathcal{D}$, B_2 peut s'écrire sous la forme $B_2 = \bigcap_{x\in E'} A_x$ où $E' \subset K$ et E' fini et où $\forall x \in E'$, $A_x \in \mathcal{A}_x$. Ainsi

$$\begin{aligned} \mathbb{P}(B_1 \cap B_2) &= \mathbb{P}\left(\bigcap_{x\in(E\cup E')} A_x \right) = \prod_{x\in(E\cup E')} \mathbb{P}(A_x) \\ &= \left(\prod_{x\in E} \mathbb{P}(A_x) \right) \left(\prod_{x\in E'} \mathbb{P}(A_x) \right) = \mathbb{P}(B_1)\mathbb{P}(B_2). \end{aligned}$$

Comme $\mathcal{C}$ est un π-système qui engendre $\sigma(\mathcal{A}_j, j \in J)$ et $\mathcal{D}$ un π-système qui engendre $\sigma(\mathcal{A}_k, k \in K)$, le théorème 3.14 permet de conclure. □

Le résultat s'étend aisément au cas de plus de deux familles de tribus.

Corollaire 5.9. *Soit $(\mathcal{A}_i)_{i\in I}$ une famille de sous-tribus de $(\Omega,\mathcal{F})$ indépendantes sous $\mathbb{P}$. Soient $(I_x)_{x\in K}$ des parties de I deux à deux disjointes. Pour $x \in K$, on note $\mathcal{B}_x$ la tribu engendrée par les $\mathcal{A}_j$, où j décrit I_x. Alors les tribus $(\mathcal{B}_x)_{x\in K}$ sont indépendantes.*

Démonstration. Ainsi qu'on l'a déjà noté, montrer qu'une famille de tribus est indépendantes, c'est montrer que chaque sous-famille finie est indépendante. Ainsi, il nous suffit de montrer le théorème dans le cas où K est fini, mettons $K = \{0,\dots,n\}$. D'après la remarque faite au chapitre 3, il suffit de montrer que pour tout k entre 1 et n, $\mathcal{B}_k = \sigma(\mathcal{A}_j, j \in I_k)$ est indépendante de la tribu engendrée par les $(\mathcal{B}_i)_{0\le i\le k-1}$, qui est aussi la tribu engendrée par les $\mathcal{A}_j$, pour $j \in \bigcup\limits_{0\le i<k} I_i$. Or ce dernier point découle directement du théorème 5.8, ce qui achève la preuve. □

Corollaire 5.10. *Soient $\mathcal{C}_1,\dots,\mathcal{C}_n$ des familles de parties mesurables de $(\Omega,\mathcal{F})$. On suppose que les $\mathcal{C}_i$ sont des π-systèmes contenant chacun Ω et que pour tout n-uplet $(A_i)_{1\le i\le n} \in \mathop{\times}\limits_{i=1}^{n} \mathcal{C}_i$, on a $\mathbb{P}\left(\bigcap\limits_{i=1}^{n} A_i\right) = \prod\limits_{i=1}^{n} \mathbb{P}(A_i)$. Alors, les tribus $\sigma(\mathcal{C}_i)$ sont indépendantes.*

Démonstration. Comme précédemment, il suffit de montrer que pour tout k avec $1 \le k \le n$, $\sigma(\mathcal{C}_k)$ est indépendante de la tribu engendrée par les $(\mathcal{C}_i)_{1\le i<k}$. Or, cette dernière tribu est engendrée par le π-système des éléments qui s'écrivent sous la forme $\bigcap\limits_{i=1}^{k-1} A_i$, avec $(A_i)_{1\le i\le k-1} \in \mathop{\times}\limits_{i=1}^{k-1} \mathcal{C}_i$, d'où le résultat avec le théorème 3.14. □

5.2.2 Vecteurs aléatoires indépendants

Théorème 5.11. *Soient $(\Omega,\mathcal{F},\mathbb{P})$ un espace probabilisé et $X_1,\dots,X_n$ des vecteurs aléatoires. Les deux propositions suivantes sont équivalentes*

1. *$X_1,\dots,X_n$ sont indépendantes.*
2. *$\mathbb{P}_{(X_1,\dots,X_n)} = \mathbb{P}_{X_1}\otimes\dots\otimes\mathbb{P}_{X_n}$*

Démonstration. Montrons $1 \Longrightarrow 2$. On pose $\mathcal{C} = \prod\limits_{i=1}^{n} \mathcal{B}(\mathbb{R}^{n_i})$.
Soit $A = A_1\times\dots\times A_n \in \mathcal{C}$. On a

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

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

Montrons $2 \Longrightarrow 1$. 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 $A = A_1\times\dots\times A_n \in \mathcal{C}$. On a alors

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

ce qui prouve l'indépendance des tribus d'après le corollaire 5.10. □

Corollaire 5.12. *Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X_1, \dots, X_n$ des vecteurs aléatoires. On suppose qu'il existe des mesures de probabilité $\mu_1, \dots, \mu_n$ telles que $\mathbb{P}_{(X_1,\dots,X_n)} = \mu_1 \otimes \cdots \otimes \mu_n$. Alors*

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

Démonstration. Soit B un borélien. On a pour tout i

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

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

5.2.3 Application : loi 0–1 de Kolmogorov

Théorème 5.13. *Soient S un ensemble infini et $(\mathcal{A}_i)_{i \in S}$ une famille de tribus indépendantes sous la loi $\mathbb{P}$. On pose*

$$\mathcal{T} = \bigcap_{\Lambda \subset S,\ \Lambda \text{ fini}} \sigma(\mathcal{A}_k; k \in S \backslash \Lambda).$$

$\mathcal{T}$ est appelée tribu de queue de la famille $(\mathcal{A}_i)_{i \in \Lambda}$. Alors, la tribu $\mathcal{T}$ est triviale, c'est-à-dire que

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

Démonstration. Posons

$$\mathcal{A} = \bigcup_{\Lambda \subset S,\ \Lambda \text{ fini}} \sigma(\mathcal{A}_k; k \in \Lambda).$$

Il n'est pas très difficile de voir que $\mathcal{A}$ est une algèbre. Montrons que l'algèbre $\mathcal{A}$ est indépendante de la tribu $\mathcal{T}$. Soient $A \in \mathcal{A}$ et $B \in \mathcal{T}$. Il existe $\Lambda \subset S$ avec Λ fini tel que $A \in \sigma(\mathcal{A}_k; k \in \Lambda)$. Or, par définition de $\mathcal{T}$, $B \in \sigma(\mathcal{A}_k; k \in \Lambda^c)$. De plus, d'après le théorème 5.8, les tribus $\sigma(\mathcal{A}_k; k \in \Lambda)$ et $\sigma(\mathcal{A}_k; k \in \Lambda^c)$ sont indépendantes, donc $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Ainsi $\mathcal{A}$ est indépendante de $\mathcal{T}$. Comme $\mathcal{A}$ est une algèbre, le théorème 3.14 assure que $\sigma(\mathcal{A})$ est indépendante de $\mathcal{T}$. Or $\mathcal{T} \subset \sigma(\mathcal{A})$, donc $\mathcal{T}$ est indépendante d'elle-même. Soit $A \in \mathcal{T}$. 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 $\mathbb{P}(A) \in \{0, 1\}$. □

Remarque 5.14. *Dans le cas où $S = \mathbb{N}$, la tribu de queue $\mathcal{T}$ coïncide avec la tribu $\mathcal{T}' = \bigcap_{n \geq 0} \sigma(\mathcal{A}_k; k \geq n)$. En effet, pour tout n, la tribu $\sigma(\mathcal{A}_k; k \geq n)$ s'écrit encore $\sigma(\mathcal{A}_k; k \in S \backslash \Lambda)$ avec $\Lambda = \{0, \dots, n-1\}$: cela donne l'inclusion $\mathcal{T} \subset \mathcal{T}'$. Réciproquement, prenons $A \in \mathcal{T}'$ et montrons que $A \in \mathcal{T}$. Prenons $\Lambda \subset S$ avec Λ fini, et posons $n = \max(S) + 1$. Par définition de $\mathcal{T}'$, $A \in \sigma(\mathcal{A}_k; k \geq n)$. Or $\sigma(\mathcal{A}_k; k \geq n) \subset \sigma(\mathcal{A}_k; k \in S \backslash \Lambda)$, donc $A \in \sigma(\mathcal{A}_k; k \in S \backslash \Lambda)$. Comme cela vaut pour tous les $\Lambda \subset S$ avec Λ fini, on a $A \in \mathcal{T}$.*

Si $(X_k)_{k \geq 0}$ est une suite de variables aléatoires indépendantes , la tribu de queue associée aux tribus $\mathcal{A}_k = \sigma(X_k)$ est donc

$$\mathcal{T} = \bigcap_{n \geq 0} \sigma(X_k; k \geq n).$$

La loi 0–1 *de Kolmogorov dit donc que cette tribu est triviale si les* $(X_k)_{k\geq 0}$ *sont des variables indépendantes.*

5.2.4 Variables aléatoires indépendantes et convolutions

Théorème 5.15. *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 $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$. La loi $\mathbb{P}_X * \mathbb{P}_Y$ 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 $(x,y) \mapsto x+y$ n'est rien d'autre que $\mathbb{P}_{X+Y}$. □

Théorème 5.16. *Soient* μ*,* ν *et* γ *trois mesures finies sur* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$*. On a*

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

et

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

Démonstration. Si l'une des trois mesures de la deuxième formule est nulle, chaque produit est nul car la mesure nulle est absorbante pour le produit de convolution ($\mu * 0 = 0 * \mu = 0$). Idem pour la première formule si l'une des deux mesures est nulle. Sinon, posons $\mathbb{P} = \frac{\mu}{\mu(\mathbb{R}^d)} \otimes \frac{\nu}{\nu(\mathbb{R}^d)} \otimes \frac{\gamma}{\gamma(\mathbb{R}^d)}$. $\mathbb{P}$ est une mesure de probabilité. Définissons sur $(\mathbb{R}^d)^3$:

$$X(x,y,z) = x;\ Y(x,y,z) = y;\ Z(x,y,z) = z;\ S = X+Y;\ T = Y+Z.$$

Par associativité de l'indépendance, X et T sont indépendantes, et de même S et Z sont indépendantes. On a donc

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

Il est alors facile de voir que $\mathbb{P}_T = \frac{1}{\nu(\mathbb{R}^d)\gamma(\mathbb{R}^d)} \nu * \gamma$ et $\mathbb{P}_S = \frac{1}{\mu(\mathbb{R}^d)\nu(\mathbb{R}^d)} \mu * \nu$, d'où $(\mu * \nu) * \gamma = \mu * (\nu * \gamma)$. 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 muni de $(+, *)$ forme un semi-anneau commutatif.

5.3 Variables aléatoires discrètes

Définition. *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 $\mathbb{P}(X \in D) = \mathbb{P}_X(D) = 1$, 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 5.17. *Soient X une variable aléatoire discrète, et D un ensemble fini ou dénombrable inclus dans $\mathbb{R}$ tel que $X(\Omega) \subset D$. 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*
$$f(x) = \begin{cases} p_x & \text{si } x \in D, \\ 0 & \text{si } x \notin D. \end{cases}$$

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 $\mu = \sum_{i \in D} p_i \delta_i$. On a alors

$$\mu(A) = \sum_{i \in D} p_i \delta_i(A) = \sum_{i \in D} p_i \, \mathbb{1}_A(i).$$

On a d'une part

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

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. $\square$

Corollaire 5.18. *Soient D un ensemble fini ou dénombrable, et $(p_i)_{i \in D}$ 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

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

$\square$

5.3.1 Fonction d'une variable aléatoire discrète

Théorème 5.19. *La loi image μ_f d'une loi discrète μ par une application f est une loi discrète.*

Démonstration. Soit D un ensemble fini ou dénombrable tel que $\mu(D) = 1$. Comme $f(D)$ est un ensemble fini ou dénombrable, il est mesurable par rapport à la tribu borélienne. En effet, on a pour tout borélien B,

$$(f \circ X)^{-1}(B) = \{\omega \in \Omega : \; X(\omega) \in f^{-1}(B) \cap X(\Omega)\},$$

où $f^{-1}(B) \cap X(\Omega)$ est fini ou dénombrable, donc borélien.
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. □

Corollaire 5.20. *Soient X une variable aléatoire discrète définie sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et f une fonction quelconque définie sur $X(\Omega)$ fini ou dénombrable. Alors la fonction Y définie par*

$$\forall \omega \in \Omega \quad Y(\omega) = f(X(\omega))$$

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

Commençons par quelques rappels. 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 $\lambda-$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. La densité f a les propriétés suivantes :

— pour tout $a, b \in \mathbb{R}, \quad a \leq b$, on a

$$\begin{aligned} \mathbb{P}(a \leq X \leq b) &= \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X \leq b) \\ &= \mathbb{P}(a < X < b) = \int_{[a,b]} f(x)\, d\lambda(x). \end{aligned}$$

— Pour tout $a \in \mathbb{R}$
$\mathbb{P}(a \leq X) = \mathbb{P}(a < X) = \int_{[a,+\infty[} f(x)\, d\lambda(x) = \int_{]a,+\infty[} f(x)\, d\lambda(x)$
et $\mathbb{P}(a \geq X) = \mathbb{P}(a > X) = \int_{]-\infty,a]} f(x)\, d\lambda(x) = \int_{]-\infty,a[} f(x)\, d\lambda(x)$.

— $\int_{\mathbb{R}} f(x)\, d\lambda(x) = 1$.

Remarquons que pour montrer qu'une fonction f positive est une densité de la variable aléatoire X, il suffit de montrer que les mesures $\mathbb{P}_X(x)$ et $f.\lambda$ coïncident sur la famille $(]-\infty, a], a \in \mathbb{R})$, car cette famille engendre les boréliens et la mesure $d\mathbb{P}_X(x)$ est une probabilité (et on utilise alors le corollaire 3.12).

5.4.2 Densités et lois marginales

Théorème 5.21. *Soient $(\Omega, \mathcal{F}, \mu)$ et $(\Omega', \mathcal{F}', \mu')$ deux espaces mesurés σ-finis. On suppose que la mesure ν sur $(\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}')$ admet une densité h par rapport à $\mu \otimes \mu'$. Alors la mesure image ν_π de ν par l'application*

$$\begin{aligned} \pi : \Omega \times \Omega' &\to \Omega \\ (x, x') &\mapsto x \end{aligned}$$

admet comme densité par rapport à μ la fonction $f(x) = \int_{\Omega'} h(x, x') d\mu'(x')$.

Démonstration. Soit $B \in \mathcal{F}$. On a d'après Tonelli

$$\begin{aligned} \nu_\pi(B) &= \nu(\pi^{-1}(B)) = \nu(B \times \Omega') = \int_{B \times \Omega'} d\nu(x, x') \\ &= \int_{B \times \Omega'} h(x, x') d(\mu \otimes \mu')(x, x') \\ &= \int_B \left(\int_{\Omega'} h(x, x') d\mu'(x') \right) d\mu(x) \\ &= \int_B f(x) d\mu(x), \end{aligned}$$

ce qui prouve le résultat. □

Théorème 5.22. *Soient 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 la relation de composition $X = \pi \circ (X, Y)$.

$$\begin{tikzcd} (\Omega, \mathcal{F}, \mathbb{P}) \arrow[rr, "{(X,Y)}"] \arrow[dr, "X"'] & & (\mathbb{R}^n \times \mathbb{R}^p, \mathcal{B}(\mathbb{R}^n) \otimes \mathcal{B}(\mathbb{R}^p), \mathbb{P}_{(X,Y)}) \arrow[dl, "\pi"] \\ & (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbb{P}_X) & \end{tikzcd}$$

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 5.23. *Soient 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, ce qui se note $h = f \otimes g$.*

Démonstration. Comme X et Y sont indépendantes sous $\mathbb{P}$, on a
$\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$. Ainsi

$$\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y = f\lambda^n \otimes g\lambda^p = (f \otimes g)\lambda^n \otimes \lambda^p = h\lambda^{n+p},$$

ce qui montre bien que (la loi de) (X,Y) admet h comme densité. □

Théorème 5.24. *Soient 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 s'écrivant $h(x,y) = f(x)g(y)$, où f et g sont des fonctions positives. Alors X et Y sont indépendantes. De plus, X admet comme densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$ la fonction*

$$x \mapsto \frac{f(x)}{\int_{\mathbb{R}^n} f(x')\, d\lambda^n(x')}$$

et Y admet comme densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^p$ la fonction

$$y \mapsto \frac{g(y)}{\int_{\mathbb{R}^p} g(y')\, d\lambda^p(y')}.$$

Démonstration. Posons $A = \int_{\mathbb{R}^n} f(x) d\lambda^n(x)$ et $B = \int_{\mathbb{R}^p} g(y) d\lambda^p(y)$. Comme f et g sont positives, on a $A \geq 0$ et $B \geq 0$. Comme h est une densité, on a par Tonelli

$$\begin{aligned} 1 &= \int_{\mathbb{R}^n \times \mathbb{R}^p} h(x,y) d(\lambda^n \otimes \lambda^p)(x,y) = \int_{\mathbb{R}^n \times \mathbb{R}^p} f(x)g(y) d(\lambda^n \otimes \lambda^p)(x,y) \\ &= \left(\int_{\mathbb{R}^n} f(x) d\lambda^n(x) \right) \left(\int_{\mathbb{R}^p} g(y) d\lambda^p(y) \right) \\ &= AB. \end{aligned}$$

On peut donc dire que A et B sont tous deux strictement positifs. Ainsi les fonctions f/A et g/B sont bien définies. Elles sont positives, et, par construction, chacune d'elle admet 1 comme intégrale par rapport à la mesure de Lebesgue. Ainsi $\mu = \frac{f}{A}\lambda^n$ et $\nu = \frac{g}{B}\lambda^p$ sont des mesures de probabilité. Comme dans le théorème précédent, la densité de $\mu \otimes \nu$ est $h(x,y) = \frac{f(x)}{A}\frac{g(y)}{B}$. Donc $h(x,y) = \frac{f(x)}{A}\frac{g(y)}{B} = \frac{f(x)g(y)}{AB}$. On en déduit que $\mathbb{P}_{(X,Y)} = \mu \otimes \nu$. Il suffit alors d'appliquer le corollaire 5.12 pour conclure. □

5.5 Variables et lois discrètes classiques

5.5.1 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

$$\mathbb{1}_A(x) = \begin{cases} 1 \text{ si } x \in A \\ 0 \text{ si } x \notin A. \end{cases}$$

$\mathbb{1}_A$ est une variable aléatoire à valeurs dans $\{0;1\}$.

5.5.2 Mesure de Dirac

On appelle mesure de Dirac au 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 5.25. *1. La variable aléatoire constante égale à x suit la loi δ_x. $\mathbb{1}_\Omega$ suit la loi δ_1 tandis que $\mathbb{1}_\varnothing$ suit la loi δ_0.*

2. 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 $\mathbb{P}(X = 0) = 1 - p$.

Remarque 5.26 (importante). *— Pour tout événement 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 $X(\omega) = \mathbb{1}_{X(\omega)=1}$. (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

$$\mathcal{P}(\Omega) \to [0,1], \qquad A \mapsto \frac{|A \cap E|}{|E|}.$$

Il s'agit du nombre de cas favorables sur le nombre de cas possibles. Ainsi, une variable aléatoire X suit la loi uniforme sur E si l'on a

$$\forall x \in E, \quad \mathbb{P}(X = x) = \frac{1}{|E|},$$

et bien entendu, pour $x \notin E$, on a $\mathbb{P}(X = x) = 0$.

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 paramètre p. Ainsi, $\mathcal{B}(n,p) = (\mathrm{Ber}(p))^{*n}$.

Théorème 5.27. *$\mathcal{B}(n,p)$ charge les entiers $\{0,\dots,n\}$. Plus précisément, on a pour tout $k \in \{0,\dots,n\}$*

$$\mathcal{B}(n,p)(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}.$$

Démonstration. Posons $\mu = \mathcal{B}(n,p)$. On a

$$\mu = (\mathrm{Ber}(p))^{*n} = ((1-p)\delta_0 + (p\delta_1))^{*n}.$$

On remarque que $\delta_1^{*k} = \delta_k$. Comme l'ensemble des mesures positives muni de (+,*) est un semi-anneau commutatif, la formule du binôme de Newton s'applique et 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 5.28. *Soient $A_1, \dots, A_n$ des 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 5.29. *X est le nombre d'événements A_k 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 5.30. *Soit $(A_i)_{i\in\mathbb{N}^*}$ une suite d'événements indépendants de même probabilité $p > 0$. On pose*

$$X(\omega) = \min\{k \in \mathbb{N}^* \quad \omega \in A_k\} \textit{ si } \{k \in \mathbb{N}^*,\ \omega \in A_k\} \neq \varnothing,$$

et $X(\omega) = +\infty$ sinon. Alors X suit la loi géométrique de paramètre p. De plus, F est continue à droite, constante sur les intervalles $[k, k+1[$ et

$$\forall k \in \mathbb{N}, \quad F_X(k) = 1 - (1-p)^k. \tag{5.1}$$

Démonstration. On a

$$\{X > k\} = \bigcap_{i\in\{1,\dots,k\}} A_i^c.$$

Comme les A_i appartiennent à $\mathcal{F}$, les A_i^c aussi, et donc puisqu'on peut écrire $\{X > k\}$ comme intersection finie d'éléments de $\mathcal{F}$, $\{X > k\}$ est dans $\mathcal{F}$, et donc, par passage au complémentaire

$$\{X \leq k\} \in \mathcal{F}.$$

En utilisant l'indépendance des (A_i), on obtient

$$\mathbb{P}(X > k) = (1-p)^k.$$

Comme

$$\{X = +\infty\} = \bigcap_{k=1}^{+\infty} \{X > k\},$$

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 pour tout $k \in \mathbb{N}^*$

$$\mathbb{P}(X = k) = \mathbb{P}(X > k-1) - \mathbb{P}(X > k) = (1-p)^{k-1} - (1-p)^k = (1-p)^{k-1}p.$$

On conclut car on a de plus

$$\forall k \in \mathbb{N} \quad F_X(k) = \mathbb{P}(X \leq k) = 1 - \mathbb{P}(X > k) = 1 - (1-p)^k.$$

□

Exemple: On lance une pièce de monnaie équilibrée jusqu'à obtention du premier "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

$$\forall k \in \mathbb{N} \quad \mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Cela définit bien une loi de probabilité car $e^{-\lambda}\frac{\lambda^k}{k!} \geq 0$ et

$$\sum_{k=0}^{+\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1.$$

On la note $\mathcal{P}(\lambda)$.

5.5.8 Loi hypergéométrique

La loi hypergéométrique $\mathcal{H}(N, n, k)$ modélise le phénomène suivant. Soit une population de N individus, composée de deux types distincts (par exemple on a n individus de taille supérieure ou égale à 1,80 m, et $N - n$ individus mesurant moins de 1,80 m). On tire au hasard k individus dans cette population. On compte ensuite le nombre d'individus possédant un certain type (par exemple mesurant plus de 1,80 m).

De manière théorique, cela s'énonce comme suit.

Proposition 5.31. *La loi hypergéométrique est la loi image de la loi uniforme sur* $\Omega = \mathcal{B}_k(\{1, \ldots, N\})$ *par l'application*

$$\begin{array}{rcl} X : \mathcal{B}_k(\{1, \ldots, N\}) & \to & \mathbb{N} \\ \omega & \mapsto & X(\omega) = |\{1, \ldots, n\} \cap \omega|. \end{array}$$

Ainsi, pour $i \in \{0, \ldots, \min(n, k)\}$*, on a*

$$\mathbb{P}(X = i) = \mathcal{H}(N, n, k)(i) = \frac{\binom{n}{i}\binom{N-n}{k-i}}{\binom{N}{k}}.$$

Démonstration. Notons $\mathbb{P}$ la loi uniforme sur Ω. On a

$$\mathcal{H}(N, n, k)(i) = \mathbb{P}(\omega \in S),$$

où $S = \{\omega \in \mathcal{B}_k(\{1,\dots,N\}); |\{1,\dots,n\} \cap \omega| = i\}$. L'application

$$\begin{array}{rcl} \mathcal{B}_i(\{1,\dots,n\}) \times \mathcal{B}_{k-i}(\{n+1,\dots,N\}) & \to & S \\ (A,B) & \mapsto & A \cup B \end{array}$$

est une bijection, donc

$$|S| = |\mathcal{B}_i(\{1,\dots,n\}) \times \mathcal{B}_{k-i}(\{n+1,\dots,N\})| = \binom{n}{i}\binom{N-n}{k-i}.$$

Comme $\mathbb{P}$ est la loi uniforme sur Ω, et $|\Omega| = \binom{N}{k}$, le résultat s'ensuit. $\square$

5.6 Lois à densité usuelles

5.6.1 Loi uniforme

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 $\mathbb{R}^d$ si elle admet la densité

$$x \mapsto \frac{1}{\lambda(K)} \mathbb{1}_K(x).$$

Loi uniforme sur un intervalle

On dit qu'une variable aléatoire X suit la loi uniforme sur l'intervalle $[a,b]$ si elle admet comme densité

$$x \mapsto \frac{1}{b-a} \mathbb{1}_{[a,b]}(x).$$

Théorème 5.32. *Soit F une fonction de $\mathbb{R}$ dans $\mathbb{R}$, croissante, continue à droite, dont la limite est nulle en $-\infty$ et vaut 1 en $+\infty$. On suppose que sur $(\Omega, \mathcal{F}, \mathbb{P})$, U est une variable aléatoire suivant la loi uniforme sur $[0,1]$. On pose*

$$\forall u \in]0,1[\quad Q^*(u) = \min\{x \in \mathbb{R} :\ 1 - F(x) \le u\}.$$

et

$$\forall u \in]0,1[\quad F^*(u) = \inf\{s \in \mathbb{R} :\ F(s) > u\}.$$

Alors $F^*(U)$ et $Q^*(u)$ sont des variables aléatoires réelles dont la fonction de répartition est F.

Démonstration. On fait seulement la preuve pour F^* ; celle pour Q^* est analogue. Notons que pour tout $u \in]0,1[$, $F^*(u) < +\infty$ car F a une limite 1 en l'infini et $F^*(u) > -\infty$ car F a une limite 0 en l'infini. Comme U prend presque sûrement ses valeurs dans $]0,1[$, la variable $F^*(U)$ est bien définie. On va calculer la fonction de répartition de $F^*(U)$, et on doit donc travailler sur l'événement $\{F^*(U) \le t\}$. Si $F(t) > U$, alors $F^*(U) \le t$ par définition de l'$\inf$. D'autre part, si $F^*(U) \le t$, alors, comme F est croissante, pour tout $t' > t$, $F(t') > U$. Donc pour tout $t' > t$, on a

$$\{F(t) > U\} \subset \{F^*(U) \le t\} \subset \{F(t') > U\}$$

et donc $\mathbb{P}(U < F(t)) \le \mathbb{P}(F^*(U) \le t) \le \mathbb{P}(U < F(t'))$. On obtient ainsi $F(t) \le \mathbb{P}(F^*(U) \le t) \le F(t')$. Cela est vrai pour tout $t' > t$. Comme F est continue à droite, on obtient ainsi le résultat en faisant tendre t' vers t. $\square$

En particulier, si F est la fonction de répartition de la loi μ, $F^*(U)$ suit la loi μ. Cela signifie que si on sait simuler une variable aléatoire suivant la loi uniforme sur $[0,1]$, on sait simuler n'importe quelle variable aléatoire réelle : c'est la simulation par méthode d'inversion. Mais ce résultat a aussi une conséquence théorique importante.

Corollaire 5.33. *Soit F une fonction de $\mathbb{R}$ dans $\mathbb{R}$, croissante, continue à droite, dont la limite est nulle en $-\infty$ et vaut 1 en $+\infty$. Alors il existe une mesure de probabilité sur $\mathbb{R}$ dont F est la fonction de répartition.*

Démonstration. Il suffit de prendre la loi de $F^*(U)$ dans le théorème précédent. □

Remarque 5.34. — *Si $a, b \in \overline{\mathbb{R}}$ sont tels que*

$$\lim_{x\to a^+} F(a) = 0 \textit{ et } \lim_{x\to b^-} F(b) = 1$$

et que F est strictement croissante sur $]a,b[$ (ce qui arrive par exemple si la loi admet une densité de la forme $\mathbb{1}_{]a,b[}g$ avec g strictement positive sur $]a,b[$), alors l'application F^* du théorème 5.32 est tout simplement la réciproque de l'application strictement croissante F. Si on sait la calculer explicitement, cela permet une simulation facile.

— *En revanche, si on cherche à simuler une loi discrète μ avec $\mu(\{x_i\}) = p_i$ pour tout $i \geq 1$, il suffit de poser $X = x_{f(U)}$, où $f(x) = \inf\{n : s_n > x\}$, avec $s_0 = 0$ et $s_n = p_1 + \cdots + p_n$.*
En effet, on a alors $\{X = x_n\} = \{f(U) = i\} = \{s_{n-1} \leq U < s_n\}$ et $\mathbb{P}(X = x_n) = \mathbb{P}(U \in [s_{n-1}, s_n[) = \lambda([s_{n-1}, s_n[) = s_n - s_{n-1} = p_n$. Noter que les x_i n'ont pas besoin d'être ordonnés.

5.6.2 Loi gaussienne

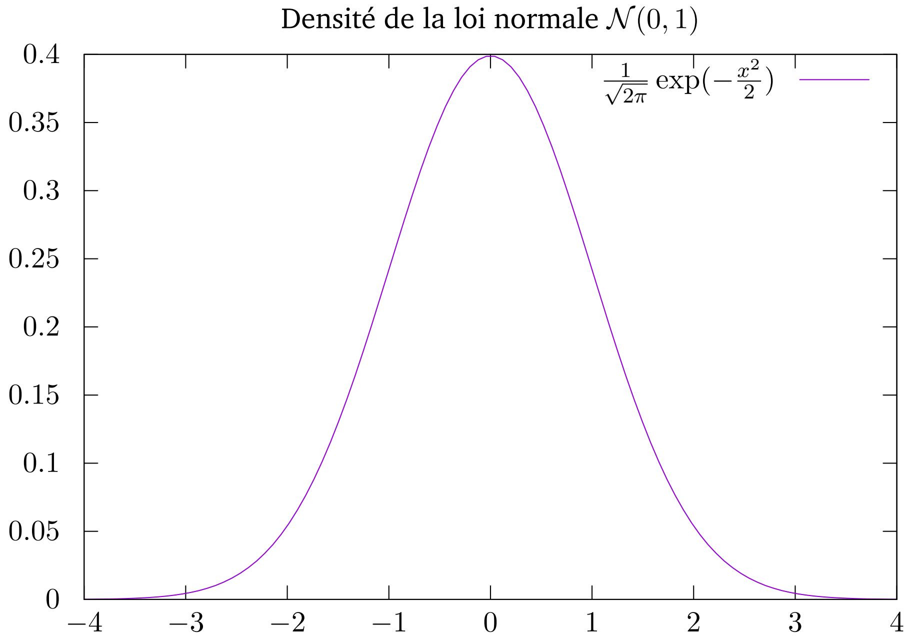
Soient $m \in \mathbb{R}$ et $\sigma^2 > 0$. On dit qu'une variable aléatoire X suit la loi gaussienne $\mathcal{N}(m, \sigma^2)$ de paramètres m et σ^2 si elle admet la densité

$$x \mapsto \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-m)^2}{2\sigma^2}).$$

On emploie également le mot "normale" à la place de "gaussienne". Ces deux termes signifient exactement la même chose.

On dit qu'une variable gaussienne est centrée lorsque $m = 0$. On dit aussi qu'une variable gaussienne est réduite lorsque $\sigma^2 = 1$. Ces deux notions seront revues plus tard dans le cas d'autres lois de probabilité.

Énonçons 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 $\sigma X + m \sim \mathcal{N}(m, \sigma^2)$.



5.6.3 Loi exponentielle

Soit $a > 0$. On dit qu'une variable aléatoire X suit la loi exponentielle de paramètre a si elle admet comme densité

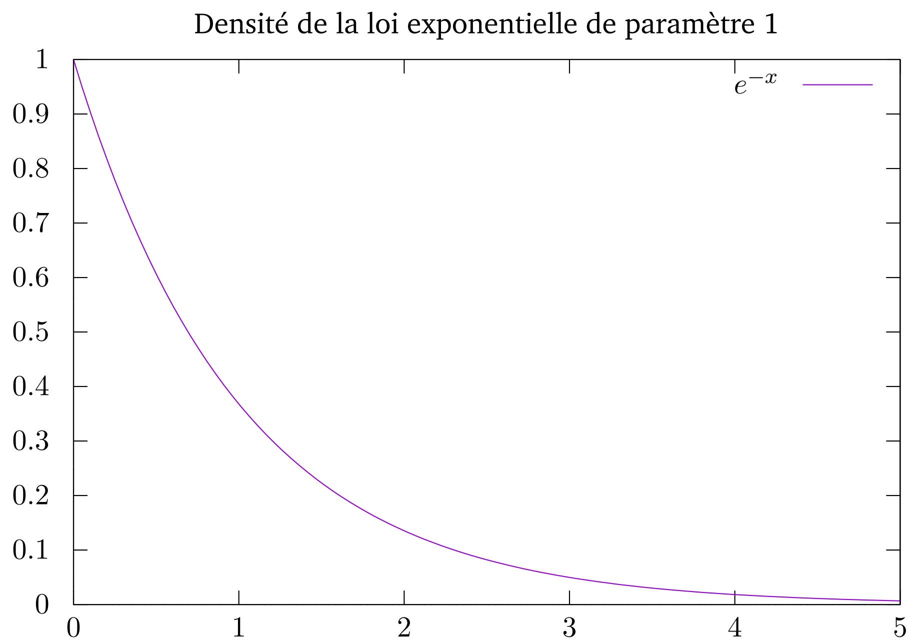
$$x \mapsto ae^{-ax}\, \mathbb{1}_{\mathbb{R}_+}(x).$$

Il est important de noter qu'on a l'identité

$$\forall t \geq 0 \quad \mathbb{P}(X > t) = \int_t^{+\infty} ae^{-ax}\, dx = e^{-at},$$

ce qui est à rapprocher de l'équation (5.1) qui caractérise la queue des lois géométriques.

La loi exponentielle est fréquemment utilisée pour modéliser la durée de vie de composants électroniques qui sont réputés ne pas subir de phénomène d'usure. Cela s'explique en particulier par la propriété d'absence de mémoire, que nous verrons en exercice corrigé.



On peut noter que la loi exponentielle est simple à simuler. En effet, la fonction de répartion de la loi exponentielle de paramètre $a > 0$ vaut $x \mapsto 1 - e^{-ax}$ sur $\mathbb{R}_+$, qui s'inverse aisément en $x \mapsto -\frac{1}{a}\log(1-x)$. Avec la méthode d'inversion, on en déduit que si U suit la loi uniforme sur $[0,1]$, alors $F^*(U) = -\frac{1}{a}\log(1-U) \sim \mathcal{E}(a)$. De même $Q^*(U) = -\frac{1}{a}\log(U) \sim \mathcal{E}(a)$, ce qui est logique car U et $1-U$ ont même loi.

5.6.4 Loi de Cauchy

Soient $a \in \mathbb{R}$, $b > 0$. La loi de Cauchy $\mathcal{C}(a,b)$ de paramètres a, b admet comme densité par rapport à la mesure de Lebesgue :

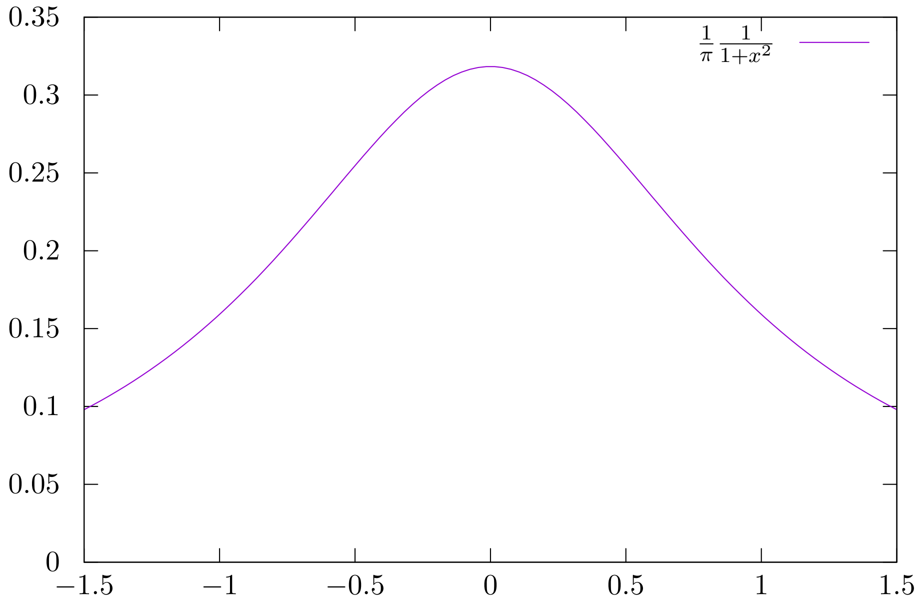
$$x \mapsto \frac{1}{\pi}\frac{b}{(x-a)^2+b^2}.$$

Exemple: Un gyrophare envoie un flash lumineux dans une direction aléatoire d'angle θ. On place un écran de longueur infinie à une distance 1 du gyrophare. On cherche la distribution de l'abscisse du point d'impact du rayon lumineux sur l'écran.
On sait que l'angle θ est une variable aléatoire uniformément distribuée sur l'intervalle $[-\pi/2, \pi/2]$. L'abscisse est alors donnée par $\tan\theta$, qui est donc une variable aléatoire dont la fonction de répartition est donnée, pour $t \in \mathbb{R}$, par

$$\begin{aligned} F(t) &= \mathbb{P}(\tan\theta \le t) = \mathbb{P}(\theta \le \arctan t) = \int_{-\infty}^{\arctan t} f_\theta(x)\,dx \\ &= \int_{-\infty}^{\arctan t} \frac{1}{\pi}\mathbb{1}_{[-\pi/2,\pi/2]}(x)\,dx = \frac{1}{\pi}\arctan t + \frac{1}{2}. \end{aligned}$$

La fonction F est de classe C^1 sur $\mathbb{R}$, de dérivée $f(t) = \frac{1}{\pi(t^2+1)}$. D'après le théorème 5.5, on en déduit que $\tan\theta$ est une variable aléatoire de densité f, donc elle suit la loi de Cauchy $\mathcal{C}(0,1)$.

Densité de la loi de Cauchy $\mathcal{C}(0,1)$



Exercice : Comment simuler la loi de Cauchy $\mathcal{C}(0,1)$?

5.6.5 Loi Gamma

Soient a et γ des réels strictement positifs. On appelle loi Gamma $\Gamma(a,\gamma)$ la loi dont la densité par rapport à la mesure de Lebesgue est

$$x \mapsto \frac{\gamma^a}{\Gamma(a)} x^{a-1} e^{-\gamma x}\, \mathbb{1}_{]0,+\infty[}(x),$$

où $\Gamma(a)$ est la valeur au point a de la fonction Γ, fonction d'Euler de seconde espèce, définie par

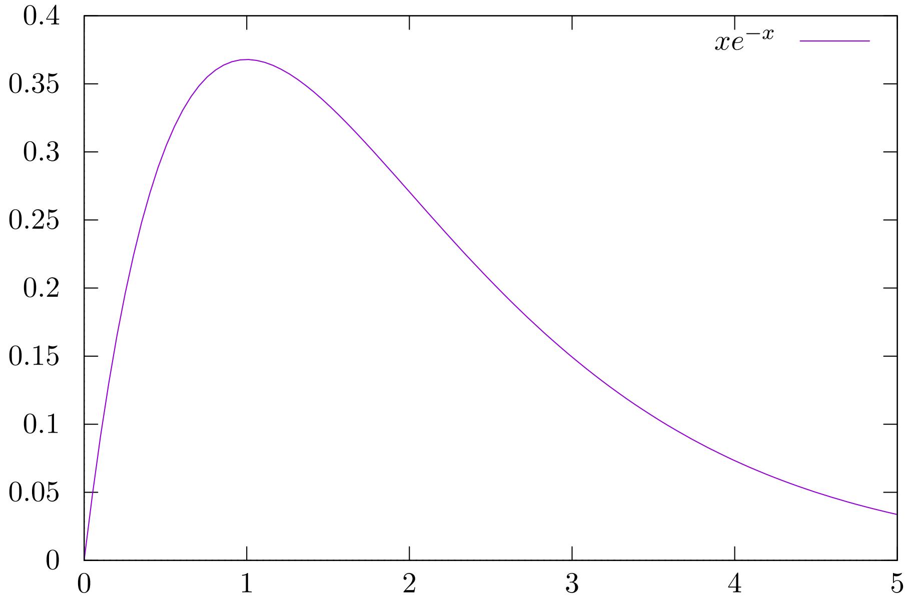
$$\Gamma(a) = \int_{\mathbb{R}+} x^{a-1} e^{-x}\, dx.$$

On dit parfois que a est le paramètre de forme et γ le paramètre d'échelle de la loi. En effet, on montrera plus loin que si $X \sim \Gamma(a,\gamma)$, alors pour tout $\mu > 0$, on a $\frac{1}{\mu}X \sim \Gamma(a,\gamma\mu)$.

On rappelle quelques propriétés classiques de la fonction Γ qui ont été démontrées en exercice dans le chapitre précédent :

- $\forall a > 0 \quad \Gamma(a+1) = a\Gamma(a)$.
- $\forall n \in \mathbb{N} \quad \Gamma(n+1) = n!$

Densité de la loi Gamma $\Gamma(2,1)$



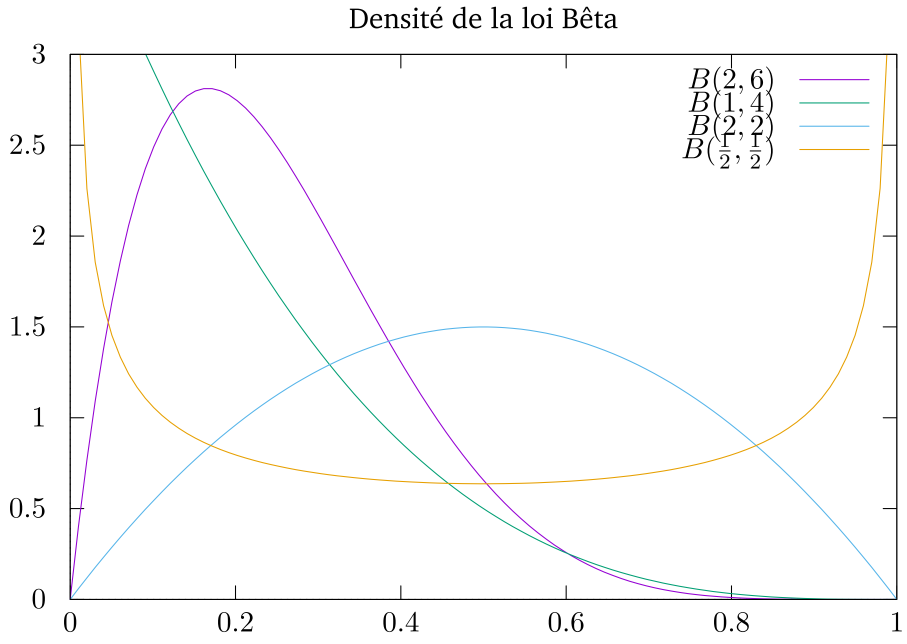
5.6.6 Loi Bêta

Soient a, b des réels strictement positifs. La densité de probabilité de la loi Bêta de paramètres a et b est :

$$x \mapsto \frac{1}{\beta(a,b)} x^{a-1}(1-x)^{b-1}\, \mathbb{1}_{[0,1]}(x),$$

où $\beta(a,b)$ est la fonction Bêta, fonction d'Euler de première espèce qui peut s'exprimer comme

$$\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\, dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$



5.7 Exercices sur les lois

5.7.1 Exercices de la série 1

Exercice 86. Soient X_1 et X_2 deux variables aléatoires indépendantes telles que

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}.$$

Posons $X_3 = X_1X_2$. Montrer que les variables X_2 et X_3 sont indépendantes, ainsi que X_1 et X_3. En revanche, montrer que X_1, X_2, X_3 ne le sont pas.

Exercice 87. La fonction de répartition F de la variable aléatoire X a pour valeurs :

$$0 \text{ si } x < 0,\ \frac{x}{4} \text{ si } 0 \leq x < 1,\ \frac{x+1}{4} \text{ si } 1 \leq x < 2,\ \frac{11}{12} \text{ si } 2 \leq x < 3,\ 1 \text{ si } x \geq 3.$$

a) Calculer $\mathbb{P}(X = i)$ pour $i = 1, 2, 3$.
b) Calculer $\mathbb{P}\left(\frac{1}{2} < X < \frac{3}{2}\right)$.

Exercice 88. Soient $F_j : \mathbb{R} \longrightarrow \mathbb{R}$, $j = 1, 2$, des fonctions données par

$$\begin{aligned} F_1(x) &:= \sum_{i=1}^{+\infty} 2^{-i} \mathbb{1}_{[\frac{1}{i},\infty[} \\ F_2(x) &:= e^{-e^{-x}}. \end{aligned}$$

Montrer qu'il s'agit de fonctions de répartition de deux probabilités sur l'espace mesurable $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$.

Posons $\mathbb{P}_i(]-\infty, x]) := F_i(x)$, $i = 1, 2$. Calculer les probabilités (pour $\mathbb{P}_1$ et $\mathbb{P}_2$) des événements suivants :

$$[1, +\infty[\ ;\ \left[\frac{1}{10}, +\infty\right[\ ;\ \{0\}\ ;\ \left[0, \frac{1}{2}\right[\ ;\ ]-\infty, 0[\ ;\ ]0, +\infty[\,.$$

Exercice 89. Soit Φ un angle uniformément distribué dans l'intervalle $\left[0, \frac{\pi}{2}\right[$. Quelle est la fonction de répartition de la variable aléatoire $y = \tan\Phi$?
Même question si Φ est uniformément distribué dans l'intervalle $\left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$.

Exercice 90. Trouver la constante c telle que la fonction cx^{-3} soit une densité sur l'intervalle $[1, +\infty[$ puis sur l'intervalle $[-2, -1[$.

Exercice 91. Soit X une variable aléatoire uniforme sur l'intervalle [1,3]. Chercher la loi de la variable aléatoire $Y = \frac{1}{2}(|X - 1| + |X|)$.

Exercice 92. Soit Y une variable aléatoire uniformément répartie sur]0,5[. Quelle est la probabilité que les racines de l'équation

$$4x^2 + 4xY + Y + 4 = 0$$

soient réelles ?

Exercice 93. Les points X et Y sont répartis uniformément et indépendamment respectivement sur les côtés AB et BC d'un triangle ABC. Soient S_{ABC} l'aire du triangle ABC et S_{XBY} celle du triangle XBY.
Trouver $\mathbb{P}(S_{ABC} > 2S_{XBY})$.

Exercice 94. Soit X le nombre de succès lors de n expériences. La probabilité d'avoir un succès dans une expérience est de p, avec $0 < p < 1$.
Trouver la valeur de X la plus probable (avec la plus grande probabilité).

Exercice 95. On dit qu'une variable aléatoire est sans mémoire si et seulement si pour tout $s, t \geq 0$

$$(*) \qquad \mathbb{P}(X > s + t \mid X > t) = \mathbb{P}(X > s) > 0.$$

1. Expliquer en quelques mots le contenu de cette égalité.
2. Montrer qu'une variable aléatoire exponentielle est sans mémoire.
3. Montrer qu'une variable aléatoire sans mémoire est distribuée exponentiellement. Indication : considérer la fonction g définie par $g(t) = \mathbb{P}(X > t)$ et commencer par exprimer, pour tout $n \in \mathbb{N}^*$, $g(n)$ et $g(1/n)$ à l'aide de $g(1)$.

Exercice 96. Soient X, Y, Z des variables aléatoires indépendantes, où X et Y suivent la loi géométrique $\mathcal{G}(1/2)$ et Z la loi uniforme sur l'ensemble $\{-1; 0; 1\}$. Calculer la loi de $X - Y$. La comparer avec la loi de ZX.

Exercice 97. Soient X et Y deux variables aléatoires indépendantes suivant la loi uniforme sur $[-1, 1]$.

1. Montrer que le vecteur (X, Y) suit la loi uniforme sur un ensemble que l'on déterminera.
2. Calculer $\mathbb{P}(1/2 < X^2 + Y^2 < 1)$.

Exercice 98. Déterminer A tel que $(x, y) \mapsto A(x^2 + y)\mathbb{1}_{[0,1]^2}(x, y)$ soit la densité d'une probabilité sur $\mathbb{R}^2$. Soit (X, Y) un vecteur aléatoire admettant cette densité. Calculer $\mathbb{P}(X > Y)$.

Exercice 99. Soient $(X_1, \ldots, X_n)$ 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 $T = \inf(X_1, \ldots, X_n)$. Pour t réel, calculer $\mathbb{P}(T > t)$. En déduire que T suit une loi exponentielle de paramètre $\lambda_1 + \cdots + \lambda_n$.

Exercice 100.

1. Soit μ une mesure de probabilité sur $\mathcal{B}(\mathbb{R}^d)$. On suppose que pour tout $A \in \mathcal{B}(\mathbb{R}^d)$, $\mu(A) \in \{0, 1\}$. Montrer qu'il existe $a \in \mathbb{R}^d$, avec $\mu(\{a\}) = 1$.

2. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles indépendantes sur $(\Omega, \mathcal{F}, \mathbb{P})$.
   On pose $X = \overline{\lim_{n\to+\infty}} X_n$. Montrer qu'il existe $a \in \overline{\mathbb{R}}$ tel que $\mathbb{P}(X = a) = 1$.

Exercice 101. Soient X et Y deux variables aléatoires réelles indépendantes. Supposons que $\mathbb{P}(X + Y = \alpha) = 1$ où α est une constante. Montrer que X et Y sont des constantes presque sûrement.

Exercice 102. 1. Montrer que deux probabilités sur $(\mathbb{N}^*, \mathcal{P}(\mathbb{N}^*))$ qui coïncident sur les ensembles de la forme $(n\mathbb{N}^*)_{n\in\mathbb{N}^*}$ sont égales.

2. Soit $s > 1$. On dit que X suit une loi Zêta de paramètre s si l'on a
$$\forall n \in \mathbb{N}^* \quad \mathbb{P}(X = n) = \frac{1}{\zeta(s)}\frac{1}{n^s},$$
   où $\zeta(s) = \sum_{n=1}^{+\infty} \frac{1}{n^s}$. Soient X, X' deux variables aléatoires indépendantes suivant les lois Zêta de paramètres respectifs $s > 1$ et $t > 1$. Montrer que $X \wedge X'$ (p.g.c.d. de X et X') suit la loi Zêta de paramètre $s + t$.
   En déduire que $\mathbb{P}(X \wedge X' = 1) = \frac{1}{\zeta(s+t)}$.

5.7.2 Exercices de la série 2

Exercice 103. Soit $s > 1$. On dit que X suit une loi Zêta de paramètre s si
$$\forall n \in \mathbb{N}^* \quad \mathbb{P}(X = n) = \frac{1}{\zeta(s)}\frac{1}{n^s},$$
où l'on a posé $\zeta(s) = \sum_{n=1}^{+\infty} \frac{1}{n^s}$. Soit donc X suivant une loi Zêta de paramètre s. On tire Y au hasard – c'est-à-dire avec équiprobabilité – entre 1 et X : $\mathbb{P}_{Y|X=x} = \mathcal{U}(\{1, \dots, x\})$ la loi uniforme sur $\{1, \dots, x\}$.

1. Pour $n, k \in \mathbb{N}^*$, calculer $\mathbb{P}(Y = k|X = n)$.
2. On pose $Z = \frac{Y}{X}$. Montrer que la fonction de répartition F_Z est strictement croissante sur $[0, 1]$.
3. Soient p, q deux entiers positifs premiers entre eux, avec $p \leq q$. Calculer $\mathbb{P}(Z = \frac{p}{q})$.
4. On rappelle que $\phi(n)$ désigne le nombre d'entiers compris entre 1 et n qui sont premiers avec n. Déduire de ce qui précède une preuve probabiliste de l'identité
$$\zeta(s+1)\sum_{n=1}^{+\infty}\frac{\phi(n)}{n^{s+1}} = \zeta(s).$$

Exercice 104. Soit $\mathcal{P}$ l'ensemble des nombres premiers. Pour $p \in \mathcal{P}$, on note $\nu_p(n)$ le plus grand entier k tel que p^k divise n. Soit X une variable aléatoire suivant la loi Zêta de paramètre s (voir exercice précédent). Montrer que les variables aléatoires $(1+\nu_p(X))_{p\in\mathcal{P}}$ sont des variables indépendantes, avec $1 + \nu_p(X) \sim \mathcal{G}(1 - \frac{1}{p^s})$.

Exercice 105. 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.

Exercice 106. Donner un exemple de deux lois distinctes sur $(\Omega, \mathcal{F})$ coïncidant sur un système $\mathcal{C}$ engendrant $\mathcal{F}$.

Exercice 107. On choisit de manière uniforme sur $[0,1]$ un réel Y. Quelle est la probabilité pour que le polynôme
$$p(x) = x^2 + x + Y$$
ait des racines réelles ? des racines distinctes ?

Exercice 108. Dans le segment $[AB]$ de longueur 1, on choisit au hasard un point M. Quelle est la probabilité pour que l'on ait $AM.MB \geq \frac{2}{9}$?

Exercice 109. Soit $(X_i)_{i\geq 1}$ des variables aléatoires indépendantes de loi uniforme sur $[0,1]$. On pose $M_n = \max(X_1,\ldots,X_n)$ et $m_n = \min(X_1,\ldots,X_n)$.

1. Déterminer la fonction de répartition de M_n. Montrer que M_n admet une densité que l'on déterminera.
2. Montrer que M_n et $1-m_n$ ont même loi.

Exercice 110. Soit X une variable aléatoire suivant la loi $\mathcal{B}(2n,1/2)$. On pose $Y = |X-n|$. Déterminer la loi de Y.

Exercice 111. Soit (X,Y) un vecteur aléatoire suivant la loi uniforme sur le rectangle $[-1,2]\times[-1,1]$. Montrer que
$$\mathbb{P}(1-Y \geq 2|X|) = \frac{1}{3}.$$

Exercice 112. *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
$$\Psi(x) \sim \frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-\frac{x^2}{2}}.$$

Exercice 113. *Encore une loi* 0–1.

1. Soit $(\mathcal{A}_n)_{n\geq 1}$ une suite de tribus indépendantes. Montrer que la tribu $\mathcal{Q} = \bigcap\limits_{k=1}^{+\infty} \sigma(\mathcal{A}_i; i \geq k)$ est triviale.
2. 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.

Exercice 114. La tradition veut que l'Épiphanie soit l'occasion de « tirer les rois » : une fève est cachée dans une galette, qui est elle-même 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 ! Quelle 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 fève de 2,7 centimètres de diamètre dans une galette de 23 centimètres de diamètre achetée ce matin.

Exercice 115. Soit n un entier naturel. On considère X une variable aléatoire exponentielle de paramètre 1 et Y une binomiale $\mathcal{B}(n,\frac{1}{2})$. 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 Rappels sur la construction de l'espérance

Définition. *Si X est une variable aléatoire réelle 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 6.1. *En toute rigueur, il faudrait écrire $\mathbb{E}_{\mathbb{P}}X$.*

Remarque 6.2. *Si X est une variable aléatoire réelle intégrable, le théorème de transfert dit que $\mathbb{E}X = \int_{\mathbb{R}} x\ d\mathbb{P}_X(x)$; ainsi l'existence de l'espérance ainsi que son éventuelle valeur ne dépendent que de la loi de 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 $X = (X_1, \ldots, X_n)$ est un vecteur aléatoire dont toutes les composantes sont intégrables, on note $\mathbb{E}X$ le vecteur $(\mathbb{E}X_1, \ldots, \mathbb{E}X_n)$.*

6.2 Propriétés élémentaires

— $\mathcal{L}^1$ est un espace vectoriel.
— $\forall A \in \mathcal{F}, \quad \mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$.
— $\forall X, Y \in \mathcal{L}^1, \quad \mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$.
— $\forall X \in \mathcal{L}^1, \forall a \in \mathbb{R}, \quad \mathbb{E}(aX) = a\mathbb{E}X$.
— La variable aléatoire X est intégrable si et seulement si $|X|$ est intégrable.
— $\forall X \in \mathcal{L}^1, \quad \mathbb{P}(X \le a) = 1 \Longrightarrow \mathbb{E}X \le a$.
— $\forall X \in \mathcal{L}^1, \quad \mathbb{P}(X \ge b) = 1 \Longrightarrow \mathbb{E}X \ge b$.
— $\forall X \in \mathcal{L}^1, \quad \mathbb{P}(X = a) = 1 \Longrightarrow \mathbb{E}X = a$.
— $\forall X \in \mathcal{L}^1, \quad \mathbb{P}(|X| \le a) = 1 \Longrightarrow \mathbb{E}|X| \le a$.
— Soient X, Y deux variables aléatoires vérifiant $0 \le X \le Y$. Si Y est intégrable, alors X est intégrable.
— $\forall X, Y \in \mathcal{L}^1, \quad \mathbb{P}(X \le Y) = 1 \Longrightarrow \mathbb{E}X \le \mathbb{E}Y$.

Définition. *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 aux inégalités classiques

6.3.1 Inégalité de Markov

Théorème 6.3. *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

$$\mathbb{E}X = \int_{\mathbb{R}} x \, d\mathbb{P}_X(x) \geq \int_{\{x \geq a\}} x \, d\mathbb{P}_X(x) \geq \int_{\{x \geq a\}} a \, d\mathbb{P}_X(x) = a\mathbb{P}(X \geq a).$$

□

6.3.2 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 6.4 (Formule de Poincaré)**.** *Pour tous événements $A_1, A_2, \dots, A_n$ sous la probabilité $\mathbb{P}$*

$$\begin{aligned} \mathbb{P}\Big(\bigcup_{i=1}^{n} A_i\Big) &= \sum_{B \in \mathcal{P}(\{1,\dots,n\}) \setminus \varnothing} (-1)^{1+|B|} \mathbb{P}(\cap_{j \in B} A_j) \qquad (6.1)\\ &= \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2}) + \cdots + \\ &\quad + \cdots + (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) + \cdots + \\ &\quad + \cdots + (-1)^{n+1} \mathbb{P}(A_1 \cap \cdots \cap A_n). \end{aligned}$$

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 nous permettant d'obtenir des encadrements de la probabilité d'une réunion.

Lemme 6.5. *Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé; $(A_x)_{x \in I}$ des événements indexés par un ensemble fini I. Pour $n \geq 1$, on pose*

$$V_n = P_n \left(\sum_{x \in I} \mathbb{1}_{A_x} - 1 \right) \mathbb{1}_{\cup_{x \in I} A_x},$$

où $(P_k)_{k \geq 0}$ est la suite 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 $n \geq k$, on a $P_k(n) = \binom{n}{k}$ tandis que $P_k(n) = 0$ pour $0 \leq n < k$. Alors

$$\forall n \in \mathbb{N}^*, \quad \mathbb{P}\left(\bigcup_{x \in I} A_x \right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{J \in \mathcal{B}_k(I)} \mathbb{P}\left(\bigcap_{x \in J} A_x \right) + (-1)^n \mathbb{E}V_n. \qquad (6.2)$$

Démonstration. Il suffit de montrer que

$$\forall n \in \mathbb{N}^* \quad \mathbb{1}_{\bigcup_{x\in I} A_x} = \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} \prod_{x\in J} \mathbb{1}_{A_x} + (-1)^n V_n. \tag{6.3}$$

Soit $\omega \in \Omega$. Si $\omega \notin \cup_{x\in I} A_x$, les deux membres de l'égalité (6.3) sont nuls. Sinon, posons $N = \sum_{x\in I} \mathbb{1}_{A_x}(\omega) = |\{x \in I; \omega \in A_x\}|$.

On a $\prod_{x\in J} \mathbb{1}_{A_x} = 1$ si et seulement si $J \subset \{x \in I; \omega \in A_x\}$.

Ainsi $\sum_{J\in\mathcal{B}_k(I)} \prod_{x\in J} \mathbb{1}_{A_x} = P_k(N)$. On doit donc montrer que

$$1 = \sum_{k=1}^{n} (-1)^{k+1} P_k(N) + (-1)^n P_n(N-1),$$

ce qui est équivalent à

$$\sum_{k=0}^{n} (-1)^k P_k(N) = (-1)^n P_n(N-1).$$

Pour conclure, deux preuves sont possibles, que nous présentons ci-dessous.

Méthode 1 : On va montrer par récurrence sur n que

$$\sum_{k=0}^{n} (-1)^k P_k(X) = (-1)^n P_n(X-1).$$

Pour $n = 0$, la relation est vérifiée. Pour passer de n à $n+1$, on écrit

$$\begin{aligned}
\sum_{k=0}^{n+1} (-1)^k P_k(X) &= \sum_{k=0}^{n} (-1)^k P_k(X) + (-1)^{n+1} P_{n+1}(X) \\
&= (-1)^n P_n(X-1) + (-1)^{n+1} P_{n+1}(X) \\
&= (-1)^{n+1} (P_{n+1}(X) - P_n(X-1)) \\
&= (-1)^{n+1} (\frac{1}{n+1} X P_n(X-1) - P_n(X-1)) \\
&= (-1)^{n+1} \frac{(X-(n+1)) P_n(X-1)}{n+1} \\
&= (-1)^{n+1} P_{n+1}(X-1).
\end{aligned}$$

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.

Ensuite, il suffit d'intégrer (6.3) pour obtenir (6.2). □

Preuve de la formule de Poincaré. En prenant $I = \{1, \dots, n\}$ dans le lemme précédent, on obtient

$$\mathbb{P}(\bigcup_{x\in I} A_x) = \sum_{k=1}^{|I|} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} \mathbb{P}(\bigcap_{x\in J} A_x),$$

car V_n est identiquement nulle. Cela démontre la formule de Poincaré. □

Théorème 6.6 (Inégalités de Bonferroni)**.** *Soient* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace probabilisé ;* $(A_x)_{x\in I}$ *des événements. Soit* $n \in \mathbb{N}^*$.

— *Si* n *est impair, on a alors*

$$\mathbb{P}(\bigcup_{x\in I} A_x) \le \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} \mathbb{P}(\bigcap_{x\in J} A_x). \tag{6.4}$$

— *Si* n *est pair, on a alors*

$$\mathbb{P}(\bigcup_{x\in I} A_x) \ge \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} \mathbb{P}(\bigcap_{x\in J} A_x). \tag{6.5}$$

Démonstration. Il suffit d'appliquer le lemme 6.5 en remarquant que V_n est une variable aléatoire positive, et que donc son espérance l'est aussi. □

6.3.3 Application de la formule de Poincaré au 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 : $d_0 = 1$ et

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

Pour i compris entre 1 et n, posons $A_i = \{\sigma(i) = i\}$. Nous cherchons à calculer

$$p_n = \mathbb{P}(\bigcap_{i=1}^{n} A_i^c) = 1 - \mathbb{P}(\bigcup_{i=1}^{n} A_i),$$

d'où, avec la formule de Poincaré

$$p_n = 1 - \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} \mathbb{P}(\bigcap_{x\in J} A_x).$$

Mais il est facile de déterminer $\bigcap_{x\in J} A_x$: 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, $\mathbb{P}(\bigcap_{x\in J} A_x) = \frac{(n-k)!}{n!}$.
Finalement,

$$\begin{aligned} p_n &= 1 - \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} \frac{(n-k)!}{n!} \\ &= 1 + \sum_{k=1}^{n} (-1)^k \binom{n}{k} \frac{(n-k)!}{n!} \\ &= \sum_{k=0}^{n} \frac{(-1)^k}{k!}. \end{aligned}$$

6.4 Théorèmes de transfert

Théorème 6.7. *Soit X 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, la variable $g(X)$ 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, on a

$$\int_{\Omega} |g(X(\omega))|\, d\mathbb{P}(\omega) = \int_{\mathbb{R}} |g(x)|\, d\mathbb{P}_X(x) = \int_{\mathbb{R}} |g(x)| f(x)\, dm(x).$$

Si cette quantité est finie, le théorème de transfert nous dit alors que

$$\int_{\Omega} g(X(\omega))\, d\mathbb{P}(\omega) = \int_{\mathbb{R}} g(x)\, d\mathbb{P}_X(x) = \int_{\mathbb{R}} g(x) f(x)\, dm(x).$$

□

6.4.1 Calcul de l'espérance d'une variable aléatoire discrète

Théorème 6.8. *Soient X 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*

$$\sum_{i \in D} |g(i)| \mathbb{P}(X = i) < +\infty.$$

Si cette somme est finie, on a alors

$$\mathbb{E}Y = \mathbb{E}g(X) = \sum_{i \in D} g(i) \mathbb{P}(X = i).$$

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 $i \mapsto \mathbb{P}(X = i)$. Il suffit donc d'appliquer le théorème 6.7 en prenant pour m la mesure de comptage sur D et pour f :

$$f(x) = \begin{cases} \mathbb{P}(X = x) & \text{si } x \in D \\ 0 & \text{si } x \notin D. \end{cases}$$

On a alors

$$\int_{\mathbb{R}} |g(x)| f(x)\, dm(x) = \sum_{x \in D} |g(x)| f(x).$$

Si cette somme est finie, on obtient ainsi

$$\int_{\mathbb{R}} g(x) f(x)\, dm(x) = \sum_{x \in D} g(x) f(x).$$

□

Corollaire 6.9. *Soient X 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*

$$\sum_{i \in D} |i|\mathbb{P}(X = i) < +\infty.$$

Si cette somme est finie, on a alors

$$\mathbb{E}X = \sum_{i \in D} i\mathbb{P}(X = i).$$

Démonstration. Il suffit d'appliquer le théorème 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 6.10. *Soient X une variable aléatoire admettant la fonction f comme densité, et g une fonction mesurable 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 6.7 avec pour m la mesure de Lebesgue. □

Corollaire 6.11. *Soit X 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 6.12 (importante)**.** *Si la densité de X est paire et que X est intégrable, alors $\mathbb{E}X = 0$.*

On a en effet

$$\mathbb{E}X = \int_{\mathbb{R}} xf(x)\, d\lambda(x),$$

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 $\mathbb{E}X = 0$.

6.4.3 Identifier une loi : technique de la fonction test

L'espérance est au cœur d'une méthode classique d'identification des lois, dite méthode de la fonction test. On commence par un résultat théorique d'identification des mesures.

Théorème 6.13. *Soient μ et ν deux mesures sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ qui donnent chacune une masse finie aux compacts de $\mathbb{R}^d$. On suppose que pour toute fonction continue à support compact f, on a $\int_{\mathbb{R}^d} f\ d\mu = \int_{\mathbb{R}^d} f\ d\nu$. Alors $\mu = \nu$.*

Démonstration. Les compacts de $\mathbb{R}^d$ forment un π-système qui engendre la tribu borélienne de $\mathbb{R}^d$ (par exemple car les pavés ouverts s'écrivent comme réunion dénombrable de pavés compacts), donc il suffit de montrer que μ et ν coïncident sur les compacts. Soit f_n la fonction de $\mathbb{R}_+$ dans $\mathbb{R}_+$ définie par $f_n(x) = (1 - nx)^+$. f est continue, vaut 1 en 0 et est nulle sur $[1/n, +\infty[$. Soit K un compact de $\mathbb{R}^d$, et posons $g_n(x) = f_n(d(x, K))$, où $d(x, K) = \inf\{d(x, y);\ y \in K\}$. g_n est continue, comme composition d'applications continues, et converge simplement vers l'indicatrice de K. Comme $|g_n| \leq \mathbb{1}_{K+\overline{B}(0,1)}$ qui est intégrable par rapport à μ et ν, le théorème de convergence dominée dit que $\int_{\mathbb{R}^d} g_n\ d\mu$ converge vers $\int_{\mathbb{R}^d} \mathbb{1}_K\ d\mu = \mu(K)$ et $\int_{\mathbb{R}^d} g_n\ d\nu$ converge vers $\int_{\mathbb{R}^d} \mathbb{1}_K\ d\nu = \nu(K)$. Vu l'hypothèse faite, on a $\int_{\mathbb{R}^d} g_n\ d\mu = \int_{\mathbb{R}^d} g_n\ d\nu$ pour tout n, donc $\mu(K) = \nu(K)$. Comme μ et ν coïncident sur les compacts, on a donc bien $\mu = \nu$. □

Corollaire 6.14. *Un vecteur aléatoire (une variable aléatoire) X suit la loi μ sur $\mathbb{R}^d$ ($\mathbb{R}$) si et seulement si toute fonction continue à support compact ϕ, on a $\mathbb{E}[\phi(X)] = \int_{\mathbb{R}^d} \phi\ d\mu$.*

Démonstration. D'après le théorème de transfert, $\mathbb{E}\phi(X) = \int \phi(x)\ d\mathbb{P}_X(x)$ et on applique le théorème précédent. □

L'usage du corollaire précédent est souvent appelé technique de la fonction test. C'est un outil commode d'identification d'une loi qui est universel, mais qui n'est pas toujours le plus rapide. Il est très efficace dans le cas de loi "hybrides", ayant à la fois une composante discrète et une composante continue.

Remarque 6.15. *Une remarque très simple, mais très importante : on applique souvent une version beaucoup plus simple de la méthode de la fonction test : si X est un vecteur aléatoire tel que $\mathbb{E}(f(X)) = \int_{\mathbb{R}^d} f\ d\mu$ pour toute fonction f mesurable positive bornée, alors $\mathbb{P}_X = \mu$. En effet, il suffit de prendre $f = \mathbb{1}_A$ avec A borélien de $\mathbb{R}^d$ pour identifier $\mathbb{P}_X$ et μ.*

Exemple: Soit X une variable suivant la loi uniforme sur $[-1, 1]$. Calculer la loi de $Y = \max(0, X)$.

On a $\phi(Y) = \phi(0)\mathbb{1}_{\{X<0\}} + \phi(X)\mathbb{1}_{\{X\geq 0\}}$, donc

$$\begin{aligned}\mathbb{E}\phi(Y) &= \mathbb{E}(\phi(0)\mathbb{1}_{\{X<0\}}) + \mathbb{E}(\phi(X)\mathbb{1}_{\{X\geq 0\}})\\ &= \phi(0)\mathbb{P}(X < 0) + \frac{1}{2}\int_{[-1,1]} \phi(x)\mathbb{1}_{\{x\geq 0\}})\ d\lambda(x)\\ &= \frac{1}{2}\phi(0) + \frac{1}{2}\int_{[0,1]} \phi(x)\ d\lambda(x)\end{aligned}$$

On reconnait là l'intégrale de ϕ par rapport à la mesure $\frac{1}{2}\delta_0 + \frac{1}{2}\lambda_{[0,1]}$.

6.5 Convexité

6.5.1 Rappels sur la convexité

On dit qu'une fonction est convexe sur l'intervalle ouvert 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 6.16. *Si $x < z < y$, alors il existe $\theta \in]0,1[$ avec $z = \theta x + (1-\theta)y$.*

Démonstration. On prend $\theta = \frac{y-z}{y-x}$. □

Lemme 6.17. *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*

$$p_x(h_1) := \frac{f(x+h_1)-f(x)}{h_1} \leq p_x(h_2) := \frac{f(x+h_2)-f(x)}{h_2}.$$

p_x admet une limite à droite et à gauche en 0 : ce sont les dérivées à droite (resp. à gauche) de f en x.

Démonstration. Si h_1 et h_2 sont positifs, on applique alors le lemme précédent avec $z = x + h_1$ et $y = x + h_2$ et on utilise la définition de la convexité.

Si h_1 et h_2 sont négatifs, on fait la même chose que dans le cas précédent, avec $x' = x + h_1$, $z' = x + h_2$, $y' = x$.

Enfin, si $h_1 < 0 < h_2$, on a alors

$$p_x(h_1) \leq p_x(\max(h_1, -h_2)) \text{ et } p_x(\min(h_2, -h_1)) \leq p_x(h_2)$$

d'après ce qui précède. Il suffit donc 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 6.18. *Soit f une fonction convexe sur l'intervalle I telle que les points x et y de I vérifient $f(x) = f(y) = 0$. Alors f est négative sur $[x, y]$, positive à l'extérieur.*

Démonstration. Si $x \leq z \leq y$ avec $z = \theta x + (1-\theta)y$, on a alors $f(z) \leq \theta f(x) + (1-\theta)f(y) \leq 0$. Sinon, l'inégalité

$$\theta f(a) + (1-\theta)f(b) \geq f(\theta a + (1-\theta)b)$$

entraîne que si de trois nombres, le terme médian et un autre ont une image nulle, alors le troisième doit avoir une image positive. □

Lemme 6.19. *On appelle corde portée par x et y la droite passant par $(x, f(x))$ et $(y, f(y))$, d'équation $\ell(t) = \frac{f(y)-f(x)}{y-x}(t-x) + f(x)$. Si f est une fonction convexe, 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 6.20. *Si f est une fonction convexe sur l'intervalle ouvert I, alors pour tous x et 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 voulue. Mais pour tout t, il existe $h > 0$ tel que t n'appartient pas à l'intervalle $]x, x+h[$, d'où le résultat. □

Théorème 6.21. *Si f est une fonction dérivable sur l'intervalle ouvert I telle que f' est croissante sur I, alors f est convexe sur I.*

Démonstration. Soient $x, y \in I$, avec $x < y$, ainsi que $\theta \in]0,1[$.
On pose $z = \theta x + (1-\theta)y$. En appliquant deux fois l'inégalité des accroissements finis, on a

$$\frac{f(x)-f(z)}{x-z} \leq f'(z) \leq \frac{f(y)-f(z)}{y-z}.$$

En prenant les membres extrémaux de cette inégalité, on a

$$\frac{f(x)-f(z)}{(1-\theta)(x-y)} \leq \frac{f(y)-f(z)}{\theta(y-x)},$$

soit

$$(f(z)-f(x))\theta \leq (1-\theta)(f(y)-f(z)),$$

ce qui donne le résultat voulu en réarrangeant les termes. □

6.5.2 Inégalité de Jensen

Théorème 6.22. *Soit X une variable aléatoire intégrable à valeurs dans l'intervalle ouvert I. Soit f une fonction convexe de I dans $\mathbb{R}$. Alors*

$$f(\mathbb{E}X) \leq \mathbb{E}f(X).$$

Notons qu'on n'exclut pas d'avoir $\mathbb{E}[f(X)] = +\infty$.

Démonstration. Soit Φ l'ensemble des fonctions affines ϕ telles que

$$\forall x \in I \quad \phi(x) \leq f(x).$$

Soit $\phi \in \Phi$. On a presque sûrement

$$\phi(X) \leq f(X).$$

On a donc $\mathbb{E}\phi(X) \leq \mathbb{E}f(X)$. Comme ϕ est une fonction affine, on voit que $\mathbb{E}\phi(X) = \phi(\mathbb{E}X)$. Ainsi,

$$\forall \phi \in \Phi \quad \phi(\mathbb{E}X) \leq \mathbb{E}f(X).$$

donc

$$\sup_{\phi \in \Phi} \phi(\mathbb{E}X) \leq \mathbb{E}f(X).$$

Prenons alors $\phi(t) = f(\mathbb{E}X) + f'_d(\mathbb{E}X)(t - \mathbb{E}X)$. D'après le lemme 6.20, $\phi \in \Phi$, ce qui donne

$$f(\mathbb{E}X) \leq \mathbb{E}f(X).$$

□

Pour retenir quel est le sens de l'égalité, nous conseillons de prendre la fonction convexe $\phi(x) = |x|$ ou $\phi(x) = x^2$.

Corollaire 6.23. *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 une variable aléatoire discrète X telle que $\mathbb{P}(X = x_i) = \theta_i$ pour tout i compris entre 1 et n. Le théorème de transfert pour une variable aléatoire discrète et l'inégalité de Jensen donnent

$$f\left(\sum_{k=1}^{n} \theta_k x_k\right) = f(\mathbb{E}X) \leq \mathbb{E}[f(x)] = \sum_{k=1}^{n} \theta_k f(x_k).$$

□

Remarque 6.24. *On peut toutefois remarquer que ce corollaire peut également se montrer simplement par récurrence sur n, à l'aide de la définition de la convexité, sans intervention des probabilités.*

6.6 Intégrale et queue de distribution

Théorème 6.25. *Soit X une variable aléatoire positive. On a*

$$\mathbb{E}X = \int_{\mathbb{R}_+} \mathbb{P}(X > t)\ d\lambda(t) = \int_{\mathbb{R}_+} \mathbb{P}(X \geq t)\ d\lambda(t).$$

Démonstration. On utilise le théorème de Tonelli :

$$\begin{aligned}
\int_{\mathbb{R}_+} \mathbb{P}(X > t)\ d\lambda(t) &= \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \mathbb{1}_{]t,+\infty[}(s)\ d\mathbb{P}_X(s)\ d\lambda(t) \\
&= \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \mathbb{1}_{[0,s[}(t)\ d\mathbb{P}_X(s)\ d\lambda(t) \\
&= \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \mathbb{1}_{[0,s[}(t)\ d\lambda(t)\ d\mathbb{P}_X(s) \\
&= \int_{\mathbb{R}_+} s\ d\mathbb{P}_X(s) \\
&= \mathbb{E}X.
\end{aligned}$$

Pour la dernière égalité on peut par exemple remarquer que les nombres $\mathbb{P}(X > t)$ et $\mathbb{P}(X \geq t)$ ne diffèrent qu'aux points t tels que $\mathbb{P}(X = t) > 0$, qui forment une famille au plus dénombrable, donc de mesure nulle. □

Corollaire 6.26. *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

Définition. *On dit qu'une variable aléatoire X admet un moment d'ordre 2 si elle est de carré intégrable, c'est-à-dire si $X^2 \in \mathcal{L}^1$.*

Remarque 6.27. *Ici encore, on peut noter que $\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2\ d\mathbb{P}_X(x)$ ne dépend que de la loi de X.*

On note $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ (ou encore $\mathcal{L}^2$) l'ensemble des variables aléatoires de carré intégrable.

Lemme 6.28. *Soient $X, Y \in \mathcal{L}^2$. Alors la variable aléatoire XY est intégrable.*

Démonstration. Pour tous les réels a, b, on a $|ab| \leq \frac{1}{2}(a^2 + b^2)$. Pour le voir, il suffit d'écrire $a^2 + b^2 + 2ab = (a + b)^2 \geq 0$, d'où $(a^2 + b^2)/2 \geq -ab$ et $a^2 + b^2 - 2ab = (a - b)^2 \geq 0$, et on a bien $(a^2 + b^2)/2 \geq ab$.
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 le résultat voulu. $\square$

En particulier, en prenant $Y = 1$, on voit que $X \in \mathcal{L}^1$ dès que $X \in \mathcal{L}^2$.

Corollaire 6.29. $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ *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. $\square$

6.7.1 Covariance et variance

Définition. *Soient X et Y deux variables aléatoires admettant chacune un moment d'ordre 2. On appelle covariance du couple (X, Y) le réel*

$$\mathrm{Covar}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)].$$

On appelle variance de X le réel positif

$$\mathrm{Var}\, X = \mathrm{Covar}(X, X) = \mathbb{E}(X - \mathbb{E}X)^2.$$

On appelle écart-type de X le réel positif

$$\sigma(X) = (\mathrm{Var}\, X)^{1/2}.$$

Définition. *On dit qu'une variable aléatoire est réduite si on a $\mathrm{Var}\, X = 1$ (ou de manière équivalente si $\sigma(X) = 1$).*

Théorème 6.30. *Les propriétés suivantes sont satisfaites :*

1. $(X, Y) \mapsto \mathrm{Covar}(X, Y)$ *est une forme bilinéaire symétrique positive.*
2. $\forall a, b \in \mathbb{R} \quad \mathrm{Covar}(X - a, Y - b) = \mathrm{Covar}(X, Y)$.
3. $\mathrm{Var}(X + Y) = \mathrm{Var}\, X + \mathrm{Var}\, Y + 2\,\mathrm{Covar}(X, Y)$.
4. $\mathrm{Covar}(X, Y) = \mathbb{E}XY - (\mathbb{E}X)(\mathbb{E}Y)$.
5. $\mathrm{Var}\, X = \mathbb{E}X^2 - (\mathbb{E}X)^2$.
6. $|\mathbb{E}XY|^2 \leq \mathbb{E}X^2\mathbb{E}Y^2$ *(inégalité de Cauchy–Schwarz).*
7. $|\mathrm{Covar}(X, Y)| \leq \sigma(X)\sigma(Y)$.

Démonstration. 1. Notons $\langle X, Y\rangle = \mathbb{E}(XY)$. Il est facile de voir que l'application $(X, Y) \mapsto \langle X, Y\rangle$ est une forme bilinéaire symétrique positive. Posons $L(X) = X - \mathbb{E}X$. L'application $X \mapsto L(X)$ est linéaire de $\mathcal{L}^2$ dans lui-même. On a $\mathrm{Covar}(X, Y) = \langle L(X), L(Y)\rangle$. Les deux observations faites ci-dessus permettent de dire que $(X, Y) \mapsto \mathrm{Covar}(X, Y)$ est une forme bilinéaire symétrique positive.

2. On a
$$\begin{aligned}\mathrm{Covar}(X-a,Y-b) &= \langle L(X-a),L(Y-b)\rangle\\ &= \langle L(X)-L(a),L(Y)-L(b)\rangle\\ &= \langle L(X)-0,L(Y)-0\rangle\\ &= \mathrm{Covar}(X,Y).\end{aligned}$$
3. On a
$$\begin{aligned}\mathrm{Var}(X+Y) &= \mathrm{Covar}(X+Y,X+Y)\\ &= \mathrm{Covar}(X,X)+2\,\mathrm{Covar}(X,Y)+\mathrm{Covar}(Y,Y)\\ &= \mathrm{Var}\,X+2\,\mathrm{Covar}(X,Y)+\mathrm{Var}\,Y.\end{aligned}$$
Pour passer de la ligne 1 à la ligne 2, on utilise le fait que la covariance est bilinéaire symétrique.
4. Comme $(X-\mathbb{E}X)(Y-\mathbb{E}Y)=XY+\mathbb{E}X\mathbb{E}Y-(\mathbb{E}X)Y-(\mathbb{E}Y)X$, on obtient
$$\begin{aligned}\mathbb{E}(X-\mathbb{E}X)(Y-\mathbb{E}Y) &= \mathbb{E}XY+\mathbb{E}(\mathbb{E}X\mathbb{E}Y)-\mathbb{E}X\mathbb{E}Y-\mathbb{E}Y\mathbb{E}X\\ &= \mathbb{E}XY+\mathbb{E}X\mathbb{E}Y-2\mathbb{E}X\mathbb{E}Y\\ &= \mathbb{E}XY-(\mathbb{E}X)(\mathbb{E}Y).\end{aligned}$$
5. Il suffit d'appliquer la formule précédente avec $X=Y$.
6. Comme $\langle\cdot,\cdot\rangle$ 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.

□

Remarque 6.31. *La covariance n'est pas définie positive. En effet, on montre facilement que* $\mathrm{Var}\,X=0$ *si et seulement si* X *est constante avec probabilité 1, mais cette constante n'est pas forcément nulle.*

Définition. *Lorsque* $\sigma(X)$ *et* $\sigma(Y)$ *sont non nuls, on définit le coefficient de corrélation entre* X *et* Y *par*
$$\rho(X,Y)=\frac{\mathrm{Covar}(X,Y)}{\sigma(X)\sigma(Y)}.$$

D'après ce qui précède, $\rho(X,Y)\in[-1;1]$. Lorsque $\mathrm{Covar}(X,Y)=0$ (ce qui implique $\rho(X,Y)=0$ si $\sigma(X)$ et $\sigma(Y)$ sont non nuls), on dit que X et Y ne sont pas corrélées .
Lorsque $\mathrm{Covar}(X,Y)\geq 0$ (ce qui implique $\rho(X,Y)\geq 0$ si $\sigma(X)$ et $\sigma(Y)$ sont non nuls), on dit que X et Y sont positivement corrélées.
Lorsque $\mathrm{Covar}(X,Y)\leq 0$ (ce qui implique $\rho(X,Y)\leq 0$ si $\sigma(X)$ et $\sigma(Y)$ sont non nuls), on dit que X et Y sont négativement corrélées.

Remarque 6.32. *On peut montrer que (cela est laissé en exercice au lecteur) :*
— $\rho(X,Y)=1$ *si et seulement s'il existe* $a>0$, $b\in\mathbb{R}$ *tels que* $\mathbb{P}(Y=aX+b)=1$,
— $\rho(X,Y)=-1$ *si et seulement s'il existe* $a<0$, $b\in\mathbb{R}$ *tels que* $\mathbb{P}(Y=aX+b)=1$.

Remarque 6.33. *Le coefficient de corrélation mesure une corrélation linéaire. Il peut être nul alors que la variable* Y *dépend fortement de* X *mais de manière non-linéaire. C'est pourquoi on ne devrait pas se passer d'une représentation graphique quand on dispose d'observations* (x_i,y_i). *À l'inverse, une forte corrélation ne doit pas être comprise comme une relation de causalité. Certaines variables n'ont aucune relation entre elles, mais donnent des coefficients de corrélation proches de 1. Cela provient souvent du fait qu'elles sont elles-mêmes influencées par une troisième variable.*

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 dit que le vecteur X admet un moment d'ordre deux et on appelle matrice de covariance de X la matrice $n \times n$ dont les coefficients sont $(\mathrm{Covar}(X_i, X_j))_{1 \le i,j \le n}$.

Notons, pour $a \in \mathbb{R}^n$, $\langle X, a \rangle$ le produit scalaire des vecteurs a et X.

Théorème 6.34. *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 6.35. *Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire admettant un moment d'ordre deux, de matrice de covariance C_X et d'espérance m_X. Soient A une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$, et b un vecteur de $\mathbb{R}^p$. Alors la variable $Y = AX + b$ admet $C_Y = AC_XA^*$ comme matrice de covariance et l'espérance de Y vaut $Am_X + b$.*

Démonstration. Soit $1 \le i \le p$. On a

$$\begin{aligned} \mathbb{E}Y_i &= \mathbb{E}(\sum_{k=1}^n a_{i,k} X_k + b_i) = \sum_{k=1}^n a_{i,k} \mathbb{E}X_k + b_i \\ &= \sum_{k=1}^n a_{i,k} m_{X_k} + b_i = (Am_X + b)_i. \end{aligned}$$

De plus, on a

$$\begin{aligned} \mathrm{Covar}(\langle Y, a \rangle, \langle Y, c \rangle) &= \mathrm{Covar}(\langle AX + b, a \rangle, \langle AX + b, c \rangle) \\ &= \mathrm{Covar}(\langle AX, a \rangle, \langle AX, c \rangle) \\ &= \mathrm{Covar}(\langle X, A^*a \rangle, \langle X, A^*c \rangle) \\ &= \langle C_X A^* a, A^* c \rangle \\ &= \langle AC_X A^* a, c \rangle. \end{aligned}$$

□

6.7.3 Espérance et indépendance

Le théorème suivant est essentiel.

Théorème 6.36. *Soient X, Y des variables aléatoires intégrables indépendantes. Alors, leur produit XY est une variable aléatoire intégrable et son espérance est*

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

Démonstration. D'après le théorème de transfert, on a

$$\mathbb{E}(|XY|) = \int_{\mathbb{R}^2} |xy| \; d\mathbb{P}_{(X,Y)}(x,y).$$

Il vient

$$\begin{aligned}\mathbb{E}(|XY|) &= \int_{\mathbb{R}^2} |xy| \; d\mathbb{P}_{(X,Y)}(x,y) = \int_{\mathbb{R}^2} |x| \cdot |y| \; d(\mathbb{P}_X \otimes \mathbb{P}_Y)(x,y) \\ &= \int_{\mathbb{R}} |x| d\mathbb{P}_X(x) \cdot \int_{\mathbb{R}} |y| \; d\mathbb{P}_Y(y) = \mathbb{E}|X| \cdot \mathbb{E}|Y| < +\infty.\end{aligned}$$

Ainsi, par le théorème de Tonelli, le produit XY est intégrable. On a de plus

$$\begin{aligned}\mathbb{E}XY &= \int_{\mathbb{R}^2} xy \; d\mathbb{P}_{(X,Y)}(x,y) = \int_{\mathbb{R}^2} xy \; d(\mathbb{P}_X \otimes \mathbb{P}_Y)(x,y) \\ &= \int_{\mathbb{R}} x d\mathbb{P}_X(x) \cdot \int_{\mathbb{R}} y \; d\mathbb{P}_Y(y) = \mathbb{E}X\mathbb{E}Y.\end{aligned}$$

□

Corollaire 6.37. *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 6.38 (importante)**.** *Des variables aléatoires peuvent être non corrélées sans être indépendantes.*

En effet, considérons deux variables aléatoires vérifiant

$$\begin{aligned}\mathbb{P}(\{X=1\} \cap \{Y=1\}) &= \mathbb{P}(\{X=1\} \cap \{Y=-1\}) \\ &= \mathbb{P}(\{X=-1\} \cap \{Y=0\}) = 1/3.\end{aligned}$$

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 de chaque colonne, soit

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 \end{pmatrix}.$$

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

$$\mathbb{E}XY = \sum_{i \in \{-1;1\}} \sum_{j \in \{-1;0;1\}} ij \mathbb{P}(\{X=i\} \cap \{Y=j\}) = 1/3 - 1/3 = 0.$$

Cela implique donc $\mathrm{Covar}(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = 0$.
Cependant, X et Y ne sont pas indépendantes car

$$0 = \mathbb{P}(\{X=1\} \cap \{Y=0\}) \neq \mathbb{P}(X=1)\mathbb{P}(Y=0) = \frac{2}{3} \times \frac{1}{3}.$$

Corollaire 6.39. *Soient X, Y deux variables aléatoires indépendantes de carré intégrable. Alors on a*

$$\mathrm{Var}(X+Y) = \mathrm{Var}\, X + \mathrm{Var}\, Y.$$

Démonstration. On a toujours $\mathrm{Var}(X+Y) = \mathrm{Var}\, X + \mathrm{Var}\, Y + 2\,\mathrm{Covar}(X,Y)$. Comme X et Y sont indépendantes, elles sont décorrélées, d'où le résultat. □

6.7.4 Inégalité de Chebychev

Théorème 6.40. *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. On a

$$\mathbb{P}(|X - \mathbb{E}X| \geq a) = \mathbb{P}(|X - \mathbb{E}X|^2 \geq a^2).$$

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. □

Remarque 6.41. *L'inégalité de Chebychev est souvent appelée "inégalité de Bienaymé–Chebychev".*

6.8 Lois images par des transformations affines

6.8.1 Exemple fondamental

Théorème 6.42. *Soient $A \in Gl_d(\mathbb{R})$ 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. Ici $O_1 = O_2 = \mathbb{R}^d$ et $T^{-1}(y) = A^{-1}(y - b)$. La différentielle en un point d'une transformation affine se confond avec l'application linéaire associée, d'où le résultat. C'est essentiellement une reformulation du corollaire 4.43. □

Voici quelques applications de ce résultat :

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 suit la loi $\Gamma(a, \gamma)$, alors pour tout $\mu > 0$, on a $\frac{1}{\mu} X \sim \Gamma(a, \gamma\mu)$.
3. Si X suit la loi exponentielle de paramètre $a > 0$, 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. Soient $\sigma > 0$ et $m \in \mathbb{R}$. On a $X \sim \mathcal{N}(m, \sigma^2) \iff \frac{X-m}{\sigma} \sim \mathcal{N}(0, 1)$.

Notons que ces remarques élémentaires permettent de simplifier des calculs théoriques et sont également intéressantes en vue de la simulation.

6.8.2 Application aux lois gaussiennes

Lemme 6.43. *Soit $X = (X_1, X_2)$ un vecteur aléatoire formé de deux variables aléatoires indépendantes suivant la loi normale $\mathcal{N}(0, 1)$.*
On pose $Y_1 = X_1 \cos\theta - X_2 \sin\theta$ et $Y_2 = X_1 \sin\theta + X_2 \cos\theta$. 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

$$y \mapsto \frac{1}{\det M} \frac{1}{2\pi} \exp(-\frac{\|M^{-1}y\|_2^2}{2}).$$

Or 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 6.44. *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 et 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

$$U_1 + U_2 = m_1 + m_2 + \sigma(X_1 \cos\theta + X_2 \sin\theta).$$

D'après le lemme précédent, $X_1 \cos\theta + X_2 \sin\theta$ suit la loi $\mathcal{N}(0, 1)$, 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 6.45. *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*

$$x \mapsto \int_{\mathbb{R}} f(x-t)g(t)\, d\lambda(t) = \int_{\mathbb{R}} g(x-t)f(t)\, d\lambda(t).$$

Si, de plus, X et Y sont à valeurs positives, alors la densité est simplement

$$x \mapsto \mathbb{1}_{\mathbb{R}_+}(x) \int_0^x f(x-t)g(t)\, dt = \mathbb{1}_{\mathbb{R}_+}(x) \int_0^x g(x-t)f(t)\, dt.$$

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

$$g(z, t) = \frac{1}{|\det A|} f(A^{-1} \begin{pmatrix} z \\ t \end{pmatrix}).$$

Comme

$$\det A = 1 \text{ et } A^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix},$$

on a donc $g(z,t) = h(z-t,t) = f(z-t)g(t)$.

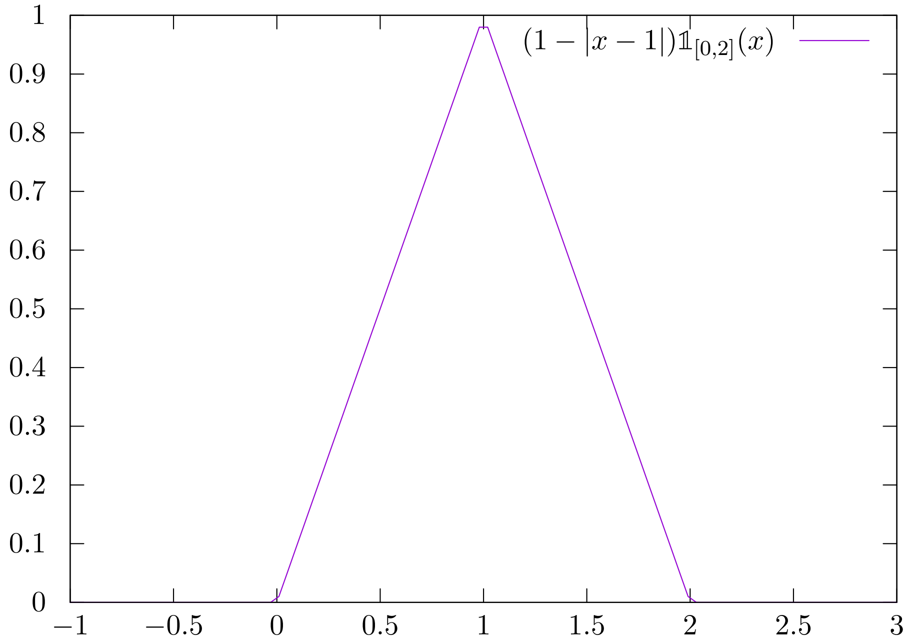
D'après le théorème 5.21, Z admet comme densité par rapport à la mesure de Lebesgue :

$$z \mapsto \int_{\mathbb{R}} g(z,t)\, d\lambda(t) = \int_{\mathbb{R}} f(z-t)g(t)\, d\lambda(t).$$

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 $0 \leq t \leq z$, ce qui n'est évidemment jamais vérifié si z est négatif.

On obtient les deux autres égalités en échangeant les rôles de X et Y. $\square$

Le graphe ci-dessous représente la densité de Z lorsque X et Y suivent toutes les deux la loi uniforme sur $[0,1]$.



Application : $\Gamma(a,\gamma) * \Gamma(b,\gamma) = \Gamma(a+b,\gamma)$

Théorème 6.46. *Soient a, b, γ strictement positifs, X et Y deux variables aléatoires indépendantes, X suivant la loi $\Gamma(a,\gamma)$ et Y la loi $\Gamma(b,\gamma)$. Alors $Z = X + Y$ suit la loi $\Gamma(a+b,\gamma)$.*

Démonstration. Pour tous a et γ strictement positifs, on note $f_{a,\gamma}$ la densité de la loi $\Gamma(a,\gamma)$, soit (rappel)

$$f_{a,\gamma}(x) = \mathbb{1}_{\mathbb{R}_+}(x)\frac{\gamma^a}{\Gamma(a)}x^{a-1}e^{-\gamma x}.$$

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

$$\begin{aligned} f_Z(x) &= \int_0^x f_{a,\gamma}(x-t)f_{b,\gamma}(t)\, dt \\ &= \int_0^x \frac{\gamma^a}{\Gamma(a)}t^{a-1}e^{-\gamma t}\frac{\gamma^b}{\Gamma(b)}(x-t)^{b-1}e^{-\gamma(x-t)}\, dt \\ &= \frac{\gamma^{a+b}e^{-\gamma x}}{\Gamma(a)\Gamma(b)}\int_0^x t^{a-1}(x-t)^{b-1}\, dt. \end{aligned}$$

On fait le changement de variable $t = \theta x$. On obtient

$$\begin{aligned} f_Z(x) &= \frac{\gamma^{a+b}e^{-\gamma x}}{\Gamma(a)\Gamma(b)} x^{a-1}x^{b-1}x \int_0^1 \theta^{a-1}(1-\theta)^{b-1}\, d\theta \\ &= \frac{\gamma^{a+b}e^{-\gamma x}}{\Gamma(a)\Gamma(b)} x^{a+b-1} \int_0^1 \theta^{a-1}(1-\theta)^{b-1}\, d\theta \\ &= K_{a,b} f_{a+b,\gamma}(x), \end{aligned}$$

où $K_{a,b} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \theta^{a-1}(1-\theta)^{b-1}\, d\theta$. Évidemment f_Z et $x \mapsto K_{a,b} f_{a+b,\gamma}(x)$ coïncident également sur $\mathbb{R}_-$ où elles sont nulles. On a donc

$$\int_{\mathbb{R}} f_Z(x)\, d\lambda(x) = \int_{\mathbb{R}} K_{a,b} f_{a+b,\gamma}(x)\, d\lambda(x) = K_{a,b} \int_{\mathbb{R}} f_{a+b,\gamma}(x)\, d\lambda(x).$$

Mais f_Z et $f_{a+b,\gamma}$ sont des densités donc leur intégrale sur $\mathbb{R}$ vaut un. On en déduit que $K_{a,b} = 1$, d'où

$$f_Z(x) = f_{a+b,\gamma}(x).$$

La densité de Z est la densité de $\Gamma(a+b,\gamma)$, donc Z suit la loi $\Gamma(a+b,\gamma)$. □

Remarque 6.47. *Comme sous-produit de cette démonstration, on a obtenu le résultat non trivial suivant :*

$$\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 \theta^{a-1}(1-\theta)^{b-1}\, d\theta.$$

On peut donc considérer qu'on a donné une preuve probabiliste de l'expression intégrale de la fonction Bêta.

6.9 Calcul des premiers moments des lois discrètes usuelles

6.9.1 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

$$\mathbb{1}_A(x) = \begin{cases} 1 \text{ si } x \in A \\ 0 \text{ si } x \notin A. \end{cases}$$

$\mathbb{1}_A$ 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 $\mathbb{1}_A^2 = \mathbb{1}_A$. On a de plus

— $\mathbb{E}\mathbb{1}_A = \mathbb{P}(A)$ et

—

$$\begin{aligned} \operatorname{Var} \mathbb{1}_A = \mathbb{E}\mathbb{1}_A^2 - (\mathbb{E}\mathbb{1}_A)^2 = \mathbb{E}\mathbb{1}_A - (\mathbb{E}\mathbb{1}_A)^2 &= \mathbb{P}(A) - \mathbb{P}(A)^2 \\ &= \mathbb{P}(A)(1-\mathbb{P}(A)). \end{aligned}$$

6.9.2 Loi binomiale

On a vu que la loi binomiale de paramètres n et p est la loi de $X = \sum_{k=1}^{n} \mathbb{1}_{A_k}$, où $A_1, \dots, A_n$ sont des 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 $\mathbb{1}_{A_k}$ sont indépendantes

$$\operatorname{Var} X = \sum_{k=1}^{n} \operatorname{Var} \mathbb{1}_{A_k} = \sum_{k=1}^{n} \mathbb{P}(A_k)(1-\mathbb{P}(A_k)) = np(1-p).$$

6.9.3 Loi géométrique

Soit X une variable aléatoire suivant une loi géométrique de paramètre $p \in]0,1]$. On a

$$\begin{aligned}\mathbb{E}X &= \sum_{k=0}^{+\infty} k\mathbb{P}(X=k) = \sum_{k=1}^{+\infty} k\mathbb{P}(X=k)\\ &= \sum_{k=1}^{+\infty} kp(1-p)^{k-1} = p\sum_{k=1}^{+\infty} k(1-p)^{k-1}\\ &= p\frac{1}{(1-(1-p))^2} = \frac{1}{p}\end{aligned}$$

et

$$\begin{aligned}\mathbb{E}[X(X-1)] &= \sum_{k=0}^{+\infty} k(k-1)\mathbb{P}(X=k) = \sum_{k=2}^{+\infty} k(k-1)\mathbb{P}(X=k)\\ &= \sum_{k=2}^{+\infty} k(k-1)p(1-p)^{k-1}\\ &= p(1-p)\sum_{k=2}^{+\infty} k(k-1)(1-p)^{k-2}\\ &= p(1-p)\frac{2}{(1-(1-p))^3} = \frac{2(1-p)}{p^2}.\end{aligned}$$

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

$$\operatorname{Var} X = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \frac{1}{p} + \frac{2(1-p)}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

6.9.4 Loi de Poisson

Soit X une variable aléatoire suivant une loi de Poisson de paramètre $\lambda > 0$. On a

$$\begin{aligned}\mathbb{E}X &= \sum_{k=0}^{+\infty} k\mathbb{P}(X=k) = \sum_{k=1}^{+\infty} k\mathbb{P}(X=k)\\ &= \sum_{k=1}^{+\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\lambda\sum_{k=1}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!}\\ &= e^{-\lambda}\lambda e^{\lambda} = \lambda.\end{aligned}$$

et

$$\begin{aligned}\mathbb{E}[X(X-1)] &= \sum_{k=0}^{+\infty} k(k-1)\mathbb{P}(X=k) = \sum_{k=2}^{+\infty} k(k-1)\mathbb{P}(X=k)\\ &= \sum_{k=2}^{+\infty} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\lambda^2\sum_{k=2}^{+\infty}\frac{\lambda^{k-2}}{(k-2)!}\\ &= e^{-\lambda}\lambda^2 e^{\lambda} = \lambda^2.\end{aligned}$$

On en déduit que $\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}X = \lambda^2 + \lambda$ et

$$\operatorname{Var} X = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

6.9.5 Loi hypergéométrique

On rappelle que la loi hypergéométrique $\mathcal{H}(N,n,k)$ est la loi image de la loi uniforme sur $\Omega = \mathcal{B}_k(\{1,\dots,N\})$ par l'application

$$\begin{array}{rcl} X : \mathcal{B}_k(\{1,\dots,N\}) & \to & \mathbb{N} \\ \omega & \mapsto & X(\omega) = |\{1,\dots,n\} \cap \omega|. \end{array}$$

On va montrer que $\mathbb{E}X = k\frac{n}{N}$ et $\operatorname{Var} X = k\frac{n}{N}(1-\frac{n}{N})\frac{N-k}{N-1}$.

Notons $\mathbb{P}$ la loi uniforme sur Ω. Par souci de lisibilité, on définit l'ensemble aléatoire A par $A(\omega) = \omega$. Ainsi

$$X = |\{1,\dots,n\} \cap A| = \sum_{i=1}^{n} \mathbb{1}_{\{i\in A\}}.$$

Pour tout $k \in \{1,\dots,n\}$, on a

$$\mathbb{E}(\mathbb{1}_{\{i\in A\}}) = \mathbb{P}(i \in A) = 1 - \frac{\binom{N-1}{k}}{\binom{N}{k}} = 1 - \frac{(N-1)!(N-k)!}{N!(N-k-1)!} = 1 - \frac{N-k}{N} = \frac{k}{N}.$$

On obtient ainsi

$$\mathbb{E}X = \sum_{i=1}^{n} \mathbb{E}\mathbb{1}_{\{i\in A\}} = \frac{nk}{N} = k\frac{n}{N}.$$

De plus, on a

$$\operatorname{Var} X = \sum_{i=1}^{n} \sum_{j=1}^{n} \operatorname{Covar}(\mathbb{1}_{\{i\in A\}}, \mathbb{1}_{\{j\in A\}}).$$

Comme pour $i = j$, on a

$$\operatorname{Covar}(\mathbb{1}_{\{j\in A\}}, \mathbb{1}_{\{j\in A\}}) = \operatorname{Var} \mathbb{1}_{\{i\in A\}} = \mathbb{P}(j \in A)(1 - \mathbb{P}(j \in A)) = \frac{k}{N}(1-\frac{k}{N}),$$

et pour $i \neq j$, on a

$$\begin{array}{rcl} \operatorname{Covar}(\mathbb{1}_{\{i\in A\}}, \mathbb{1}_{\{j\in A\}}) & = & \operatorname{Covar}(1-\mathbb{1}_{\{i\in A\}}, 1-\mathbb{1}_{\{j\in A\}}) \\ & = & \operatorname{Covar}(\mathbb{1}_{\{i\notin A\}}, \mathbb{1}_{\{j\notin A\}}) \\ & = & \mathbb{P}(i \notin A, j \notin A) - \mathbb{P}(i \notin A)\mathbb{P}(j \notin A) \\ & = & \dfrac{\binom{N-2}{k}}{\binom{N}{k}} - (\dfrac{N-k}{N})^2 \\ & = & \dfrac{(N-k)(N-k-1)}{N(N-1)} - (\dfrac{N-k}{N})^2 \\ & = & -\dfrac{1}{N-1}\dfrac{k}{N}(1-\dfrac{k}{N}), \end{array}$$

on en déduit que

$$\begin{array}{rcl} \operatorname{Var} X & = & n \times \dfrac{k}{N}\left(1-\dfrac{k}{N}\right) + n(n-1) \times \left(-\dfrac{1}{N-1}\dfrac{k}{N}(1-\dfrac{k}{N})\right) \\ & = & n\dfrac{k}{N}\left(1-\dfrac{k}{N}\right)\left(1-\dfrac{n-1}{N-1}\right) \\ & = & n\dfrac{k}{N}\left(1-\dfrac{k}{N}\right)\dfrac{N-n}{N-1} \\ & = & k\dfrac{n}{N}\left(1-\dfrac{n}{N}\right)\dfrac{N-k}{N-1}. \end{array}$$

Remarque 6.48. *Une variable suivant loi hypergéométrique a la même espérance qu'une loi binomiale $\mathcal{B}(k, \frac{n}{N})$ et sa variance ne diffère de celle de cette binomiale que d'un facteur $\frac{N-k}{N-1}$.*

6.10 Calcul des premiers moments des lois à densité usuelles

6.10.1 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

$$\mathbb{E}Y = \mathbb{E}\frac{a+b}{2} + \frac{b-a}{2}\mathbb{E}X = \frac{a+b}{2} \quad \text{et} \quad \operatorname{Var} Y = (\frac{b-a}{2})^2 \operatorname{Var} X = \frac{(b-a)^2}{12}.$$

6.10.2 Loi gaussienne

Soit X une variable aléatoire suivant la loi gaussienne $\mathcal{N}(0,1)$. On rappelle que la densité de X est

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

Lemme 6.49. *Soit g une fonction dérivable sur $\mathbb{R}$ telle qu'il existe $A > 0$ et $c > 0$ tels que*

$$\forall x \in \mathbb{R} \quad |g(x)| + |g'(x)| \le A\exp(+c|x|).$$

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

$$(g(x)f(x))' = (g'(x) - xg(x))f(x).$$

On a donc pour tout $a,b \in \mathbb{R}$,

$$g(b)f(b) - g(a)f(a) = \int_a^b g'(x)f(x)\ dx - \int_a^b xg(x)f(x)\ dx.$$

Les hypothèses faites sur g et g' assurent l'intégrabilité sur $\mathbb{R}$ de $g'f$ et gf. Comme, de plus $\lim\limits_{a\to-\infty} g(a)f(a) = \lim\limits_{b\to+\infty} g(b)f(b) = 0$, 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 $Y = m + \sigma X$, on sait alors 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.

6.10.3 Loi Gamma

Soit X une variable aléatoire suivant la loi Gamma $\Gamma(a,\gamma)$. Alors, X admet des moments de tout ordre, avec pour tout $\alpha > -a$, on a

$$\mathbb{E}X^\alpha = \gamma^{-\alpha}\frac{\Gamma(a+\alpha)}{\Gamma(a)}.$$

En particulier $\mathbb{E}X = \frac{a}{\gamma}$ et $\operatorname{Var} X = \frac{a}{\gamma^2}$.

Démonstration. Pour tous a et γ strictement positifs, on note $f_{a,\gamma}$ la densité de la loi $\Gamma(a,\gamma)$, soit (rappel)

$$f_{a,\gamma}(x) = \mathbb{1}_{\mathbb{R}_+}(x)\frac{\gamma^a}{\Gamma(a)}x^{a-1}e^{-\gamma x}.$$

D'après le théorème de transfert, on a

$$\mathbb{E}[X^\alpha] = \int_{\mathbb{R}_+} x^\alpha f_{a,\gamma}(x)\,dx = \int_{\mathbb{R}_+} \gamma^{-\alpha}\frac{\Gamma(a+\alpha)}{\Gamma(a)}f_{a+\alpha,\gamma}(x)\,dx = \gamma^{-\alpha}\frac{\Gamma(a+\alpha)}{\Gamma(a)}.$$

Ainsi $\mathbb{E}X = \gamma^{-1}\frac{\Gamma(a+1)}{\Gamma(a)} = \frac{a}{\gamma}$, $\mathbb{E}[X^2] = \gamma^{-2}\frac{\Gamma(a+2)}{\Gamma(a)} = \frac{a(a+1)}{\gamma^2}$ et
$\operatorname{Var} X = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \frac{a(a+1)}{\gamma^2} - \frac{a^2}{\gamma^2} = \frac{a}{\gamma^2}$. □

6.10.4 Loi exponentielle

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 $\operatorname{Var} X = \frac{1}{\lambda^2}$.

6.10.5 Loi Bêta

Soient a, b des réels strictement positifs. On rappelle que la densité de probabilité de la loi Bêta de paramètres a et b est $x \mapsto \frac{1}{\beta(a,b)}x^{a-1}(1-x)^{b-1}\mathbb{1}_{[0,1]}(x)$. Ainsi, on a pour $k > 0$

$$\begin{aligned}\mathbb{E}[X^k] &= \frac{1}{\beta(a,b)}\int_0^1 x^{a+k-1}(1-x)^{b-1}\,dx = \frac{\beta(a+k,b)}{\beta(a,b)} \\ &= \frac{\Gamma(a+k)\Gamma(b)}{\Gamma(a+b+k)}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} = \frac{\Gamma(a+k)}{\Gamma(a)}\frac{\Gamma(a+b)}{\Gamma(a+b+k)}.\end{aligned}$$

En particulier

$$\mathbb{E}[X] = \frac{a\Gamma(a)}{\Gamma(a)}\frac{\Gamma(a+b)}{(a+b)\Gamma(a+b)} = \frac{a}{a+b}$$

et $\mathbb{E}[X^2] = \frac{a(a+1)}{(a+b)(a+b+1)}$, d'où l'on déduit

$$\operatorname{Var} X = \frac{ab}{(a+b)^2(a+b+1)}.$$

6.10.6 Loi 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

$$\mathbb{E}|X|^k = \int \frac{1}{\pi} \frac{b|x|^k}{(x-a)^2 + b^2}\, dx = +\infty,$$

donc la loi de Cauchy n'admet pas de moment d'ordre 1, ni a fortiori de moment d'ordre supérieur.

6.11 Exercice détaillé : polynômes de Bernstein

Nous présentons ici une preuve probabiliste de la convergence des polynômes de Bernstein.
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 de degré n associé à f :

$$B_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} \ .$$

Pour tout $x \in [0,1]$ on se donne une suite (X_k) de variables de Bernoulli indépendantes de même paramètre x. On note $S_n = \sum_{k=1}^n X_k$.

1. Déterminer la moyenne $\mathbb{E}[f(\frac{S_n}{n})]$.
2. Soit pour tout $\varepsilon > 0$, le réel $\delta(\varepsilon)$ défini par

$$\delta(\varepsilon) = \sup\{|f(x) - f(y)| : x, y \in [0,1] \text{ et } |x-y| \leq \varepsilon\} \ .$$

   (a) Démontrer que $\delta(\varepsilon)$ tend vers 0 avec ε.

   (b) Démontrer que

$$\sup_{x\in[0,1]} |B_n(x) - f(x)| \leq \delta(\varepsilon) + \frac{\|f\|_\infty}{2n\varepsilon^2} \ .$$

   En déduire que la suite des polynômes B_n converge vers f uniformément sur $[0,1]$.

Solution

1. Notons déjà que la loi de S_n est la loi binomiale de paramètres (n,x). On peut donc appliquer le théorème de transfert à S_n et la fonction $x \mapsto f(\frac{x}{n})$, ce qui nous donne

$$\begin{aligned}\mathbb{E} f(\frac{S_n}{n}) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \mathbb{P}(S_n = k) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} \\ = B_n(x).\end{aligned}$$

2. (a) Comme $[0,1]$ est compact et f est continue, f est uniformément continue et on obtient immédiatement la convergence de $\delta(\varepsilon)$ vers 0 lorsque ε tend vers 0.

   (b) On découpe l'intégrale suivante en deux parties

$$\begin{aligned} & B_n(x) - f(x) \\ = \ & \int_\Omega \left[f\left(\tfrac{S_n}{n}\right) - f(x)\right]\, d\mathbb{P} \\ = \ & \int_{|\frac{S_n}{n} - x| \leq \varepsilon} \left[f\left(\tfrac{S_n}{n}\right) - f(x)\right]\, d\mathbb{P} + \int_{|\frac{S_n}{n} - x| > \varepsilon} \left[f\left(\tfrac{S_n}{n}\right) - f(x)\right]\, d\mathbb{P}. \end{aligned}$$

   Ainsi, on a par l'inégalité triangulaire, pour tout $\varepsilon > 0$

$$\begin{aligned} & |B_n(x) - f(x)| \\ \leq \ & \delta(\varepsilon)\mathbb{P}\left(\left|\frac{S_n}{n} - x\right| \leq \varepsilon\right) + \int_{|\frac{S_n}{n} - x| > \varepsilon} \left|f\left(\frac{S_n}{n}\right) - f(x)\right|\, d\mathbb{P} \\ \leq \ & \delta(\varepsilon) + 2\|f\|_\infty \mathbb{P}\left(\left|\frac{S_n}{n} - x\right| > \varepsilon\right). \end{aligned}$$

Il suffit d'appliquer l'inégalité de Chebychev à cette dernière probabilité pour obtenir

$$\mathbb{P}\left(\left|\frac{S_n}{n}-x\right|>\varepsilon\right) \le \frac{x(1-x)}{n\varepsilon^2} \le \frac{1}{4n\varepsilon^2}.$$

On en déduit donc la relation cherchée.

Pour tout $x \in [0,1]$ et $\varepsilon > 0$, on a $|B_n(x) - f(x)| \le \delta(\varepsilon) + \|f\|_\infty \frac{1}{2n\varepsilon^2}$. On en déduit successivement que

$$\forall \varepsilon > 0 \quad \forall n \ge 1 \quad \|B_n - f\|_\infty \le \delta(\varepsilon) + \|f\|_\infty \frac{1}{2n\varepsilon^2},$$

puis

$$\forall \varepsilon > 0 \qquad \varlimsup_{n\to+\infty} \|B_n - f\|_\infty \le \delta(\varepsilon),$$

et enfin $\varlimsup\limits_{n\to+\infty} \|B_n - f\|_\infty \le 0$, ce qui donne le résultat voulu.

6.12 Exercices sur l'espérance

6.12.1 Exercices de la série 1

Exercice 116. On rappelle que pour x réel, $\{x\}$ désigne la partie fractionnaire de x : $x = \lfloor x \rfloor + \{x\}$.
Soient $\alpha \in \mathbb{R}$ et X une variable aléatoire sur $(\Omega, \mathcal{F}, \mathbb{P})$ suivant la loi uniforme sur $[0,1]$. Montrer que $Y = \{X + \alpha\}$ a la même loi que X.

Exercice 117. Soit X une variable aléatoire de carré intégrable (c'est-à-dire que $\mathbb{E}(X^2) < +\infty$).

1. Trouver la constante m telle que $\mathbb{E}[(X-m)^2] = \inf\limits_{x\in\mathbb{R}} \mathbb{E}[(X-x)^2]$.
2. On dit que λ est une médiane de X si

$$\mathbb{P}(X \ge \lambda) \ge 1/2 \text{ et } \mathbb{P}(X \le \lambda) \ge 1/2.$$

   Montrer qu'il existe toujours au moins une médiane.
   On suppose que X suit une loi de Bernoulli de paramètre $1/2$. Trouver les médianes de X. Faire de même si X suit une loi de Bernoulli de paramètre $p \ne 1/2$.

Exercice 118. Soit X une variable aléatoire de densité $f(x) = \frac{\mathbb{1}_{[0,1]}(x)}{(1+x)\log 2}$. Montrer que $\{\frac{1}{X}\} = \frac{1}{X} - \lfloor \frac{1}{X} \rfloor$ a même loi que X.

Exercice 119. Soit X une variable aléatoire uniforme sur l'intervalle $[-1,1]$. Posons $Y = |4X^2 - 1|$. Calculer $\mathbb{E}(Y),\ \mathbb{E}(X+Y),\ \mathbb{E}[(Y+4X^2)^3]$.

Exercice 120.
1. On suppose que $X_1, \dots, X_n$ sont des variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$.
   Calculer $\mathbb{E}\left(\max\{X_1, \dots, X_n\}\right)$.
2. On suppose que $Y_1, \dots, Y_n$ sont des variables aléatoires indépendantes suivant la loi exponentielle de paramètre 1.
   Calculer $\mathbb{E}\left(\min\{Y_1, \dots, Y_n\}\right)$.

Exercice 121. *Une preuve probabiliste d'un théorème d'Erdös.*
On dit que $A \subset \mathbb{Z}$ est sans-somme s'il n'existe pas $x, y, z \in A$ avec $x + y = z$.

Soit A une partie de $\mathbb{Z}$ non vide ne contenant pas 0. On va montrer qu'il existe $B \subset A$ tel que $|B| > |A|/3$ et B est sans-somme.

1. Montrer qu'il y a une infinité de nombres premiers de la forme $3k+2$. On notera désormais p un nombre premier de la forme $3k+2$ tel que $A \subset [-p/3, p/3] \backslash \{0\}$.
2. Soit B_0 l'ensemble des classes de $[k+1, \dots, 2k+1]$ dans $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$. Montrer que B_0 est sans-somme.
3. On rappelle que $\mathbb{F}_p$ est un corps : l'ensemble des éléments non nuls de $\mathbb{F}_p$ est aussi l'ensemble des éléments inversibles, noté $\mathbb{F}_p^*$. Soit X suivant la loi uniforme sur $\mathbb{F}_p^*$. Notons π_A la restriction à A de la projection canonique de $\mathbb{Z}$ dans $\mathbb{F}_p$ (la classe modulo p) et $A' = \pi_A(A)$. Vérifier que $B' = A' \cap (X.B_0)$ est un ensemble aléatoire sans-somme .
4. Montrer que $\mathbb{E}[|B'|] = |A'| . \frac{k+1}{3k+1} > |A| \times \frac{1}{3}$.
5. En déduire que $\mathbb{P}(|B'| > |A|/3) > 0$.
6. Conclure.

Exercice 122. Combien y a-t-il de nombres à huit chiffres (les chiffres sont pris dans $0, 1, \dots, 9$) dont la somme des chiffres est égale à 40 ?

Indication : si on note A_0 les solutions dans $\mathbb{N}^8$ de $x_1 + \dots + x_8 = 40$ et A_i les éléments de A_0 avec $x_i \geq 10$, on pourra par exemple remarquer que $A_1 \cap A_2 \cap \dots \cap A_i$ est en bijection avec l'ensemble des solutions de
$y_1 + \dots + y_8 = 40 - 10i$ dans $\mathbb{N}^8$.

Exercice 123.
1. Soient X et Y deux variables aléatoires réelles indépendantes et identiquement distribuées. Montrer que

$$\mathbb{P}(|X - Y| \leq 2) \leq 5\mathbb{P}(|X - Y| \leq 1).$$

Indication : remarquer que $\mathbb{P}(|X - Y| \leq 2) \leq \mathbb{P}(|\lfloor X \rfloor - \lfloor Y \rfloor| \leq 2)$ et $\mathbb{P}(\lfloor X \rfloor = \lfloor Y \rfloor) \leq \mathbb{P}(|X - Y| \leq 1)$. On est ainsi ramené à l'étude de variables discrètes.

2. Soit C la plus petite constante telle que tout couple (X, Y) de variables indépendantes de même loi vérifie

$$\mathbb{P}(|X - Y| \leq 2) \leq C\mathbb{P}(|X - Y| \leq 1).$$

Montrer que $3 \leq C \leq 5$.
Indication : considérer deux variables X et Y indépendantes suivant chacune la loi uniforme sur $\{2, 4, 6, \dots, 2N\}$.

Exercice 124. *Loi multinomiale.*

1. Soit X une variable aléatoire de loi concentrée sur $\{1, \dots, k\}$:

$$\mathbb{P}(X = j) = p_j, \ \ 1 \leq j \leq k.$$

Soit $Z^j = \mathbb{1}_{\{X=j\}}$ et $s = (s_1, \dots, s_k) \in \mathbb{R}_+^k$. Calculer :

$$\mathbb{E}\left[s_1^{Z^1} s_2^{Z^2} \dots s_k^{Z^k}\right].$$

2. Soit $X_1, \dots, X_n$ un n-échantillon de la loi précédente (n variables indépendantes de même loi). Posons $N^j = \sum_{i=1}^{n} \mathbb{1}_{\{X_i=j\}}$.
Calculer $\mathbb{E}\left[s_1^{N^1} s_2^{N^2} \dots s_k^{N^k}\right]$.
En déduire, pour $a_1, \dots, a_k$ entiers de somme n :

$$\mathbb{P}(N^1 = a_1, \dots, N^k = a_k) = \binom{n}{a_1, a_2, \dots, a_k} p_1^{a_1} \dots p_k^{a_k},$$

où les $\binom{n}{a_1,a_2,\ldots,a_k}$ sont les coefficients qui apparaissent dans la formule du multinôme $(X_1+\cdots+X_k)^n = \sum \binom{n}{a_1,a_2,\ldots,a_k} \prod_{i=1}^k X_i^{a_i}$, où la sommation a lieu sur les k-uplets d'entiers naturels de somme n (voir l'annexe de dénombrement). On dit que le vecteur $N=(N^1,\ldots,N^k)$ suit la loi multinomiale de paramètre $(p_1,\ldots,p_k)$ et d'ordre n.

Exercice 125. Soient X et Y deux variables définies sur $(\Omega,\mathcal{A},\mathbb{P})$, ne pouvant prendre que deux valeurs distinctes. Montrer que X et Y sont indépendantes si et seulement si $\mathbb{E}(XY)=\mathbb{E}(X)\mathbb{E}(Y)$.
Montrer que ce résultat ne s'étend pas au cas où X,Y prennent plus de deux valeurs.

Exercice 126. Soient $X_1,\ldots,X_n$ des variables aléatoires indépendantes d'espérance 0 et de variance finie. Soit $S_n=X_1+\cdots+X_n$. Montrer l'inégalité de Kolmogorov : pour tout $x>0$

$$\mathbb{P}\left(\max_{1\le k\le n}|S_k|\ge x\right)\le x^{-2}\sum_{1\le i\le n}\mathrm{Var}(X_i).$$

Indication : considérer les événements disjoints

$$A_k=\bigcap_{j<k}\{|S_j|<x\}\cap\{|S_k|\ge x\}$$

et commencer par montrer la minoration

$$\mathbb{E}(S_n^2)\ge \sum_{k=1}^n \mathbb{E}\left(\mathbb{1}_{A_k}\left(S_k^2+2S_k(S_n-S_k)\right)\right).$$

Utiliser ensuite l'inégalité de Chebychev, $\mathbb{P}(A_k)\le x^{-2}\mathbb{E}(\mathbb{1}_{A_k}S_k^2)$.

Exercice 127. Soient X et Y deux variables aléatoires indépendantes identiquement distribuées et centrées. Montrer que

$$\mathbb{E}(|X-Y|)\ge\mathbb{E}(|X|).$$

Indication : montrer que l'application $y\mapsto\mathbb{E}(|X-y|)$ est convexe.

Exercice 128. Soient X et Y deux variables aléatoires indépendantes suivant la loi exponentielle $\mathcal{E}(a)$, avec $a>0$. Calculer $\mathbb{E}(\max(X,Y))$ et $\mathbb{E}(\min(X,Y))$.

Exercice 129. Soient $X_1,\ldots,X_n$ des variables aléatoires indépendantes de loi exponentielle de paramètre $a>0$: $\mathbb{P}(X_k\ge x)=e^{-ax}$ pour $x\ge 0$. Posons $M=\inf\{X_1,\ldots,X_n\}$.

1. Calculer $\mathbb{P}(M\ge x)$ pour tout $x\in\mathbb{R}$. En déduire la loi de M.
2. Pour $t\ge 0$, on pose $N_t=\mathrm{card}(k\in\{1,\ldots,n\}|X_k\ge t)$. Déterminer la loi, l'espérance et la variance de N_t (si X_k est la durée de vie du k-ième individu, N_t est le nombre de survivants au temps t).

Exercice 130. La variable aléatoire X est uniformément répartie sur l'intervalle $[0,3]$. Soit $Y=\max\{2,X\}$. Trouver la fonction de répartition et l'espérance de Y. La variable aléatoire Y admet-elle une densité ?

Exercice 131. Soient X et Y des variables aléatoires indépendantes de loi $\mathcal{N}(0,1)$. Déterminer la loi du couple $(X/Y,Y)$ puis celle de X/Y. Les variables X/Y et Y sont-elles indépendantes ?

Exercice 132. Un marchand de journaux reçoit chaque jour la visite d'un nombre aléatoire de clients représenté pour un jour fixé par une variable aléatoire entière positive N. Nous supposons connue la loi de N (grâce par exemple à une statistique portant sur les ventes des jours précédents). Le marchand se propose d'optimiser le nombre k de journaux qu'il commande à l'éditeur, compte-tenu que :

1. chaque journal vendu lui rapporte un bénéfice a,
2. chaque journal invendu lui vaut une perte b,
3. chaque client insatisfait lui coûte une somme c, dont le montant représente (en terme monétaire) le risque de perdre ce client définitivement.

Si G_k représente le gain aléatoire du marchand de journaux, déterminer k afin de rendre $\mathbb{E}(G_k)$ maximale.

Exercice 133. *Volume de la boule unité de* $(\mathbb{R}^n, \|.\|_p)$.
On rappelle que pour $x \in \mathbb{R}^n$, on note $\|x\|_p = (\sum_{k=1}^n |x_k|^p)^{1/p}$.

1. Soient $Y_1, \dots, Y_n$ des variables aléatoires indépendantes suivant la loi $\Gamma(\frac{1}{p}, 1)$. On pose $X_k = Y_k^{1/p}$ et $S = X_1^p + \dots + X_n^p$. Montrer que X_1 admet la densité $x \mapsto \frac{e^{-x^p} \mathbb{1}_{\mathbb{R}_+}(x)}{\Gamma(\frac{1}{p}+1)}$ et que $S \sim \Gamma(n/p, 1)$.
2. Soit $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ une fonction mesurable positive. Montrer que
$$\int_{\mathbb{R}_+^n} \phi(\|x\|_p^p)\, d\lambda^{\otimes n}(x) = \mathbb{E}[\psi(S)], \text{ où } \psi(x) = \Gamma\left(\frac{1}{p} + 1\right)^n \phi(x) e^x.$$
En déduire que
$$\int_{\mathbb{R}^n} \phi(\|x\|_p^p)\, d\lambda^{\otimes n}(x) = \frac{2^n \Gamma(\frac{1}{p} + 1)^n}{\Gamma(n/p)} \int_{\mathbb{R}_+} u^{\frac{n}{p}-1} \phi(u)\, du.$$
3. Montrer que le volume de la boule unité de $(\mathbb{R}^n, \|.\|_p)$ est $\frac{2^n \Gamma(\frac{1}{p}+1)^n}{\Gamma(\frac{n}{p}+1)}$.
4. Soit $n \geq 2$. On pose $T_n = \{(x_1, \dots, x_n) \in \mathbb{R}_+^n ; x_1 + \dots + x_n \leq 1\}$.
Calculer $\displaystyle\int_{T_n} \frac{d\lambda^{\otimes n}(x)}{x_1 + \dots + x_n}$.

Exercice 134. Soit (X, Y) un couple de variables aléatoires de densité de probabilité :
$$g(x,y) = \begin{cases} \dfrac{1}{2} x\, y \;\text{ si } (x,y) \in D \\ \\ 0 \text{ si } (x,y) \notin D \end{cases}$$
où D est le premier quart du disque centré en 0 de rayon 2. Les variables X et Y sont-elles indépendantes ? Déterminer la densité de probabilité de la variable $R = \sqrt{X^2 + Y^2}$, puis celle de $Z = R^2$.

Exercice 135. Soient n et k des entiers avec $1 \leq k \leq n$. On pose
$$\Omega = \left\{ x \in \{0,1\}^n ; \sum_{i=1}^n x_i = k \right\}.$$
On note $\mathbb{P}$ la loi uniforme sur Ω. On note $X = (X_1, \dots, X_n)$ le vecteur aléatoire représentant les composantes d'un élément de Ω

1. Soient i et j deux entiers entre 1 et n distincts, $a, b \in \{0,1\}$. Montrer qu'il y a une bijection entre
$$\Omega_{1,2}^{a,b} = \{x \in \Omega ; x_1 = a \text{ et } x_2 = b\}$$
et
$$\Omega_{i,j}^{a,b} = \{x \in \Omega ; x_i = a \text{ et } x_j = b\}.$$
En déduire que les vecteurs (X_1, X_2) et (X_i, X_j) ont même loi.

2. Montrer que pour tout i, X_i suit la loi de Bernoulli de paramètre k/n. Donner la variance de X_i.
3. Pour tout entier r avec $0 \leq r \leq n$, on pose $S_r = \sum_{i=1}^{r} X_i$. Calculer l'espérance de S_r. Montrer qu'à n fixé, il existe un polynôme P_n de degré 2 tel que pour tout r entre 0 et n, $\operatorname{Var} S_r = P_n(r)$.
4. Montrer que $\operatorname{Var}(X_1 + \cdots + X_n) = 0$. En déduire la variance de S_r.
5. Proposer une expérience basée sur un tirage sans remise qui puisse se modéliser à l'aide de la variable S_r.

Exercice 136. *Loi Bêta de deuxième espèce – formule des compléments.*

1. Soient X, Y des variables aléatoires positives indépendantes admettant les densités respectives f_X, f_Y. Montrer que $V = \frac{X}{Y}$ admet la densité
$$v \mapsto \left(\int_{]0,+\infty[} f_X(uv) f_Y(u) u \, d\lambda(u) \right) \mathbb{1}_{\{v>0\}}.$$
2. Dans le cas où $X \sim \Gamma(a, \gamma)$ et $Y \sim \Gamma(b, \gamma)$, avec $a, b > 0$, montrer que $V = \frac{X}{Y}$ admet la densité
$$v \mapsto \left(\frac{1}{B(a,b)} \frac{v^{a-1}}{(1+v)^{a+b}} \right) \mathbb{1}_{\{v>0\}}, \text{ où } B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$
Cette loi est appelée *loi Bêta prime* ou *loi Bêta de deuxième espèce* de paramètres a et b. En déduire que
$$B(a,b) = \int_0^{+\infty} \frac{v^{a-1}}{(1+v)^{a+b}} \, dv.$$
3. *Cette question utilise la théorie de la variable complexe.*
Soit $n \geq 2$ un entier. Montrer que
$$B\left(\frac{1}{n}, 1 - \frac{1}{n}\right) = n \int_0^{+\infty} \frac{dx}{1+x^n} = \frac{\pi}{\sin(\pi/n)}.$$
(On pourra appliquer la formule des résidus à un chemin fermé reliant 0, R, $R\omega^2$, avec $\omega = \exp(i\pi/n)$.) En déduire que pour tout nombre complexe z avec $0 <$ Re $z < 1$, on a
$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin \pi z}.$$
C'est la formule des compléments.

Exercice 137. *Nombres de Bell et loi de Poisson.*
Le n-ième nombre de Bell, noté B_n, est le nombre de partitions de l'ensemble $\{1, \ldots, n\}$.

1. À chaque fonction f de $\{1, \ldots, n\}$ dans $\{1, \ldots, p\}$, on associe une partition $P(f)$, formée par les ensembles antécédents par f de chaque point de Im f. Une partition Π étant donnée, formée de $N(\Pi)$ blocs, combien existe-t-il de fonctions f telles que $P(f) = \Pi$?
2. En déduire que si l'on pose $P_0 = 1$, $P_1 = X$, $P_i(X) = X(X-1)$, ..., $P_n(X) = X(X-1)\ldots(X-i+1)$, on a l'identité entre polynômes
$$X^n = \sum_{\Pi \text{ Partition de } \{1,\ldots,n\}} P_{N(\Pi)}(X).$$

3. Soit X une variable aléatoire suivant la loi de Poisson de paramètre 1. Pour p entier naturel, calculer $\mathbb{E}(P_p(X))$.
4. En déduire que pour tout n, on a $B_n = \mathbb{E}(X^n)$, puis établir la formule de Dobiński :

$$B_n = \frac{1}{e}\sum_{k=0}^{+\infty} \frac{k^n}{k!}.$$

5. Montrer que la fonction génératrice exponentielle des nombres de Bell, définie par $\phi(x) = \sum_{n=0}^{+\infty} \frac{B_n}{n!} x^n$, vaut

$$\phi(x) = \exp(\exp(x) - 1).$$

6. À l'aide de la formule de Cauchy, en déduire que $B_n \le n! \frac{e^{n-1}}{(\log n)^n}$.

Remarque : prolongeant les idées de la dernière question, la méthode du col permet de donner un équivalent de B_n . Voir par exemple De Bruijn [3], page 102, ou Flajolet–Sedgewick [5], page 560.

Exercice 138. Soient X et Y deux variables aléatoires réelles indépendantes de même loi μ avec $\mu(]0,+\infty[) = 1$. On pose $Z = X/Y$ et $Z' = \log Z$.

1. On suppose que $X' = \log X$ admet un moment d'ordre deux. Montrer que

$$\operatorname{Var} X' = \frac{1}{2}\mathbb{E}(Z'^2) = \mathbb{E}\left(Z'^2 \mathbb{1}_{\{Z'>0\}}\right) = 2\int_0^{+\infty} t\mathbb{P}(Z' > t)\, dt.$$

2. On suppose que X suit la loi exponentielle de paramètre 1. Montrer que

$$\operatorname{Var}(\log X) = \int_0^{+\infty} \frac{2t}{1+e^t}\, dt = \frac{\pi^2}{6}.$$

   Remarque : on peut noter que $G = -\log X$ suit la loi de Gumbel, c'est-à-dire que pour tout t réel, $\mathbb{P}(G \le t) = \exp(-e^{-t})$. On a ainsi calculé la variance de la loi de Gumbel.

Exercice 139. Soient X et Y deux variables indépendantes, où X suit la loi $\mathcal{N}(0,1)$ et Y la loi Gamma $\Gamma(\frac{n}{2}, \frac{1}{2})$ [1].
On dit alors que $T = \frac{X}{\sqrt{Y/n}} = X\sqrt{n/Y}$ suit la loi de Student à n degrés de liberté et on la note t_n. Montrer que t_n est une loi à densité ; la déterminer.

Exercice 140. *Une preuve probabiliste d'un développement eulérien.*

1. Soient $s > 1$ et $(N_i)_{i\ge1}$ une suite de variables aléatoires indépendantes telles que $1+N_i \sim \mathcal{G}(1-p_i^{-s})$, où $(p_i)_{i\ge1}$ est la suite des nombres premiers. Soit M le nombre d'indices i tels que $N_i \ne 0$. Montrer que $\mathbb{E}(M) < +\infty$, puis que $X = \prod_{i=1}^{+\infty} p_i^{N_i}$ suit la loi Zêta de paramètre s.
2. Une fonction $\phi : \mathbb{N}^* \to \mathbb{C}$ est dit multiplicative si est vérifiée la relation $\phi(pq) = \phi(p)\phi(q)$ dès que p et q sont premiers entre eux. Si c'est vrai pour tous les couples $(p,q) \in (\mathbb{N}^*)^2$, alors la fonction est dite complètement multiplicative.
   Soit ϕ une fonction multiplicative bornée. Montrer que

$$\sum_{n=1}^{+\infty} \frac{\phi(n)}{n^s} = \prod_{i=1}^{+\infty}\left(\sum_{j=0}^{+\infty} p_i^{-sj}\phi(p_i^j)\right).$$

   En particulier, si ϕ est complètement multiplicative, montrer que

$$\sum_{n=1}^{+\infty} \frac{\phi(n)}{n^s} = \prod_{i=1}^{+\infty}(1 - p_i^{-s}\phi(p_i))^{-1}.$$

1. appelée aussi loi du χ^2 à n degrés de libertés

Exercice 141. *Processus de Poisson.*
Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi exponentielle de paramètre λ. On pose $S_0 = 0$ et, pour $n \geq 1$, $S_n = \sum_{k=1}^{n} X_k$. Pour $t > 0$, on définit $N_t = \sup\{n \geq 0; S_n < t\}$.

1. Montrer que pour tout entier n strictement positif, la variable aléatoire S_n suit la loi Gamma $\Gamma(n, \lambda)$.
2. Soient $n \geq 1, t > 0$. Montrer que $\mathbb{P}(S_n < t) = g_n(\lambda t)$, où on a posé
$$g_n(t) = \frac{1}{\Gamma(n)} \int_0^t x^{n-1} e^{-x} \, dx,$$
puis que $g_{n+1}(t) = -\frac{1}{\Gamma(n+1)} e^{-t} t^n + g_n(t)$.
3. Montrer que $\mathbb{P}(N_t = n) = \mathbb{P}(S_n < t) - \mathbb{P}(S_{n+1} < t)$. En déduire que N_t suit la loi de Poisson de paramètre λt.

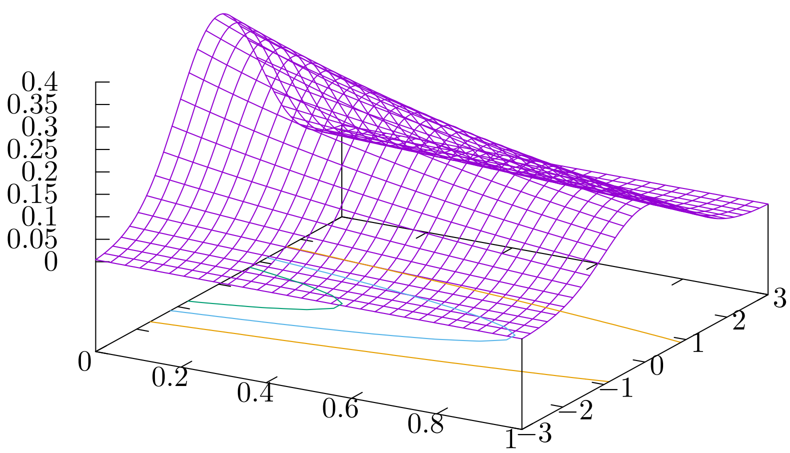
On dit alors que $(N_t)_{t\geq 0}$ est un processus de Poisson d'intensité λ. Noter que ce procédé permet de simuler une variable aléatoire suivant la loi de Poisson à partir d'un générateur de loi uniforme sur $[0, 1]$, puisque la loi exponentielle se simule simplement par la méthode d'inversion.

6.12.2 Exercices de la série 2

Exercice 142. 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 ?

Exercice 143. 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 ?



Exercice 144. 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 que $\mathbb{1}_{A\Delta B} = (\mathbb{1}_A - \mathbb{1}_B)^2$. En déduire que

$$\mathbb{P}(A\Delta B) = \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A \cap B).$$

Exercice 145. Soient A, B deux éléments observables. Montrer que

$$|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \leq \sqrt{\mathbb{P}(A)\mathbb{P}(B)}.$$

Exercice 146. Calculer $\mathbb{E}(\sin X)$, où $\mathbb{P}(X = \frac{\pi}{2}) = \frac{1}{6}$, $\mathbb{P}(X = \frac{\pi}{3}) = \frac{1}{3}$ et $\mathbb{P}(X = \frac{\pi}{6}) = \frac{1}{2}$.

Exercice 147.

1. Soit X une variable aléatoire de carré intégrable. On note σ^2 sa variance et m son espérance. Montrer que pour tout a réel, on a $\sigma^2 \leq \mathbb{E}(X - a)^2$.
2. Soient $a_1 \leq a_2 \leq \cdots \leq a_n$. On pose

$$m = \frac{a_1 + \cdots + a_n}{n}.$$

Montrer que

$$\frac{1}{n}\sum_{k=1}^{n}(a_k - m)^2 \leq \frac{(a_n - a_1)^2}{4}.$$

Exercice 148. 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 $\mathbb{E}(\sigma(X)) \geq \mathbb{E}X$.

Exercice 149. On suppose que $Y = \log X$ vérifie $Y \sim \mathcal{N}(m, \sigma^2)$ (on dit alors que X est log-normale). Calculer $\mathbb{E}X$ et $\operatorname{Var} X$.

Exercice 150. 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 ?

Exercice 151. Soit X une variable aléatoire suivant la loi uniforme sur $[0, 1]$. Déterminer la loi de $Z = -\log(1 - X)$.

Exercice 152. Soit X une variable aléatoire de densité f. Montrer que la variable aléatoire $|X|$ admet comme densité

$$x \mapsto \mathbb{1}_{\mathbb{R}_+}(x)(f(x) + f(-x)).$$

Exercice 153. Soit X une variable aléatoire positive de densité f. Montrer que la variable aléatoire $X^{1/2}$ admet comme densité

$$x \mapsto \mathbb{1}_{\mathbb{R}_+}(x)2xf(x^2).$$

Exercice 154. Soit X une variable aléatoire normale centrée réduite. Montrer que la variable aléatoire X^2 est à densité et la déterminer.

Exercice 155. 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).

Exercice 156. Soit $(X_n)_{n\geq 1}$ 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 $\mathbb{E}[X_iX_j] = b(i-j)$ quels que soient i et j dans $\mathbb{N}$.

1. Montrer que b est paire, puis exprimer simplement la variance de $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ en fonction des $b(i)$.
2. Montrer que si $b(i) < 0$ pour tout i non nul, alors la série de terme général $b(i)$ est convergente, avec $\sum_{i=1}^{+\infty}(-b(i)) \leq \frac{b(0)}{2}$.

Exercice 157. Soient n, r deux entiers tels que $1 \leq r \leq n$. On prend r nombres distincts au hasard dans $\{1, \dots, n\}$ et on note X le plus petit de ces r nombres.

1. Quelles valeurs peut prendre X ? Montrer que pour $k \in \{0, n-r\}$, on a

$$\mathbb{P}(X > k) = \frac{\binom{n-k}{r}}{\binom{n}{r}}.$$

2. En déduire que

$$\mathbb{E}X = \frac{\binom{n+1}{r+1}}{\binom{n}{r}} = \frac{n+1}{r+1}.$$

Exercice 158. Soient X et Y deux variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$.

1. 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.
2. Pour quelles valeurs de α la variable aléatoire $\frac{1}{|X-Y|^\alpha}$ est-elle intégrable ? Lorsqu'elle l'est, calculer sa valeur.

Exercice 159. *Lois de Poisson : inégalité de Le Cam.*
Soient $(p_1, \dots, p_n) \in [0,1]^n$, $Y_1, \dots, Y_1, Z_1, \dots, Z_n$ des variables aléatoires indépendantes telles que $Y_i \sim \mathcal{P}(p_i)$ et $Z_i \sim \mathrm{Ber}(1-(1-p_i)e^{p_i})$. On pose $X_i = \max\left(\mathbb{1}_{\{Y_i\geq 1\}}, Z_i\right)$, $X = \sum_{i=1}^{n} X_i$ et $Y = \sum_{i=1}^{n} Y_i$.

1. Quelle est la loi de $(X_1, \dots, X_n)$?
2. Montrer que pour tout i entre 1 et n, on a $\mathbb{P}(Y_i \neq X_i) \leq p_i^2$.
3. (a) Quelle est la loi de Y ?
   (b) Montrer que si $X_1, \dots, X_n$ sont des variables aléatoires indépendantes suivant respectivement la loi de Bernoulli de paramètre p_i, on a alors pour tout $A \subset \mathbb{N}$:

$$|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| \leq \sum_{i=1}^{n} p_i^2.$$

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 conclut que $A \subset \Omega$. » Si A est inclus dans Ω, on dit que A est un sous-ensemble, ou 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 indexée 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 indexée 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 $x \in \underset{i\in I}{\cap} A_i$, la preuve ressemblera donc à « Soit $i \in I$ ... (raisonnement) ... Donc $x \in A_i$. Comme i est quelconque, on a donc $x \in \underset{i\in I}{\cap} A_i$. »

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 D dans A. 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 est 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. On note $\mathfrak{S}(\Omega)$l'ensemble des permutations de Ω, et simplement $\mathfrak{S}_n$ pour l'ensemble des permutations de $\{1,\dots,n\}$.

Un ensemble Ω est 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 $\{1,\dots,n\}$.

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 $p \in \{0,\dots,n\}$, on note $\mathcal{B}_p(\Omega)$l'ensemble des parties de Ω de cardinal p. Par exemple $\mathcal{B}_2(\{a,b,c\}) = \{\{a,b\},\{b,c\},\{a,c\}\}$. On note de plus $\mathcal{P}(\Omega)$l'ensemble des parties de Ω, quel que soit leur cardinal. Par exemple $\mathcal{P}(\{a,b,c\}) = \{\varnothing,\{a\},\{b\},\{c\},\{a,b\},\{b,c\},\{a,c\},\{a,$

Soient 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.
Soit $f : A \to D$ une fonction, où A et D sont deux ensembles finis. Si $|A| = |D|$, alors f est injective si et seulement si f est surjective si et seulement si f est bijective.

Un ensemble Ω est dit *dénombrable* s'il existe une bijection entre Ω et $\mathbb{N}$.

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 cœur) le nombre d'éléments. La section suivante énoncera un certain nombre de résultats qu'il faut connaître.

A.3.2 Principe d'indépendance

Il s'agit juste de la formule

$$|A \times B| = |A| \cdot |B| \quad .$$

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 $A = \bigcup\limits_{i\in I} A_i$ et $i \neq j \Longrightarrow A_i \cap A_j = \varnothing$.[1]
On a alors

$$|A| = \sum_{i\in I} |A_i|.$$

1. Certains auteurs imposent en plus que les A_i soient tous non-vides. Il nous semble que cette condition supplémentaire a plus d'inconvénients que d'avantages.

Le résultat élémentaire suivant peut souvent être utile.

Théorème A.1. *Soient Ω un ensemble quelconque, I un ensemble d'indices fini ou dénombrable et $(\Omega_i)_{i\in I}$ une partition de Ω. Alors, les ensembles $(A\cap\Omega_i)_{i\in I}$ forment une partition de A.*

Démonstration. Posons $A_i = A\cap\Omega_i$. Comme $\Omega = \bigcup_{i\in I}\Omega_i$, on a

$$A = A\cap\Omega = \bigcup_{i\in I}(A\cap\Omega_i) = \bigcup_{i\in I} A_i.$$

D'autre part, pour $i\neq j$, on a $A_i\cap A_j\subset\Omega_i\cap\Omega_j=\varnothing$, d'où $A_i\cap A_j=\varnothing$. □

Lemme A.2. *Soit $\phi : D\to A$ une application surjective. Alors les ensembles $(\phi^{-1}(\{a\}))_{a\in A}$ forment une partition de D.*

La preuve de ce résultat est laissée en exercice au lecteur.

A.3.4 Lemme des bergers

Le lemme suivant peut également être utile

Lemme A.3 (des bergers)**.** *Soit ϕ une application* surjective *de D dans A. On suppose qu'il existe un entier $a\geq 1$ tel que*

$$\forall y\in A\quad |\phi^{-1}(\{y\})| = |\{x\in D;\phi(x)=y\}| = a$$

(autrement dit si tout élément de A admet exactement a antécédents), on a

$$|A| = \frac{|D|}{a}.$$

Démonstration. On applique le principe de partition avec $I=A$. Si l'on pose, pour $y\in A$, $D_y=\{x\in D;\phi(x)=y\}$, 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 dû à la procédure prétendument 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 à 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 A.4. *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

$$n! = n \times (n-1) \times \cdots \times 1.$$

Remarque A.5. $n!$ *se lit* "factorielle n" *ou* "n factorielle".

Exemple: Un professeur doit faire passer dans la journée cinq 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

Proposition A.6. *On pose $|A| = n$ et $|D| = p$. En vertu de la remarque faite en A.2, il existe une injection de D dans A si et seulement si $p \leq n$. Alors, le nombre d'injections de D dans 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 $\{1, \dots, p\}$. Avec un peu de réflexion, on 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. □

Remarque A.7.
- *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 cardinal est une bijection.*

Exemple: 3500 personnes se présentent au concours de l'Agrégation de Mathématiques. 300 places sont mises au 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 $|A| = 3500$). 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

Proposition A.8. *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 $\phi(f) = \text{Image}(f) = \{f(k); k \in \{1,\dots,p\}\}$.

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 au 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

Proposition A.9. *Le nombre total de parties de Ω est $|\mathcal{P}(\Omega)| = 2^{|\Omega|}$.*

Démonstration. Il suffit de remarquer que l'application

$$\mathcal{P}(\Omega) \to \{0;1\}^{\Omega}, \quad A \mapsto \mathbb{1}_A$$

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 alphabétiques 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 A.10. *Soient n et p des entiers. Si $n \geq p$, alors il existe exactement $\binom{n}{p}$ applications strictement croissantes de $\{1,\dots,p\}$ dans $\{1,\dots,n\}$. 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 croissante de p éléments pris dans $\{1,\dots,n\}$ revient à choisir une partie de $\{1,\dots,n\}$ possédant p éléments, puis à les ordonner avec l'ordre naturel. Or $\binom{n}{p}$ est, par définition, le nombre de parties de $\{1,\dots,n\}$ 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 $\{1,\dots,7\}$ dans $\{1,\dots,70\}$ 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 A.11. *Pour n et p des entiers vérifiant $n \geq p$, il existe exactement $\binom{n}{p}$ p-uplets $(x_1,\dots,x_p) \in (\mathbb{N}\backslash\{0\})^p$ solutions de l'inéquation :*

$$x_1 + x_2 + \cdots + x_p \leq n. \tag{A.1}$$

Démonstration. Il suffit de remarquer que l'application

$$(x_1,\dots,x_p) \quad \mapsto \quad (x_1, x_1+x_2, x_1+x_2+x_3, \dots, x_1+x_2+\cdots+x_p)$$

réalise une bijection entre l'ensemble des solutions recherchées de l'inéquation et l'ensemble des suites strictement croissantes de p éléments à valeurs dans $\{1,\dots,n\}$. □

Théorème A.12. *Pour n, p des entiers vérifiant $n \geq p$, il existe exactement $\binom{n-1}{p-1}$ p-uplets $(x_1,\dots,x_p) \in (\mathbb{N}\backslash\{0\})^p$ solutions de l'équation*

$$x_1 + x_2 + \cdots + x_p = n. \tag{A.2}$$

Démonstration. Il suffit de remarquer que les solutions $(x_1,\dots,x_p) \in (\mathbb{N}\backslash\{0\})^p$ 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 A.13. *Pour n, p des entiers positifs tels que $p \geq 1$, il existe exactement $\binom{n+p-1}{p-1}$ p-uplets $(x_1,\dots,x_p) \in \mathbb{N}^p$ solutions de l'équation*

$$x_1 + x_2 + \cdots + x_p = n. \tag{A.4}$$

Démonstration. Il suffit de remarquer que l'application

$$(x_1,\dots,x_p) \quad \mapsto (x_1+1,\dots,x_p+1)$$

réalise une bijection entre les solutions $(x_1,\dots,x_p) \in \mathbb{N}^p$ de l'équation (A.4) et les solutions $(x_1,\dots,x_p) \in (\mathbb{N}\backslash\{0\})^p$ de l'équation

$$x_1 + x_2 + \cdots + x_p = n+p. \tag{A.5}$$

□

Exemple: Quatre listes se présentent aux élections 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 de l'équation

$$x_1 + x_2 + x_3 + x_4 = 9,$$

où 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 A.14. *Soient n, p des entiers positifs.*
Il existe exactement $\binom{n+p}{p}$ p-uplets $(x_1,\dots,x_p) \in \mathbb{N}^p$ solutions de 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 les ensembles $A_1, A_2, \ldots, A_n$, on a

$$\begin{aligned}
\left| \bigcup_{i=1}^{n} A_i \right| &= \sum_{B \in \mathcal{P}(\{1,\ldots,n\}) \setminus \varnothing} (-1)^{1+\text{Card}(B)} \left| \bigcap_{j \in B} A_j \right| \\
&= \sum_{i=1}^{n} |A_i| - \sum_{1 \le i_1 < i_2 \le n} |A_{i_1} \cap A_{i_2}| + \cdots + \\
&\quad + \cdots + (-1)^{k+1} \sum_{1 \le i_1 < i_2 < \ldots < i_k \le n} |A_{i_1} \cap \ldots \cap A_{i_k}| + \cdots + \\
&\quad + \cdots + (-1)^{n+1} |A_1 \cap \ldots \cap A_n|.
\end{aligned} \tag{A.7}$$

Exemple: Pour $n = 3$, on a

$$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_2 \cap A_3| - |A_1 \cap A_3| + |A_1 \cap A_2 \cap A_3|.$$

On pourrait prouver la formule par récurrence sur n, mais c'est plutôt lourd. On préférera une preuve probabiliste (voir par exemple Garet–Kurtzmann).

A.7 Développement d'un produit de sommes

A.7.1 Développement d'un produit dans un anneau

Dans un anneau quelconque, on a l'identité très utile

$$\prod_{i=1}^{n} \left(\sum_{j=1}^{m} X_{i,j} \right) = \sum_{\phi \in \mathcal{F}(\{1,\ldots,n\},\{1,\ldots,m\})} \prod_{i=1}^{n} X_{i,\phi(i)}.$$

A.7.2 Formule du multinôme

En particulier si l'anneau est commutatif et si X_{ij} ne dépend pas de i, on a

$$\left(\sum_{j=1}^{m} X_j \right)^n = \sum_{\phi \in \mathcal{F}(\{1,\ldots,n\},\{1,\ldots,m\})} \prod_{i=1}^{n} X_{\phi(i)}.$$

Ainsi, si l'on note $\Psi^n(a_1, \ldots, a_m)$, l'ensemble des applications ϕ de $\{1, \ldots, n\}$ dans $\{1, \ldots, m\}$ telles que pour tout $i \in \{1, \ldots, m\}$, on a $|\phi^{-1}(i)| = a_i$ et si l'on note $\binom{n}{a_1, a_2, \ldots, a_m} = |\Psi^n(a_1, \ldots, a_m)|$, on obtient en regroupant les termes la formule du multinôme :

$$\left(\sum_{j=1}^{m} X_j \right)^n = \sum_{(a_1,\ldots,a_m)} \binom{n}{a_1, a_2, \ldots, a_m} \prod_{k=1}^{m} X_k^{a_k},$$

où la sommation a lieu sur les m-uplets d'entiers naturels de somme n.
Notons que d'après le théorème A.13, la somme comporte $\binom{n+m-1}{m-1}$ termes.
Si m est égal à 2, on retrouve simplement la formule du binôme, et on a $\binom{a_1+a_2}{a_1,a_2} = \binom{a_1+a_2}{a_1} = \binom{a_1+a_2}{a_2}$.

Calcul des coefficients du multinôme

Pour calculer $\binom{n}{a_1,a_2,\ldots,a_m}$, considérons l'application de $\Psi^n(a_1,\ldots,a_m)$ dans $\Psi^n(a_1,\ldots,a_{m-2},a_{m-1}+a_m)$ qui à ϕ associe la fonction $\phi' = \phi \wedge (m-1)$: on remplace chaque occurence de m par $m-1$.
L'image réciproque de $\phi' \in \Psi^n(a_1,\ldots,a_{m-2},a_{m-1}+a_m)$ est formée des fonctions ϕ qui coïncident avec ϕ' pour les x tels que $\phi'(x) < m-1$, et qui valent $m-1$ ou m pour les points x tels que $\phi'(x) = m-1$, avec la condition supplémentaire que parmi ces $a_{m-1}+a_m$ points, il doit y en avoir a_{m-1} tels que $\phi(x) = m-1$ et a_m tels que $\phi(x) = m$. Il est aisé de voir qu'il y a $\binom{a_m+a_{m-1}}{a_m}$ telles fonctions. Le lemme des bergers nous dit alors que

$$\binom{n}{a_1,a_2,\ldots,a_m} = \binom{a_m+a_{m-1}}{a_m}\binom{n}{a_1,a_2,\ldots,a_{m-2},a_{m-1}+a_m}.$$

On peut remarquer que $\binom{a_m+a_{m-1}}{a_m} = \binom{a_m+a_{m-1}}{a_{m-1},a_m} = \frac{(a_m+a_{m-1})!}{a_{m-1}!a_m!}$. On établit alors aisément par récurrence sur m que

$$\binom{n}{a_1,a_2,\ldots,a_m} = \frac{n!}{a_1!a_2!\ldots a_m!}.$$

A.8 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 considère l'ensemble Ω des suites de n chiffres (les chiffres sont pris dans $\{0,1,\ldots,9\}$). Combien vaut $|\Omega|$? Combien y-a-il de chiffres comportant un nombre pair de zéros ?

Annexe B

Indications des exercices

B.1 Solutions sur les compléments

Indication 1 Si la comparaison à des suites de référence et le critère $\frac{u_{n+1}}{u_n}$ ne permettent pas de conclure très vite, le passage au logarithme est souvent une bonne idée pour l'étude des formes produit.

Indication 2 Penser à la décomposition en éléments simples et aux formes télescopiques.

Indication 3 Noter que $((x_n) \to \ell) \iff ((x_{2n}) \to \ell)$ et $((x_{2n+1}) \to \ell)$.

Indication 4 On pourra penser à exprimer certaines sommes infinies comme des supremums de sommes finies.

Indication 5 On pourra utiliser le critère de comparaison séries/intégrales pour les fonctions positives.

Indication 7 Il faut se souvenir qu'une suite à valeurs dans $\mathbb{N}$ qui converge est constante à partir d'un certain rang.

Indication 8

1. Deviner la valeur de la limite supérieure, exhiber une sous-suite appropriée afin de minorer la limite supérieure, puis majorer a_n par une suite convergente.
2. Dans le cas particulier, on peut, comme précédemment, utiliser des majorations.
3. Dans le cas général, il faut revenir aux définitions.

Indication 9 On peut faire la remarque simple suivante, très utile dans les problèmes d'inf et de sup : pour a, b dans $\overline{\mathbb{R}}$

$$(a = b) \iff (\forall x \in \mathbb{R} \quad (x > a) \iff (x > b)).$$

Indication 10 On peut utiliser le résultat de la première question de l'exercice précédent.

Indication 11

1. À partir d'un certain rang, $a_n \ldots$et $b_n \ldots$donc $a_n + b_n \ldots$
2. On pourra comparer l'ensemble des valeurs d'adhérence de (b_n) et l'ensemble des valeurs d'adhérences de $(a_n + b_n)$.

Indication 12

1. On peut faire une comparaison avec une intégrale ou remarquer que $\log(n) - \log(n-1) \sim \frac{1}{n}$.
2. Dans la somme représentant $S_n - f(0)H_n$, traiter séparément les k tels que $k/n \leq \alpha$ et les autres.

3. Choisir g telle que $\frac{g(x)}{\sin x}$ se prolonge en une fonction continue sur $[0,1]$.

Indication 13 Effectuer une transformation d'Abel avec $R_n = \sum_{k=n+1}^{+\infty} \frac{u_k}{k^x}$.

Indication 14 1. Commencer par remarquer que $u_{kn+r} \leq u_{kn} + u_r$, puis procéder par récurrence sur n.

2. Commencer par montrer que pour tout k, $\overline{\lim\limits_{n\to+\infty}} \frac{u_n}{n} \leq \frac{u_k}{k}$.

3. $\inf\limits_{n\to+\infty} \frac{u_n}{n} \leq \underline{\lim\limits_{n\to+\infty}} \frac{u_n}{n} \leq \overline{\lim\limits_{n\to+\infty}} \frac{u_n}{n}$.

4. On pourra considérer $u_n = \log |||A^n|||$.

5. On pourra considérer $u_n = \log |A_n|$ et construire une injection de A_{n+p} dans $A_n \times A_p$.

Indication 15 Pour n assez grand, $(1-\varepsilon)a \leq (a_n)^n \leq (1+\varepsilon)a$.

B.2 Exercices sur la théorie de la mesure

Indication 16 Pour montrer que $\mathcal{A}$ est stable par réunion finie, on considèrera séparément le cas où tous les ensembles de la réunion sont finis et le cas où il en existe au moins un infini.

Indication 17 Pour montrer que deux tribus sont égales, on procède fréquemment par double inclusion. Pour montrer que $\mathcal{A} \subset \mathcal{B}$, on choisit souvent $\mathcal{C} \subset \mathcal{A}$ avec $\sigma(C) = \mathcal{A}$, et on montre que $\mathcal{C} \subset \mathcal{B}$ par des opérations ensemblistes. Des suites comme $(a-1/n)_{n\geq 1}$ ou $(b-1/n)_{n\geq 1}$ permettent classiquement d'écrire des fermés comme intersections dénombrable d'ouverts, ou des ouverts comme réunion dénombrables de fermés.

Indication 18 Vérifier les définitions des ensembles ainsi définis ; exploiter le fait qu'une indicatrice ne peut prendre que deux valeurs.

Indication 19 On pourra penser à utiliser le résultat de l'exercice corrigé 4 sur les sommes doubles de termes positifs.

Indication 20 1. Si la tribu $\mathcal{A}$ la tribu engendrée par $A_1, \dots, A_n$, on sait tout ce que la tribu $\mathcal{A}$ peut dire de ω dès qu'on sait auxquels Ω_i il appartient et dans lesquels il n'est pas.

2. (a) Pour montrer la stabilité par réunion dénombrable, il sera intéressant de distinguer deux cas.
   (b) À l'aide de la première question de l'exercice, on peut montrer que tout élément de $\mathcal{F}$ s'écrit sous la forme $A_K = \bigcup\limits_{k\in K} \Omega_k$, où K est une partie de J.
   (c) Pour $A \in \mathcal{G}$ tel que $x \in A$, on peut regarder $A \cap \Omega_{i_0}$.

3. On pourra procéder par l'absurde.

Indication 21 Il faudra traduire le fait d'être (ou de ne pas être) de Cauchy de manière ensembliste. Il est aussi utile de noter que si $(A_\varepsilon)_{\varepsilon>0}$ est une suite croissante d'ensembles, alors $\cap_{\varepsilon>0} A_\varepsilon = \cap_{n\geq 1} A_{1/n}$, ou que si $(B_\varepsilon)_{\varepsilon>0}$ est une suite décroissante d'ensembles, alors $\cup_{\varepsilon>0} B_\varepsilon = \cup_{n\geq 1} B_{1/n}$.

Indication 22
1. Considérer S' l'image de S par $s \mapsto \{s\}$.
2. On peut remarquer que $U = \underset{q' \in \mathbb{Q} \cap [0,1]}{\cup} (q' + S') \subset [0, 2[$.
3. Penser à la mesure de Lebesgue.

Indication 23
1. Contredire la σ-additivité ou le théorème de continuité séquentielle croissante.
2. On peut utiliser le résultat de l'exercice 19 (ou refaire la preuve dans ce cas particulier).

Indication 24 Vérifier les hypothèses du théorème de continuité séquentielle décroissante.

Indication 25 Penser à introduire $B \backslash A$.

Indication 26 Revoir la définition de la mesurabilité d'une application entre deux tribus. On pourra noter que si $F(x) = -x$, on a $h = h \circ F$ et $g = F \circ g \circ F$.

Indication 27 Remarquer les égalités successives suivantes :
$f^{-1}(A) = f^{-1}(A \cap \{0, 1\}) = f^{-1}(A \cap \{0\}) \cup f^{-1}(A \cap \{1\})$.

Indication 28 Poser $F' = f(\Omega)$ et revenir à la définition d'une tribu engendrée par une application.

Indication 29 Relire la définition d'une mesure, puis vérifier les propriétés.

Indication 30
1. Remarquer que $\{f < g\} = \cup_x \{f < x < g\}$ et utiliser la séparabilité de $\mathbb{R}$.
2. Que signifie "$A = B$" presque partout ?

Indication 31 Pour les deux premières questions, relire le cours, le résultat est $\frac{11}{6}$. Pour la deuxième question, penser aux sommes de Riemann.

Indication 32 Pour la deuxième question, on pourra d'abord observer que certaines des conditions des axiomes sont vérifiées sans hypothèse supplémentaire sur f.

Indication 33 On procèdera par double inclusion. On rappelle que si $\mathcal{C} \subset \mathcal{D}$, alors $\sigma(\mathcal{C}) \subset \sigma(\mathcal{D})$.

Indication 34 $\mathbb{Q}$ est dense dans $\mathbb{R}$, ainsi chaque réel est limite d'une suite croissante et d'une suite décroissante.

Indication 35
1. Exprimer C en fonction de A_2 et A_5.
2. Exprimer B en fonction des $(A_p)_{p \in \mathcal{P}}$.

Indication 36 Pour F fermé de E, on peut poser $d_F(x) = \inf\{d(x, y); y \in F\}$.

Indication 37
1. On rappelle que si $x \in \overline{A}$, tout ouvert contenant x contient un élément de A.
2. On pourra montrer que pour tout $y \in O$, il existe $x \in \mathbb{Q}^x d$ et $r \in \mathbb{Q}_*^+$ avec $\{x\} \subset B(y, r) \subset O$.
3. Considérer la réunion de tous les ouverts de mesure nulle.

B.3 Exercices sur le formalisme probabiliste

Indication 38 On peut noter que si $A, B \subset \Omega$, alors
$A = B \iff \mathbb{1}_A = \mathbb{1}_B \iff 5\forall x \in \Omega \quad \mathbb{1}_A(x) = 1 \iff \mathbb{1}_B(x) = 1.$

Indication 39 On peut commencer par montrer que si on remplace un seul élément de la famille par son complémentaire, on a encore une famille d'événements indépendants.

Indication 40
1. Revenir à la définition d'une limite infinie et utiliser la divergence de la série harmonique.
2. Appliquer le principe de partition.
3. Penser au théorème de continuité séquentielle décroissante.
4. Vérifier la définition de l'indépendance pour une famille d'événements.
5. Égaler deux expressions de $\mu_s(\{1\})$.
6. Utiliser la première question et l'identité qu'on vient de démontrer.
7. Noter que les $(p_i)_{i\geq n}$ sont des nombres premiers dépassant n, donc ne pouvant diviser n.
8. On montrera que $\mu(\{n\}) = 0$ pour tout $n \geq 1$, ce qui mène à une contradiction.

Indication 41
1. Dans les problèmes de modélisation, la première loi à laquelle il faut songer (quitte à l'écarter) est souvent une loi uniforme.
2. On peut penser au lemme des bergers et au principe de partition.
3. Calculer Φ^{-1}.
4. Combiner les deux précédentes questions.
5. Penser au critère spécial des séries alternées.

Indication 42
1. $A_d = d\mathbb{N} \cap \Omega_n$.
2. On déterminera un entier d tel que $\bigcap_{i=1}^{r} A_{d_i} = A_d$.
3. On remarquera que deux nombres sont premiers entre eux si et seulement si ils n'ont pas de diviseur premier commun.
4. Deux méthodes sont possibles : utiliser la formule du crible (formule de Poincaré) ou utiliser le résultat de l'exercice 40. Cette dernière méthode est utilisée dans le sujet du capes 2003.

Indication 43 On pourra poser $p_n = \mu(\{n\})$ et $r_n = \mu([n, +\infty[)$.

Indication 44 Remarquer qu'une union dénombrable d'événements est de probabilité nulle si et seulement si chacun est de probabilité nulle.

Indication 45
1. Appliquer la formule du crible. Pour la deuxième égalité, remarquer que $\Phi : B \mapsto \prod_{i\in B} p_i$ réalise une bijection de $\mathcal{P}_f(\mathbb{N}^*)$ dans A.
2. Bien sûr $\sum_{k\in A} \frac{1}{k^2} \leq \sum_{k=1}^{+\infty} \frac{1}{k^2} < +\infty$. On conclut à l'aide d'un raisonnement classique sur les séries à paramètre : on se fixe M tel que $\sum_{k\in A; k>M} \frac{1}{k^2} \leq \varepsilon$, puis on coupe la somme en deux. On a

$$\left|\mathbb{P}_n(A) - \sum_{k\in A} \frac{\mu(k)}{k^2}\right| \leq \varepsilon + \sum_{k\in A; k\leq M} \left|\frac{1}{n}\lfloor\frac{n}{k^2}\rfloor - \frac{1}{k^2}\right|,$$

d'où $\overline{\lim_{n\to+\infty}} |\mathbb{P}_n(A) - \sum_{k\in A} \frac{\mu(k)}{k^2}| \leq \varepsilon$. Pour un énoncé général, on pourra voir également le théorème 4.35.

3. Pour la première égalité, on pourra procéder comme dans l'exercice corrigé 41. Pour la deuxième, on utilisera une formule établie dans l'exercice 41. Enfin, la continuité de $s \mapsto \zeta(s)$ sur $[2, +\infty[$ nous donne la valeur de la densité de Dirichlet : $\frac{1}{\zeta(2)}$.
4. Si on pose $N_k(E) = |E \cap \{1, \ldots, k\}|$, comme
$$\mu_s(E) = \frac{1}{\zeta(s)} \sum_{k=1}^{+\infty} \frac{\mathbb{1}_E(k)}{k^s} = \frac{1}{\zeta(s)} \sum_{k=1}^{+\infty} \frac{N_k(E) - N_{k-1}(E)}{k^s},$$
on peut faire une transformation d'Abel et on obtient que pour tout ensemble E, on a $\mu_s(E) = \sum_{k=1}^{+\infty} \frac{N_k(E)}{\zeta(s)} \left(\frac{1}{k^s} - \frac{1}{(k+1)^s}\right)$. Ensuite on a
$$\mu_s(E) - \ell = \mu_s(E) - \ell\mu_s(\mathbb{N}^*) = \sum_{k=1}^{+\infty} \frac{N_k(E) - k\ell}{\zeta(s)} \left(\frac{1}{k^s} - \frac{1}{(k+1)^s}\right).$$
Alors, si M est choisi tel que $|N_k(E) - k\ell| \leq \varepsilon k = \varepsilon N_k(\mathbb{N}^*)$ pour $k \geq M$, on a
$$\begin{aligned} & \sum_{k=M}^{+\infty} \frac{|N_k(E) - k\ell|}{\zeta(s)} \left(\frac{1}{k^s} - \frac{1}{(k+1)^s}\right) \\ \leq\ & \varepsilon \sum_{k=M}^{+\infty} \frac{|N_k(\mathbb{N}^*)|}{\zeta(s)} \left(\frac{1}{k^s} - \frac{1}{(k+1)^s}\right) \\ \leq\ & \varepsilon\mu_s(\mathbb{N}^*) = \varepsilon. \end{aligned}$$
Pour conclure, il faut utiliser que $\lim_{s\to1^+} \zeta(s) = +\infty$.
5. Comme A a une densité naturelle, elle coïncide avec sa densité de Dirichlet et vaut
$$\sum_{k\in A} \frac{\mu(k)}{k^2} = \frac{1}{\zeta(2)} = \frac{6}{\pi^2},$$
ce qui nous donne une jolie identité [1].
6. Comme $\mathbb{P}_k(n\mathbb{N}^*) = \frac{\lfloor \frac{k}{n} \rfloor}{k}$, il est facile de voir que $\frac{1}{n}$ est la densité naturelle (et donc de Dirichlet) de $n\mathbb{N}^*$. Pour autant, cela n'apporte pas de contradiction au résultat final de l'exercice 41 car la densité naturelle et la densité de Dirichlet ne sont pas des mesures de probabilité.

Indication 46 On pourra conditionner par la valeur prise par l'*ensemble* des trois nombres tirés au sort.

Indication 47 Faire en sorte que le résultat soit nul.

1. Cette formule est un cas particulier de l'identité $\sum_{k\in\mathbb{N}^*} \frac{\mu(k)}{k^s} = \frac{1}{\zeta(s)}$, qui peut se déduire de la question 3 de l'exercice 41 et de la formule du crible. Elle admet également une preuve plus directe, que nous ébauchons : on forme le produit $\left(\sum_{k=1}^{+\infty} \frac{\mu(k)}{k^2}\right)\left(\sum_{\ell=1}^{+\infty} \frac{1}{\ell^2}\right)$. En regroupant les termes (k, ℓ) tels que $k\ell = n$, on obtient
$$\left(\sum_{k=1}^{+\infty} \frac{\mu(k)}{k^s}\right)\left(\sum_{\ell=1}^{+\infty} \frac{1}{\ell^s}\right) = \sum_{n=1}^{+\infty} \frac{1}{n^s} \left(\sum_{d|n} \mu(d)\right).$$
Or, un résultat d'arithmétique bien connu dit que $\sum_{d|n} \mu(d) = \delta_1(n)$; ainsi le produit des deux séries fait 1, ce qui est le résultat voulu. Pour plus de détails, voir par exemple le chapitre 4 de l'ouvrage de Bordellès [2].

B.4 Exercices sur les intégrales

Indication 48 On pourra noter que $\{\frac{1}{x}\} = \lim_{n\to+\infty} \{\frac{1}{x}\}\mathbb{1}_{]1/n,1]}(x)$.

Indication 49 On peut poser $F_n(\lambda) = \int_0^n e^{-\lambda t} f(t)\, dt$ et montrer, à l'aide d'une intégration par parties, que $(F_n(\lambda))$ vérifie un critère de Cauchy uniforme.

Indication 50
1. Pour tout t dans $\mathbb{R}$, on pourra montrer que l'événement $\{Y < t\}$ est mesurable.
2. Penser au théorème de transfert.
3. Noter que $Z - Y \geq 0$.

Indication 51
1. On peut montrer que pour $n \geq 1$ et $x \geq 0$, $\log(n+x) = \log(n) + \log(1 + x/n)$.
2. Minorer $(1+x)^n$.
3. Pour $n \geq 3$, $\frac{1+nx}{(1+x)^n} \leq \frac{3}{(1+x)^2}$.

Indication 52
1. Penser à Tonelli, et à x fixé, considérer l'intégrale entre 0 et $x/2$.
2. Appliquer Tonelli. Noter que le cas $p = -1$ se traite séparément.
3. On peut la comparer à la première intégrale. L'équivalence des normes en dimension finie peut guider les calculs.
4. Même remarque que pour la question précédente.

Indication 53
1. $\cos^2 = 1 - \sin^2$ et intégrer par parties.
2. Calculer $W_0 W_1$.

Indication 54
1. Appliquer le théorème de convergence dominée.
2. Faire (au moins) un changement de variable.

Indication 55 Faire un changement de variable.

Indication 56
1. Bien contrôler le reste.
2. Revoir les techniques usuelles du calcul des primitives.

Indication 57 Fubini est ton ami.

Indication 58
1. Première technique : on coupe l'intégrale en deux. La partie entre 0 et 1 ne pose pas de difficulté ; pour la deuxième, on peut la réécrire à l'aide d'une intégration par parties. Deuxième technique : écrire les intégrales comme limites des intégrales de 0 à n et établir un critère de Cauchy uniforme.
2. On pourra commencer par montrer que pour $\lambda \geq \delta > 0$, on a $\phi(\lambda) = \phi(\delta)\exp(\int_\delta^\lambda \frac{\alpha}{i-u}\, du)$.
3. Faire un changement de variable affine.
4. Utiliser un théorème de convergence dominée.
5. Faire un changement de variables.

Indication 59 Une intégration par parties permet de vérifier le critère de Cauchy. La convergence se montre classiquement à l'aide d'une intégration par parties. On pourra réécrire l'intégrale de $\frac{|\sin t|}{t}$ entre $n\pi$ et $(n+1)\pi$ comme une intégrale sur $[0,\pi]$ de manière à en déterminer un équivalent de la forme $\frac{K}{n}$ avec une intégrale à paramètre.

Indication 60 1. Première technique : on coupe l'intégrale en deux. La partie entre 0 et 1 ne pose pas de difficulté ; pour la deuxième, on peut la réécrire à l'aide d'une intégration par parties. Deuxième technique : écrire les intégrales comme limites des intégrales de 0 à n et établir un critère de Cauchy uniforme.

2. Commencer par montrer la dérivabilité sur $]a, +\infty[$, avec $a > 0$.
3. On pourra montrer que F a une limite nulle en l'infini.

Indication 61 Penser à la formule d'intégration des fonctions radiales ; utiliser les symétries.

Indication 62 1. Faire une intégration par parties.

2. Utiliser la parité, puis un changement de variables.
3. Développer le cosinus en série entière. Faire attention pour intervertir la somme et l'intégrale.
4. Prendre $X \sim \mathcal{N}(0, 1)$.

Indication 63 On pourra remarquer que $\frac{1}{e^x-1} = \frac{e^{-x}}{1-e^{-x}}$, puis penser à son développement en série.

Indication 64 1. On peut écrire $\int_{2n}^{+\infty} x^n e^{-x}\, dx = \int_n^{+\infty} (x+n)^n e^{-(n+x)}\, dx$.

2. Remarquer que $\Gamma(n+1) \sim \int_0^{2n} x^n e^{-x}\, dx$, puis faire un changement de variable affine.
3. On traitera séparément suivant le cas où $x > 0$ ou $x < 0$. Dans les deux, on pourra penser à un développement en série.
4. Utiliser le théorème de convergence dominée.

Indication 65 1. Remarquer que $(S \cap H^-)_z = S_z \cap H_z^-$. Regarder le dessin peut aussi aider.

2. Utiliser le théorème de Tonelli (on peut encore regarder le dessin !)
3. On peut utiliser des symétries du modèle pour réduire les calculs. Ensuite on appliquera Fubini.
4. Noter que S est la réunion disjointe de $S \cap H^+$ et $S \cap H^-$,

Indication 66 1. Pour $x \geq \delta$ et $t > 1$, on peut écrire

$$th(x) = h(x) + (t-1)h(x) \leq h(x) + (t-1)h(\delta).$$

2. Faire le changement de variable $x = u^\beta$.
3. Faire un changement de variable affine.
4. Penser au théorème de convergence dominée.

Indication 67 On peut appliquer la technique de la fonction test et le théorème de Tonelli.

Indication 68 On note $f_{a,b}$ la fonction affine par morceaux, valant 1 avant a, 0 après b. On peut remarquer que $\mathbb{1}_{]-\infty,a]} \leq f_{a,a+1/n} \leq \mathbb{1}_{]-\infty,a+1/n]}$.

Indication 69 Utiliser la transformation d'Abel.

Indication 70 On peut trouver une constante A telle que $g \leq A|f|$.

Indication 71 1. Vérifier les axiomes.

2. Commencer par le cas où f est étagée.

Indication 72 Utiliser le théorème de convergence monotone.

Indication 73 Utiliser le théorème de convergence dominée.

Indication 74 On pourra prendre $A = \{f > 0\}$.

Indication 75
1. Procéder par récurrence.
2. Faire un développement en série.
3. On peut développer avec la formule du binôme d'une part, et d'autre part remarquer que $\int_0^1 (1-x)^n \ln(x)\, dx = \int_0^1 x^n \ln(1-x)\, dx$. Une intégration par partie donne alors le résultat, à condition que la primitive soit choisie judicieusement.

Indication 76 Remarquer que $\log x + \frac{1}{x} \geq 0$ et développer $e^x - 1$ en série.

Indication 77 Développer en série et utiliser le théorème de convergence dominée.

Indication 78 On pourra remarquer que $\frac{1}{e^x-1} = \frac{e^{-x}}{1-e^{-x}}$.

Indication 79 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.

Indication 80 Une intégration en x pour commencer donnera

$$J = 4\sum_{k=1}^{+\infty} \frac{1}{(2k+1)^2}.$$

Pour intégrer en y, noter que $1 + 2xy + y^2 = (y+x)^2 + (1-x^2)$ et faire le changement de variable $u = \dfrac{y+x}{\sqrt{1-x^2}}$.

Indication 81 On pourra intégrer des fonctions par rapport à la mesure de comptage et utiliser le théorème de convergence dominée.

Indication 82 On pourra procéder par récurrence sur n, ou commencer par prolonger la fonction Γ.

Indication 83 Si on pose $\phi(x) = -\frac{1}{x^2}\log\frac{\sin x}{x}$ et

$$f_n(y) = \mathbb{1}_{[0,\pi\sqrt{n}]}(y)e^{-y^2\phi(y/\sqrt{n})}\, dy,$$

on a

$$\int_0^\pi (\frac{\sin x}{x})^n\, dx = \frac{1}{\sqrt{n}}\int_{\mathbb{R}_+} f_n(y)\, d\lambda(y).$$

On peut noter que cet exercice est aussi un cas particulier de la méthode de Laplace (voir exercice 73). Ici, il n'est pas nécessaire de procéder à un découpage, comme dans le cas général. L'intégrale de Wallis classique de l'exercice 58 peut, de la même manière, être traitée avec le changement de variables approprié et le théorème de convergence dominé.

Indication 84 Il faut d'abord développer $\frac{1}{1+e^t}$ sous forme d'une série alternée en prenant bien garde à contrôler le reste. Pour la deuxième égalité, on sépare classiquement les termes pairs des termes impairs.

Indication 85 Pour la première identité, on peut noter que

$$\int_1^n \frac{\{x\}}{x^\alpha}\, dx = \int_0^1 xS_{n,\alpha}(x)\, dx \text{ avec } S_{n,\alpha}(x) = \sum_{k=1}^{n-1}\frac{1}{(k+x)^\alpha}.$$

Pour la deuxième, on pourra démontrer la continuité de $\alpha \mapsto \int_1^{+\infty} \frac{\{x\}}{x^\alpha}\, dx$ au point $\alpha = 2$.

B.5 Exercices sur les lois

Indication 86 On pourra commencer par déterminer la loi de X_3.

Indication 87 Noter que $\mathbb{P}(a < X \leq b) = F(b) - F(a)$ et que les masses des points sont données par les sauts de discontinuité.

Indication 88 Pour montrer que les F_i sont des fonctions de répartition, il faut vérifier qu'elles sont continues à droites, croissantes et satisfont

$\lim_{n\to-\infty} F_i(n) = 0$ et $\lim_{n\to+\infty} F_i(n) = 1$.

Indication 89 Transformer la condition sur y en condition sur Φ.

Indication 90 Il s'agit de déterminer la constante de normalisation.

Indication 91 Penser aux valeurs effectivement prises par X.

Indication 92 Introduire le discriminant.

Indication 93 On pourra éventuellement introduire l'angle entre les droites (AB) et (BC).

Indication 94 Reconnaître une loi binomiale. Former le quotient des probabilités respectives de deux entiers consécutifs.

Indication 95 Pour la dernière question, il pourra être utile de se souvenir qu'une fonction de répartition est continue à droite.

Indication 96 Pour déterminer la loi de $X - Y$, partitionner suivant la valeur de X. Il peut être intéressant de noter que cette loi est symétrique.

Indication 97

1. On pourra par exemple calculer la densité du couple (X, Y).
2. Se ramener à un calcul d'aire.

Indication 98 Penser au théorème de Tonelli.

Indication 99 On rappelle que la fonction de répartition caractérise la loi des variables aléatoires réelles.

Indication 100

1. On pourra commencer par exhiber un compact de mesure strictement positive, puis raisonner par l'absurde.
2. Commencer par traiter le cas où la suite (X_n) prend ses valeurs dans un ensemble borné.

Indication 101 Si X n'est pas constante, il existe a tel que $\mathbb{P}(X \leq a) > 0$ et $\mathbb{P}(X > a) > 0$. Idem pour Y (avec b). De là, on peut en déduire que si, en plus, X et Y sont indépendantes, alors $X + Y$ n'est pas constante.

Indication 102

1. Il suffit de montrer que ces ensembles forment un π-système qui engendre la tribu.
2. Calculer $\mathbb{P}_{X\wedge X'}(n\mathbb{N}^*)$ et appliquer la question précédente.

Indication 103

1. Penser à discuter suivant les positions relatives de n et k.
2. $\mathbb{Q}$ est dense dans $\mathbb{R}$.

3. Utiliser le principe de partition.
4. Écrire $\mathbb{P}(D) = 1$, avec D bien choisi.

Indication 104 Calculer $\mathbb{P}(1 + \nu_{p_1}(X) > k_1, \dots, 1 + \nu_{p_n}(X) > k_n)$.

Indication 105 Prendre X et Y deux variables indépendantes suivant chacune une loi de Bernoulli de paramètre $1/2$ et poser $Z = |X - Y|$.
Solution : $\mathcal{C} = \{\{X = 0\}, \{Y = 0\}\}$ et $\mathcal{D} = \{Z = 0\}$

Indication 106 On propose de prendre $\Omega = \{0,1\} \times \{0,1\}$, $\mathcal{F} = \mathcal{P}(\omega)$, $\mathcal{C} = \{\{(0,0),(0,1)\}, \{(0,0),(1,0)\}\}$, ainsi que $\mathbb{P} = \frac{1}{2}(\delta_0 + \delta_1) \otimes \frac{1}{2}(\delta_0 + \delta_1)$, $\mathbb{P}' = \frac{1}{2}(\delta_{(0,0)} + \delta_{(1,1)})$.

Indication 107 Traduire les événements considérés en fonction de Y.

Indication 108 Poser $x = AM$ et résoudre l'inéquation.

Indication 109 1. Dire que le maximum de n nombres ne dépasse pas x, c'est dire que chacun ne dépasse pas x.
2. Remarquer que $1 - m_n = \max(1 - X_1, \dots, 1 - X_n)$.

Indication 110 À k fixé, il faut déterminer les valeurs de X qui sont telles que $Y = k$.

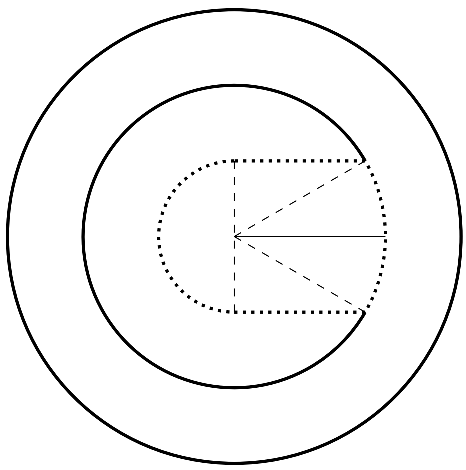
Indication 111 Représenter graphiquement les domaines considérés.

Indication 112 On pourra écrire $e^{-\frac{x^2}{2}} = \frac{1}{x}(xe^{-\frac{x^2}{2}})$ afin de procéder à une intégration par parties.

Indication 113 1. On pourra montrer que $\mathcal{Q}$ est la tribu de queue associée à la famille $(\mathcal{A}_n)_{n\geq 1}$.
2. 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, $A = \bigcap_{n \geq n_0} \bigcup_{k \geq n} A_k$.

Indication 114 Si on pose $\alpha = 2 \arcsin \frac{r}{1-r}$, on doit trouver par exemple

$$p = \frac{1 - \cos\alpha}{4} + \frac{\alpha + \sin\alpha}{2\pi}.$$



Indication 115 Commencer par calculer la fonction de répartition.

B.6 Exercices sur les esperances

Indication 116 On peut utiliser la méthode de la fonction test, ou calculer la fonction de répartition.

Indication 117

1. Faire apparaître un polynôme du second degré.
2. Considérer une borne des réels vérifiant une des deux conditions.

Indication 118 On pourra montrer que pour toute fonction H mesurable bornée, on a $\mathbb{E}[H(\{\frac{1}{X}\})] = \mathbb{E}[H(X)]$.

Indication 119 Appliquer le théorème de transfert.

Indication 120 Penser à utiliser la fonction de queue.

Indication 121

1. Par l'absurde, considérer le produit des nombres premiers congrus à -1 modulo 6.
2. Procéder par l'absurde.
3. Reprendre la définition d'un ensemble sans-somme.
4. Écrire $|B'| = \sum_{b\in\mathbb{F}_p} \mathbb{1}_{b\in B'}$.
5. Raisonner par contraposée.
6. Considérer $\pi_A^{-1}(B')$.

Indication 122 On pourra appliquer la formule du crible et revoir les résultats en annexe sur le nombre de solutions d'équations en entiers.

Indication 123

1. À l'aide de l'inégalité de Cauchy–Schwarz, on peut majorer $\sum\limits_{k\in\mathbb{Z}} \mathbb{P}(\lfloor X\rfloor - \lfloor Y\rfloor = \ell, \lfloor Y\rfloor = k)$ pour tout $\ell \in \mathbb{Z}$.
2. On sait déjà que $C \leq 5$. Les exemples proposés servent à minorer C.

Indication 124

1. On peut noter que $s_1^{Z^1} s_2^{Z^2} \dots s_k^{Z^k}$ est une fonction de X.
2. Réécrire le produit à l'aide des vecteurs $Z_1, \dots, Z_n$. On pourra utiliser la formule du multinôme.

Indication 125 Pour la première question, on peut se ramener au cas où les variables suivent des lois de Bernoulli.

Indication 126 On pourra remarquer que pour tout $k \leq n$, les variables $\mathbb{1}_{A_k} S_k$ et $S_n - S_k$ sont indépendantes.

Indication 127 On pourra établir que $\mathbb{E}|X - Y| = \mathbb{E}(f(Y))$.

Indication 128 Pour $\min(X, Y)$, penser à utiliser la fonction de queue.

Indication 129

1. Penser à utiliser la fonction de queue.
2. Faire apparaître une somme de Bernoulli.

Indication 130 Pour trouver la fonction de répartition, transformer les conditions sur Y en conditions sur X. Pour l'espérance, on pourra par exemple appliquer le théorème de transfert.

Indication 131

1. Utiliser un C^1-difféomorphisme.

2. On peut noter que X/Y n'est pas de carré intégrable.

Indication 132 On peut remarquer que $G_{k+1} - G_k = -b + (a+b+c)\mathbb{1}_{\{N>k\}}$.

Indication 133
1. Noter que $y \mapsto y^{1/p}$ est une transformation C^1.
2. On pourra faire apparaître la densité de $(X_1, \dots, X_n)$.
3. Bien choisir ϕ.
4. Symétriser l'expression.

Indication 134 Pour disqualifier l'indépendance, observer le support de la loi. Faire ensuite un changement de variables polaire.

Indication 135
1. Considérer l'application qui échange la coordonnée 1 et la coordonnée i.
2. On pourra remarquer que X_i est une variable de Bernoulli.
3. On pourra remarquer et justifier que si $i \neq j$, (X_1, X_2) a même loi que (X_i, X_j).
4. Remarquer que P_n s'annule en 0 et en n.
5. Penser à un tirage de boules indiscernables dans une urne.

Indication 136
1. Penser à la technique de la fonction test.
2. L'intégrale d'une densité fait toujours...
3. Pour la première égalité, faire un changement de variable. Penser ensuite au principe des zéros isolés.

Indication 137
1. Les partitions étant fixées, il faut juste déterminer les images de chaque classe.
2. Il suffit d'identifier les deux polynômes sur des entiers p suffisamment nombreux. On pourra penser à partitionner un ensemble.
3. Appliquer le théorème de transfert.
4. Même remarque, et penser à utiliser la question 2.
5. Prendre $x > 0$ pour pouvoir inverser les sommes à loisir.
6. Il faut appliquer la formule de Cauchy pour retrouver $\frac{B_n}{n!}$ à l'aide de l'intégrale le long d'un cercle dont on optimisera le rayon.

Indication 138
1. Noter que Z' et $-Z'$ ont même loi. Pour la dernière égalité, on pourra penser à l'expression de l'espérance d'une variable positive à l'aide de la fonction de queue.
2. Pour calculer la fonction de queue de Z', on peut penser au théorème de transfert. Pour le calcul de l'intégrale, il est naturel de développer la fonction à intégrer. L'identité $\frac{2t}{1+e^{-t}} = \frac{2t}{e^t-1} - \frac{4t}{e^{2t}-1}$ peut (éventuellement) simplifier les choses. On peut noter que cette question ressemble aux exercices 69 et 92.

Indication 139 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.

Indication 140
1. On a $M = \sum\limits_{i=1}^{+\infty} \mathbb{1}_{\{N_i \neq 0\}}$. Pour $n \in \mathbb{N}^*$, décomposer n en produit de facteurs premiers.
2. Calculer de deux manières différentes $\mathbb{E}(\phi(X))$.

Indication 141
1. Revoir les propriétés de convolution des lois Gamma (ou refaire la preuve !).
2. Un simple changement de variable et une intégration par parties.
3. Remarquer que $\{N_t = n\} = \{S_n < t \le S_{n+1}\}$.

Indication 142 Si X désigne le nombre de fois où l'on a obtenu le nombre choisi, le gain est $X - \mathbb{1}_{\{X=0\}}$.

Indication 143 Bien observer que X et Y sont indépendantes.

Indication 144 Une probabilité est l'espérance d'une indicatrice.

Indication 145 Interpréter le membre de gauche comme la valeur absolue d'une covariance.

Indication 146 Appliquer le théorème de transfert.

Indication 147
1. On montrera que $\mathbb{E}(X-a)^2 - \sigma^2 = (m-a)^2$.
2. Prendre $a = \frac{a_1+a_n}{2}$.

Indication 148 Effectuer une transformation d'Abel.

Indication 149 Appliquer le théorème de transfert, et penser à la forme canonique des polynômes du second degré.

Indication 150 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.

Indication 151 Fonction de répartition, ou transformation : on a le choix !

Indication 152 On peut raisonner en terme de loi.

Indication 153 S'inspirer de l'exercice précédent et utiliser une transformation.

Indication 154 Appliquer l'exercice précédent.

Indication 155 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 $|x-y| = \sqrt{S^2 - 4P}$.

Indication 156
1. On trouve la variance $b(0) + \sum_{i=1}^{n-1} 2b(i)(1 - i/n)$.
2. Remarquer qu'une variance est toujours positive. On peut alors par exemple appliquer le lemme de Fatou ou procéder de manière plus élémentaire.

Indication 157
1. On prendra $\Omega = \mathcal{B}_r(\{1,\dots,n\})$.
2. Utiliser les relations entre l'espérance et la queue de distribution, puis utiliser la relation de récurrence du triangle de Pascal.

Indication 158
1. On pourra remarquer qu'une rotation est une application linéaire isométrique.
2. Remarquer que $|X-Y| = \sqrt{2}U$.

Indication 159
1. On trouvera que les X_i sont des variables de Bernoulli indépendantes, avec $X_i \sim \mathrm{Ber}(p_i)$.
2. $\mathbb{P}(Y_i \neq X_i) = \mathbb{P}(Y_i \ge 2) + \mathbb{P}(Y_i = 0, Z_i = 1)$.
3. (a) Grâce aux propriétés de convolution des lois de Poisson, on trouvera que $Y \sim \mathcal{P}(p_1 + \dots + p_n)$.
   (b) Remarquer que $|\mathbb{1}_{\{X\in A\}} - \mathbb{1}_{\{Y\in A\}}| \le \mathbb{1}_{\{X\neq Y\}}$ et aussi que $\mathbb{P}(X \in A) - \mathbb{P}(Y \in A) = \mathbb{E}[\mathbb{1}_{\{X\in A\}} - \mathbb{1}_{\{Y\in A\}}]$.

Bibliographie

[1] P. Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995.

[2] O. Bordellès. *Thèmes d'arithmétique*. Ellipses, Paris, 2006.

[3] N. G. de Bruijn. *Asymptotic methods in analysis*. Dover Publications Inc., New York, third edition, 1981.

[4] M. Fekete. Über die Verteilung der Würzeln bei gewissen algebraischen Gleichungen mit ganzzähligen Koeffizienten. *Math. Z.*, 17(1) :228–249, 1923.

[5] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.

[6] O. Garet and A. Kurtzmann. *De l'intégration aux probabilités*. Ellipses, Paris, 2nde edition, 2019.

[7] Olivier Garet. *Probabilités et processus stochastiques*. auto-édité, distribué par Amazon, 2017.

[8] Jean-Pierre Kahane. *Séries de Fourier absolument convergentes*. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 50. Springer-Verlag, Berlin-New York, 1970.

[9] G. Pólya and G. Szegő. *Problems and theorems in analysis. I*. Classics in Mathematics. Springer-Verlag, Berlin, 1998.

[10] W. Rudin. *Analyse réelle et complexe*. Masson, Paris, 1980.