Olivier GARET – Aline KURTZMANN

De l'intégration aux probabilités

Deuxième édition augmentée

Avant-propos

« Pouvez-vous nous conseiller un bon livre en probabilités, qui reprenne bien les bases ? » Cette question, les auteurs de ce livre l'ont entendue de nombreuses fois ! Et pourtant, répondre à la question n'est pas facile, car en réalité, l'expérience montre que les difficultés sont bien plus souvent liées à une défaillance du lien entre le cours d'intégration et ses applications en probabilités, qu'à un défaut de compréhension des probabilités elles-mêmes.

Ainsi avons-nous fait le choix d'écrire un cours complet (intégration, probabilités, statistiques) qui favorise les allers-retours entre l'analyse et les probabilités : évidemment, les résultats d'analyse permettent de construire la théorie des probabilités, mais en retour, la théorie des probabilités fournit des preuves très élégantes, parfois éclairantes, de résultats d'analyse ou de théorie des nombres.

Savoir des mathématiques, c'est d'abord savoir faire des mathématiques. Les étudiants le sentent bien, qui nous réclament bien souvent des exercices corrigés. Ainsi, une large place a-t-elle été consacrée aux exercices : à la fin de chaque chapitre, on trouvera deux séries d'exercices. Pour la première série, les solutions détaillées sont données en fin d'ouvrage. Pour les deux séries, on trouvera, également à la fin, des pistes ou des indications de solution. Outre cela, pour certains chapitres, un ou plusieurs exercices corrigés particulièrement importants sont détaillés dans le chapitre. Les choix des exercices sont très variés, des grands classiques aux créations plus originales, afin de contribuer à la fois à l'acquisition des méthodes et à la progression de la culture mathématique du lecteur.

Tout en gardant la trame d'un cours de Licence classique, nous avons voulu écrire un livre contenant des notions avancées, qui sont habituellement effleurées faute de temps. Disons-le sans détour : le contenu de l'enseignement classique de L3 est trop riche pour être assimilé dans le temps qu'il faut pour valider une quinzaine d'ECTS. Très souvent, nous avons rencontré des étudiants de M1 ou préparant l'agrégation mis en difficulté lorsqu'ils doivent utiliser seuls des notions de base du programme de L3 qu'ils ont insuffisamment pratiquées, ou dont la nature véritable a été masquée par l'emploi de théorèmes sophistiqués. On trouvera dans ce livre peu d'outils théoriques hors du programme de L3, qui reste la référence (pas d'espérance conditionnelle, pas de chaînes de Markov). Mais tout en restant fidèle à ce corpus, on va souvent un peu plus loin, proposant parfois à côté des théorèmes classiques des extensions plus originales, que le lecteur – et cela lui est clairement signalé – ne doit pas mémoriser en première intention, mais qui lui permettent de voir les notions de base mises en action. Par ailleurs, l'expérience nous a montré qu'un certain nombre de résultats classiques de probabilités sont tombés dans un trou noir pédagogique : plus enseignés en Licence (faute de temps), pas enseignés en M1 (car trop éloignés du corpus traditionnel de M1), ils refont leur apparition en M2 ou en préparation à l'agrégation, où l'enseignant s'étonne de ces lacunes ... sans avoir toujours le temps de les combler. C'est par exemple le cas des notions de tension ou d'équi-intégrabilité. Aussi ces résultats trouvent-ils naturellement refuge dans cet ouvrage.

Remerciements. Les auteurs tiennent à remercier leurs relecteurs : Gilles Auriol, Aurélien Deya, Jean-Sébastien Giet, Jean-Baptiste Gouéré, Samuel Herrmann, Dominique Lépingle, Emmanuel Lesigne, Régine Marchand, Joseph Ngatchou-Wandji, Ivan Nourdin, Hervé Queffelec, Karim Ramdani, Bernard Roynette, Charles Suquet et Samy Tindel, qui ont fait un travail formidable. Un immense merci à Bernard Roynette qui s'est acquitté de la lourde tâche de relire la totalité du manuscrit avec une incroyable rapidité et nous a fait part de ses nombreuses remarques constructives. Merci également à Monique Pontier qui a su nous donner très vite une référence utile, là où plus d'une personne consultée avait séché...

Enfin, il est impossible au premier auteur de cet ouvrage de terminer l'écriture d'un livre sur les bases de l'intégration et des probabilités sans avoir une pensée affectueuse pour Raymond Moché qui lui en a transmis le goût par ses cours, tout à la fois limpides et exigeants.

Sur la deuxième édition. La deuxième édition a bénéficié des retours de nombreux lecteurs, nos étudiants bien sûr, mais également les forumeurs du site `http://www.les-mathematiques.net/`. Nous remercions particulièrement Éloi Mehr et Nicolas Schaeffer, qui ont non seulement débusqué un nombre impressionnant de coquilles, mais nous ont aussi fait des suggestions d'amélioration tout à fait pertinentes. Ainsi quelques théorèmes admis dans la première édition ont-ils pu être prouvés sans faire grossir le corpus central. Dans la partie analyse, les principales nouveautés concernent le calcul de volumes d'une part, la fonction Gamma et les fonctions spéciales associées d'autre part. Pour les probabilités, on a accordé une plus grande importance aux simulations et à la méthode de Monte-Carlo dans le cours, ainsi qu'à la distribution Zêta dans les exercices, ce qui donne lieu à des applications à la théorie des nombres, qui, nous l'espérons, enthousiasmeront le lecteur. De manière générale, l'essentiel de l'accroissement du volume provient des exercices, car progresser en mathématiques, c'est d'abord les pratiquer. Ainsi, on trouvera ici plus de 300 exercices dont plus de 150 sont corrigés.

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 Grand O, petit o : des amis fidèles

Nous commençons ce livre par plusieurs rappels sur des notions relatives à la comparaison des grandeurs que le lecteur a sans doute déjà rencontrées. Mais notre expérience d'enseignants nous a appris que comme toutes les notions et notations puissantes, elles demandent un certain temps pour être appréhendées. Et que celui pour lequel elles ne restent qu'un vague bruit de fond finit par être bien malheureux.

1.1.1 La notation grand O

La notation O est utilisée pour majorer l'ordre de grandeur d'une quantité. Si S est un ensemble quelconque, f et g des fonctions définies sur S réelles (ou complexes ou vectorielles), on dira que

$$f(x) \quad = \quad O(g(x)) \quad x \in S \tag{1.1}$$

s'il existe un réel $A > 0$ tel que $\forall x \in S$, $|f(x)| \leq A|g(x)|$. Ainsi, cette notation exprime qu'il n'y a pas nécessité, pour les calculs qui sont à effectuer, de garder trace de la valeur précise de A : seule son existence compte. Ce premier usage de la notation O n'est pas si fréquent dans la littérature : nous l'empruntons à De Bruijn [6] car il nous semble très éclairant pour appréhender le second usage de la notation O, beaucoup plus courant.

Si S est une partie de $\mathbb{R}$ telle que $\sup\limits_{x \in S} x = +\infty$, on dit que

$$f(x) \quad = \quad O(g(x)) \quad \text{lorsque } x \to +\infty \tag{1.2}$$

s'il existe un réel a tel que

$$f(x) \quad = \quad O(g(x)) \quad \forall x \in S \cap]a, +\infty[. \tag{1.3}$$

Autrement dit, l'assertion (1.2) est équivalente à

$$\exists (a, A) \in \mathbb{R}^2, \quad \forall x \in S, \quad x > a \Longrightarrow |f(x)| \leq A|g(x)|.$$

Moralement, les conditions (1.1) et (1.2) sont "presque équivalentes". Évidemment (1.1) entraîne (1.2), mais quand on a (1.2), on n'est pas très loin d'avoir (1.1). En particulier, elles sont équivalentes si $S = \mathbb{N}$ et si g ne s'annule pas sur $\mathbb{N}$, ou si $S = [0, +\infty[$, f et g sont continues sur S et si g ne s'annule pas sur S. En effet, dans le premier cas $\frac{f}{g}$ est bornée sur l'ensemble fini $\mathbb{N} \cap [0, a]$, et dans le second cas $\frac{f}{g}$ est continue sur le compact $[0, a]$, donc est bornée sur ce compact.

Si f et g sont définies sur $S =]0, c]$ (respectivement $S =]-c, c[\setminus\{0\}$), on dit également que

$$f(x) \quad = \quad O(g(x)) \quad x \to 0 \tag{1.4}$$

s'il existe un réel a tel que

$$f(x) = O(g(x)) \quad x \in S\cap]0, a[\quad (\text{respectivement } x \in S\cap]-a, a[). \tag{1.5}$$

On remarque que si on a $f = O(g)$ au voisinage d'un point ou de l'infini, et si g est de limite nulle en ce point, alors f est aussi de limite nulle.

1.1.2 La notation petit o

La notation o exprime une relation de négligeabilité : on dit que

$$f(x) \quad = \quad o(g(x)) \quad \text{lorsque } x \to +\infty \tag{1.6}$$

si, pour tout $\varepsilon > 0$, il existe un réel a tel que

$$\forall x \in S\cap]a, +\infty[, \quad |f(x)| \quad \leq \quad \varepsilon |g(x)|. \tag{1.7}$$

La notation o est plus connue que la notation O. Paradoxalement, elle est aussi moins efficace pour la raison suivante : o et O sont souvent utilisés dans les développements limités en l'origine de fonctions classiques. Les fonctions classiques étant de classe C^∞, on obtient plus (ou autant) d'information avec autant de termes non-nuls dans le développement pour les O que pour les o.
Exemple: Comparer, en 0, $\sin x = x + o(x^2)$ avec $\sin x = x + O(x^3)$. Supposons que l'on veuille montrer que $\lim\limits_{x \to 0} \frac{x - \sin x}{x^{2,5}} = 0$. De la première écriture, on tire $x - \sin x = o(x^2)$, d'où $\frac{x - \sin x}{x^{2,5}} = o(1/\sqrt{x})$, ce qui ne permet pas de conclure : il est sans intérêt ici d'être négligeable devant une quantité de limite infinie. En revanche de la deuxième écriture, on tire $x - \sin x = O(x^3)$, d'où l'on déduit que $\frac{x - \sin x}{x^{2,5}} = O(\sqrt{x})$, ce qui permet de conclure puisque x a évidemment une limite nulle en 0. Pour obtenir le même résultat avec des petits o, il faut aller un cran plus loin : $\sin x = x - \frac{x^3}{6} + o(x^3)$.

1.1.3 Équivalence de deux fonctions, de deux suites

On dit que les fonctions (ou les suites) f et g sont équivalentes en truc (truc étant un point ou l'infini) si $f(x) - g(x) = o(f(x))$ quand x tend vers truc. On écrit alors $f(x) \underset{x \to \text{truc}}{\sim} g(x)$. Il n'est pas très difficile de démontrer que cette relation est bien une relation d'équivalence.

Cependant, l'usage direct de la définition est rarement le bon moyen de travailler avec les équivalents. Il est souvent plus utile de remarquer que si f et g ne s'annulent pas, f et g sont équivalents en truc si et seulement si $f(x)/g(x) \to 1$ quand x tend vers truc.

On remarque que grand O, petit o, et équivalents sont tous trois des affaires de quotient : quotient qui reste borné, qui tend vers 0, ou qui tend vers 1, ce qui explique que ces trois relations se comportent très bien avec le produit.
Exemple: Étude de la suite $u_{n+1} = \sin u_n$, avec $u_0 \in]0, 1]$.
Comme $0 < \sin x < x$ pour $x \in]0, 1]$, la suite est à valeurs dans $]0, 1]$ et décroissante, minorée, donc convergente. La limite est 0 car la limite est un point fixe de la fonction continue $x \mapsto \sin x$ sur $[0, 1]$, et 0 est bien la seule solution.
Étudions $\frac{1}{u_{n+1}^2} - \frac{1}{u_n^2} = g(u_n)$, avec $g(x) = \frac{1}{\sin^2 x} - \frac{1}{x^2} = \frac{1}{x^2}\left(\frac{x^2}{\sin^2 x} - 1\right)$.

Au voisinage de 0, $\sin x = x - \frac{x^3}{6} + O(x^5)$, donc $\frac{\sin x}{x} = 1 - \frac{x^2}{6} + O(x^4)$ [1], d'où $\frac{x}{\sin x} = 1 + \frac{x^2}{6} + O(x^4)$, puis $\frac{x^2}{\sin^2 x} = 1 + \frac{x^2}{3} + O(x^4)$, soit $g(x) = \frac{1}{3} + O(x^2)$. Comme (u_n) est de limite nulle, $g(u_n)$ tend vers $1/3$, donc par le théorème de Cesàro

$$\frac{1}{n}\left(\frac{1}{u_n^2} - \frac{1}{u_0^2}\right) = \frac{1}{n}\sum_{k=0}^{n-1}\left(\frac{1}{u_{k+1}^2} - \frac{1}{u_k^2}\right) = \frac{1}{n}\sum_{k=0}^{n-1} g(u_k)$$

tend vers $1/3$ quand n tend vers l'infini. Ainsi, $\frac{1}{nu_n^2} \sim \frac{1}{3}$ soit $u_n \sim \sqrt{\frac{3}{n}}$. L'usage du O permet de préciser l'équivalent, sans qu'il soit nécessaire de pousser plus loin le développement en série du sinus. En réinjectant dans la formule asymptotique l'équivalent trouvé, on a $g(u_k) = \frac{1}{3} + O(u_k^2) = \frac{1}{3} + O(1/k)$, d'où

$$\frac{1}{u_n^2} - \frac{1}{u_0^2} = \sum_{k=0}^{n-1}\left(\frac{1}{3} + O(1/k)\right) = \frac{n}{3} + O\left(\sum_{k=0}^{n-1}\frac{1}{k}\right).$$

1. C'est le 2, indice du second coefficient non-nul dans le développement de $\frac{\sin x}{x}$, qui explique le choix "miraculeux" de la suite auxiliaire $\frac{1}{u_{n+1}^2} - \frac{1}{u_n^2}$. À titre d'exercice, le lecteur pourra étudier la récurrence $u_{n+1} = \log(1 + u_n)$.

Il est classique que $\sum_{k=0}^{n-1} \frac{1}{k} \underset{n\to+\infty}{\sim} \log n$. On en déduit :

$$\frac{1}{u_n^2} = \frac{n}{3} + O(\log n) = \frac{n}{3}\left(1 + O\left(\frac{\log n}{n}\right)\right), \text{ puis } u_n = \sqrt{\frac{3}{n}}\left(1 + O\left(\frac{\log n}{n}\right)\right).$$

1.2 Convergence de séries et d'intégrales

Soit $(a_n)_{n\in\mathbb{N}}$ une suite réelle ou complexe. On lui associe la suite (S_n) définie par $S_0 := a_0$, $S_1 := a_0 + a_1$, $S_n := a_0 + \cdots + a_n$.

Le terme S_n est la somme partielle d'ordre n de la série de terme général a_n. On dit que la série de terme général a_n converge si la suite (S_n) a une limite S. Dans ce cas, on définit le reste d'ordre n de la série par

$$R_n := S - S_n = \sum_{k=n+1}^{+\infty} a_k.$$

Proposition 1.1. *Soit a_n le terme général d'une série convergente. Alors :*

$$\lim_{n\to+\infty} a_n = 0 \ (\textit{et} \ \lim_{n\to+\infty} R_n = 0).$$

Remarque 1.2. *La réciproque est fausse. En effet, si on considère par exemple $a_n := 1/n$ pour $n \geq 1$, on a $\lim_{n\to+\infty} \frac{1}{n} = 0$ et pourtant $\sum_{n=1}^{+\infty} a_n = \infty$.*

Exemple: La série géométrique. Soit $q \in \mathbb{C}; q \neq 1$. Si l'on considère la série de terme général $a_n = q^n$, on a $S_n = \frac{1-q^{n+1}}{1-q}$. Si $|q| < 1$, alors S_n tend vers $\frac{1}{1-q}$ lorsque n tend vers l'infini, et la série converge. Sinon le terme général ne tend pas vers 0 (car $|q|^n \geq 1$) et donc la série diverge.

1.2.1 Séries à termes positifs

Définition. *La série de terme général a_n est une série à termes positifs si pour tout $n \geq 0$, on a $a_n \geq 0$.*

Proposition 1.3. *Si la série de terme général a_n est une série à termes positifs, alors elle converge si et seulement si la suite des sommes partielles est majorée.*

Démonstration. Une suite croissante majorée converge ; une suite croissante non majorée tend vers $+\infty$. □

Théorème 1.4 (de comparaison)**.** *Soient a_n et b_n les termes généraux de deux séries à termes positifs, tels que $a_n \leq b_n$ pour tout n. On a :*

1. *si la série de terme général b_n converge, alors celle de terme général a_n converge ;*
2. *si la série de terme général a_n diverge, alors celle de terme général b_n diverge.*

Ce théorème est en fait un corollaire de la proposition 1.3.
On dit que a_n est équivalent à b_n lorsque $n \to +\infty$, ce que l'on note $a_n \sim b_n$, si $\lim\limits_{n\to+\infty} \frac{a_n}{b_n} = 1$ (où il existe N tel que $b_n \neq 0$ pour $n > N$).

Proposition 1.5. *Soient a_n et b_n les termes généraux de deux séries à termes positifs telles que, à l'infini, $a_n \sim b_n$. Alors la série de terme général a_n est de même nature que la série de terme général b_n. Si elles sont convergentes, on a l'équivalence des restes $\sum\limits_{k=n+1}^{+\infty} a_k \sim \sum\limits_{k=n+1}^{+\infty} b_k$ tandis que si elles sont divergentes, on a l'équivalence des sommes partielles $\sum\limits_{k=1}^{n} a_k \sim \sum\limits_{k=1}^{n} b_k$.*

Démonstration. Il existe n_0 tel que pour $n \geq n_0$, on a $\frac{2}{3}a_n \leq b_n \leq \frac{3}{2}a_n$, ce qui entraîne que les séries sont de même nature. Soit $\varepsilon > 0$. Il existe N tel que $(1-\varepsilon)a_n \leq b_n \leq (1+\varepsilon)a_n$ pour $n \geq N$. Si les séries sont convergentes, on a pour $n \geq N$ $(1-\varepsilon) \sum\limits_{k=n+1}^{+\infty} a_k \leq \sum\limits_{k=n+1}^{+\infty} b_k \leq (1+\varepsilon) \sum\limits_{k=n+1}^{+\infty} a_k$, ce qui donne l'équivalence des restes. Si les séries sont divergentes, on a l'équivalence $\dfrac{\sum\limits_{k=1}^{n} b_k}{\sum\limits_{k=1}^{n} a_k} \sim \dfrac{\sum\limits_{k=N}^{n} b_k}{\sum\limits_{k=N}^{n} a_k}$. Or $(1-\varepsilon) \leq \dfrac{\sum\limits_{k=N}^{n} b_k}{\sum\limits_{k=N}^{n} a_k} \leq 1+\varepsilon$ pour $n \geq N$, donc pour n assez grand $1-2\varepsilon \leq \dfrac{\sum\limits_{k=1}^{n} b_k}{\sum\limits_{k=1}^{n} a_k} \leq 1+2\varepsilon$, ce qui est le résultat voulu. □

Exemple: La série de Riemann. Soit $\alpha > 0$. La série de terme général $a_n = n^{-\alpha} - (n+1)^{-\alpha}$ converge vers 1, car

$$S_n = \sum_{k=1}^{n} \left(\frac{1}{k^\alpha} - \frac{1}{(k+1)^\alpha} \right) = 1 - \frac{1}{(n+1)^\alpha}.$$

Or $a_n = n^{-\alpha}(1 - (1+1/n)^{-\alpha}) \sim \alpha n^{-(\alpha+1)}$, donc pour tout $\alpha > 0$, la série de terme général $\alpha n^{-(\alpha+1)}$ converge, et donc la série de terme général $n^{-(\alpha+1)}$ converge aussi. De même, la série de terme général $a_n = \log(n+1) - \log(n)$ diverge, car $S_n = \log(n+1)$ tend vers $+\infty$ (lorsque $n \to +\infty$). Mais

$$\log(n+1) - \log(n) = \log(n(1+1/n)) - \log n = \log(1+1/n) \sim \frac{1}{n}$$

ce qui implique que la série de terme général $1/n$ diverge. Pour $\alpha < 1$, on a $\frac{1}{n^\alpha} \geq \frac{1}{n}$, donc la série de terme général $\frac{1}{n^\alpha}$ diverge. Finalement, la série de terme général $\frac{1}{n^\alpha}$ pour $n \geq 1$ (dite "série de Riemann") converge si et seulement si $\alpha > 1$.

1.2.2 Convergences et divergences triviales

Proposition 1.6 (Règle de Cauchy). *Soit $(a_n)_{n\in\mathbb{N}}$ une suite de réels positifs.*

1. *S'il existe $0 < q < 1$ et $N > 0$ tels que $0 \leq a_n^{1/n} < q$ pour $n \geq N$, alors la série de terme général a_n converge.*
2. *Si $a_n^{1/n} \geq 1$ pour une infinité de termes, alors la série de terme général a_n diverge.*

Démonstration. 1) Comparaison avec une série géométrique.
2) Le terme général ne tend pas vers 0. □

Corollaire 1.7. *Soit a_n le terme général d'une série à termes positifs telle que $\lim\limits_{n\to+\infty} a_n^{1/n} = \ell$. Si $\ell < 1$, alors la série converge et elle diverge si $\ell > 1$.*

Proposition 1.8 (Règle de d'Alembert). *Soit a_n le terme général d'une série à termes strictement positifs. Alors :*

1. *s'il existe $0 < q < 1$ tel que $\frac{a_{n+1}}{a_n} < q$ pour $n \geq N$, alors la série de terme général a_n converge.*
2. *s'il existe $N \geq 1$ tel que $\frac{a_{n+1}}{a_n} > 1$ pour $n \geq N$, alors la série de terme général a_n diverge.*

Démonstration. Dans le premier cas, on a $a_n \leq a_N q^{n-N}$ pour $n \geq N$. Dans le second cas, on a $a_n \geq a_N$ pour $n \geq N$, et le terme général ne tend pas vers 0 car on a une suite strictement croissante. □

Corollaire 1.9. *Soit a_n le terme général d'une série à termes strictement positifs telle que $\lim\limits_{n\to+\infty} \frac{a_{n+1}}{a_n} = \ell$. Si $0 < \ell < 1$, alors la série converge et elle diverge si $\ell > 1$.*

Il faut retenir que les règles de Cauchy et de d'Alembert sont des règles qui permettent de comparer le terme général de la série avec des suites géométriques.

1.2.3 Critère de Cauchy

Soit (a_n) une suite réelle ou complexe.
La série de terme général a_n converge si et seulement si

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} : \quad N < n < p,\ \left| \sum_{k=n+1}^{p} a_k \right| < \varepsilon.$$

Remarquons que ce résultat est vrai car $\mathbb{R}$ est complet.

1.2.4 Séries absolument convergentes

Proposition 1.10. *Si la série de terme général a_n est absolument convergente (c'est-à-dire la série de terme général $|a_n|$ converge), alors elle est convergente.*

Démonstration. C'est une conséquence immédiate du critère de Cauchy. □

Si on veut montrer qu'une série est convergente, la première question à se poser est : est-elle absolument convergente ? Ce n'est que si la réponse est négative qu'il faut se tourner vers des outils plus complexes.

1.2.5 Outils pour les séries semi-convergentes

Définition. *On dit qu'une série est semi-convergente si elle est convergente, mais pas absolument convergente.*

On dit que la série de terme général a_n est une série à termes alternés si pour tout n, $a_n a_{n+1} \leq 0$.

Proposition 1.11 (Critère spécial des séries alternées)**.** *Soit a_n le terme général d'une série alternée telle que $(|a_n|)_n$ décroît et tend vers 0. Alors la série de terme général a_n converge et de plus le reste $R_N = \sum_{k=N+1}^{+\infty} a_k$ a le même signe que a_{N+1} et vérifie $|R_N| \leq |a_{N+1}|$.*

Démonstration. On suppose que le terme général s'écrit $a_n = (-1)^n b_n$, avec $b_n \geq 0$. On a alors

$$S_{2n+2} - S_{2n} = (-1)^{2n+2} b_{2n+2} + (-1)^{2n+1} b_{2n+1} = b_{2n+2} - b_{2n+1} \leq 0$$

et $S_{2n} \geq S_{2n-1} = \sum_{k=0}^{n-1} (b_{2k} - b_{2k+1}) \geq 0$. Ainsi (S_{2n}) est une suite décroissante, minorée par 0, donc converge vers une limite ℓ, avec $0 \leq \ell \leq S_0 = b_0$. Comme b_n tend vers 0, $S_{2n+1} = S_{2n} - b_{2n+1}$ converge aussi vers ℓ. Finalement (S_n) converge vers ℓ qui a le même signe que le premier terme de la série et est plus petit en valeur absolue. En remplaçant la suite $(a_n)_{n\geq 0}$ par $(a_{N+n})_{n\geq 1}$, on obtient le résultat voulu pour le reste. □

Proposition 1.12 (Critère de Dirichlet)**.** *Soient (a_n) une suite à valeurs réelles ou complexes telle que les sommes partielles de la série de terme général a_n sont bornées et (f_n) une suite décroissante de réels positifs de limite nulle. Alors la série de terme général $a_n f_n$ converge.*

Démonstration. Posons $S_n = \sum\limits_{k=0}^{n} a_k$ pour $n \geq 0$. Pour $n \geq 1$, on a

$$\begin{aligned}\sum_{k=n}^{p} a_k f_k &= \sum_{k=n}^{p} (S_k - S_{k-1}) f_k = \sum_{k=n}^{p} S_k f_k - \sum_{k=n}^{p} S_{k-1} f_k \\ &= \sum_{k=n}^{p} S_k f_k - \sum_{k=n-1}^{p-1} S_k f_{k+1} \\ &= -S_{n-1} f_n + S_p f_p + \sum_{k=n}^{p-1} S_k (f_k - f_{k+1}).\end{aligned}$$

Cette écriture se nomme une transformation d'Abel. Il s'agit tout simplement d'une intégration par parties discrète.

Maintenant en posant $M = \sup\{|S_k|; k \in \mathbb{N}\}$, on a

$$\begin{aligned}\left|\sum_{k=n}^{p} a_k f_k\right| &\leq Mf_n + Mf_n + \sum_{k=n}^{p-1} M|f_k - f_{k+1}| \\ &= 2Mf_n + \sum_{k=n}^{p-1} M(f_k - f_{k+1}) \leq 3Mf_n.\end{aligned}$$

□

Le critère spécial des séries alternées est un cas particulier du critère de Dirichlet (on écrit $a_n = (-1)^n |a_n|$, ou $a_n = (-1)^{n+1}|a_n|$).

Il est très important de savoir effectuer une transformation d'Abel. On verra en exercice 13 une variante où la transformation est effectuée sur les restes, plutôt que sur les sommes partielles [2].

1.2.6 Bref rappel sur l'intégrale de Riemann

Le présent paragraphe ne vise pas à refaire toute la théorie de l'intégration de Riemann, mais juste à préciser les statuts relatifs de l'intégrale de Riemann et de ses extensions "impropres".

2. À l'inverse du critère spécial des séries alternées, le critère de Dirichlet n'est ni au programme des concours des grandes écoles, ni à celui des concours de recrutement des professeurs de l'enseignement secondaire. Il est donc très important de savoir refaire la preuve, d'autant plus que la terminologie n'est pas bien fixée.

Pour les usages courants, l'intégrale dite de Riemann étudie essentiellement les fonctions continues (ou continues par morceaux) sur un intervalle compact. Ainsi, on sait presque tout sur l'intégrale de Riemann si l'on retient que à tout intervalle compact $[a,b]$, à toute fonction continue par morceaux sur $[a,b]$, on peut associer un nombre que l'on note $\int_a^b f(x)dx$, de telle manière que l'application $(a,b,f) \mapsto \int_a^b f(x)dx$ vérifie :

— $f \mapsto \int_a^b f(x)dx$ est linéaire
— $f \geq 0$ sur $[a,b]$ entraîne $\int_a^b f(x)dx \geq 0$.
— $f \mapsto \int_a^b 1\ dx = b-a$
— $\int_a^c f(x)\ dx = \int_a^b f(x)\ dx + \int_b^c f(x)\ dx$

Ces propriétés suffisent à retrouver la convergence des sommes de Riemann des fonctions continues vers l'intégrale. En effet, si l'on a

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

et si l'on pose $S_n = \sum_{k=1}^{n} (a_k - a_{k-1}) f(x_{k-1})$, on obtient alors

$$S_n - \int_a^b f(x)\ dx = \sum_{k=1}^{n} \int_{a_{k-1}}^{a_k} \left(f(x_{k-1}) - f(x)\right)\ dx,$$

d'où

$$\begin{aligned} |S_n - \int_a^b f(x)\ dx| &\leq \sum_{k=1}^{n} \int_{a_{k-1}}^{a_k} |f(x_{k-1}) - f(x)|\ dx \\ &\leq \sum_{k=1}^{n} \int_{a_{k-1}}^{a_k} D_n\ dx = (b-a)D_n, \end{aligned}$$

avec

$$D_n = \omega_f(e_n) \text{ où } \omega_f(\delta) = \sup_{\substack{x,y\in[a,b] \\ |x-y|\leq\delta}} |f(x)-f(y)| \text{ et } e_n = \max_{1\leq i\leq n} (a_i - a_{i-1}).$$

Comme f est continue sur le compact $[a,b]$, elle y est uniformément continue, donc $\omega_f(\delta)$ a une limite nulle lorsque $\delta \to 0$. Ainsi, on obtient la convergence de S_n vers $\int_a^b f(x)\ dx$ dès que D_n est de limite nulle. En particulier, on a le résultat classique

$$\lim_{n\to+\infty} \frac{b-a}{n} \sum_{k=0}^{n-1} f\left(a + (b-a)\frac{k}{n}\right) = \int_a^b f(t)\ dt$$

pour toute fonction continue sur le compact $[a,b]$.

Malheureusement, l'intégrale de Riemann n'est pas équipée pour traiter de l'intégrale des fonctions continues sur un intervalle ouvert, bornée ou non. On fabrique alors une rustine, appelée “intégrale impropre” : si f est une fonction continue sur $[a, b[$ (avec éventuellement $b = +\infty$) telle que la fonction $x \mapsto \int_a^x f(t)\, dt$ admette une limite L quand x tend vers b, on dit alors que L est l'intégrale de f entre a et b et on note $\int_a^b f(t)\, dt = L$. Comme pour les séries, il est fréquent que l'on ne sache pas déterminer la limite. En revanche, l'existence de la limite peut être obtenue par

— un argument de monotonie : si f est positive et $\int_a^y f(x)\, dx \leq M < +\infty$ pour tout $y \in [a, b[$, alors $\int_a^b f(x)\, dx$ existe,

— le critère de Cauchy : si $\lim\limits_{y \to b} \sup\limits_{y \leq s \leq t < b} |\int_s^t f(x)\, dx| = 0$, alors $\int_a^b f(x)\, dx$ existe.

Ces principes permettent de dégager des critères de convergence analogues à ceux pour les séries. Nous donnons ci-après un unique résultat, renvoyant le lecteur à un cours de deuxième année pour une vue plus exhaustive.

Proposition 1.13. *Soient f et g des fonctions continues définies sur $[a, +\infty[$. On suppose de plus que g est positive, que l'intégrale $\int_a^{+\infty} g(t)\, dt$ est convergente et qu'en l'infini, on a $f = O(g)$. Alors, l'intégrale $\int_a^{+\infty} f(t)\, dt$ est convergente et lorsque x tend vers l'infini, on a $\int_x^{+\infty} f(t)\, dt = O(\int_x^{+\infty} g(t)\, dt)$.*

Démonstration. Comme $f = O(g)$, il existe des constantes a et A telles que $|f(x)| \leq Ag(x)$ pour $x \geq a$. Soit $\varepsilon > 0$. Comme $\int_a^{+\infty} g(t)\, dt$ est convergente, on peut trouver $a' > a$ tel que $\sup\limits_{y \leq s \leq t < b} |\int_s^t g(x)\, dx| \leq \frac{\varepsilon}{A}$ pour $y \geq a'$. On en déduit que $\sup\limits_{y \leq s \leq t < b} |\int_s^t f(x)\, dx| \leq \varepsilon$ pour $y \geq a'$, ce qui donne la convergence de l'intégrale, d'après le critère de Cauchy. Pour $x \geq a$, on a

$$\left| \int_x^{+\infty} f(t)\, dt \right| \leq \int_x^{+\infty} Ag(t)\, dt = A \int_x^{+\infty} g(t)\, dt,$$

ce qui est l'inégalité voulue. □

1.2.7 Lien série-intégrale

Proposition 1.14. *Soient $f : [a, +\infty[\to \mathbb{R}_+$ une fonction positive et (x_n) une suite à valeurs dans $[a, +\infty[$ telle que $x_0 := a$ et $\lim\limits_{n \to +\infty} x_n = +\infty$.*

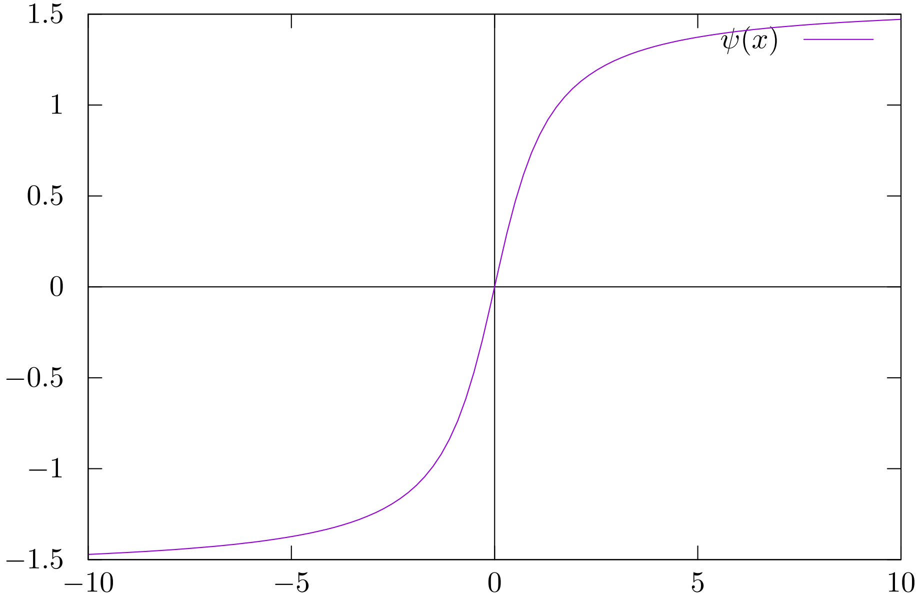
Si on pose $u_n := \int_{x_n}^{x_{n+1}} f(x)\, dx$, alors la série de terme général u_n est de même nature que l'intégrale $\int_a^{+\infty} f(x)\, dx$ et $\sum\limits_{n=0}^{+\infty} u_n = \int_a^{+\infty} f(x) dx$.

Corollaire 1.15. *Soit $f : \mathbb{R}_+ \to \mathbb{R}_+$ une fonction positive et décroissante. Alors la série de terme général $f(n)$ et l'intégrale $\int_0^{+\infty} f(x)\, dx$ sont de même nature.*

1.3 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 $\psi(-\infty) = -\pi/2$.

Homéomorphisme de $(\overline{\mathbb{R}}, d)$ dans $[-\pi/2, \pi/2]$



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 1.16. *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 1.17. *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 1.18. *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 $+\infty$. $\square$

1.4 Limite supérieure

1.4.1 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 $v_n = \sup_{k\geq n} a_k$ 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 $w_n = \inf_{k\geq n} a_k$ 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 1.19. *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 1.20. *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 $\sup_{n \geq \phi(k)+1} \psi(a_n) = \psi\left(\sup_{n \geq \phi(k)+1} a_n\right)$, 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 1.21.

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

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

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 $\overline{\lim_{n \to +\infty}} \; a_n$ 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 1.22. *Une suite* (a_n) *à valeurs dans* $\overline{\mathbb{R}}$ *converge si et seulement si* $\underline{\lim}_{n\to+\infty} a_n = \overline{\lim}_{n\to+\infty} a_n$. *On a alors* $\underline{\lim}_{n\to+\infty} a_n = \overline{\lim}_{n\to+\infty} a_n = \lim_{n\to+\infty} a_n$.

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

Passons au cas où $\underline{\lim}_{n\to+\infty} a_n = \overline{\lim}_{n\to+\infty} a_n = \ell \in \mathbb{R}$. 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 $\ell \in \overline{\mathbb{R}}$, $\underline{\lim}_{n\to+\infty} a_n$ et $\overline{\lim}_{n\to+\infty} a_n$ sont égales à ℓ puisque ce sont des valeurs d'adhérence de $(a_n)_n$ et que ℓ est la seule valeur d'adhérence de la suite $(a_n)_n$. □

Théorème 1.23. *Soient* (u_n) *et* (u'_n) *deux suites vérifiant* $u_n \leq u'_n$ *pour tout* n. *On a* $\overline{\lim}_{n\to+\infty} u_n \leq \overline{\lim}_{n\to+\infty} u'_n$ *et* $\underline{\lim}_{n\to+\infty} u_n \leq \underline{\lim}_{n\to+\infty} u'_n$.

Démonstration. Pour tout n, 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 1.24. *Soit* $\ell \in \mathbb{R}$. *On suppose que pour tout* $\varepsilon > 0$,

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

Alors, (u_n) *converge vers* ℓ.

Démonstration. On a

$$\ell - \varepsilon \leq \varliminf_{n\to+\infty} u_n \leq \varlimsup_{n\to+\infty} u_n \leq \ell + \varepsilon.$$

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

$$\ell \leq \varliminf_{n\to+\infty} u_n \leq \varlimsup_{n\to+\infty} u_n \leq \ell.$$

Le résultat découle du théorème 1.22. □

1.4.2 Limites supérieures, inférieures d'ensembles

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

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

et

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

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 $\omega \in \varlimsup_{n\to+\infty} \{\omega : X_n(\omega) \geq M - \varepsilon\}$. On en déduit que

$$\left\{\omega : \varlimsup_{n\to+\infty} X_n(\omega) \geq M\right\} = \bigcap_{\varepsilon>0} \varlimsup_{n\to+\infty} \{\omega : X_n(\omega) \geq M - \varepsilon\}. \qquad (1.8)$$

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 $\omega \in \varliminf_{n\to+\infty} \{\omega : X_n(\omega) < M + \varepsilon\}$. On en déduit que

$$\left\{\omega : \varlimsup_{n\to+\infty} X_n(\omega) \leq M\right\} = \bigcap_{\varepsilon>0} \varliminf_{n\to+\infty} \{\omega : X_n(\omega) < M + \varepsilon\}. \qquad (1.9)$$

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

$$\left\{\omega : \underline{\lim_{n\to+\infty}} X_n(\omega) \geq M\right\} = \bigcap_{\varepsilon>0} \underline{\lim_{n\to+\infty}} \{\omega : X_n(\omega) > M - \varepsilon\}. \quad (1.10)$$

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

$$\left\{\omega : \overline{\lim_{n\to+\infty}} X_n(\omega) > M\right\} = \bigcup_{\varepsilon>0} \overline{\lim_{n\to+\infty}} \{\omega : X_n(\omega) \geq M + \varepsilon\} \quad (1.11)$$

et

$$\left\{\omega : \underline{\lim_{n\to+\infty}} X_n(\omega) < M\right\} = \bigcup_{\varepsilon>0} \overline{\lim_{n\to+\infty}} \{\omega : X_n(\omega) \leq M - \varepsilon\}. \quad (1.12)$$

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 $p \in \mathbb{N}^*$.

1.5 Compléments sur la compacité

1.5.1 Le procédé diagonal d'extraction

Théorème 1.25. *Soit $(\psi_n)_{n\geq 1}$ une suite d'applications avec $\psi_n : X \to E_n$, où E_n est une partie de $\mathbb{R}^d$. On suppose que pour tout $n \geq 1$, $\psi_n(X)$ est relativement compact dans E_n. Alors, pour toute suite $(x_n)_{n\geq 1}$ à valeurs dans X, il existe une sous-suite $(x_{\phi(n)})_{n\geq 1}$ telle que pour tout i, $\psi_i(x_{\phi(n)})$ converge.*

Démonstration. Soit $(x_n)_{n\geq 1}$ une suite à valeurs dans X. On dit qu'une partie infinie I de $\mathbb{N}$ a la propriété P_i si la suite $(\psi_i(x_n))_{n\in I}$ converge. L'hypothèse de relative compacité dit que pour toute partie infinie I de $\mathbb{N}$, il existe une partie J infinie avec $J \subset I$ et telle que $P_i(J)$ est vérifiée.

On construit alors par récurrence une suite décroissante de parties infinies de $\mathbb{N}$ comme suit. Posons $I_0 = \mathbb{N}$. Pour tout $i \geq 0$, prenons une partie infinie $I_{i+1} \subset I_i$ telle que $P_{i+1}(I_{i+1})$ est vérifiée. Notons $\phi(n)$ le n-ième élément de I_n dans l'ordre croissant. Comme $I_{n+1} \subset I_n$, on a $\phi(n+1) > \phi(n)$. Soit $i \geq 1$. Pour tout $n \geq i$, on a $x_{\phi(n)} \in I_n \subset I_i$. Comme la suite $(\psi_i(x_{\phi(n)}))_{n\geq i}$ est une suite extraite de $(\psi_i(x_k))_{k\in I_i}$, c'est aussi une suite convergente, donc la suite $(\psi_i(x_{\phi(n)}))_{n\geq 1}$ converge. □

1.5.2 Le théorème de Dini-Polyà

Théorème 1.26. *Soit (f_n) une suite d'applications croissantes définies sur $[a,b]$ et convergeant simplement vers la fonction continue f. Alors f_n converge uniformément vers f sur $[a,b]$.*

Ce théorème est aussi appelé deuxième théorème de Dini.

Démonstration. Comme f est uniformément continue sur $[a,b]$ qui est compact, il existe n_0 tel que $|f(x)-f(y)| \leq \varepsilon/3$ dès que $|x-y| \leq \frac{b-a}{n_0}$. Posons, pour k compris entre 0 et n_0 : $x_k = a + k\frac{b-a}{n_0}$: on peut trouver $n_1 \geq n_0$ tel que $|f(x_k) - f_n(x_k)| \leq \varepsilon/3$ pour tout $n \geq n_1$ et tout k compris entre 0 et n_0. Pour x compris entre x_k et x_{k+1}, et $n,p \geq n_1$, on a

$$f_n(x) - f_p(x) \leq f_n(x_{k+1}) - f_p(x_k) \leq 2\varepsilon/3 + f(x_{k+1}) - f(x_k) \leq \varepsilon.$$

Ainsi $\sup\limits_{n,p\geq n_1} \sup\limits_{x\in[a,b]} |f_n(x)-f_p(x)| \leq \varepsilon$, ce qui montre bien qu'on a un critère de Cauchy uniforme pour la suite f_n. □

On en propose également une version un peu plus forte, qui sera utile en probabilités.

Théorème 1.27. *Soit $(f_n)_{n\geq 1}$, f des applications croissantes définies sur $[a,b]$. On suppose de plus que f est continue et qu'il existe D dense dans $[a,b]$ avec $\{a,b\} \subset D \subset [a,b]$ tel que pour tout $x \in D$, $f_n(x)$ tend vers $f(x)$. Alors f_n converge uniformément vers f sur $[a,b]$.*

Démonstration. Soit $\varepsilon > 0$. Par continuité de f et densité de D, pour tout x de $[a,b]$, il existe un voisinage $[c_x, d_x]$ de x dans $[a,b]$ tel que c_x et d_x soient dans D et que $f(d_x) - f(c_x) \leq \varepsilon/3$. Comme $[a,b]$ est compact, on peut extraire de cette famille une famille finie de voisinages $[c_1,d_1]$,…$[c_n,d_n]$.
Maintenant, il existe N tel que pour tout $x \in \{c_1,\dots,c_n,d_1,\dots,d_n\}$ et tout $k \geq N$, on a $|f_k(x) - f(x)| \leq \varepsilon/3$.

Prenons maintenant $x \in [a,b]$ quelconques et $k \geq N$: il existe p entre 1 et n avec $x \in [c_p,d_p]$. Comme f et f_k sont croissantes, $f(x) \in [f(c_p), f(d_p)]$ et $f_k(x) \in [f_k(c_p), f_k(d_p)] \subset [f(c_p) - \frac{\varepsilon}{3}, f(d_p) + \frac{\varepsilon}{3}]$. Ainsi $f(x)$ et $f_k(x)$ sont dans un intervalle de longueur ne dépassant pas ε, donc $|f_k(x) - f(x)| \leq \varepsilon$.

□

1.6 Exercices d'analyse

1.6.1 Exercices corrigés

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

$\rightarrow$ indication $\rightarrow$ solution

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

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

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

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

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

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

$\rightarrow$ indication $\rightarrow$ solution

Exercice 3. On suppose que la suite $(u_n)_{n\geq 0}$ est telle que $u_n \rightarrow 0$ lorsque $n \rightarrow +\infty$ 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 $\left(\log\left(1+\frac{(-1)^n}{n+2}\right)\right)_{n\geq 0}$.

$\rightarrow$ indication $\rightarrow$ solution

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

$\rightarrow$ indication $\rightarrow$ solution

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 $a > 1$.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 6. *Valeurs d'adhérence de certaines suites.*

Rappel : pour x réel, on note $\{x\}$ la partie fractionnaire de x (c'est-à-dire $x - \lfloor x \rfloor$).

1. Soit $(s_n)_{n \geq N}$ une suite croissante de réels de limite $+\infty$. On suppose que $x \in]0, 1[$ est tel que pour tout $n \geq N$, $s_{n+1} - s_n < \min(x, 1 - x)$. Montrer qu'il existe un entier $n \geq N$ tel que $s_{n+1} - s_n = \{s_{n+1}\} - \{s_n\}$ et $\{s_n\} \leq x \leq \{s_{n+1}\}$.
2. Soit $(s_n)_{n \geq 1}$ une suite croissante de réels de limite $+\infty$ et telle que $\lim(s_{n+1} - s_n) = 0$. Montrer que l'ensemble des valeurs d'adhérence de la suite $(\{s_n\})_{n \geq 1}$ est l'intervalle $[0, 1]$.
3. Soit z un nombre complexe de module 1 qui n'est pas une racine de l'unité. Montrer que l'ensemble des valeurs d'adhérence de la suite $(z^n)_{n \geq 1}$ est le cercle unité $\mathbb{U}$ tout entier.
4. Pour θ irrationnel, on considère les suites $(\cos(2n\pi\theta))$ et $(\sin(2n\pi\theta))$. Quelles en sont les valeurs d'adhérence ? En déduire les limites supérieure et inférieure de chacune de ces suites.

$\rightarrow$ indication $\rightarrow$ solution

1.6.2 Exercices non corrigés

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. $\rightarrow$ indication

Exercice 8.

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

   Déterminer $\varliminf\limits_{n \to +\infty} a_n$ et $\varlimsup\limits_{n \to +\infty} a_n$.
2. Soit f une fonction continue croissance de $\mathbb{R}$ dans $\mathbb{R}$. Donner une expression simple de $\varliminf\limits_{n \to +\infty} f(a_n)$ et $\varlimsup\limits_{n \to +\infty} f(a_n)$.

3. Même question lorsque (a_n) est remplacée par une suite dont la limite supérieure est 1 et la limite supérieure est -1

$\rightarrow$ indication

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

$\rightarrow$ indication

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 $\overline{\mathbb{R}}$. $\rightarrow$ indication

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

$\rightarrow$ indication

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
$$\overline{\lim}_{n\to+\infty} \frac{|S_n - f(0)H_n|}{H_n} \leq \varepsilon.$$
Conclure.
3. Donner un équivalent de $\displaystyle\sum_{k=1}^{n} \frac{1}{\sin \frac{k}{n}}$ lorsque n tend vers l'infini.

$\rightarrow$ indication

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. $\to$ indication

Exercice 14. *Suites sous-additives (lemme de Fekete)* (voir [15]).
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 $\overline{\lim_{n\to+\infty}} \frac{u_{kn+r}}{nk+r} \leq \frac{u_k}{k}$.
2. Montrer que $\overline{\lim_{n\to+\infty}} \frac{u_n}{n} \leq \inf_{k\geq 1} \frac{u_k}{k}$.
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.

$\to$ indication

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 $\lim_{n\to+\infty} (pa_n + qb_n)^n$. $\to$ indication

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 $\bigcap_{i=1}^{+\infty} A_i \in \mathcal{A}$.

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

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

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 $\varnothing \in \bigcap_{\mathcal{A}\in T} \mathcal{A} = \mathcal{B}$.

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.

— 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 $A^c \in \bigcap\limits_{\mathcal{A} \in T} \mathcal{A} = \mathcal{B}$.

— Soit $(A_i)_{i \in I}$ une famille dénombrable d'éléments de $\mathcal{B}$. On doit montrer que $\bigcup\limits_{i \in I} A_i \in \mathcal{B}$. Soit $\mathcal{A} \in T$. Comme les A_i sont dans $\mathcal{A}$ et $\mathcal{A}$ est une tribu, on a $\bigcup\limits_{i \in I} A_i \in \mathcal{A}$. Ceci étant vrai pour tout $\mathcal{A} \in T$, on obtient que $\bigcup\limits_{i \in I} A_i \in \bigcap\limits_{\mathcal{A} \in T} \mathcal{A} = \mathcal{B}$.

$\square$

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

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

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

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 $f^{-1}(O) = \bigcup_{i \in I} f_i^{-1}(O)$. 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. $\square$

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

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

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

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

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 $\mathcal{T}$. $\square$

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

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

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 $\Omega = \bigcup_{i \in I'} A_i$ et $\Omega' = \bigcup_{j \in J'} B_j$. 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 $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{A}$, on a
$$\mu\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} \mu(A_i).$$

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 $\Omega = \bigcup_{i=1}^{+\infty} A_i$, 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 $\mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu(A_i)$.*
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 $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{A}$, on a $\mu\left(\bigcup_{i=1}^{+\infty} A_i\right) \leq \sum_{i=1}^{+\infty} \mu(A_i)$.*
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
$$\begin{aligned}
\mu(A \cup B) &= \mu(A \backslash B) + \mu(B \backslash A) + \mu(A \cap B) \\
\mu(A \cup B) &= (\mu(A \backslash B) + \mu(A \cap B)) + (\mu(B \backslash A) + \mu(A \cap B)) \\
&\quad -\mu(A \cap B) \\
&= \mu(A) + \mu(B) - \mu(A \cap B)
\end{aligned}$$
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 $n \geq 2$, $B_n = A_n \backslash \left(\overset{n-1}{\underset{i=1}{\cup}} B_i \right)$. Par construction, les $(B_n)_{n \geq 1}$ sont deux à deux disjoints. De plus, on peut montrer par récurrence sur n que
$$\forall n \geq 1 \quad \overset{n}{\underset{i=1}{\cup}} A_i = \overset{n}{\underset{i=1}{\cup}} B_i.$$
On en déduit
$$\overset{+\infty}{\underset{i=1}{\cup}} A_i = \overset{+\infty}{\underset{i=1}{\cup}} B_i.$$
Donc
$$\mu\left(\overset{+\infty}{\underset{i=1}{\cup}} A_i \right) = \mu\left(\overset{+\infty}{\underset{i=1}{\cup}} B_i \right) = \sum_{i=1}^{+\infty} \mu(B_i) \leq \sum_{i=1}^{+\infty} \mu(A_i).$$
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 :
$$\forall n \geq 1 \quad A_n = \overset{n}{\underset{i=1}{\cup}} B_i$$
et
$$A = \overset{+\infty}{\underset{i=1}{\cup}} A_i = \overset{+\infty}{\underset{i=1}{\cup}} B_i.$$
Ainsi,
$$\begin{aligned} \mu(A) &= \mu\left(\overset{+\infty}{\underset{i=1}{\cup}} B_i \right) = \sum_{i=1}^{+\infty} \mu(B_i) = \lim_{n \to +\infty} \sum_{i=1}^{n} \mu(B_i) \\ &= \lim_{n \to +\infty} \mu\left(\overset{n}{\underset{i=1}{\cup}} B_i \right) = \lim_{n \to +\infty} \mu(A_n). \end{aligned}$$
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, $\bigcup\limits_{i=1}^{+\infty} A_i$ est infini et la somme $\sum\limits_{i=1}^{+\infty} C(A_i)$ a une infinité de termes qui dépassent 1 donc encore une fois $\sum\limits_{i=1}^{+\infty} C(A_i) = +\infty = C\left(\bigcup\limits_{i=1}^{+\infty} A_i\right)$. 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 $A = \underset{n\geq 1}{\cup} A_n$ 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 $\mu\left(A \setminus \overset{n}{\underset{k=1}{\cup}} A_k\right) \leq \varepsilon/2$. On a

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

□

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 $a \le b$, $\mu([a,b]) = \lambda([a-t, b-t]) = b-a$, 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

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

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. $\sup\limits_{n\geq 1} f_n$
2. $\inf\limits_{n\geq 1} f_n$
3. $\varlimsup\limits_{n\to+\infty} f_n$
4. $\varliminf\limits_{n\to+\infty} f_n$
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 $f = \sup\limits_{n\geq 1} f_n$. 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 $\inf\limits_{n\geq 1} f_n = -\sup\limits_{n\geq 1} (-f_n)$ 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. $\varlimsup\limits_{n\to+\infty} f_n = \inf\limits_{n\geq 1} g_n$, avec $g_n = \sup\limits_{k\geq n} f_k$. La mesurabilité des (g_n) provient du point 1 ; on applique alors le point 2.

4. Preuve analogue, ou $\varliminf\limits_{n\to+\infty} f_n = -\varlimsup\limits_{n\to+\infty} (-f_n)$.

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.

$\square$

2.4 Exercices de théorie de la mesure

2.4.1 Exercices corrigés

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

$\to$ indication $\to$ solution

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).

$\to$ indication $\to$ solution

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

$$\varliminf_{n\to+\infty} A_n \in \mathcal{A}, \quad \varlimsup_{n\to+\infty} A_n \in \mathcal{A}, \text{ et } \varliminf_{n\to+\infty} A_n \subset \varlimsup_{n\to+\infty} A_n.$$

2. Montrer que

$$\varlimsup_{n\to+\infty} \mathbb{1}_{A_n} - \varliminf_{n\to+\infty} \mathbb{1}_{A_n} = \mathbb{1}_{\{\varlimsup_{n\to+\infty} A_n \setminus \varliminf_{n\to+\infty} A_n\}}.$$

$\to$ indication $\to$ solution

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 $C = \bigcap_{n\geq 1} A_n$.

$\to$ indication $\to$ solution

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 $\mathbb{R}$.)

$\to$ indication $\to$ solution

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

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

$\to$ indication $\to$ solution

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 $\mathcal{B}(\mathbb{R}) \neq \mathcal{P}(\mathbb{R})$.

$\to$ indication $\to$ solution

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

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

$\to$ indication $\to$ solution

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 $\mu\left(\bigcap_{n=1}^{+\infty} A_n\right) = \lim_{n \to +\infty} \mu(A_n)$? Expliquer.

$\to$ indication $\to$ solution

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 $\mu(A \cap C) = \mu(B \cap C)$.
$\rightarrow$ indication $\rightarrow$ solution

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.

$\rightarrow$ indication $\rightarrow$ solution

2.4.2 Exercices non corrigés

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 $\sigma(f, g)$. $\rightarrow$ indication

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. $\rightarrow$ indication

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 $(\Omega, \mathcal{F})$. $\rightarrow$ indication

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.

$\rightarrow$ indication

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 $\mu_n([n, +\infty[)$. $\rightarrow$ indication

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.
$\rightarrow$ indication

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 $\mathcal{A} \times \mathcal{B}$. $\rightarrow$ indication

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

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

$\rightarrow$ indication

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.

$\rightarrow$ indication

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). $\rightarrow$ indication

Exercice 37. *Lemme de Doob.*
Soient X et Y deux applications mesurables de $(\Omega, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On veut montrer que Y est $\sigma(X)$-mesurable si et seulement si il existe une application mesurable f de $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ dans lui-même telle que $Y = f \circ X$.

1. Traiter le sens “facile” : si $Y = f \circ X$, alors ...
2. Traiter la réciproque, lorsque Y est une fonction simple (étagée).
3. Passer au cas général.

$\rightarrow$ indication

Exercice 38. *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. $\rightarrow$ indication

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 $\mathbb{P}\left(\overset{+\infty}{\underset{i=1}{\cup}} A_i \right) = \sum_{i=1}^{+\infty} \mathbb{P}(A_i)$.

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.*

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 ».

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* $\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i)$.
2. $\forall A, B \in \mathcal{F}, \quad (A \cap B = \varnothing) \Longrightarrow (\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B))$.
3. $\forall A \in \mathcal{F}, \quad \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
4. $\forall A, B \in \mathcal{F}, \quad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
5. $\forall A, B \in \mathcal{F}, \quad \mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.
6. $\forall A, B \in \mathcal{F}, \quad (A \subset B) \Longrightarrow (\mathbb{P}(A) \leq \mathbb{P}(B))$.
7. $\forall A, B \in \mathcal{F}, \quad \mathbb{P}(A \cap B) \leq \min(\mathbb{P}(A), \mathbb{P}(B))$.
8. $\forall A, B \in \mathcal{F}, \quad \mathbb{P}(A \cup B) \geq \max(\mathbb{P}(A), \mathbb{P}(B))$.
9. *Pour toute suite* $(A_i)_{i\geq 1}$ *d'éléments de* $\mathcal{F}$, $\mathbb{P}\left(\bigcup_{i=1}^{+\infty} A_i\right) \leq \sum_{i=1}^{+\infty} \mathbb{P}(A_i)$.
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*

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

$\mathbb{P}(A|B)$ 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 $\mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) > 0$. Alors*

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

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 $\{0, \dots, n\}$, il suffit de montrer que pour tout $n \geq 1$, les tribus $\mathcal{A}_0, \dots, \mathcal{A}_n$ 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. □

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

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

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 :

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

□

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 $\mathcal{L}$,* $\bigcup\limits_{n \geq 1} A_n \in \mathcal{L}$.

On peut déjà remarquer que si $\mathcal{A} \subset \mathcal{P}(\Omega)$ est à la fois un λ-système et un π-système, alors $\mathcal{A}$ est une tribu. En effet, soit $(A_n)_{n \geq 1}$ une suite d'éléments de $\mathcal{A}$. Si on pose $A'_n = \bigcup_{k=1}^{n} A_k = \Omega \backslash \left(\bigcap_{k=1}^{n} A_k^c \right)$, 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 $A = \bigcup_{n \geq 1} A_n = \bigcup_{n \geq 1} A'_n \in \mathcal{A}$, 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. Notons $\mathcal{L}'$ l'intersection de tous les λ-systèmes contenant $\mathcal{P}$. En s'inspirant de la preuve de l'existence de la tribu engendrée, on voit sans peine que $\mathcal{L}'$ est un λ-système, contenu dans tous les λ-systèmes contenant $\mathcal{P}$. Ainsi, pour montrer que $\sigma(\mathcal{P}) \subset \mathcal{L}$, il suffit de montrer que $\sigma(\mathcal{P}) \subset \mathcal{L}'$. Pour cela, il suffit de prouver que $\mathcal{L}'$ est une tribu puisque $\mathcal{L}'$ contient $\mathcal{P}$ et que $\sigma(\mathcal{P})$ est la plus petite des tribus qui contiennent $\mathcal{P}$. D'après la remarque précédente, il suffit de montrer que $\mathcal{L}'$ est un π-système. Pour $E \in \mathcal{L}'$, notons $\mathcal{L}_E = \{A \in \mathcal{L}'; A \cap E \in \mathcal{L}'\}$. Montrons que pour tout $E \in \mathcal{L}'$, $\mathcal{L}_E$ est un λ-système. On a :

— $\Omega \in \mathcal{L}'$ et $\Omega \cap E = E \in \mathcal{L}'$, donc $\Omega \in \mathcal{L}_E$

— Si A, B sont dans $\mathcal{L}_E$ avec $A \subset B$, alors on a

$$(B \backslash A) \cap E = (B \cap E) \backslash (A \cap E) \in \mathcal{L}'$$

car $\mathcal{L}'$ est un λ-système, et $B \backslash A \in \mathcal{L}'$, donc $B \backslash A \in \mathcal{L}_E$.

— De même, si les A_i forment une suite croissante d'éléments de $\mathcal{L}_E$, alors $\bigcup_{n \geq 1} A_n \in \mathcal{L}'$ et $\left(\bigcup_{n \geq 1} A_n \right) \cap E = \left(\bigcup_{n \geq 1} (A_n \cap E) \right) \in \mathcal{L}'$ car $\mathcal{L}'$ est un λ-système, donc $\bigcup_{n \geq 1} A_n \in \mathcal{L}_E$.

Comme $\mathcal{L}_E$ est un λ-système qui contient $\mathcal{P}$, on a $\mathcal{L}_E$ contient $\mathcal{L}'$, ce qui veut dire que pour tout $E \in \mathcal{L}'$ et $F \in \mathcal{L}'$, on a $E \cap F \in \mathcal{L}'$, ce qui montre que $\mathcal{L}'$ est un π-système et achève donc la preuve. □

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 corrigés

Exercice 39. 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 $\mathbb{1}_{A\cap B}, \mathbb{1}_{A^c}, \mathbb{1}_{A\setminus B}, \mathbb{1}_{A\Delta B}$ 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 $\mathbb{1}_{A\cup B} = 1 - (1 - \mathbb{1}_A)(1 - \mathbb{1}_B)$.
$\to$ indication $\to$ solution

Exercice 40. 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.
$\to$ indication $\to$ solution

Exercice 41. 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.
   Cette mesure, appelée loi Zêta de paramètre s, reviendra plusieurs fois dans cet ouvrage en lien avec la théorie des nombres.
   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.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 42. 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.

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. 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.

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 $\frac{n!}{e}$.

→ indication → solution

Exercice 43. *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 $\frac{\phi(n)}{n} = \prod\limits_{i \in I} (1 - 1/p_i)$.

→ indication → solution

Exercice 44. *Valeurs d'adhérence de la suite $\phi(n)/n$.*

1. Montrer que pour toute série divergente positive dont le terme général (u_n) tend vers 0, et pour tout $\ell > 0$, on peut extraire une sous-série (u_{n_k}) telle que $\sum_{k=1}^{+\infty} u_{n_k} = \ell$.
2. On note $\phi(n)$ l'indicatrice d'Euler (voir exercice 43). Montrer que l'ensemble des valeurs d'adhérences de la suite $(\phi(n)/n)_{n\geq 1}$ est l'intervalle $[0,1]$ tout entier.

→ indication → solution

3.6.2 Exercices non corrigés

Exercice 45. Déterminer toutes les mesures de probabilité μ sur $\mathcal{B}(\mathbb{R})$ vérifiant $\mu(\mathbb{N}^*) = 1$ et pour tout $n \in \mathbb{N}^*$, $\mu(\{n\}) = \mu([n+1, +\infty[)$. → indication

Exercice 46. 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

$$B = \varliminf_{n\to+\infty} A_n = \bigcup_{n=1} \bigcap_{k=n}^{+\infty} A_k.$$

Montrer que $\mathbb{P}(A) = 0$ si et seulement si $\mathbb{P}(B) = 0$ → indication

Exercice 47. *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

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

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
$$\mathbb{P}_n(A) = \frac{1}{n} \sum_{B\subset\{1,\dots,N\}} (-1)^{|B|} \lfloor \frac{n}{\prod_{i\in B} p_i^2} \rfloor = \frac{1}{n} \sum_{k\in A} \mu(k) \lfloor \frac{n}{k^2} \rfloor,$$
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 $\Phi : B \mapsto \prod_{i\in B} p_i$ 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 $\mu_s(A) = \prod_{i=1}^{+\infty} \left(1 - \frac{1}{p_i^{2s}}\right) = \frac{1}{\zeta(2s)}$. 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 $\mu_s(E) = \frac{1}{\zeta(s)} \sum_{k=1}^{+\infty} \frac{\mathbb{1}_E(k)}{k^s}$. 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.

$\to$ indication

Exercice 48. Une enquête effectuée parmi les nouveaux adhérents d'un parti politique a montré que les femmes représentaient $40,55$% des nouveaux adhérents. $20,4$% des nouvelles militantes sont enseignantes, tandis que seulement $12,81$% des nouveaux militants de sexe masculin sont enseignants. Parmi les enseignants qui militent nouvellement à ce parti, quelle est la proportion de femmes ? $\to$ indication

Exercice 49. On mélange $n(n \geq 6)$ paires de chaussetttes et l'on tire au hasard six chaussettes. On considère les événements suivants : $E_1 = \{$ obtenir une seule paire $\}$, $E_2 = \{$ obtenir au moins une paire $\}$ et $E_3 = \{$ obtenir trois paires $\}$.

En supposant que tous les ensembles de six chaussettes ont la même probabilité d'être tirés, calculer $\mathbb{P}(E_1), \mathbb{P}(E_2), \mathbb{P}(E_3)$. $\to$ indication

Exercice 50. 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. $\rightarrow$ indication

Exercice 51. *Problème du scrutin.*
Une élection a lieu entre deux candidats A et B. Le premier candidat A obtient a voix et le second B obtient b voix avec $a > b$.

1. Représenter le dépouillement des bulletins à l'aide d'un chemin dans $\mathbb{R}^2$ partant de $(0, 0)$ arrivant à (a, b) constitué uniquement de segments de longueur 1, parallèles à l'axe Ox ou Oy, orientés dans le sens croissant. En déduire un modèle équiprobable concernant le dépouillement.
2. Quelle est la probabilité pour qu'au cours du dépouillement,
   — le premier bulletin soit en faveur de B ?
   — A et B se retrouvent à un instant à égalité ? (Indication : distinguer suivant le premier bulletin.)
   — A ait toujours strictement plus de voix que B ?

$\rightarrow$ indication

Exercice 52. Donner un exemple de trois événements A_1, A_2, A_3 qui ne sont pas indépendants et pour lesquels

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

$\rightarrow$ indication

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*

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.

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.*

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*

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

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*

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

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, $g = \lim_{n\to+\infty} g_n = \underline{\lim}_{n\to+\infty} f_n$. 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

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

Mais d'après le théorème de convergence monotone $\int g_n \, d\mu$ converge vers $\int g \, d\mu$, ce qui est le résultat voulu. Dans cette inégalité, les termes peuvent valoir $+\infty$. □

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 $\int f \, d\mu$.*

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

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

et

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

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.

soit
$$\int g\, d\mu + \int f\, d\mu \leq \int g\, d\mu + \varliminf_{n\to+\infty} \int f_n\, d\mu$$
et
$$\int g\, d\mu - \int f\, d\mu \leq \int g\, d\mu - \varlimsup_{n\to+\infty} \int f_n\, d\mu.$$
Comme $|\int g\, d\mu| < +\infty$, on peut simplifier et on obtient
$$\int f\, d\mu \leq \varliminf_{n\to+\infty} \int f_n\, d\mu$$
et
$$\varlimsup_{n\to+\infty} \int f_n\, d\mu \leq \int f\, d\mu$$
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 $(A_n)_{n\geq 1}$ 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

$$\mu(\Omega_i) \inf_{\Omega_i} f \le \mu(\Omega_i) f(x).$$

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

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

d'où en passant au supremum sur toutes les partitions $\int f \, d\mu \le f(x)$. 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

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

D'où en sommant sur I

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

Cependant, par le théorème de Fubini, on a

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

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 $f = f_1 + i f_2$ 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 $(\int f_n\, d\mu)_{n\geq 1}$ 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 $|f_n| \leq g$. □

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 $t \in T$:*

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

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. □

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 $\sum_{k=2}^{n} (u_k - u_{k-1}) = u_n - u_1$, 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

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

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 $\lim_{n\to+\infty} (H_n - \log n) = \gamma$, cela nous donne $\lim_{n\to+\infty} I_n = -\gamma$.

Or, d'après la question précédente, $\lim_{n\to+\infty} I_n = \int_0^{+\infty} e^{-t} \log t \, dt$, qui d'après la question 2, est égale à $\Gamma'(1)$.
On obtient donc l'identité $\gamma = -\Gamma'(1)$.

4.7.4 Holomorphie d'une intégrale dépendant d'un paramètre

Cette sous-section peut être omise sans dommage par un lecteur non-familier de l'analyse complexe.

Théorème 4.21. *Soient $(\Omega, \mathcal{F}, \mu)$ un espace mesuré et O un ouvert de $\mathbb{C}$. Soit $f(x,z)$ une fonction de deux variables définie sur $\Omega \times O$. On suppose que pour tout $z \in O$, la fonction $x \mapsto f(x,z)$ est mesurable par rapport à $\mathcal{F}$. On suppose que pour tout compact K inclus dans O, il existe une fonction g_K intégrable par rapport à μ telle que pour tout $z \in K$.*

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

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

Alors $F(z) = \int_\Omega f(x,z) \, d\mu(x)$ définit une fonction holomorphe sur O avec

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

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

Démonstration. Commençons par un argument d'analyse complexe. Montrons que pour tout $n \geq 1$ et pour tout compact K inclus dans O, il existe une fonction $g_{n,K}$ intégrable par rapport à μ telle que

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

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

Soit K un compact, $n \geq 1$. Un raisonnement classique de compacité donne l'existence d'un $r > 0$ tel que $K + \overline{B}(0,r) \subset O$ (où $\overline{B}(0,r)$ est la boule fermée centrée en l'origine de rayon r). Notons que $K + \overline{B}(0,r) \subset O$ est également un compact. Pour tout z dans K, on a, pour μ-presque tout x, l'identité

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

Ainsi $|\frac{\partial^n}{\partial^n z} f(x,z)| \le \frac{n!}{r^n} g_{K+\overline{B}(0,r)}(x)$, ce qui donne le résultat voulu en prenant comme fonction majorante $g_{n,K} = n! r^{-n} g_{K+\overline{B}(0,r)}$.

Passons maintenant à la preuve de l'identité et de la formule pour $n = 1$. Soit $z_0 \in O$. Prenons r tel que la boule fermée de centre z_0 et de rayon r soit incluse dans O. On prend $K = \overline{B}(z_0, r)$.
Posons $F_{\theta,x}(r) = f(x, z_0 + re^{i\theta})$. On a, pour μ-presque tout x, l'égalité (vectorielle)

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

soit

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

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

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

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

Application à la fonction Gamma

Théorème 4.22. *La fonction* Γ

$$z \mapsto \int_{[0,+\infty[} t^{z-1} e^{-t}\, d\lambda(t)$$

définit une fonction holomorphe sur $\{a + ib; (a,b) \in]0,+\infty[\times \mathbb{R}\}$. *Ses dérivées sont données par*

$$\Gamma^{(n)}(z) = \int_{[0,+\infty[} (\log t)^n t^{z-1} e^{-t}\, d\lambda(t)$$

Elle vérifie la relation fonctionnelle $\Gamma(z+1) = z\Gamma(z)$*, ce qui permet de la prolonger en une fonction holomorphe sur* $\mathbb{C}\backslash\{-n; n \in \mathbb{N}\}$.

Démonstration. On prend $O = \{a + ib; (a,b) \in]0,+\infty[\times \mathbb{R}\}$. Soit K un compact inclus dans O. On pose $a(K) = \inf\{\text{Re}\,(z); z \in K\}$. Comme $z \mapsto \text{Re}\,(z)$ est continue, l'infimum est atteint et on a $a(K) > 0$.
On pose maintenant $M(K) = \sup\{\text{Re}\,(z); z \in K\}$, puis on considère la fonction $g_K(x) = t^{a(K)-1} e^{-t} + t^{M(K)-1} e^{-t}$. On a déjà vu dans l'exercice sur la

fonction Gamma réelle que g_K était intégrable, et on a pour tout $z \in K$ et tout $t > 0$:

$$|t^{z-1}e^{-t}| \leq g_K(t).$$

Avec le théorème précédent, cela nous donne l'holomorphie de la fonction Gamma complexe. Pour $z \in O$, une intégration par parties donne

$$\int_0^M e^{-t}t^z \, dt = [-t^z e^{-t}]_0^M + z\int_0^M e^{-t}t^{z-1} \, dt,$$

d'où $\Gamma(z+1) = z\Gamma(z)$ en faisant tendre M vers l'infini.
Posons $\Gamma_0(z) = \Gamma(z)$, puis pour $k \geq 0$, $\Gamma_{k+1}(z) = \frac{\Gamma_k(z+1)}{z}$.
On montre aisément par récurrence que Γ_k définit une fonction holomorphe sur $H_k = \{a+ib; a > -k\}\backslash\{0; -1; -2; \ldots; -k\}$. De manière explicite, on a sur H_k :

$$\Gamma_k(z) = \frac{\Gamma(z+k)}{(z+k-1)(z+k-2)\ldots z}.$$

Cependant pour z dans H_k, on a encore pour tout entier naturel non nul ℓ :

$$\begin{aligned}\Gamma_{k+\ell}(z) = \frac{\Gamma(z+k)(z+k)(z+1)\ldots(z+k+\ell-1)}{(z+k+\ell-1)\ldots z} &= \frac{\Gamma(z+k)}{(z+k-1)\ldots z} \\ &= \Gamma_k(z).\end{aligned}$$

Finalement, les fonctions Γ_k coïncident et définissent une fonction holomorphe sur $\cup_{k\geq 1} H_k = \mathbb{C}\backslash\{-n; n \in \mathbb{N}\}$. $\square$

Les conséquences sont aussi intéressantes dans le domaine réel : on en déduit directement que la restriction de Γ à $]0, +\infty[$ est C^∞, avec l'expression des dérivées. En particulier, pour tout $x > 0$

$$\Gamma''(x) = \int_{[0,+\infty[} (\log t)^2 t^{x-1}e^{-t} \, d\lambda(t) > 0,$$

et Γ est donc strictement convexe sur $]0, +\infty[$, et Γ' est strictement croissante. Comme $\Gamma(1) = \Gamma(2) = 1$, Γ' s'annule en un unique point c de $]1, 2[$: Γ est strictement décroissante sur $]0, c[$, et strictement croissante sur $]c, +\infty[$.

Comme $\Gamma(n+1) = n!$, on a $\lim_{x\to+\infty} \Gamma(x) = +\infty$.
L'inégalité $\Gamma(x) \geq \int_0^1 t^{x-1}e^{-t} \, dt \geq \frac{1}{ex}$ donne $\lim_{x\to 0^+} \Gamma(x) = +\infty$.

Exercice. Posons $\phi(z) = \int_{\mathbb{R}_+} \frac{e^{-x}}{x-z} \, d\lambda(x)$ et montrons que ϕ est holomorphe sur $\mathbb{C}\backslash\mathbb{R}_+$. Soit K un compact de $\mathbb{C}\backslash\mathbb{R}_+$: la fonction $z \mapsto d(z, \mathbb{R}_+)$ est continue sur K, donc y atteint son minimum, noté ε_K. Comme K ne rencontre pas $\mathbb{R}_+$, on a $\varepsilon_K > 0$. On peut donc appliquer le théorème avec $g_K(x) = \frac{e^{-x}}{\varepsilon_K}$. Il s'agit en fait de la transformée de Stieljès de la fonction e^{-x}.

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 4.23. *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 4.24. *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 4.25. *Un tel borélien N n'est pas nécessairement unique.*

Théorème 4.26. *Toute mesure σ−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 4.27. *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 [32]).*

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 4.28. *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 4.29. *Pour tout $A \in \mathcal{X} \otimes \mathcal{Y}$, $x \in X$ et $y \in Y$, on note*

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

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

Théorème 4.31. *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 4.32. *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 4.33 (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 4.29.

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 4.31 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. $\square$

Théorème 4.34 (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*

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

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

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

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

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

□

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 $|f(x,t)| \le \frac{1-\cos t}{t} e^{-xt}$, d'où pour tout $t > 0$,

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

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 $0 \le F(x) \le \int_{]0,+\infty[} |f(x,t)|\, d\lambda(t)$, 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 4.35. *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 4.36 (Série à paramètre)**.** *Soit une famille de nombres réels $a(k, n)$ pour $k \geq 1, n \geq 1$ entiers. On suppose qu'il existe une suite de nombres réels positifs $(c_k)_{k\geq 1}$ avec les propriétés :*

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

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

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

Alors les deux séries $s_n = \sum_{k=1}^{+\infty} a(k, n)$ et $s = \sum_{k=1}^{+\infty} a(k, \infty)$ convergent absolu-

ment 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 4.36 par des moyens élémentaires (avec des ε) est également un exercice très instructif que nous vous conseillons vivement.

Théorème 4.37. *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 4.38. *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 4.39. *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. $\square$

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 4.40. *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 :

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

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. $\square$

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

Théorème 4.41. *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 $\mathbb{R} \times \mathbb{R}^{\{2,\dots,d\}}$, on prend $\mu = \lambda^{d-1}$ et $f(x) = \alpha\langle x, e_j\rangle e_i$. $\square$

Théorème 4.42. *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 $M = \text{diag}(\lambda, 1, \dots, 1)$, on vérifie facilement la formule lorsque A est un pavé : on a deux mesures qui coïncident sur un π-système qui engendre la tribu, elles sont donc égales. Passons au cas où M est inversible. On peut alors écrire $M = \text{diag}(\det M, 1, \dots, 1)N$, où $N \in SL_d(\mathbb{R})$. Ainsi

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

Reste le cas où M 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. $\square$

Corollaire 4.43. *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 4.31. 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

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

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.

4.12.2 Exercice : la fonction Bêta

Rappelons que la fonction *Gamma* $\Gamma : \mathbb{R}_+^* \to \mathbb{R}$ est définie pour tout $x > 0$ par $\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}\, dt$.

1. Montrer que quels que soient $x, y > 0$, on a
$$\Gamma(x)\Gamma(y) = \int_{\mathbb{R}^2} (G \circ T)(t_1, t_2)\, d(\lambda \otimes \lambda)(t_1, t_2),$$
où $G(t, s) = t^{x-1}(s-t)^{y-1}e^{-t}\, \mathbb{1}_{\{0 \le t \le s\}}$ et $T(t_1, t_2) = (t_1, t_1 + t_2)$.
2. Montrer que
$$\Gamma(x)\Gamma(y) = \int_{\mathbb{R}^2} G(t, s)\, d(\lambda \otimes \lambda)(s, t).$$
3. En déduire que
$$\beta(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)},$$
où la fonction *Bta* $\beta : \mathbb{R}_+^* \times \mathbb{R}_+^* \to \mathbb{R}$ est définie pour tous $x > 0$, $y > 0$ par $\beta(x, y) = \int_0^1 \theta^{x-1}(1-\theta)^{y-1}\, d\theta$.

Solution

1. D'après le théorème de Tonelli, on a
$$\Gamma(x)\Gamma(y) = \int_{\mathbb{R}^2} t_1^{x-1}t_2^{y-1}e^{-(t_1+t_2)}\, \mathbb{1}_{\mathbb{R}_+}(t_1)\, \mathbb{1}_{\mathbb{R}_+}(t_2)\, d(\lambda \otimes \lambda)(t_1, t_2).$$
Il n'est pas difficile de vérifier que
$$t_1^{x-1}t_2^{y-1}e^{-(t_1+t_2)}\, \mathbb{1}_{\mathbb{R}_+}(t_1)\, \mathbb{1}_{\mathbb{R}_+}(t_2) = (G \circ T)(t_1, t_2),$$
ce qui donne le résultat voulu.
2. D'après le théorème de transfert pour les fonctions mesurables positives,
$$\int_{\mathbb{R}^2} G \circ T(t_1, t_2)\, d(\lambda \otimes \lambda)(t_1, t_2) = \int_{\mathbb{R}^2} G(t, s)\, d\mu(s, t),$$

où μ est la mesure image de $\lambda \otimes \lambda$ par T. Or l'application T est de déterminant 1, donc laisse invariante la mesure $\lambda \otimes \lambda$.
On a donc $\mu = \lambda \otimes \lambda$, d'où

$$\Gamma(x)\Gamma(y) = \int_{\mathbb{R}^2} G(t,s)\ d(\lambda \otimes \lambda)(s,t).$$

3. D'après le théorème de Tonelli, on a

$$\int_{\mathbb{R}^2} G(t,s)\ d(\lambda \otimes \lambda)(s,t) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} G(t,s)\ d\lambda(t) \right)\ d\lambda(s).$$

À s fixé, calculons $\int_{\mathbb{R}} G(t,s)\ d\lambda(t)$. Si $s < 0$, l'intégrale est évidemment nulle. Sinon,

$$\int_{\mathbb{R}} G(t,s)\ d\lambda(t) = e^{-s} \int_0^s t^{x-1}(s-t)^{y-1}\ dt.$$

En posant $t = \theta s$, on a facilement

$$\int_{\mathbb{R}} G(t,s)\ d\lambda(t) = e^{-s}s^{x+y-1} \int_0^1 \theta^{x-1}(1-\theta)^{y-1}\ d\theta = e^{-s}s^{x+y-1}\beta(x,y).$$

Finalement, on obtient le résultat voulu

$$\Gamma(x)\Gamma(y) = \int_{\mathbb{R}_+} e^{-s}s^{x+y-1}\beta(x,y)\ d\lambda(s) = \beta(x,y)\Gamma(x+y).$$

Corollaire 4.44. *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. □

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

4.12.3 Changement de variables C^1

Théorème 4.45. *Soient U, U' deux ouverts de $\mathbb{R}^d$, ϕ un C^1-difféomorphisme de U dans U'. Soit f une application mesurable définie sur U'. Alors f est intégrable sur U' si et seulement si $f \circ \phi(.) \times |\det D.\ \phi|$ est intégrable sur U et*

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

En particulier, pour A borélien de U, on a

$$\int_{\phi(A)} f(y)\ d\lambda(y) = \int_A f(\phi(x)) \times |\det D_x\ \phi|\ d\lambda(x).$$

Remarque 4.46. *La quantité $\det D_x\ \phi$ est appelée déterminant jacobien (ou plus simplement Jacobien) de ϕ au point x.*

Nous avons choisi d'admettre ce théorème. Sa démonstration, difficile, est basée sur le calcul différentiel et nous semble assez éloignée des techniques que ce cours se propose d'enseigner. On pourra trouver une démonstration dans les ouvrages cités en référence.

Application : calcul de l'intégrale de Gauss

On prend $U =]0, +\infty[\times]0, 2\pi[$, $U' = \mathbb{R}^2 \backslash (\mathbb{R}_+ \times \{0\})$ et $f(x, y) = \exp(-\frac{x^2+y^2}{2})$. On fait le changement de variable polaire : $\phi(r, \theta) = (r\cos\theta, r\sin\theta)$. D'un côté, on a

$$\int_{U'} f(x, y)\ d(\lambda \otimes \lambda)(x, y) = \int_{\mathbb{R}^2} f(x, y)\ d(\lambda \otimes \lambda)(x, y) = I^2,$$

avec $I = \int_{\mathbb{R}} \exp(-\frac{x^2}{2})\ d\lambda(x)$, où la dernière égalité vient du théorème de Tonelli. De l'autre, on a

$$|\det D_{r,\theta}\ \phi| = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r,$$

d'où

$$\int_{]0,+\infty[\times]0,2\pi[} e^{-\frac{r^2}{2}} r\ d(\lambda \otimes \lambda)(r, \theta) = \int_{]0,+\infty[} (2\pi) r e^{-\frac{r^2}{2}}\ d\lambda(r) = 2\pi.$$

Pour la dernière égalité, on a remarqué que $-e^{-r^2/2}$ est une primitive de $re^{-\frac{r^2}{2}}$. On a donc $I^2 = 2\pi$, soit

$$I = \int_{\mathbb{R}} \exp(-\frac{x^2}{2})\, d\lambda(x) = \sqrt{2\pi}.$$

Application : mesure image par un C^1-difféomorphisme

Corollaire 4.47. *Soient O_1 et O_2 deux ouverts de $\mathbb{R}^d, d \geq 1$. On suppose que T est un C^1-difféomorphisme de O_1 dans O_2. Soit maintenant μ_1 une mesure positive sur $\mathbb{R}^d$ telle que $\mu_1(\mathbb{R}^d \backslash O_1) = 0$ et admettant une densité f_1 par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. Alors, la mesure image de μ_1 par T admet comme densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ la fonction f_2 définie par*

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

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

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

ce qui donne le résultat voulu. □

4.12.4 Intégration des fonctions radiales

Théorème 4.48. *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 4.49. 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 4.50. *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

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

En faisant le quotient et en appliquant le théorème précédent, on obtient le résultat voulu. □

Remarque 4.51. *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 4.52. *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 $\mu(\Omega_i) \leq \mu(f \geq \inf\{f(\omega);\omega \in \Omega_i\}) \leq \mu(f > 0) = 0$. Cela entraîne que pour toute partition finie, $I((\Omega_i)_{i\in I}, f) = 0$, d'où la nullité de l'intégrale. □

Lemme 4.53. *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

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

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

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

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. 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.

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.

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

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

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 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

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

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 corrigés

Exercice 53. 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]
$\rightarrow$ indication $\rightarrow$ solution

Exercice 54. *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 $\mathbb{R}_+$. $\rightarrow$ indication $\rightarrow$ solution

Exercice 55. *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 $N(x) = \sum_{k=0}^{+\infty} \mathbb{1}_A(T^k(x))$ 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.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 56.

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

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 [28], Partie II, chapitre 1, paragraphe 5.

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

$\rightarrow$ indication $\rightarrow$ solution

Exercice 57. 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 $p < 0$.
3. $(x^2 - y^2)/(x^2 + y^2)^2$.
4. $(x - y)/(x + y)^3$.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 58. *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. $\rightarrow$ indication $\rightarrow$ solution

Exercice 59. *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 :
$$\int_0^{+\infty} e^{-t^2/2}\, dt = \frac{1}{2}\sqrt{2\pi} = \sqrt{\pi/2}.$$

$\rightarrow$ indication $\rightarrow$ solution

Exercice 60. *Calcul de* $\Gamma(1/2)$.
Connaissant la valeur de l'intégrale de Gauss, montrer que
$$\Gamma(1/2) = \sqrt{\pi}.$$

$\rightarrow$ indication $\rightarrow$ solution

Exercice 61. 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

$$\sum_{n=0}^{+\infty} \frac{(-1)^n}{3n+1} = \frac{3\log 2 + \sqrt{3}\pi}{9}.$$

$\to$ indication $\to$ solution

Exercice 62. *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 $\displaystyle\int_0^{+\infty} \frac{\sin x}{x}\, dx$.

$\to$ indication $\to$ solution

Exercice 63. *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
$$\int_0^{+\infty} \cos(u^2)\,du \text{ et } \int_0^{+\infty} \sin(u^2)\,du.$$

→ indication → solution

Exercice 64. *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 $\int_0^{n\pi} \frac{|\sin t|}{t}\,dt \sim \frac{2}{\pi}\log n$.

→ indication → solution

Exercice 65. *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 $\int_0^{+\infty} \frac{\sin t}{t}\,dt$.

→ indication → solution

Exercice 66. 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 $a=b=c=1$.

→ indication → solution

Exercice 67. *Théorème de Minkowski. Application : le théorème des 4 carrés.*
Soit d un entier, avec $d \geq 2$. Pour $u_1,\dots,u_d$ vecteurs de $\mathbb{R}^d$, on rappelle que $\det(u_1,\dots,u_d)$ est le déterminant de la matrice dont les colonnes sont les coordonnées de $u_1,\dots,u_d$ dans la base canonique.
Une partie A de $\mathbb{R}^d$ est dite *convexe* si quels que soient $a,b\in A$ et $\theta\in[0,1]$, on a $\theta a+(1-\theta)b\in A$. Une partie A de $\mathbb{R}^d$ est dite symétrique par rapport à l'origine si pour tout $a\in A$, $-a\in A$.

Si $(u_1, \ldots, u_d)$ sont des vecteurs de $\mathbb{R}^d$ formant une partie libre, l'ensemble

$$S = \left\{ \sum_{k=1}^{d} z_i u_i ; z \in \mathbb{Z}^d \right\}$$

est appelé un réseau.

Le but de l'exercice est de montrer le théorème suivant, dû à Minkowski. Soit A un convexe de $\mathbb{R}^d$ symétrique par rapport à l'origine. Si le réseau S basé sur $u_1, \ldots, u_d$ est tel que $\lambda^d(A) > 2^d |\det(u_1, \ldots, u_d)| > 0$, alors A rencontre le réseau S en un point différent de $(0, \ldots, 0)$.

1. On pose $T = [0, 1[^d$. Montrer que les ensembles $\left((\frac{1}{2}A) \cap (s + T)\right)_{s \in \mathbb{Z}^d}$ forment une partition de $\frac{1}{2}A$.
2. Montrer que si $\lambda^d 2(A) > 2^d$, alors les ensembles $(\frac{1}{2}A + s)_{s \in \mathbb{Z}^d}$ ne sont pas deux-à-deux disjoints.
   En déduire qu'alors A intersecte $\mathbb{Z}^d \backslash \{(0, \ldots, 0)\}$.
3. Montrer le théorème voulu dans le cas général.
4. (a) Soit p un nombre premier. Calculer le nombre de carrés dans $\mathbb{F}_p$. En déduire qu'il existe des entiers r et s tels que $r^2 + s^2 + 1$ soit divisible par p.

   (b) Soit A l'image de $\mathbb{Z}^d$ par l'application de matrice

   $$\begin{pmatrix} p & 0 & r & s \\ 0 & p & s & -r \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

   Montrer que le carré de la norme euclidienne d'un point de A est divisible par p. Montrer que A rencontre la boule euclidienne ouverte de rayon $\sqrt{2p}$ ailleurs qu'en zéro. (On rappelle que le volume de la boule unité de $\mathbb{R}^4$ est $\frac{\pi^2}{2}$.)

   (c) On rappelle l'identité de Lagrange :
   si $X = bc' - b'c$, $Y = ca' - c'a$, $Z = ab' - a'b$, $T = aa' + bb' + cc'$, alors
   $X^2 + Y^2 + Z^2 + T^2 = (a^2 + b^2 + c^2) \cdot (a'^2 + b'^2 + c'^2)$. Montrer que tout entier s'écrit comme somme de 4 carrés d'entiers.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 68. *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.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 69. *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

$$\zeta(s)\Gamma(s) = \int_0^{+\infty} t^{s-1}\frac{e^{-t}}{1-e^{-t}}\;dt.$$

$\rightarrow$ indication $\rightarrow$ solution

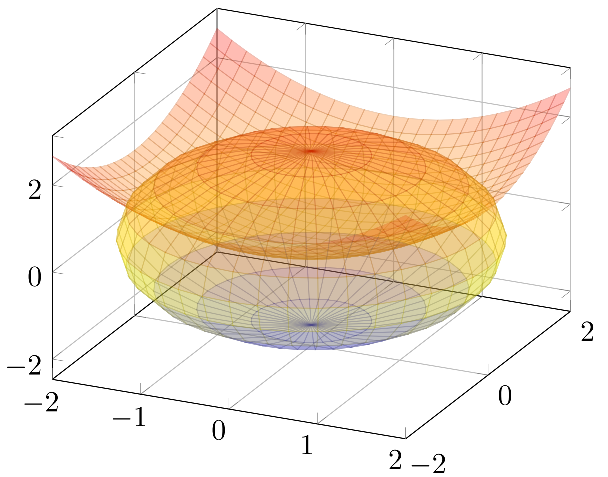
Exercice 70. *Formule de Stirling.*

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

$\rightarrow$ indication $\rightarrow$ solution

Exercice 71. 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.

$\to$ indication $\to$ solution

Exercice 72. *Solide de révolution - centre de gravité.*

Pour θ réel, on note $M_\theta = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}$. M_θ est la matrice de la rotation d'angle θ autour de l'axe des y.

1. On prend $0 \leq \alpha < \beta \leq 2\pi$. Montrer que $T : (x,y,\theta) \mapsto M_\theta \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$ réalise un C^1-difféomorphisme de $]0,+\infty[\times\mathbb{R}\times]\alpha,\beta[$ sur son image.
2. Soit V un borélien de $]0,+\infty[\times\mathbb{R}$. On note $R = T(V\times]\alpha,\beta[)$. R est un volume de révolution engendré par la rotation de la surface $V \times \{0\}$ autour de l'axe des y entre les angles α et β. Montrer que
$$\lambda^3(R) = (\beta - \alpha) \int_V x \, d\lambda^2(x,y).$$
3. Si V est un ensemble mesurable borné de $\mathbb{R}^n$ de mesure non-nulle, on appelle alors centre de gravité de V le point $G = (m_1, \dots, m_n)$ défini par $m_i = \frac{1}{\lambda^d(V)} \int_V x_i \, d\lambda^d(x_1, \dots, x_d)$. Démontrer le théorème de Guldin : la mesure du volume engendré par la révolution d'un élément de surface plane autour d'un axe situé dans son plan et ne le coupant pas est égale au produit de l'aire de la surface par la longueur de la circonférence décrite par son centre de gravité. En déduire la position du centre de gravité d'un quart de disque.
4. Montrer que si $\lambda^2(V) > 0$, alors les coordonnées du centre de gravité de $W = T(V \times [0,2\pi])$ sont
$$\left(0, \frac{\int_V xy \, d\lambda^2(x,y)}{\int_V x \, d\lambda^2(x,y)}, 0\right).$$

$\to$ indication $\to$ solution

Exercice 73. *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
$$I(t) = \int_0^{+\infty} g(x) e^{th(x)} \, d\lambda(x) \sim A e^{h(0)t} (ct)^{-\frac{\alpha+1}{\beta}} K_{\alpha,\beta},$$
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.

→ indication → solution

Exercice 74. 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 $\mathbb{R}^n$. → indication → solution

Exercice 75. *Convergence uniforme de séries trigonométriques.*

1. Pour un entier naturel non nul n et un réel t, on pose
   $$F_n(t) = \sum_{k=1}^n \frac{\sin(kt)}{k}.$$
   Montrer l'existence d'une constante M telle que
   $$\forall n \geq 1 \quad \forall t \in \mathbb{R} \quad |F_n(t)| \leq M.$$
2. Soit $(\varepsilon_n)_{n\geq 1}$ une suite réelle décroissante de limite nulle. Montrer que la série de terme général $\varepsilon_k \frac{\sin(kt)}{k}$ converge uniformément.
3. Donner un exemple de série trigonométrique convergeant uniformément, mais pas normalement.

Pour d'autres jolis résultats sur les séries trigonométriques, on pourra se reporter à Kahane [22].

→ indication → solution

4.14.2 Exercices non corrigés

Exercice 76. 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 $\mu = \nu$. → indication

Exercice 77. 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 $a \iff b, a \iff c, d \implies c, (c\&b) \implies d$. $\to$ indication

Exercice 78. 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. $\to$ indication

Exercice 79. *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

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

$\to$ indication

Exercice 80. 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

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

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

Exercice 81. 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. $\to$ indication

Exercice 82. 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.
$\to$ indication

Exercice 83. 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
$$\sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{(k+1)^2} = \frac{H_{n+1}}{n+1}, \text{ où } H_n = \sum_{k=1}^{n} \frac{1}{k}.$$

→ indication

Exercice 84. Démontrer que
$$\int_0^1 (e^x - 1)(\log x + \frac{1}{x}) \, dx = \sum_{n=1}^{+\infty} \frac{n^2+n+1}{n(n+1)(n+1)!}.$$

→ indication

Exercice 85. Calculer $\displaystyle\int_0^1 \frac{1}{1+x^2} \, dx$. En déduire $\displaystyle\sum_{n=0}^{+\infty} \frac{(-1)^n}{2n+1}$. → indication

Exercice 86. Démontrer que $\displaystyle\int_0^{+\infty} \frac{\sin ax}{e^x - 1} \, dx = \sum_{n=1}^{+\infty} \frac{a}{n^2 + a^2}$. → indication

Exercice 87. Démontrer que $\displaystyle\int_0^1 \frac{(x \log x)^2}{1+x^2} \, dx = 2 \sum_{n=0}^{+\infty} \frac{(-1)^n}{(2n+3)^3}$. → indication

Exercice 88. *Calcul de* $\sum\limits_{n\geq 1} \frac{1}{n^2}$ *avec Fubini.*
Calculer de deux façons différentes :
$$J = \int_{-1}^{1} \int_{-1}^{1} \frac{1}{1+2xy+y^2} \, dx \, dy.$$

→ indication

Exercice 89. Montrer que pour tout $z \in \mathbb{C}$, $(1+z/n)^n$ tend vers $\exp(z)$ lorsque n tend vers l'infini. → indication

Exercice 90. *Régularité de la fonction Gamma.*
Démontrer que la fonction Γ est de classe $\mathcal{C}^\infty$ sur $]0, +\infty[$ et que pour tout entier $n \geq 1$,
$$\Gamma^{(n)}(x) = \int_0^{+\infty} e^{-t} (\log t)^n t^{x-1} \, dt.$$

→ indication

Exercice 91. *Intégrale de Wallis de deuxième espèce.*
Trouver par convergence dominée l'équivalent de l'intégrale de Wallis de seconde espèce :

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

$\to$ indication

Exercice 92. *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 $\eta(s) = (1 - \frac{1}{2^{s-1}})\zeta(s)$. $\to$ indication

Exercice 93. 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. $\to$ indication

Exercice 94.

1. Soit ϕ une application de $\mathbb{N}$ dans $\mathbb{R}_+$. Montrer que l'on a

$$\int_{\mathbb{R}^d} \phi(\text{Ent}(\|x\|_\infty)) \, d\lambda^d(x) = \sum_{n=0}^{+\infty} \left((2n+2)^d - (2n)^d\right) \phi(n).$$

2. Soit $\alpha > 0$. À quelle condition la fonction $x \mapsto \frac{1}{\|x\|_2^\alpha}$ est elle intégrable sur le complémentaire de la boule unité ?
3. Montrer que l'application $f : x \mapsto \frac{x}{\|x\|^2}$ réalise un C^1 difféomorphisme de $\{x \in \mathbb{R}^d; 0 < \|x\|_2 < 1\}$ sur $\{x \in \mathbb{R}^d; 1 < \|x\|_2\}$ et que sa différentielle est

$$h \mapsto Df_x.h = \frac{\|x\|_2^2 h - 2\langle x, h\rangle x}{\|x\|_2^4}.$$

4. Soit $\alpha > 0$. À quelle condition la fonction $x \mapsto \frac{1}{\|x\|_2^\alpha}$ est elle intégrable sur la boule unité ?

$\to$ indication

Exercice 95. Pour $x \geq 0$, on pose $R(x) = \int_x^{+\infty} \frac{\sin u}{u}\, du$. Le but de l'exercice est de démontrer l'identité

$$\int_0^{+\infty} \frac{R(x)}{\sqrt{x}}\, dx = \sqrt{2\pi}.$$

1. Montrer que $R(x)$ est bien définie pour $x \geq 0$ et qu'on a en l'infini $R(x) = O(1/x)$. En déduire que R est bornée.
2. $\frac{R(x)}{\sqrt{x}}$ est-elle intégrable par rapport à la mesure de Lebesgue sur $\mathbb{R}_+$?
3. Pour $\lambda \geq 0$, On pose $R_\lambda(x) = \int_x^{+\infty} \frac{\sin u}{u} e^{-\lambda u}\, du$.
   (a) Justifier l'existence de $R_\lambda(x)$, puis montrer que pour tout $x > 0$ et tout $\lambda > 0$, on a
   $$|R_\lambda(x) - R(x)| \leq |\cos x| \frac{1 - e^{-\lambda x}}{x} + \int_x^{+\infty} \frac{\Psi(\lambda u)}{u^2}\, du,$$
   où on a posé $\Psi(x) = |e^{-x}(1 + x) - 1|$.
   (b) Montrer qu'il existe $M < +\infty$ telle que $\Psi(x) \leq M$ pour tout $x \geq 0$. En déduire que pour tout $x > 0$, $\lim_{\lambda \to 0} R_\lambda(x) = R(x)$.
   (c) Montrer qu'il existe $L > 0$ telle que $\Psi(x) \leq Lx^2$ pour tout $x \geq 0$. Établir les inégalités
   $$\begin{aligned} \forall \lambda > 0 \quad \forall x \geq 1 \quad |R_\lambda(x) - R(x)| \quad &\leq \frac{1 + 2M}{x} \\ \forall \lambda > 0 \quad \forall x \leq 1 \quad |R_\lambda(x) - R(x)| \quad &\leq \lambda + L\lambda^2 + 2M. \end{aligned}$$
   Indication :
   $$\int_x^{+\infty} \frac{\Psi(\lambda u)}{u^2}\, du = \int_x^1 \frac{\Psi(\lambda u)}{u^2}\, du + \int_1^{+\infty} \frac{\Psi(\lambda u)}{u^2}\, du.$$
   (d) Montrer que pour tout $\lambda > 0$, $\frac{R_\lambda(x)}{\sqrt{x}}$ est intégrable par rapport à la mesure de Lebesgue sur $\mathbb{R}_+$, puis que
   $$\int_0^{+\infty} \frac{R(x)}{\sqrt{x}}\, dx = \lim_{\lambda \to 0} \int_0^{+\infty} \frac{R_\lambda(x)}{\sqrt{x}}\, dx.$$
4. Montrer que pour tout $\lambda > 0$, on a
   $$\int_0^{+\infty} \frac{R_\lambda(x)}{\sqrt{x}}\, dx = 2 \int_0^{+\infty} \frac{\sin u}{\sqrt{u}} e^{-\lambda u}\, du.$$
5. Montrer finalement que
   $$\int_0^{+\infty} \frac{R(x)}{\sqrt{x}}\, dx = \sqrt{2\pi}.$$
   (On rappelle que $\int_0^{+\infty} \frac{\sin u}{\sqrt{u}}\, du = \sqrt{\frac{\pi}{2}}$.)

→ indication

Exercice 96. *Expression intégrale de* $\Gamma'(x)/\Gamma(x)$.
Le but de cet exercice est de donner une expression intégrale de la fonction spéciale digamma (appelée aussi fonction psi), qui est la dérivée logarithmique de la fonction Gamma.
On note ψ la fonction définie sur $]0,+\infty[$ par la relation

$$\psi(x)=\int_0^{+\infty}\left(\frac{e^{-t}}{t}-\frac{e^{-xt}}{1-e^{-t}}\right)\,dt.$$

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 ψ est bien définie.
2. Montrer que pour tout entier $n\geq 1$,
$$H_n=\int_0^{+\infty}\frac{e^{-t}-e^{-(n+1)t}}{1-e^{-t}}\,dt.$$
3. Pour $u>0$, on pose
$$F(u)=\int_0^{+\infty}\frac{e^{-t}-e^{-ut}}{t}\,dt.$$
Vérifier que F est bien définie sur $]0,+\infty[$, puis montrer qu'elle est dérivable. En déduire la valeur de F.
4. On pose
$$J_n=\int_0^{+\infty}e^{-nt}\left(\frac{1}{t}-\frac{1}{1-e^{-t}}\right)\,dt.$$
Déterminer la limite de J_n quand n tend vers l'infini.
5. En déduire que $\psi(1)=-\gamma$, où γ est la constante d'Euler.
6. On pose $\beta(x,y)=\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ (c'est la fonction Bêta). Montrer que
$$\forall t>0,\quad \lim_{\varepsilon\to 0}\left(\frac{1}{\varepsilon}-\beta(\varepsilon,t-\varepsilon)\right)=-\Gamma'(1)+\frac{\Gamma'(t)}{\Gamma(t)}.$$
7. On rappelle la représentation intégrale de la fontion Bêta :
$$\beta(a,b)=\int_0^1\theta^{a-1}(1-\theta)^{b-1}\,d\theta.$$
Montrer que
$$\frac{\Gamma'(t)}{\Gamma(t)}-\frac{\Gamma'(1)}{\Gamma(1)}=\int_0^1\frac{1-\theta^{t-1}}{1-\theta}\,d\theta.$$
8. Montrer que $\frac{\Gamma'(t)}{\Gamma(t)}=\psi(t)$.

$\rightarrow$ indication

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 ω, $X(\omega) \in \{0,1\}$. 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

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

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 $\mathbb{R}$.*
— $\lim_{t\to-\infty} F_X(t) = 0$ *et* $\lim_{t\to+\infty} F_X(t) = 1$.
— *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 $s \le t$, on a $]-\infty, s] \subset]-\infty, t]$, d'où

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

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

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

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

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

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

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

— Soient $t \in \mathbb{R}$, $(t_n)_{n\ge1}$ une suite décroissante convergeant vers t. Posons, pour $n \ge 1$, $A_n =]-\infty, t_n]$. On a $A_{n+1} \subset A_n$ et $\bigcap_{n=1}^{+\infty} A_n =]-\infty, t]$, 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 $(H_i) \Longrightarrow (H_{i+1})$ 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.

Théorème de Helly

Théorème 5.6 (Théorème de Helly). *De toute suite $(F_n)_{n\geq 1}$ de fonctions de répartition, on peut extraire une sous-suite $(F_{n_k})_{k\geq 1}$ telle qu'il existe une fonction F croissante continue à droite avec*

$$F_{n_k}(x) \to F(x)$$

en chaque point de continuité de F.

Démonstration. À l'aide du procédé diagonal d'extraction, on commence par extraire une suite $(n_k)_{k\geq 1}$ telle que $F_{n_k}(x)$ converge en tout point x rationnel. On note $G(x)$ la limite obtenue. C'est une fonction définie sur $\mathbb{Q}$, et croissante. On définit alors

$$F(x) = \inf\{G(r); r \in \mathbb{Q}\cap]x, +\infty[\}.$$

Il est encore clair que F est croissante. Montrons que F est continue à droite. Soient $x \in \mathbb{R}$ et $\varepsilon > 0$. Par définition de F, il existe $r > x$, où r est un rationnel tel que $G(r) < F(x) + \varepsilon$. Maintenant, on a

$$\forall y \in [x, r[\quad F(x) \leq F(y) \leq G(r) < F(x) + \varepsilon,$$

ce qui montre bien que F est continue à droite. Reste à montrer que F_{n_k} converge vers F en chaque point de continuité de F. Soit x un point de continuité de F et soit $\varepsilon > 0$. On peut trouver η tel que $|F(x) - F(y)| \leq \varepsilon$ en tout y de $[x - \eta, x + \eta]$. Comme $\mathbb{Q}$ est dense dans $\mathbb{R}$, on peut trouver des rationnels r et s tels que $x - \eta \leq r \leq x \leq s \leq x + \eta$. On a pour tout $k \geq 1$:

$$F_{n_k}(r) \leq F_{n_k}(x).$$

On en déduit que

$$F(r) = \lim_{k\to+\infty} F_{n_k}(r) = \varliminf_{k\to+\infty} F_{n_k}(r) \leq \varliminf_{k\to+\infty} F_{n_k}(x),$$

ce qui implique que pour tout $\varepsilon > 0$

$$F(x) - \varepsilon \leq \varliminf_{k \to +\infty} F_{n_k}(x).$$

On obtient finalement que

$$F(x) \leq \varliminf_{k \to +\infty} F_{n_k}(x).$$

De la même manière, on montre que

$$\varlimsup_{k \to +\infty} F_{n_k}(x) \leq F(x).$$

Finalement, on a

$$F(x) \leq \varliminf_{k \to +\infty} F_{n_k}(x) \leq \varlimsup_{k \to +\infty} F_{n_k}(x) \leq F(x),$$

ce qui prouve bien que

$$\lim_{k \to +\infty} F_{n_k}(x) = F(x).$$

□

Remarque 5.7. *Attention, nous n'avons pas montré ici que* $\lim_{x \to +\infty} F(x) = 1$ *et donc F n'est pas forcément une fonction de répartition. (F est la « fonction de répartition » d'une mesure positive de masse totale inférieure ou égale à 1, qui n'est une probabilité que si la masse vaut exactement 1.) Voici d'ailleurs un exemple de suite de fonctions de répartitions dont la limite n'est pas une fonction de répartition. Il suffit de considérer la variable X_n de loi $\mathbb{P}_{X_n} = \delta_n$. On a alors $F_{X_n}(x) = \mathbb{1}_{\{x \geq n\}}$ qui ne converge pas vers une fonction de répartition.*

Exemple : l'ensemble triadique de Cantor

Cet exemple, difficile et à la frontière du thème de ce livre (mais c'est tout de même très joli), s'adresse au lecteur averti. Nous conseillons donc au lecteur peu familier avec la théorie de Lebesgue et les probabilités de ne pas lire cette section.

On part de $K_0 = [0, 1]$. On lui enlève le tiers médian ouvert pour fabriquer $K_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$. On refait le même procédé et on enlève à nouveau le tiers médian pour fabriquer $K_2 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right]$, et ainsi de suite. Plus formellement, on a la récurrence $K_{n+1} = f_1(K_n) \cup f_2(K_n)$, avec $f_1(x) = \frac{x}{3}$ et $f_2(x) = \frac{x+2}{3}$. À l'étape n, l'ensemble K_n est constitué de 2^n

intervalles, chacun de longueur 3^{-n}. L'ensemble de Cantor est $K = \bigcap_{n=0}^{+\infty} K_n$.

Notre but est de montrer que l'ensemble de Cantor est de mesure de Lebesgue nulle (alors qu'il n'est pas dénombrable). De plus, nous verrons que la probabilité associée à K n'est pas une mesure à densité (alors que sa fonction de répartition est presque partout dérivable). En effet, posons pour tout $n \geq 0$

$$F_n(x) = \begin{cases} 0 & \text{si} \quad x \leq 0, \\ \left(\frac{3}{2}\right)^n \int_0^x \mathbb{1}_{K_n}(t)\, d\lambda(t) & \text{si} \quad x \in [0,1]. \end{cases}$$

Nous allons montrer ici que $(F_n)_{n \geq 0}$ est une suite de fonctions de répartition continues, convergeant uniformément vers une fonction de répartition dérivable presque partout, mais qui n'est pas égale à l'intégrale de sa dérivée.

Tout d'abord, montrons que, pour tout $n \geq 0$, F_n est une fonction de répartition. On remarque facilement que $x \mapsto F_n(x)$ est croissante. On a de plus :

1. $\lim\limits_{x \to +\infty} F_n(x) = \left(\frac{3}{2}\right)^n \int_0^1 \mathbb{1}_{K_n}(t)\, d\lambda(t) = \left(\frac{3}{2}\right)^n 3^{-n} 2^n = 1$ car l'ensemble K_n est composé de 2^n intervalles de longueur 3^{-n}.

2. $\lim\limits_{x \to -\infty} F_n(x) = 0$ par définition de F_n.

De plus, par continuité de l'intégrale, F_n est continue. Donc F_n est bien une fonction de répartition continue.

Notons maintenant I un intervalle quelconque de K_n. Comme K_{n+1} est composé d'intervalles de longueur $3^{-(n+1)}$, I contient par définition le double d'intervalles de K_{n+1} que de ceux de K_n. Cela suffit à prouver l'égalité

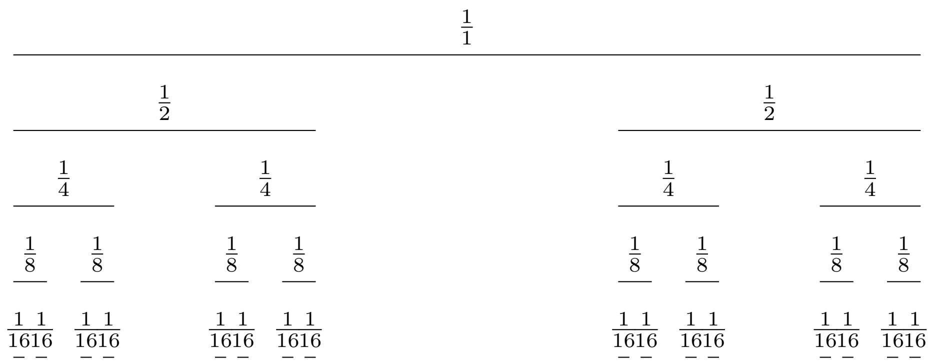
$$\int_I \left(\frac{3}{2}\right)^n \mathbb{1}_{K_n}(t)\, d\lambda(t) = \int_I \left(\frac{3}{2}\right)^{n+1} \mathbb{1}_{K_{n+1}}(t)\, d\lambda(t).$$

Prouvons maintenant que la mesure de Lebesgue de l'ensemble de Cantor est nulle (donc K est négligeable). Il est évident que $\lambda(K_0) = 1$ et on a pour tout $n \geq 0$, $\lambda(K_{n+1}) = \frac{2}{3}\lambda(K_n)$. Cela implique donc, par récurrence immédiate, que $\lambda(K_n) = \left(\frac{2}{3}\right)^n$. De plus, K est l'intersection d'ensembles décroissants et on obtient par le théorème de continuité séquentielle

$$\lambda(K) = \lambda\left(\bigcap_{n=0}^{+\infty} K_n\right) = \lim_{n \to +\infty} \lambda(K_n) = 0.$$

L'ensemble de Cantor K est bien négligeable pour la mesure de Lebesgue.

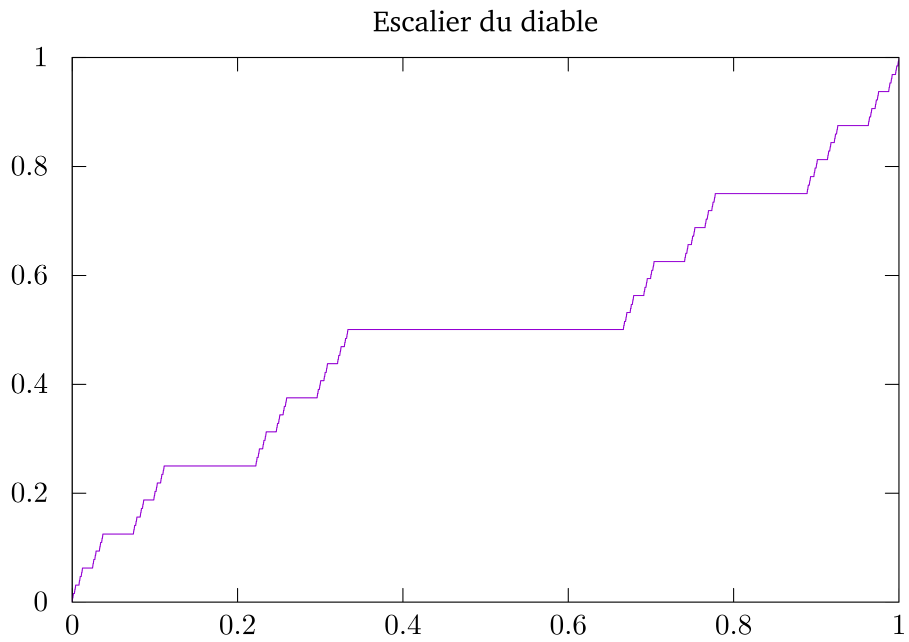
Voici le graphe de l'ensemble de Cantor, avec la répartition des masses de chaque segment :



Le passage de n à $n+1$ réorganise la répartition de la masse à l'intérieur des composantes de K_n. Aucune masse n'est rajoutée ailleurs puisque la suite (K_n) est décroissante. En conséquence, pour $x \in \overline{K_n^c}$ et $k \geq n$, on a $F_k(x) = F_n(x)$. Deux cas sont possibles, soit l'intégrale se découpe en somme d'intégrales sur des intervalles de K_n et alors la différence est nulle, soit la borne x est au "milieu d'un intervalle de K_n" (car si $x \notin K_n$, alors on est dans le premier cas). Dans ce dernier cas, comme $\int_I \left(\frac{3}{2}\right)^n \mathbb{1}_{K_n}(t)\, dt = 2^{-n}$ et que I contient deux intervalles de K_{n+1}, on trouve que la différence entre les deux fonctions de répartition est inférieure à $2 \cdot 2^{-n} = 2^{-(n-1)}$. Ainsi, on obtient par le critère de Cauchy que $(F_n)_{n \geq 1}$ converge uniformément vers une fonction F, qui est donc continue. Le graphe de F s'appelle "l'escalier du diable". Il s'agit d'une fonction continue croissante non-constante, mais presque partout constante. Voici ci-contre le graphe de F :

Montrons que F est constante par morceaux sur K^c. On sait déjà que pour tout $n \geq 1$, pour tout $x \in \overline{K_n^c}$ et $k \geq n$, on a $F_k(x) = F_n(x)$. En passant à la limite, cela montre bien que F est constante par morceaux. De plus, F est croissante car pour tout n, F_n est une fonction de répartition, donc croissante. Ainsi, F est dérivable presque partout et $F'(x) = 0$ presque partout. Pour conclure, remarquons que $F(1) - F(0) = \lim\limits_{n \to +\infty} (F_n(1) - F_n(0))$. Comme $F_n(1) = 1$ et $F_n(0) = 0$, on en déduit immédiatement que $F(1) - F(0) = 1$.

Finissons par montrer que la probabilité associée à F n'admet pas de densité. D'après le théorème de différentiation de Lebesgue (voir le théorème B.12 en annexe), s'il y avait une densité, elle coïnciderait en λ-presque tout x avec $\lim_{h \to 0} \frac{F(x+h)-F(x-h)}{2h}$, qui est presque partout nulle d'après ce qui précède. Nous aboutissons donc à une contradiction et on conclut que la probabilité associée à F n'est pas à densité.



Remarque 5.8. *On a donc exhibé ici un exemple de fonction presque partout dérivable, qui n'est pas égale à l'intégrale de sa dérivée, même si cette dernière est intégrable. Une telle fonction F est appelée fonction singulière.*

Remarque 5.9. *On peut exhiber simplement une variable aléatoire réelle X prise au hasard entre 0 et 1 dont la fonction de répartition est l'escalier de Cantor, encore appelé escalier du diable. Il suffit de tirer au hasard les chiffres successifs (0, 1 ou 2) du développement en base trois de X de manière un peu spéciale, à savoir par des tirages indépendants équiprobables restreints à 0 ou 2, le chiffre 1 étant exclu. Cela sera repris dans l'exercice non-corrigé 246 du chapitre 10.*

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

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

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

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

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 5.10. *De nombreux théorèmes et exercices de ce cours commenceront ainsi : "soit $(X_n)_{n\geq 1}$ 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 5.11. *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 5.12. *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 5.13. *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 5.12, ce qui achève la preuve. □

Corollaire 5.14. *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 5.15. *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. *$\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$*

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

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

Ainsi $\mathbb{P}_{(X_1, \dots, X_n)}$ et $\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$ 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 $A = A_1 \times \cdots \times A_n \in \mathcal{C}$. On a alors

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

ce qui prouve l'indépendance des tribus d'après le corollaire 5.14. □

Corollaire 5.16. *Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X_1, \ldots, X_n$ des vecteurs aléatoires. On suppose qu'il existe des mesures de probabilité $\mu_1, \ldots, \mu_n$ telles que $\mathbb{P}_{(X_1,\ldots,X_n)} = \mu_1 \otimes \cdots \otimes \mu_n$. Alors*

1. *$X_1, \ldots, X_n$ 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

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

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

5.2.3 Application : loi 0–1 de Kolmogorov

Théorème 5.17. *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 5.12, 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 5.18. *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 $\Lambda = \{0, \ldots, n-1\}$: 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 5.19. *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 5.20. *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 5.21. *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. □

Corollaire 5.22. *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

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

□

5.3.1 Fonction d'une variable aléatoire discrète

Théorème 5.23. *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 5.24. *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 5.25. *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*

$$\begin{array}{rcl} \pi : \Omega \times \Omega' & \to & \Omega \\ (x, x') & \mapsto x & \end{array}$$

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 5.26. *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 5.27. *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 5.28. *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 5.16 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 5.29. 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 5.30 (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 5.31. *$\mathcal{B}(n, p)$ charge les entiers $\{0, \ldots, n\}$. Plus précisément, on a pour tout $k \in \{0, \ldots, n\}$*

$$\mathcal{B}(n, p)(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}.$$

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 5.32. *Soient $A_1, \ldots, A_n$ 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 5.33. *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 5.34. *Soit $(A_i)_{i \in \mathbb{N}^*}$ une suite d'événements indépendants de même probabilité $p > 0$. On pose*

$$X(\omega) = \min\{k \in \mathbb{N}^* \quad \omega \in A_k\} \text{ si } \{k \in \mathbb{N}^*, \, \omega \in A_k\} \neq \varnothing,$$

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

$$\{X > k\} = \bigcap_{i \in \{1, \ldots, k\}} A_i^c.$$

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

$$\forall k \in \mathbb{N} \quad F_X(k) = \mathbb{P}(X \leq k) = 1 - \mathbb{P}(X > k) = 1 - (1-p)^k.$$

$\square$

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 5.35. *La loi hypergéométrique est la loi image de la loi uniforme sur $\Omega = \mathcal{B}_k(\{1, \dots, N\})$ par l'application*

$$\begin{array}{rcl} X : \mathcal{B}_k(\{1, \dots, N\}) & \to & \mathbb{N} \\ \omega & \mapsto & X(\omega) = |\{1, \dots, n\} \cap \omega|. \end{array}$$

Ainsi, pour $i \in \{0, \dots, \min(n, k)\}$, 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. □

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 5.36. *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 $\mathbb{P}(U < F(t)) \leq \mathbb{P}(F^*(U) \leq t) \leq \mathbb{P}(U < F(t'))$. On obtient ainsi $F(t) \leq \mathbb{P}(F^*(U) \leq t) \leq F(t')$. Cela est vrai pour tout $t' > t$. Comme F est continue à droite, on obtient ainsi le résultat en faisant tendre t' vers t. □

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 5.37. *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 5.38.

— *Si $a, b \in \overline{\mathbb{R}}$ sont tels que*

$$\lim_{x \to a^+} F(a) = 0 \text{ et } \lim_{x \to b^-} F(b) = 1$$

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 5.36 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

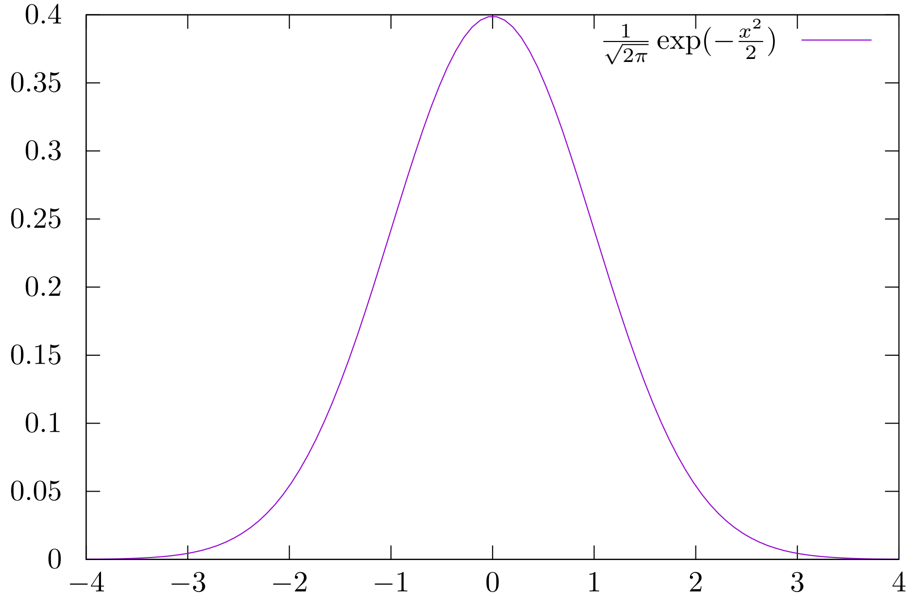
$$x \mapsto \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-m)^2}{2\sigma^2}).$$

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 $\sigma X + m \sim \mathcal{N}(m, \sigma^2)$.

Densité de la loi normale $\mathcal{N}(0, 1)$



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

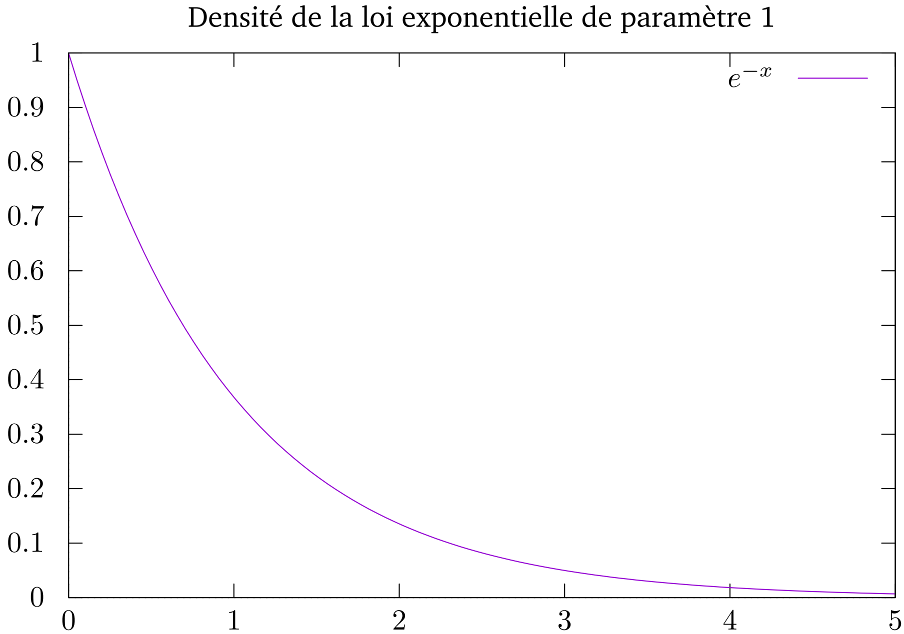
$$x \mapsto ae^{-ax} \, \mathbb{1}_{\mathbb{R}_+}(x).$$

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 :

$$x \mapsto \frac{1}{\pi} \frac{b}{(x-a)^2 + b^2}.$$

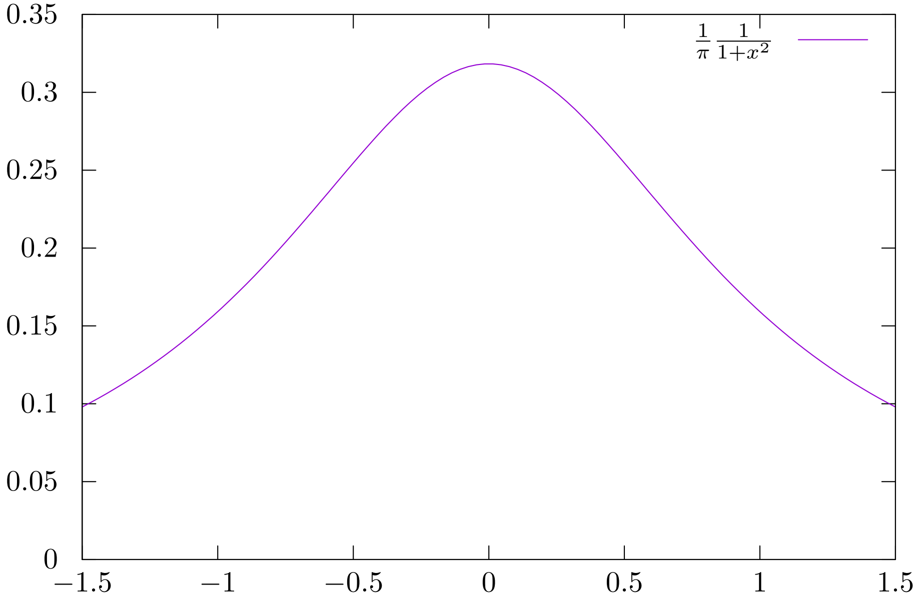
Exemple: Un gyrophare envoie un flash lumineux dans une direction aléatoire d'angle θ. On place un écran de longueur infinie à une distance 1 du gyrophare. On cherche la distribution de l'abscisse du point d'impact du rayon lumineux sur l'écran.
On sait que l'angle θ est une variable aléatoire uniformément 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

$$\begin{aligned} F(t) &= \mathbb{P}(\tan\theta \leq t) = \mathbb{P}(\theta \leq \arctan t) = \int_{-\infty}^{\arctan t} f_\theta(x)\,dx \\ &= \int_{-\infty}^{\arctan t} \frac{1}{\pi}\,\mathbb{1}_{[-\pi/2,\pi/2]}(x)\,dx = \frac{1}{\pi}\arctan t + \frac{1}{2}. \end{aligned}$$

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

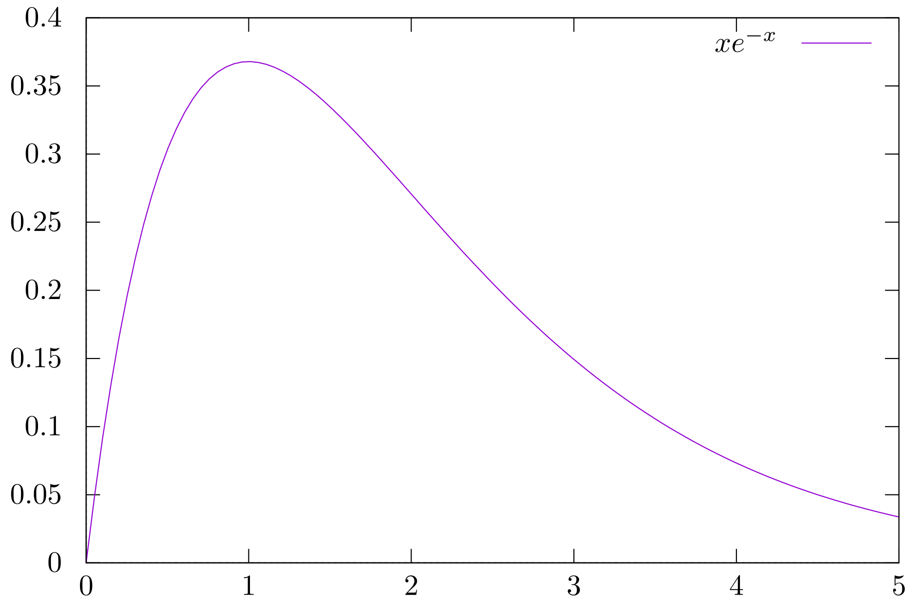
$$\Gamma(a) = \int_{\mathbb{R}+} x^{a-1} e^{-x}\,dx.$$

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 $\Gamma(2, 1)$



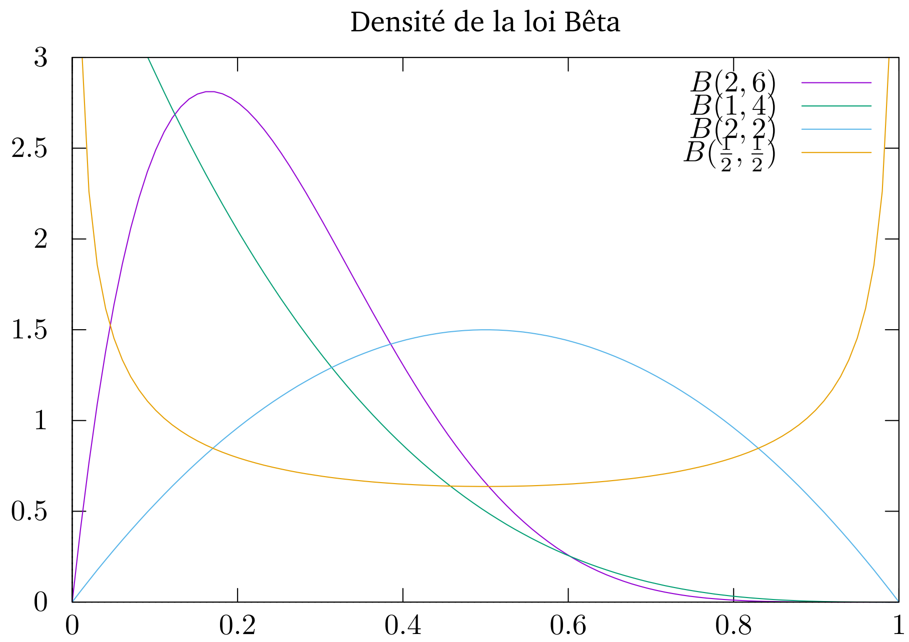
5.6.6 Loi Bêta

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

$$\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} \, dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$



5.6.7 Exemple

On lance une fléchette sur une cible circulaire de rayon 1. Le joueur est suffisamment loin de la cible pour que l'on suppose le point d'impact M de la fléchette uniformément distribué (on n'observe que les lancers qui atteignent la cible). Les coordonnées cartésiennes de M sont à valeurs dans le disque unité $B(0,1) = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$. On suppose qu'elles constituent un couple aléatoire $M = (X,Y)$ de loi uniforme sur le disque, soit de densité

$$f_{(X,Y)}(x,y) = \frac{1}{\pi}\, \mathbb{1}_{\{x^2+y^2\leq 1\}}.$$

Par la formule des marginales, on voit immédiatement que l'abscisse est distribuée selon la densité marginale

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x,y)\, dy = \frac{2}{\pi}\sqrt{1-x^2}\, \mathbb{1}_{[-1,1]}(x).$$

De même, on trouve que la densité de Y est $f_Y(y) = \frac{2}{\pi}\sqrt{1-y^2}\, \mathbb{1}_{[-1,1]}(y)$.
Il est naturel de paramétrer M par ses coordonnées polaires

$$R = \sqrt{X^2+Y^2} \in [0,1] \quad \text{et} \quad \Theta \in [0, 2\pi[, \text{ avec } \tan\Theta = \frac{Y}{X}.$$

Pour $0 \le r \le 1$ et $0 \le \theta < 2\pi$, on définit le domaine $C(r,\theta)$ par

$$\{R \le r,\ \Theta \le \theta\} = \{M \in C(r,\theta)\}.$$

On voit facilement que $C(r,\theta) = rC(1,\theta)$, d'où $\lambda^2(C(r,\theta)) = r^2\lambda^2(C(1,\theta))$. On a $\mathbb{P}(R \le r) = P(M \in C(r,2\pi)) = \frac{\lambda^2(C(r,2\pi))}{\lambda^2(B(0,1))} = \frac{\lambda^2(C(r,2\pi))}{\lambda^2(C(1,2\pi))} = r^2$. On a donc la fonction de répartition de R : on en déduit facilement que R admet la densité $f_R(r) = 2r\ \mathbb{1}_{[0,1]}(r)$. Par ailleurs, $\mathbb{P}(\Theta \le \theta) = \frac{\lambda^2(C(1,\theta))}{\lambda^2(B(0,1))} = \frac{\lambda^2(C(1,\theta))}{\pi}$, d'où

$$\mathbb{P}(R \le r,\ \Theta \le \theta) = \frac{\lambda^2(C(r,\theta))}{\lambda^2(B(0,1))} = \frac{r^2\lambda^2(C(1,\theta))}{\lambda^2(B(0,1))} = \mathbb{P}(R \le r)\mathbb{P}(\Theta \le \theta).$$

Ainsi, les variables aléatoires R et Θ sont indépendantes. Reste à calculer la loi de Θ, ou, de manière équivalente, à déterminer l'aire d'un secteur angulaire. Il est intuitif que l'angle représente la proportion du disque occupée par le secteur angulaire, soit $\lambda^2(C(1,\theta)) = \frac{\theta}{2\pi}\pi$. Montrons-le rigoureusement.

Posons $C'(\theta) = \{(x,y) \in \mathbb{R}^2 \backslash \{0,0\}; y \le x \tan\theta\}$. Le disque unité épointé en l'origine est réunion disjointe de n secteurs angulaires qui sont isométriques à $C'(2\pi/n)$. On a donc $\lambda^2(C'(2\pi/n)) = \frac{\pi}{n}$. Par des réunions disjointes de secteurs isométriques, on obtient $\lambda^2(C'(2k\pi/n)) = \frac{k\pi}{n}$, ou, de manière équivalente $\mathbb{P}(\Theta \in [0, 2k\pi/n[) = \frac{k\pi}{n}$.
Comme les ensembles de la forme $[0, 2k\pi/n[$ forment un Π-système qui engendre la tribu borélienne de $[0,2\pi]$, on peut conclure que Θ suit la loi uniforme sur $[0,2\pi]$ et que $\lambda^2(C(1,\theta)) = \pi\mathbb{P}(\Theta \le \theta) = \pi\frac{\theta}{2\pi} = \frac{\theta}{2}$.

5.6.8 Simulation par rejet

Reprenons l'exemple précédent. On se demande maintenant comment simuler une loi uniforme sur le disque unité $\{x^2 + y^2 \le 1\}$. L'idée ici est de tirer des points uniformément dans le carré unité, et de recommencer jusqu'à en obtenir un dans le disque. La loi ainsi obtenue est la loi d'un point tiré uniformément dans le carré conditionnellement à être dans le disque, ce qui est encore la loi uniforme sur le disque. Ce raisonnement se généralise de la façon suivante.

Proposition 5.39. *Soit f la densité d'une loi sur $\mathbb{R}^d$ que l'on souhaite simuler. On suppose qu'il existe $k > 0$ et une loi de densité g (sur $\mathbb{R}^d$) facile à simuler telles que*

$$\forall x \in \mathbb{R}^d \quad f(x) \le kg(x).$$

Soient U et Z deux variables aléatoires indépendantes. On suppose que U suit la loi uniforme sur $[0,1]$ et que Z est de densité g. On pose $V = kUg(Z)$. Alors, la loi *de Z conditionnellement à l'événement $\{V < f(Z)\}$ a pour densité f, c'est-à-dire que pour tout borélien A de $\mathbb{R}^d$, on a $\mathbb{P}(Z \in A | V < f(Z)) = \int_A f(z)\, d\lambda^d(z)$.*

Démonstration. Notons que, comme f et g sont des densités, on a nécessairement $k \geq 1$. Notons aussi que pour tout $z \in \mathbb{R}^d$, $\frac{f(z)}{kg(z)} \leq 1$. On a tout d'abord

$$\begin{aligned}\mathbb{P}(V < f(Z)) = \mathbb{P}(kUg(Z) < f(Z)) &= \int_{\mathbb{R}^d} g(z) \int_0^1 \mathbb{1}_{\{kug(z)<f(z)\}}\, du\, d\lambda^d(z) \\ &= \int_{\mathbb{R}} g(z) \frac{f(z)}{kg(z)}\, d\lambda^d(z) = \frac{1}{k}.\end{aligned}$$

De plus, pour $A \in \mathcal{B}(\mathbb{R}^d)$, on a

$$\begin{aligned}\mathbb{P}(\{Z \leq t\} \cap \{V < f(Z)\}) &= \int_{\mathbb{R}^d} \mathbb{1}_A(z) g(z) \int_0^1 \mathbb{1}_{\{kug(z)<f(z)\}} du\, d\lambda^d(z) \\ &= \int_{\mathbb{R}^d} \mathbb{1}_A(z) g(z) \frac{f(z)}{kg(z)}\, d\lambda^d(z) = \frac{1}{k} \int_A f(z)\, d\lambda^d(z).\end{aligned}$$

Donc $\mathbb{P}(Z \in A | V < f(Z)) = \int_A f(z)\, d\lambda^d(z)$, donc la loi conditionnelle de Z sachant que $\{V < f(Z)\}$ a bien pour densité f. □

Ceci permet de construire un algorithme de simulation par rejet.

Supposons ici qu'on a à disposition une fonction `simulg` qui simule une variable aléatoire de densité g, ainsi qu'une fonction `rand` qui simule une variable aléatoire suivant la loi uniforme sur $[0, 1]$.

En syntaxe Scilab, l'algorithme s'écrit alors comme suit :

```
function z=simulf //simule par rejet une va de densite f
  u=rand(); z=simulg; v=k*u*g(z);
  while (v>=f(z));  u=rand(); z=simulg; v=k*u*g(z); end;
```

Démontrons ce résultat. Notons N le nombre de tests faits par cette fonction. N est une variable aléatoire, à valeurs dans $\mathbb{N}^* \cup \{+\infty\}$. Notons $(U_n)_{n\geq 1}$ la suite des appels à la fonction `rand()`, et $(Z_n)_{n\geq 1}$ la suite des appels à la fonction `simulg`. Toutes ces variables aléatoires sont supposées indépendantes, les premières de loi uniforme sur $[0, 1]$, les secondes de densité g. On pose $V_n = kU_n g(Z_n)$ et on note X la sortie de la fonction, c'est-à-dire

$$X = V_N \text{ avec } N = \inf\{n \geq 1; V_n < f(Z_n)\}.$$

Par indépendance, N suit la loi géométrique de paramètre $1/k$, ce qui assure la terminaison presque sûre du programme.

Soit A un borélien de $\mathbb{R}^d$. On a alors par la démonstration précédente

$$\mathbb{P}(X \in A \text{ et } N = 1) \;=\; \mathbb{P}(V_1 < f(Z_1) \text{ et } Z_1 \in A) = \frac{1}{k} \int_A f(z)\, d\lambda^d(z).$$

Soit maintenant $i \geq 2$. Par indépendance, et comme précédemment :

$$\begin{aligned}
\mathbb{P}(X \in A \text{ et } N = i) &= \mathbb{P}\begin{pmatrix} V_1 \geq f(Z_1),\ \dots,\ V_{i-1} \geq f(Z_{i-1}) \\ V_i < f(Z_i),\ Z_i \in A \end{pmatrix} \\
&= \left(\prod_{j=1}^{i-1} \mathbb{P}(V_j \geq f(Z_j)) \right) \mathbb{P}(V_i < f(Z_i),\ Z_i \in A) \\
&= \left(1 - \frac{1}{k}\right)^{i-1} \frac{1}{k} \int_A f(z)\, d\lambda^d(z).
\end{aligned}$$

Finalement comme $k \geq 1$, on obtient

$$\begin{aligned}
\mathbb{P}(X \in A) &= \sum_{i=1}^{+\infty} \mathbb{P}(X \in A \text{ et } N = i) \\
&= \sum_{i=1}^{+\infty} \left(1 - \frac{1}{k}\right)^{i-1} \frac{1}{k} \int_A f(z)\, dz = \int_A f(z)\, d\lambda^d(z),
\end{aligned}$$

On conclut donc que la densité de X est bien f.

5.7 Loi 0–1 de Hewitt et Savage

Cette section peut être omise en première lecture.

5.7.1 Le théorème de Hewitt et Savage sur l'espace canonique

Dans cette section, on note Π_i la projection canonique de $\mathbb{R}^{\mathbb{N}}$ qui, à $\omega = (\omega_i)_{i \geq 0}$ associe $\Pi_i(\omega) = \omega_i$.

On note $\mathcal{B}$ la plus petite tribu sur $\mathbb{R}^{\mathbb{N}}$ qui rend mesurable les Π_i : $\mathcal{B} = \sigma((\Pi_n)_{n \geq 0})$. De manière équivalente c'est la tribu engendrée par les ensembles de la forme $\Pi_i^{-1}(A)$ où i décrit $\mathbb{N}$ et A l'ensemble des boréliens de $\mathbb{R}$.

Pour μ une probabilité quelconque sur $\mathbb{R}$, on admet l'existence d'une mesure de probabilité $\mathbb{P}$ sur $\mathcal{B}$ telle que sous la probabilité $\mathbb{P}$, les Π_i sont des variables indépendantes de loi μ. On note $\mu^{\otimes \mathbb{N}}$ cette loi.

Soit τ une permutation de $\mathbb{N}$. On appelle support de τ l'ensemble des $i \in \mathbb{N}$ tels que $\tau(i) \neq i$. On note Σ^* l'ensemble des permutations à support fini. Pour une permutation à support fini τ, on note T_τ l'application de $\mathbb{R}^{\mathbb{N}}$ dans lui-même qui à x associe $x \circ \tau$.

Notons que pour tout i dans $\mathbb{N}$ et tout A, on a

$$T_\tau^{-1}(\Pi_i^{-1}(A)) = (\Pi_i \circ T_\tau)^{-1}(A) = (\Pi_{\tau(i)})^{-1}(A) \in \mathcal{B}.$$

Ainsi T_τ est $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}) - (\mathbb{R}^{\mathbb{N}}, \mathcal{B})$ mesurable.

On définit
$$\mathcal{I}_\Sigma = \{A \in \mathcal{B} : \ \forall \tau \in \Sigma^* \quad T_\tau^{-1}(A) = A\}.$$

Théorème 5.40. *Soit μ une probabilité quelconque sur $\mathbb{R}$. Posons $\mathbb{P} = \mu^{\otimes \mathbb{N}}$. Alors pour tout $A \in \mathcal{I}_\Sigma$, on a $\mathbb{P}(A) = 0$ ou 1.*

Démonstration. Soit $A \in \mathcal{I}_\Sigma$. Soit $\varepsilon > 0$. D'après le théorème de Hahn, il existe une suite $(A_n)_{n\geq 1}$ de parties vérifiant, pour tout n, $A_n \in \sigma(\Pi_0, \dots, \Pi_{n-1})$ et telle que $A \subset \tilde{A} = \bigcup\limits_{n=1}^{+\infty} A_n$ et $\mathbb{P}(\tilde{A}) \leq \mathbb{P}(A) + \varepsilon$. Notons τ_n la permutation définie par $\tau_n(i) = 2n - 1 - i$ si $0 \leq i < 2n$ et $\tau_n(i) = i$ pour $i \geq 2n$. Il est aisé de voir que T_{τ_n} préserve $\mathbb{P}$ et que A_n et $T_{\tau_n}^{-1}(A_n)$ sont indépendantes : on a
$$\mathbb{P}(A_n \cap T_{\tau_n}^{-1}(A_n)) = \mathbb{P}(A_n)\mathbb{P}(T_{\tau_n}^{-1}(A_n)) = \mathbb{P}(A_n)^2.$$
Comme $A \subset \tilde{A}$, on trouve donc $T_{\tau_n}^{-1}(A) \subset T_{\tau_n}^{-1}(\tilde{A})$, soit $A \subset T_{\tau_n}^{-1}(\tilde{A})$. Finalement $A \subset \tilde{A} \cap T_{\tau_n}^{-1}(\tilde{A})$, d'où
$$\mathbb{P}(A) \leq \mathbb{P}(\tilde{A} \cap T_{\tau_n}^{-1}(\tilde{A})) = \sup_{n\geq 1} \mathbb{P}(A_n \cap T_{\tau_n}^{-1}(A_n)) = \sup_{n\geq 1} \mathbb{P}(A_n)^2 \leq (\mathbb{P}(A) + \varepsilon)^2.$$
Comme ε peut être pris arbitrairement petit, cela donne $\mathbb{P}(A) \leq \mathbb{P}(A)^2$, d'où $\mathbb{P}(A) \in \{0, 1\}$. □

Remarque 5.41. *Il est possible d'utiliser le théorème 2.21 à la place du théorème de prolongement de Hahn.*

Corollaire 5.42. *Soit $(\Pi_n)_{n\geq 0}$ une suite de variables aléatoires indépendantes identiquement distribuées de loi μ. On pose pour tout k,*
$S_k = \Pi_0 + \Pi_1 + \dots + \Pi_k$ et $\mathcal{Q} = \bigcap\limits_{n\geq 1} \sigma((S_i)_{i\geq n})$. Alors on a pour tout $A \in \mathcal{Q}$, $\mathbb{P}(A) = 0$ ou 1.

Démonstration. Soient $A \in \mathcal{Q}$ et $\tau \in \Sigma^*$. Soit n tel que le support de τ est inclus dans $\{0, \dots n\}$. On a $S_n(T_\sigma(\omega)) = S_n(\omega)$ (les termes de la somme sont juste permutés), d'où $S_k(T_\tau(\omega)) = S_k(\omega)$ pour $k \geq n$. Soit $k \geq n$. Tout élément B de $\sigma(S_k)$ s'écrit $S_k^{-1}(B')$, où B' est un borélien de $\mathbb{R}$. Comme $S_k \circ T_\tau = S_k$, on a
$$T_\tau^{-1}(B) = T_\tau^{-1}(S_k^{-1}(B')) = (S_k \circ T_\tau)^{-1}(B') = S_k^{-1}(B') = B,$$
ce qui montre que pour $k \geq n$, $\sigma(S_k)$ est incluse dans la tribu des événements invariants par T_τ. Par suite, $\sigma((S_k)_{k\geq n})$ est incluse dans la tribu des événements invariants par T_τ. Comme $A \in \sigma((S_k)_{k\geq n})$, A est invariant par T_τ. Comme on a choisi τ quelconque, on a $A \in \mathcal{I}_\Sigma$, ce qui montre que $\mathbb{P}(A) \in \{0, 1\}$ d'après la loi du 0–1 de Hewitt et Savage. □

5.7.2 Loi d'un processus

Dans le paragraphe précédent, nous avons travaillé sur un espace probabilisé abstrait un peu spécial : $\mathbb{R}^{\mathbb{N}}$ que l'on appelle parfois l'espace canonique ou espace des trajectoires. Nous allons voir maintenant comment relier cet espace abstrait aux suites de variables aléatoires que l'on rencontre couramment.

Théorème 5.43. *Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$. L'application $X : \omega \mapsto (X_i(\omega))_{i\geq 0}$ est une application $(\Omega, \mathcal{F}) - (\mathbb{R}^{\mathbb{N}}, \mathcal{B})$ mesurable. La loi image $\mathbb{P}_X$ de $\mathbb{P}$ par X est appelée loi de la suite (ou du processus) $(X_n)_{n\geq 0}$*

Démonstration. Comme les ensembles de la forme $\Pi_i^{-1}(A)$, avec A borélien, engendrent $\mathcal{B}$, il suffit de montrer que pour B de la forme $B = \Pi_i^{-1}(A)$, avec A borélien, on a $X^{-1}(B) \in \mathcal{F}$. Or

$$X^{-1}(B) = X^{-1}(\Pi_i^{-1}(A)) = (\Pi_i \circ X)^{-1}(A) = X_i^{-1}(A),$$

qui est bien dans $\mathcal{F}$ puisque X_i est une variable aléatoire. □

Théorème 5.44. *Soient $(X_n)_{n\geq 0}$ des variables aléatoires définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ et $(X'_n)_{n\geq 0}$ des variables aléatoires définies sur $(\Omega', \mathcal{F}', \mathbb{P}')$. On suppose que pour tout n, la loi de $(X_0, \dots, X_n)$ sous $\mathbb{P}$ est la même que la loi de $(X'_0, \dots, X'_n)$ sous $\mathbb{P}'$. Alors $(X_n)_{n\geq 0}$ et $(X'_n)_{n\geq 0}$ ont même loi.*

Démonstration. On veut identifier $\mathbb{P}_X$ et $\mathbb{P}'_{X'}$. Or pour identifier deux probabilités, il suffit de les identifier sur un π-système qui engendre la tribu (ici $\mathcal{B}$). Or les ensembles qui peuvent s'écrire sous la forme

$$\bigcap_{i=0}^{n} \Pi_i^{-1}(A_i)$$

pour un certain n et une certaine suite $(A_i)_{0\leq i\leq n}$ de boréliens forment manifestement un tel π-système. Comme

$$\begin{aligned}\mathbb{P}_X(\bigcap_{i=0}^{n} \Pi_i^{-1}(A_i)) &= \mathbb{P}(X^{-1}(\bigcap_{i=0}^{n} \Pi_i^{-1}(A_i)) \\ &= \mathbb{P}(\bigcap_{i=0}^{n} \{X_i \in A_i) = \mathbb{P}_{(X_0,\dots,X_n)}(A_0 \times \cdots \times A_n),\end{aligned}$$

on a alors le résultat voulu. □

Corollaire 5.45. *Soient $(X_n)_{n\geq 0}$ des variables aléatoires indépendantes identiquement distribuées de loi commune μ. Alors $\mathbb{P}_X = \mu^{\otimes\mathbb{N}}$.*

Démonstration. Il suffit de noter que $\mathbb{P}_{(X_0,\dots,X_n)} = \mu^{\otimes(n+1)} = \mu^{\otimes\mathbb{N}}_{(\Pi_0,\dots,\Pi_n)}$ par définition de $\mu^{\otimes\mathbb{N}}$. □

Exemple : Soient $(X_n)_{n\geq 0}$ des variables aléatoires indépendantes suivant la loi de Bernoulli de paramètre p. Alors quel que soit k entier naturel, les variables

$$Y_0 = \sum_{n=0}^{+\infty} \frac{X_n}{2^n} \text{ et } Y_k = \sum_{n=0}^{+\infty} \frac{X_{k+n}}{2^n}$$

ont même loi. Le résultat est très naturel, mais sa preuve rigoureuse est plutôt difficile sans l'aide du formalisme que nous avons introduit. Avec ce formalisme, la preuve est très simple : définissons

$$\begin{array}{rcl} f : [0,1]^{\mathbb{N}} & \to & \mathbb{R} \\ (x_n)_{n\geq 0} & \mapsto & \displaystyle\sum_{n=0}^{+\infty} \frac{x_n}{2^n} \end{array}$$

f est $([0,1]^{\mathbb{N}}, \mathcal{F}) - (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable comme limite d'applications mesurables. On a $Y_k = f((X_{k+n})_{n\geq 0})$ donc $\mathbb{P}_{Y_k}$ est la loi image de la loi du processus $(X_{k+n})_{n\geq 0}$ par f : comme la loi du processus $(X_{k+n})_{n\geq 0}$ est indépendante de k (c'est $\mathrm{Ber}(p)^{\otimes\mathbb{N}}$), on a le résultat voulu.

5.8 Exercices sur les lois

5.8.1 Exercices corrigés

Exercice 97. 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.

→ indication → solution

Exercice 98. 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 $\mathbb{P}\left(\frac{1}{2} < X < \frac{3}{2}\right)$. → indication → solution

Exercice 99. Soient $F_j : \mathbb{R} \longrightarrow \mathbb{R}$, $j = 1, 2$, des fonctions données par

$$\begin{array}{rcl} F_1(x) & := & \displaystyle\sum_{i=1}^{+\infty} 2^{-i} \mathbb{1}_{[\frac{1}{i},\infty[} \\ F_2(x) & := & e^{-e^{-x}}. \end{array}$$

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 :

$$[1, +\infty[\,;\ \left[\frac{1}{10}, +\infty\right[\,;\ \{0\}\,;\ \left[0, \frac{1}{2}\right[\,;\ ]-\infty, 0[\,;\ ]0, +\infty[\,.$$

$\rightarrow$ indication $\rightarrow$ solution

Exercice 100. 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 $\left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$.
$\rightarrow$ indication $\rightarrow$ solution

Exercice 101. Trouver la constante c telle que la fonction cx^{-3} soit une densité sur l'intervalle $[1, +\infty[$ puis sur l'intervalle $[-2, -1[$.
$\rightarrow$ indication $\rightarrow$ solution

Exercice 102. Soit X une variable aléatoire uniforme sur l'intervalle [1,3]. Chercher la loi de la variable aléatoire $Y = \frac{1}{2}(|X-1| + |X|)$. $\rightarrow$ indication $\rightarrow$ solution

Exercice 103. 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 ?
$\rightarrow$ indication $\rightarrow$ solution

Exercice 104. 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 $\mathbb{P}(S_{ABC} > 2S_{XBY})$.
$\rightarrow$ indication $\rightarrow$ solution

Exercice 105. 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).
$\rightarrow$ indication $\rightarrow$ solution

Exercice 106. 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 $g(1)$.

→ indication → solution

Exercice 107. 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. → indication → solution

Exercice 108. 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 $\mathbb{P}(1/2 < X^2 + Y^2 < 1)$.

→ indication → solution

Exercice 109. 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 $\mathbb{P}(X > Y)$. → indication → solution

Exercice 110. Soient $(X_1, \dots, X_n)$ des variables aléatoires indépendantes telles que pour tout i entre 1 et n, X_i suit la loi exponentielle $\mathcal{E}(\lambda_i)$.
On note $T = \inf(X_1, \dots, X_n)$. Pour t réel, calculer $\mathbb{P}(T > t)$. En déduire que T suit une loi exponentielle de paramètre $\lambda_1 + \dots + \lambda_n$.
Pour aller plus loin, on pourra se reporter à l'exercice 127

→ indication → solution

Exercice 111.

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 $\mathbb{P}(X = a) = 1$.

→ indication → solution

Exercice 112. 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. → indication → solution

Exercice 113. 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 $\mathbb{P}(X \wedge X' = 1) = \frac{1}{\zeta(s+t)}$.

$\to$ indication $\to$ solution

5.8.2 Exercices non corrigés

Exercice 114. 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 $p \le q$. 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

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

$\to$ indication

Exercice 115. 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 $1+\nu_p(X)\sim\mathcal{G}(1-\frac{1}{p^s})$. $\to$ indication

Exercice 116. 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.

$\to$ indication

Exercice 117. Donner un exemple de deux lois distinctes sur $(\Omega,\mathcal{F})$ coïncidant sur un système $\mathcal{C}$ engendrant $\mathcal{F}$.

$\to$ indication

Exercice 118. 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 ? $\to$ indication

Exercice 119. Dans le segment $[AB]$ de longueur 1, on choisit au hasard un point M. Quelle est la probabilité pour que l'on ait $AM.MB\ge\frac{2}{9}$?

$\to$ indication

Exercice 120. Soit $(X_i)_{i\ge1}$ des variables aléatoires indépendantes de loi uniforme sur $[0,1]$. On pose $M_n=\max(X_1,\dots,X_n)$ et $m_n=\min(X_1,\dots,X_n)$.

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.

$\to$ indication

Exercice 121. Soit X une variable aléatoire suivant la loi $\mathcal{B}(2n,1/2)$. On pose $Y=|X-n|$. Déterminer la loi de Y.

$\to$ indication

Exercice 122. Soit (X,Y) un vecteur aléatoire suivant la loi uniforme sur le rectangle $[-1,2]\times[-1,1]$. Montrer que

$$\mathbb{P}(1-Y\ge 2|X|)=\frac{1}{3}.$$

$\to$ indication

Exercice 123. Pour n entier strictement positif, on note $A_n=n\mathbb{N}^*$. Notons $\mathcal{P}$ l'ensemble des nombres premiers positifs et $\mathcal{T}$ la sous-tribu de $(\mathbb{N}^*,\mathcal{B}(\mathbb{N}^*))$ engendrée par les $(A_p)_{p\in\mathcal{P}}$. Pour $\omega\in\mathbb{N}^*$, on pose

$$F(\omega)=\prod_{p\in\mathcal{P};d\text{ divise }\omega} p.$$

1. Montrer que $\mathcal{T} = \sigma(X)$.
2. Déterminer le plus petit ensemble $\mathcal{T}$-mesurable contenant 1980.
3. Montrer que A_n est $\mathcal{T}$-mesurable si et seulement si n n'est divisible par aucun carré.
4. On munit $(\mathbb{N}^*, \mathcal{B}(\mathbb{N}^*))$ de la mesure de probabilité ζ de paramètre s, c'est-à-dire que
$$\forall n \in \mathbb{N}^* \quad \mathbb{P}(\{n\}) = \frac{1}{\zeta(s)} \frac{1}{n^s},$$
où l'on a posé
$$\zeta(s) = \sum_{n=1}^{+\infty} \frac{1}{n^s}.$$
Montrer que $\mathbb{P}(A_n) = \frac{1}{n^s}$. À quelle condition les événements A_n et A_m sont-ils indépendants sous la loi $\mathbb{P}$?
5. Soit $\mathcal{N} = \{\omega \in \mathbb{N}^*; \quad A_\omega \text{ est } \mathcal{T} - \text{mesurable}\}$.
Montrer que $\mathcal{N} = \bigcap_{p \in \mathcal{P}} A^c_{p^2}$, puis que $\mathbb{P}(\mathcal{N}) > 0$.

$\rightarrow$ indication

Exercice 124. *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
$$\Psi(x) \sim \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-\frac{x^2}{2}}.$$
$\rightarrow$ indication

Exercice 125. *Encore une loi* 0–1.

1. Soit $(\mathcal{A}_n)_{n\geq 1}$ une suite de tribus indépendantes. Montrer que la tribu $\mathcal{Q} = \bigcap_{k=1}^{+\infty} \sigma(\mathcal{A}_i; i \geq k)$ 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.

$\rightarrow$ indication

Exercice 126. 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.

$\rightarrow$ indication

Exercice 127. *Première sonnerie.*

Soient $(X_1, \dots X_n)$ des variables aléatoires indépendantes telles que pour tout i compris entre 1 et n, X_i suit la loi exponentielle $\mathcal{E}(\lambda_i)$.

On note $T = \inf(X_1, \dots, X_n)$ et $N = \inf\{i \geq 1; X_i = m\}$.

1. Montrer que $\mathbb{P}(\exists (i,j) \in \mathbb{N}^2 \quad 1 \leq i < j \leq n; X_i = X_j) = 0$.
2. Montrer que pour tout i compris entre 1 et n

$$\mathbb{P}(T > t, N = i) = \mathbb{P}\left(\bigcap_{\substack{1 \leq j \leq n \\ j \neq i}} \{t < X_i < X_j\} \right)$$

$$= \int_{\mathbb{R}^n} \mathbb{1}_{]t,+\infty[}(x_i) \prod_{\substack{1 \leq j \leq n \\ j \neq i}} \mathbb{1}_{]x_i,+\infty[}(x_j) \prod_{j=1}^{n} \lambda_j e^{-\lambda_j x_j} \, d\lambda^{\otimes n}(x_1, \dots, x_n)$$

3. On pose $\lambda = \sum_{j=1}^{n} \lambda_j$. Montrer que $\mathbb{P}(T > t, N = i) = \frac{\lambda_i}{\lambda} \exp(-\lambda t)$.
4. Montrer que T et N sont indépendantes et préciser leurs lois.

$\rightarrow$ indication

Exercice 128. 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.

$\rightarrow$ indication

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 $X = (X_1, \dots, X_n)$ est un vecteur aléatoire dont toutes les composantes sont intégrables, on note $\mathbb{E}X$ le vecteur $(\mathbb{E}X_1, \dots, \mathbb{E}X_n)$.*

6.2 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.
— $\forall X \in \mathcal{L}^1, \quad \mathbb{P}(X \le a) = 1 \Longrightarrow \mathbb{E}X \le a$.
— $\forall X \in \mathcal{L}^1, \quad \mathbb{P}(X \ge b) = 1 \Longrightarrow \mathbb{E}X \ge b$.

— $\forall X \in \mathcal{L}^1, \quad \mathbb{P}(X = a) = 1 \Longrightarrow \mathbb{E}X = a.$
— $\forall X \in \mathcal{L}^1, \quad \mathbb{P}(|X| \leq a) = 1 \Longrightarrow \mathbb{E}|X| \leq a.$
— Soient X, Y deux variables aléatoires vérifiant $0 \leq X \leq Y$. Si Y est intégrable, alors X est intégrable.
— $\forall X, Y \in \mathcal{L}^1, \quad \mathbb{P}(X \leq Y) = 1 \Longrightarrow \mathbb{E}X \leq \mathbb{E}Y.$

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 $\mathbb{P}$*

$$\begin{aligned}
\mathbb{P}\Big(\bigcup_{i=1}^{n} A_i\Big) &= \sum_{B \in \mathcal{P}(\{1,\dots,n\}) \setminus \varnothing} (-1)^{1+|B|} \mathbb{P}(\cap_{j \in B} A_j) \qquad (6.1)\\
&= \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2}) + \dots + \\
&\quad + \dots + (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) + \dots + \\
&\quad + \dots + (-1)^{n+1} \mathbb{P}(A_1 \cap \dots \cap A_n).
\end{aligned}$$

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

$$\forall n \in \mathbb{N}^*, \quad \mathbb{P}\left(\bigcup_{x\in I} A_x\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} \mathbb{P}\left(\bigcap_{x\in J} A_x\right) + (-1)^n \mathbb{E} V_n. \tag{6.2}$$

Démonstration. Il suffit de montrer que

$$\forall n \in \mathbb{N}^* \quad \mathbb{1}_{\bigcup\limits_{x\in I} A_x} = \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} \prod_{x\in J} \mathbb{1}_{A_x} + (-1)^n V_n. \tag{6.3}$$

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 $\prod\limits_{x\in J} \mathbb{1}_{A_x} = 1$ si et seulement si $J \subset \{x \in I; \omega \in A_x\}$.

Ainsi $\sum\limits_{J\in\mathcal{B}_k(I)} \prod\limits_{x\in J} \mathbb{1}_{A_x} = P_k(N)$. On doit donc montrer 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

$$\begin{aligned}
\sum_{k=0}^{n+1}(-1)^k P_k(X) &= \sum_{k=0}^{n}(-1)^k P_k(X) + (-1)^{n+1}P_{n+1}(X)\\
&= (-1)^n P_n(X-1) + (-1)^{n+1}P_{n+1}(X)\\
&= (-1)^{n+1}(P_{n+1}(X) - P_n(X-1))\\
&= (-1)^{n+1}(\frac{1}{n+1}XP_n(X-1) - P_n(X-1))\\
&= (-1)^{n+1}\frac{(X-(n+1))P_n(X-1)}{n+1}\\
&= (-1)^{n+1}P_{n+1}(X-1).
\end{aligned}$$

Méthode 2 : De manière équivalente, il faut montrer que

$$\sum_{k=0}^{n}(-1)^{n-k} P_k(N) = P_n(N-1).$$

Mais on reconnaît en $\sum_{k=0}^{n}(-1)^{n-k}P_k(N)$ le coefficient en x^n de la série entière sur $B(0,1)$:

$$\begin{aligned}
\left(\sum_{k=0}^{+\infty}P_k(N)x^k\right)\left(\sum_{k=0}^{+\infty}(-1)^k x^k\right) &= \left(\sum_{k=0}^{N}\binom{N}{k}x^k\right)\frac{1}{1+x}\\
&= (1+x)^N\frac{1}{1+x} = (1+x)^{N-1}
\end{aligned}$$

dont le coefficient en x^n est précisément $P_n(N-1)$, d'où l'identification des termes.

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

$$\mathbb{P}(\underset{x\in I}{\cup} A_x) = \sum_{k=1}^{|I|}(-1)^{k+1}\sum_{J\in\mathcal{B}_k(I)}\mathbb{P}(\underset{x\in J}{\cap} A_x),$$

car V_n est identiquement nulle. Cela démontre la formule de Poincaré. □

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*

$$\mathbb{P}(\bigcup_{x\in I} A_x) \leq \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} \mathbb{P}(\bigcap_{x\in J} A_x). \qquad (6.4)$$

— *Si n est pair, on a alors*

$$\mathbb{P}(\bigcup_{x\in I} A_x) \geq \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} \mathbb{P}(\bigcap_{x\in J} A_x). \qquad (6.5)$$

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

$$\int_{\Omega} g(X(\omega)) \, d\mathbb{P}(\omega) = \int_{\mathbb{R}} g(x) \, d\mathbb{P}_X(x) = \int_{\mathbb{R}} g(x) f(x) \, dm(x).$$

$\square$

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 :

$$f(x) = \begin{cases} \mathbb{P}(X = x) & \text{si } x \in D \\ 0 \quad \text{si } x \notin D. \end{cases}$$

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. □

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. □

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 $\operatorname{Covar}(X, Y) = \langle L(X), L(Y)\rangle$. Les deux observations faites ci-dessus permettent de dire que $(X, Y) \mapsto \operatorname{Covar}(X, Y)$ est une forme bilinéaire symétrique positive.
2. On a
$$\begin{aligned}\operatorname{Covar}(X-a, Y-b) &= \langle L(X-a), L(Y-b)\rangle \\ &= \langle L(X) - L(a), L(Y) - L(b)\rangle \\ &= \langle L(X) - 0, L(Y) - 0\rangle \\ &= \operatorname{Covar}(X, Y).\end{aligned}$$
3. On a
$$\begin{aligned}\operatorname{Var}(X+Y) &= \operatorname{Covar}(X+Y, X+Y) \\ &= \operatorname{Covar}(X, X) + 2\operatorname{Covar}(X, Y) + \operatorname{Covar}(Y, Y) \\ &= \operatorname{Var} X + 2\operatorname{Covar}(X, Y) + \operatorname{Var} Y.\end{aligned}$$
Pour passer de la ligne 1 à la ligne 2, on utilise le fait que la covariance est bilinéaire symétrique.
4. 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* $\operatorname{Var} X = 0$ *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*
$$\rho(X, Y) = \frac{\operatorname{Covar}(X, Y)}{\sigma(X)\sigma(Y)}.$$

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 $(\mathrm{Covar}(X_i, X_j))_{1 \leq i,j \leq n}$.

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

$\operatorname{Covar}(\langle X, e_i\rangle, \langle X, e_j\rangle) = \operatorname{Covar}(X_i, X_j)$. 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

$$\begin{aligned} \operatorname{Covar}(\langle Y, a\rangle, \langle Y, c\rangle) &= \operatorname{Covar}(\langle AX + b, a\rangle, \langle AX + b, c\rangle) \\ &= \operatorname{Covar}(\langle AX, a\rangle, \langle AX, c\rangle) \\ &= \operatorname{Covar}(\langle X, A^*a\rangle, \langle X, A^*c\rangle) \\ &= \langle C_XA^*a, A^*c\rangle \\ &= \langle AC_XA^*a, c\rangle. \end{aligned}$$

□

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 4.44. □

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 $g(z,t) = h(z-t, t) = f(z-t)g(t)$.

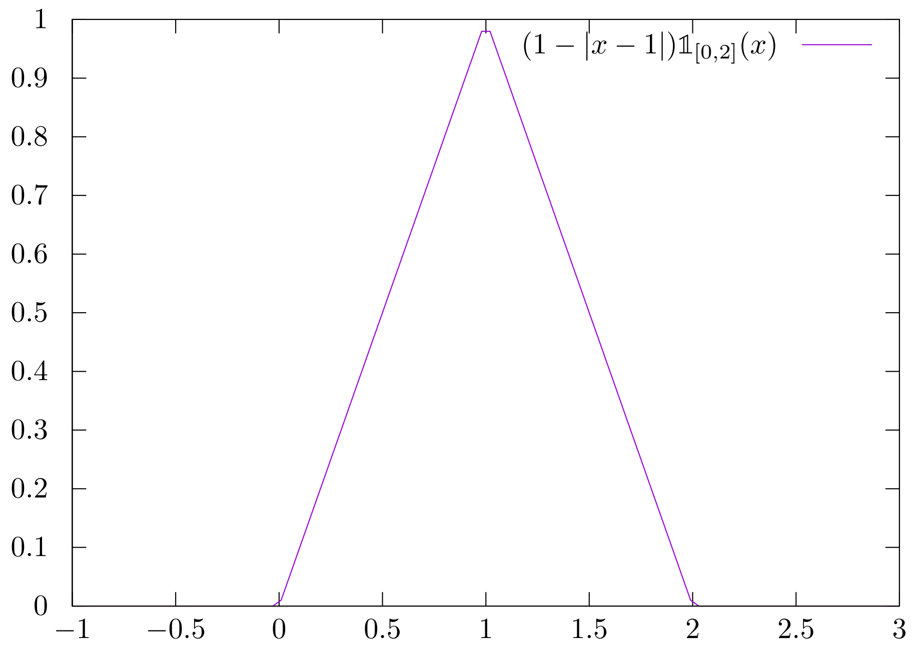
D'après le théorème 5.25, Z admet comme densité par rapport à la mesure de Lebesgue :

$$z \mapsto \int_{\mathbb{R}} g(z,t)\ d\lambda(t) = \int_{\mathbb{R}} f(z-t)g(t)\ d\lambda(t).$$

Dans le cas où X et Y sont à valeurs positives, il suffit de remarquer que $f(z-t)$ est nul si z dépasse t et que $g(t)$ est nul si t est négatif. Ainsi, $f(z-t)g(t)$ ne peut être non nul que pour z vérifiant $0 \le t \le z$, ce qui n'est évidemment jamais vérifié si z est négatif.

On obtient les deux autres égalités en échangeant les rôles de X et Y. □

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 Loi image par un C^1-difféomorphisme

Le premier réflexe à avoir est d'utiliser le corollaire 4.47, qui est bien sûr très utile. Le paragraphe suivant décrit des stratégies lorsque l'application n'est pas immédiate.

6.9.1 Compléments méthodologiques

Cette sous-section peut être omise en première lecture.

Il arrive que l'on ait besoin de calculer l'image d'une mesure de probabilité μ à densité par une transformation qui est C^1 (ou presque partout C^1), mais qui ne réalise pas un C^1-difféomorphisme de l'ensemble de départ vers l'ensemble d'arrivée. Il y a deux raisons qui peuvent empêcher une application T de classe C^1 de réaliser un C^1-difféomorphisme entre l'ensemble de départ O et l'ensemble d'arrivée :

— il existe $x \in O$ tel que D_xT est non inversible.
— T est non injectif.

Dans tous les cas, on commence par enlever les points où T n'est pas C^1 (si du moins cet ensemble est de mesure nulle, sinon ce texte ne peut pas grand chose pour vous). Ainsi pour une loi à densité sur $\mathbb{R}^2$ la transformation

$$(x,y) \mapsto (\min(x,y), |x-y|)$$

est tout à fait recevable puisqu'elle est C^1 sur l'ouvert $\mathbb{R}^2 \backslash \{(x,x); x \in \mathbb{R}^2\}$.

— Le premier écueil n'est pas très dangereux : en effet, dans la plupart des cas $\{x \in O; \det D_xT = 0\}$ est de mesure nulle.
Il suffit alors de remplacer O par $O\backslash\{x \in O; \det D_xT = 0\}$ (en effet $\{x \in O; \det D_xT = 0\}$ est un fermé donc $O\backslash\{x \in O; \det D_xT = 0\}$ est bien un ouvert), et on peut appliquer le théorème du cours.
  — Un exemple dans $\mathbb{R}$
  Soit $X \sim \mathcal{N}(0,1)$. On veut calculer la loi de $Y = X^3$.
  Ici, on a évidemment $O = \mathbb{R}$ et $T(x) = x^3$. Un calcul immédiat nous donne $\{x \in O; \det D_xT = 0\} = \{x \in \mathbb{R}; 3x^2 = 0\} = \{0\}$; on prend donc $\mathbb{R}\backslash\{0\}$ comme ouvert de départ. T est C^1 sur $\mathbb{R}\backslash\{0\}$, sa différentielle y est partout inversible et T est injective sur $\mathbb{R}\backslash\{0\}$ (puisqu'elle l'est sur $\mathbb{R}$). T réalise donc un C^1-difféomorphisme de T sur son image, qui est $\mathbb{R}\backslash\{0\}$. On trouve ainsi qu'une densité de Y est

$$y \mapsto \frac{1}{\sqrt{2\pi}} \frac{1}{3} \frac{1}{|y|^{\frac{2}{3}}} \exp\left(-\frac{|y|^{\frac{2}{3}}}{2}\right) \mathbb{1}_{\mathbb{R}\backslash\{0\}}(y),$$

  qui est presque partout égale à

$$y \mapsto \frac{1}{\sqrt{2\pi}} \frac{1}{3} \frac{1}{|y|^{\frac{2}{3}}} \exp\left(-\frac{|y|^{\frac{2}{3}}}{2}\right),$$

  dont l'écriture est tout de même plus simple.
  — Un exemple dans $\mathbb{R}^2$
  Soient X et Y indépendantes suivant toutes deux la loi $\mathcal{N}(0,1)$. On veut calculer la loi du couple $(U,V) = (X^3, XY)$.
  On a ici $O = \mathbb{R}^2$ et $T(x) = (x^3, xy)$. On a

$$\{x \in O; \det D_xT = 0\} = \{(x,y) \in \mathbb{R}; 3x^3 = 0\} = \{0\} \times \mathbb{R}.$$

  On prend donc $\mathbb{R}^2\backslash(\{0\} \times \mathbb{R})$ comme ouvert de départ. T est C^1 sur $\mathbb{R}^2\backslash(\{0\} \times \mathbb{R})$, sa différentielle y est partout inversible et T est injective sur $\mathbb{R}\backslash\{0\}$, de réciproque $(u,v) \mapsto T^{-1}(u,v) = (\sqrt[3]{u}, \frac{v}{\sqrt[3]{u}})$. T réalise donc un C^1-difféomorphisme de T sur son image, qui est $(\mathbb{R}\backslash\{0\})^2$.

On trouve ainsi qu'une densité de (U,V) est

$$(u,v) \mapsto \frac{1}{2\pi}\frac{1}{3|u|}\exp\left(-\frac{|u|^{\frac{2}{3}}+v^2|u|^{-\frac{2}{3}}}{2}\right)\mathbb{1}_{(\mathbb{R}\backslash\{0\})^2}(u,v),$$

qui est presque partout égale à

$$(u,v) \mapsto \frac{1}{2\pi}\frac{1}{3|u|}\exp\left(-\frac{|u|^{\frac{2}{3}}+v^2|u|^{-\frac{2}{3}}}{2}\right).$$

— Le cas où T est C^1, mais non injective sur O est plus délicat. Voici comment procéder.
On commence par déterminer $N = \{x \in O; \det D_xT = 0\}$. Si N est de mesure nulle, on commence, comme précédemment, par remplacer O par $O' = O\backslash N$. Ensuite on découpe O' suivant ses composantes connexes. Dans la suite, nous supposerons que

$$O' = \underset{i\in I}{\cup} O_i,$$

où les O_i' sont des connexes et où I est dénombrable. (Nous n'avons jamais rencontré d'exemple où $O'\backslash N$ a une infinité non dénombrable de composantes connexes, en fait en général $|I| = 1$ ou 2.)
Nous allons maintenant montrer que si pour tout $i \in I$, T est injective sur O_i, alors on est capable de déterminer la loi image. Faisons donc cette hypothèse, et notons ν la mesure image de μ par T. On a pour tout borélien A

$$\nu(A) = \mu(T^{-1}(A)) = \sum_{i\in I}\mu(T^{-1}(A)\cap O_i) = \sum_{i\in I}\mu(O_i)\mu_i(T^{-1}(A)),$$

où l'on a posé

$$\mu_i(B) = \frac{\mu(B\cap O_i)}{\mu(O_i)} = \mu(B|O_i).$$

Ainsi, notre problème est maintenant, pour tout i, de trouver la loi image ν_i de μ_i par T. Mais nous pouvons dorénavant appliquer le théorème classique car
— μ_i admet la densité $x \mapsto \frac{1}{\mu(O_i)}f(x)\mathbb{1}_{O_i}(x)$, où f est la densité de μ.
— T est C^1 sur O_i et sa différentielle y est partout inversible (par définition de O').
— T est injective sur O_i (par hypothèse).
Ainsi, on obtient que la loi image de μ par T admet la densité

$$y \mapsto \sum_{i\in I} f(T_i^{-1}y)|\det(DT_i^{-1})_y|\mathbb{1}_{T(O_i)}(y)$$

où T_i^{-1} représente la réciproque de la restriction de T_i à O_i.

Remarque 6.48. *En dimension 1, l'injectivité de T sur les composantes connexes est gratuite, en vertu du théorème de Rolle.*

— Un exemple dans $\mathbb{R}$
  X suit la loi uniforme sur $[-1, 2]$ et $Y = X^2$. On a $O =]-1, 2[$ et $N = \{0\}$, de sorte que $O_1 =]-1, 0[$ et $O_2 =]0, 2[$. On trouve la densité
$$y \mapsto \frac{1}{2\sqrt{y}}\frac{1}{3}\, \mathbb{1}_{]0,1[}(y) + \frac{1}{2\sqrt{y}}\frac{1}{3}\, \mathbb{1}_{]0,4[}(y).$$

— Un exemple dans $\mathbb{R}^2$
  Soient X et Y deux variables aléatoires indépendantes suivant la loi uniforme sur $[0, 1]$. On pose $S = X + Y$ et $P = XY$. On veut déterminer la loi de (S, P). Ici, les points problématiques sont $N = \{(x, y) \in \mathbb{R}^2 ; x = y\}$, donc $O_1 = \{(x, y) \in \mathbb{R}^2 ; 0 < x < y < 1\}$ et $O_2 = \{(x, y) \in \mathbb{R}^2 ; 0 < y < x < 1\}$. Ici la symétrie axiale $s(x, y) = (y, x)$ d'axe $x = y$ laisse μ invariante (c'est-à-dire que μ est la mesure image de μ par s), envoie O_1 sur O_2, et vérifie de plus $T \circ s = T$. On a
$$\begin{aligned}
\nu_2(A) &= \frac{\mu(T^{-1}A \cap O_2)}{\mu(O_2)} = \frac{\mu((T \circ s)^{-1}(A) \cap s^{-1}(O_1))}{\mu(O_2)} \\
&= \frac{\mu(s^{-1}(T^{-1}(A)) \cap s^{-1}(O_1))}{\mu(s^{-1}(O_1))} = \frac{\mu_s(T^{-1}(A) \cap O_1)}{\mu_s(O_1)} \\
&= \frac{\mu(T^{-1}(A) \cap O_1)}{\mu(O_1)} = \nu_1(A).
\end{aligned}$$

Il s'ensuit que $\nu = \mu(O_1)\nu_1 + \mu(O_2)\nu_2 = \nu_1 = \nu_2$.
Dans les problèmes comme celui-ci où une symétrie entre en jeu, on peut parfois rédiger la solution d'une manière un peu différente en factorisant l'application T.
Ici, si on pose $(X', Y') = (\max(X, Y), \min(X, Y))$, on remarque sans peine que $(S, P) = T(X', Y')$. On peut donc décomposer le problème en deux étapes. D'abord calculer la loi de (X', Y') (on trouvera que c'est la loi uniforme sur O_2), puis calculer la loi image de la loi de (X', Y') par l'application T. Ce dernier calcul se fait sans difficulté théorique, puisqu'on a vu que T réalise un C^1-difféomorphisme de O_2 sur son image. Notons qu'on n'échappe pas à l'emploi d'arguments de symétrie, puisqu'on en a besoin pour calculer la loi de (X', Y'). Cette deuxième approche a comme avantage de mieux séparer les difficultés du problème. Gardons toutefois à l'esprit que ceci ne peut être érigé en méthode : ici les choses se passent très bien car la symétrie s laisse invariante à la fois l'application T et la mesure de départ.

Remarque 6.49. *Si on n'a pas de chance, T peut ne pas être injective sur les composantes connexes de O'. Dans ce cas, il faut essayer de trouver un fermé de mesure nulle telle que T soit injective sur chaque composante de $O'' = O' \backslash F$ et remplacer O' par O''. Là, il n'y a pas vraiment de méthode générale.*
Exemple: Soit $T(x, y) = (x^2 - y^2, 2xy)$. Le seul point où la différentielle n'est pas inversible est l'origine, mais bien sûr T n'est pas injective sur l'ouvert connexe $\mathbb{R}^2 \backslash \{(0, 0)\}$, puisque $T(-x, -y) = T(x, y)$. On constate que $F = \{(x, 0); x \in \mathbb{R}\}$ convient.

6.9.2 Un exemple : l'algorithme de Box–Muller

Dans la sous-section 6.8.2, on a déjà vu l'intérêt de noter que le produit de deux gaussiennes $\mathcal{N}(0, 1)$ est une loi invariante par rotation. Cela laisse penser que la position radiale dans $\mathbb{R}^2$ et l'angle par rapport à l'axe des x sont des variables indépendantes. On va voir que c'est le cas, et on va utiliser ce fait pour obtenir une méthode de simulation de variables gaussiennes : l'algorithme de Box–Muller. On fait donc un changement de variable polaire, ce qui nécessite, comme on l'a vu à la sous-section 4.12.3, de prendre quelques précautions. On prend $U =]0, +\infty[\times]0, 2\pi[$, $U' = \mathbb{R}^2 \backslash (\mathbb{R}_+ \times \{0\})$ et on pose $\phi(r, \theta) = (r \cos\theta, r \sin\theta)$. ϕ est clairement de classe C^1, réalise une bijection de U dans U'. La différentielle $|\det D_{r,\theta}\ \phi| = r$ ne s'annulant pas sur U, on a bien un C^1-difféomorphisme.
Si $(X, Y) \sim \mathcal{N}(0, 1) \otimes \mathcal{N}(0, 1)$, on a $\mathbb{P}((X, Y) \in U') = 1$, donc on peut poser $(R, \Theta) = \phi^{-1}(X, Y)$ et, avec le corollaire 4.47, le vecteur (R, Θ) admet la densité

$$(r, \theta) \mapsto f_{X,Y}(\phi(r, \theta))|\det D_{(r,\theta)}\phi| \mathbb{1}_U(r, \theta).$$

Comme $f_{X,Y}(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$, la densité de (R, Θ) est

$$(r, \theta) \mapsto \frac{1}{2\pi} \mathbb{1}_{]0,2\pi[}(\theta) r e^{-\frac{r^2}{2}} \mathbb{1}_{]0,+\infty[}.$$

Ainsi, R et Θ sont indépendantes. Θ suit clairement la loi uniforme sur $[0, 2\pi]$. Quant à R, c'est une variable positive avec

$$Q(t) = \mathbb{P}(X > t) = \int_{]t,+\infty[} r e^{-\frac{r^2}{2}}\ d\lambda(r) = e^{-t^2/2}.$$

Q est strictement croissante sur $]0, +\infty[$, avec $Q^{-1}(u) = \sqrt{-2 \log u}$. Grâce à la remarque 5.38, il s'ensuit que si U et V sont deux variables aléatoires indépendantes suivant la loi uniforme sur $[0, 1]$, le vecteur $(G^{-1}(U), 2\pi V)$ a la loi de (R, Θ), et par suite $\phi((G^{-1}(U), 2\pi V))$ a la loi de $\phi(R, \theta)$, c'est-à-dire la loi de (X, Y). On vient de montrer que si U et V sont deux variables aléatoires indépendantes suivant la loi uniforme sur $[0, 1]$, alors les variables

$\sqrt{-2\log U}\cos(2\pi V)$ et $\sqrt{-2\log U}\sin(2\pi V)$ sont des gaussiennes $\mathcal{N}(0,1)$ indépendantes : c'est l'algorithme de Box-Muller pour la simulation des gaussiennes.

6.10 Calcul des premiers moments des lois discrètes usuelles

6.10.1 Indicatrice d'un événement

On rappelle que pour $A \subset \Omega$, l'application $\mathbb{1}_A$ (appelée indicatrice de A) est définie sur Ω par

$$\mathbb{1}_A(x) = \begin{cases} 1 \text{ si } x \in A \\ 0 \text{ si } x \notin A. \end{cases}$$

$\mathbb{1}_A$ 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

—
$$\begin{aligned} \operatorname{Var} \mathbb{1}_A = \mathbb{E}\mathbb{1}_A^2 - (\mathbb{E}\mathbb{1}_A)^2 = \mathbb{E}\mathbb{1}_A - (\mathbb{E}\mathbb{1}_A)^2 &= \mathbb{P}(A) - \mathbb{P}(A)^2 \\ &= \mathbb{P}(A)(1 - \mathbb{P}(A)). \end{aligned}$$

6.10.2 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

$$\operatorname{Var} X = \sum_{k=1}^{n} \operatorname{Var} \mathbb{1}_{A_k} = \sum_{k=1}^{n} \mathbb{P}(A_k)(1 - \mathbb{P}(A_k)) = np(1-p).$$

6.10.3 Loi géométrique

Soit X une variable aléatoire suivant une loi géométrique de paramètre $p \in]0,1]$. On a

$$\begin{aligned}
\mathbb{E}X &= \sum_{k=0}^{+\infty} k\mathbb{P}(X=k) = \sum_{k=1}^{+\infty} k\mathbb{P}(X=k) \\
&= \sum_{k=1}^{+\infty} kp(1-p)^{k-1} = p\sum_{k=1}^{+\infty} k(1-p)^{k-1} \\
&= p\frac{1}{(1-(1-p))^2} = \frac{1}{p}
\end{aligned}$$

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

$$\operatorname{Var} X = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \frac{1}{p} + \frac{2(1-p)}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

6.10.4 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

$$\operatorname{Var} X = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

6.10.5 Loi hypergéométrique

On rappelle que la loi hypergéométrique $\mathcal{H}(N,n,k)$ est la loi image de la loi uniforme sur $\Omega = \mathcal{B}_k(\{1,\dots,N\})$ par l'application

$$\begin{aligned} X : \mathcal{B}_k(\{1,\dots,N\}) &\to \mathbb{N} \\ \omega &\mapsto X(\omega) = |\{1,\dots,n\} \cap \omega|. \end{aligned}$$

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

$$\begin{aligned}
\operatorname{Covar}(\mathbb{1}_{\{i\in A\}}, \mathbb{1}_{\{j\in A\}}) &= \operatorname{Covar}(1-\mathbb{1}_{\{i\in A\}}, 1-\mathbb{1}_{\{j\in A\}}) \\
&= \operatorname{Covar}(\mathbb{1}_{\{i\notin A\}}, \mathbb{1}_{\{j\notin A\}}) \\
&= \mathbb{P}(i\notin A, j\notin A) - \mathbb{P}(i\notin A)\mathbb{P}(j\notin A) \\
&= \frac{\binom{N-2}{k}}{\binom{N}{k}} - (\frac{N-k}{N})^2 \\
&= \frac{(N-k)(N-k-1)}{N(N-1)} - (\frac{N-k}{N})^2 \\
&= -\frac{1}{N-1}\frac{k}{N}(1-\frac{k}{N}),
\end{aligned}$$

on en déduit que

$$\begin{aligned}
\operatorname{Var} X &= n\times\frac{k}{N}\left(1-\frac{k}{N}\right) + n(n-1)\times\left(-\frac{1}{N-1}\frac{k}{N}(1-\frac{k}{N})\right) \\
&= n\frac{k}{N}\left(1-\frac{k}{N}\right)\left(1-\frac{n-1}{N-1}\right) \\
&= n\frac{k}{N}\left(1-\frac{k}{N}\right)\frac{N-n}{N-1} \\
&= k\frac{n}{N}\left(1-\frac{n}{N}\right)\frac{N-k}{N-1}.
\end{aligned}$$

Remarque 6.50. *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 $\frac{N-k}{N-1}$.*

6.11 Calcul des premiers moments des lois à densité usuelles

6.11.1 Loi uniforme sur un segment

Soit X une variable aléatoire suivant la loi uniforme sur $[-1,1]$. La densité de X est donc

$$x \mapsto \frac{1}{2}\mathbb{1}_{[-1,1]}(x).$$

On a donc

$$\mathbb{E}X = \int_{-1}^{1}\frac{1}{2}x\,dx = 0$$

et

$$\mathbb{E}X^2 = \int_{-1}^{1}\frac{1}{2}x^2\,dx = \frac{1}{3}.$$

Comme X est centrée, on a $\operatorname{Var} X = \mathbb{E}X^2$.

Passons au cas général. On pose $Y = \frac{a+b}{2} + \frac{b-a}{2}X$. X suit la loi uniforme sur $[-1,1]$ si et seulement Y suit la loi uniforme sur $[a,b]$. On a alors

$$\mathbb{E}Y = \mathbb{E}\frac{a+b}{2} + \frac{b-a}{2}\mathbb{E}X = \frac{a+b}{2} \quad \text{et} \quad \operatorname{Var} Y = (\frac{b-a}{2})^2 \operatorname{Var} X = \frac{(b-a)^2}{12}.$$

6.11.2 Loi gaussienne

Soit X une variable aléatoire suivant la loi 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 6.51. *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 $\lim_{a \to -\infty} g(a)f(a) = \lim_{b \to +\infty} g(b)f(b) = 0$, on en déduit que

$$0 = \int_{\mathbb{R}} g'(x)f(x)\, d\lambda(x) - \int_{\mathbb{R}} xg(x)f(x)\, d\lambda(x),$$

soit $\mathbb{E}[g'(X)] = \mathbb{E}[Xg(X)]$. □

En prenant $g(x) = x$, on obtient l'existence d'un moment d'ordre 2 avec $\mathbb{E}[X^2] = 1$. D'autre part, l'existence d'un moment d'ordre 2 implique celle d'un moment d'ordre 1. Comme la densité de X est paire, on en déduit que $\mathbb{E}X = 0$. On a donc $\operatorname{Var} X = \mathbb{E}X^2 = 1$.

Passons au cas général. Si $Y = m + \sigma X$, 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.

6.11.3 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
$\operatorname{Var} X = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \frac{a(a+1)}{\gamma^2} - \frac{a^2}{\gamma^2} = \frac{a}{\gamma^2}$. □

6.11.4 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 $\operatorname{Var} X = \frac{1}{\lambda^2}$.

6.11.5 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

$$\operatorname{Var} X = \frac{ab}{(a+b)^2(a+b+1)}.$$

6.11.6 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.

6.12 Exercices détaillés

6.12.1 Loi de Dirichlet

Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes avec pour tout i entre 1 et n : $X_i \sim \Gamma(a_i, 1)$. On pose $a^* = \sum\limits_{i=1}^{n} a_i$, et, pour tout $i \in \{1, \dots, n\}$, $a_i^* = a^* - a_i$. On définit $S = \sum\limits_{k=1}^{n} X_k$ et $Y_i = \frac{X_i}{S}$.

On dit que $(Y_1, \dots, Y_n)$ suit la loi de Dirichlet de paramètre $(a_1, \dots, a_n)$.

1. Pour tout i entre 1 et n, déterminer la loi du couple $(X_i, S - X_i)$, puis la loi du couple $(X_i/S, S)$. Montrer que Y_i suit une loi Bêta de paramètre (a_i, a_i^*).
2. Déterminer la loi du n-uplet $(Y_1, \dots, Y_{n-1}, S)$.
   Montrer que $(Y_1, \dots, Y_{n-1})$ admet la densité

   $$(x_1, \dots, x_{n-1}) \mapsto \frac{1}{\tilde{B}(a)} \left(\prod_{i=1}^{n-1} x_i^{a_i - 1} \right) \left(1 - \sum_{i=1}^{n-1} x_i \right)^{a_n - 1} \mathbb{1}_{\Delta}(x_1, \dots, x_{n-1}),$$

   où l'on a posé

   $$\tilde{B}(a) = \frac{\prod_{i=1}^{n} \Gamma(a_i)}{\Gamma(\sum_{i=1}^{n} a_i)} \text{ et } \Delta = \{(\theta_1, \dots, \theta_{n-1}) \in]0, +\infty[^{n-1}; \sum_{i=1}^{n-1} \theta_i < 1\}.$$

3. Soient $X'_1, \dots, X'_n$ des variables aléatoires indépendantes avec pour tout i entre 1 et n : $X'_i \sim \Gamma(a_i+b_i, 1)$. On pose $S' = \sum\limits_{k=1}^{n} X'_k$.

   Montrer que $\mathbb{E}[\prod\limits_{i=1}^{n} Y_i^{b_i}] = \frac{\prod\limits_{i=1}^{n} \Gamma(a_i+b_i)}{\prod\limits_{i=1}^{n} \Gamma(a_i)} \mathbb{E}\left(S'^{-\sum\limits_{i=1}^{n} b_i}\right)$. En déduire que
$$\mathbb{E}[\prod_{i=1}^{n} Y_i^{b_i}] = \frac{\tilde{B}(a+b)}{\tilde{B}(a)}.$$

Solution

1. $S - X_i = \sum_{j\neq i} X_j$ donc X_i et $S - X_i$ sont indépendantes. Comme $\sum_{j\neq i} X_j$ est la somme de variables aléatoires indépendantes suivant la loi $\Gamma(a_i, 1)$, $\sum_{j\neq i} X_j$ suit la loi $\Gamma(a_i^*, 1)$.
   Finalement la loi de $(X_i, S - X_i)$ est $\Gamma(a_i, 1) \otimes \Gamma(a_i^*, 1)$.
   L'application $T : (x, y) \mapsto (\frac{x}{x+y}, x + y)$ réalise un C^1-difféomorphisme de $(\mathbb{R}_+^*)^2$ dans $]0, 1[\times\mathbb{R}_+^*$. Elle admet comme réciproque l'application $T^{-1}(\theta, s) = (\theta s, (1 - \theta)s)$. On a $|DT^{-1}_{(\theta,s)}| = \begin{vmatrix} s & \theta \\ -s & 1-\theta \end{vmatrix} = s$. Comme la densité de $(X_i, S - X_i)$ est $\frac{1}{\Gamma(a_i)\Gamma(a_i^*)} x^{a_i-1} y^{a_i^*-1} e^{-(x+y)} \mathbb{1}_{\mathbb{R}_+^*}(x)\mathbb{1}_{\mathbb{R}_+^*}(y)$, on obtient alors la densité de (Y'_i, S) :
$$\begin{aligned}(\theta, s) \quad \mapsto \quad & \frac{1}{\Gamma(a_i)\Gamma(a_i^*)} s(\theta s)^{a_i-1}((1-\theta)s)^{a_i^*-1} e^{-s} \mathbb{1}_{]0,1[}(\theta)\mathbb{1}_{\mathbb{R}_+^*}(s) \\ & = \frac{\theta^{a_i-1}(1-\theta)^{a_i^*-1}\mathbb{1}_{]0,1[}(\theta)}{B(a_i, a_i^*)} \frac{1}{\Gamma(a_i + a_i^*)} s^{a_i+a_i^*-1} e^{-s} \mathbb{1}_{\mathbb{R}_+^*}(s).\end{aligned}$$
   On reconnaît alors la densité de Bêta$(a_i, a_i^*) \otimes \Gamma(a_i + a_i^*, 1)$, ce qui nous dit que Y_i et S sont indépendantes et suivent respectivement les lois Bêta(a_i, a_i^*) et $\Gamma(a_i + a_i^*, 1)$. Remarquons qu'on a utilisé le fait que nous savions que la constante de renormalisation $B(a_i, a_i^*)$ est égale à $\frac{\Gamma(a_i^*)\Gamma(a_i)}{\Gamma(a_i^*+a_i)}$, ce qui nous permet d'écrire tout de suite les densités espérées, mais nous aurions tout aussi bien pu nous en passer et utiliser le théorème 5.28, qui donne également le résultat voulu et permet de retrouver la formule exprimant $B(.,.)$ à l'aide de la fonction Γ.

2. On utilise cette fois la transformation
$$(x_1, \dots, x_n) \mapsto \left(\frac{x_1}{s(x)}, \dots, \frac{x_{n-1}}{s(x)}, s(x)\right),$$
   avec $s(x) = \sum_{i=1}^{n}(x_i)$. Elle réalise un C^1-difféomorphisme de $(\mathbb{R}_+^*)^n$

dans $\Delta \times \mathbb{R}_+^*$, et sa réciproque est

$$(\overline{\theta}, s) = (\theta_1, \dots, \theta_{n-1}, s) \mapsto (\theta_1 s, \dots \theta_{n-1} s, (1 - \sum_{i=1}^{n-1} \theta_i) s).$$

On a

$$|DT^{-1}_{(\overline{\theta},s)}| = \begin{vmatrix} s & 0 & \dots & 0 & \theta_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \\ 0 & \dots & 0 & s & \theta_{n-1} \\ -s & \dots & \dots & -s & 1 - \sum_{i=1}^{n-1} \theta_i \end{vmatrix} = \begin{vmatrix} s & 0 & \dots & 0 & \theta_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \\ 0 & \dots & 0 & s & \theta_{n-1} \\ 0 & \dots & \dots & 0 & 1 \end{vmatrix}$$
$$= s^{n-1}.$$

Pour l'égalité entre les deux déterminants, on a remplacé la dernière ligne par la somme de toutes les lignes. Ainsi, on obtient la densité

$$\prod_{i=1}^n \frac{1}{\Gamma(a_i)} s^{n-1} \prod_{i=1}^{n-1} (\theta s)^{a_i - 1} ((1 - \sum_{i=1}^{n-1} \theta_i) s)^{a_n - 1} e^{-s} \ \mathbb{1}_\Delta(\overline{\theta}) \ \mathbb{1}_{\mathbb{R}_+^*}(s),$$

que l'on peut réécrire en

$$\frac{1}{\tilde{B}(a)} (\prod_{i=1}^{n-1} x_i^{a_i - 1})(1 - \sum_{i=0}^{n-1} x_i)^{a_n - 1} \ \mathbb{1}_\Delta(x_1, \dots, x_{n-1}) \frac{e^{-s}}{\Gamma(a^*)} s^{a^* - 1} \ \mathbb{1}_{\mathbb{R}_+^*}(s).$$

Cela montre que le vecteur aléatoire $(Y_1, \dots, Y_{n-1})$ est indépendant de S et a la densité voulue.

3. Posant $b^* = \sum_{i=1}^n b_i$, on a $\prod_{i=1}^n Y_i^{b_i} = \frac{\prod_{i=1}^n X_i^{b_i}}{(\sum_{i=1}^n X_i)^{b^*}}$, donc avec le théorème de transfert

$$\left(\prod_{i=1}^n \Gamma(a_i)\right) \mathbb{E}[\prod_{i=1}^n Y_i^{b_i}] = \int_{\mathbb{R}_+^n} \frac{\prod_{i=1}^n x_i^{b_i}}{(\sum_{i=1}^n x_i)^{b^*}} \prod_{i=1}^n x_i^{a_i - 1} e^{-x_i} \ d\lambda^{\otimes n}(x).$$

D'autre part, toujours par le théorème de transfert, on a

$$\left(\prod_{i=1}^n \Gamma(a_i + b_i)\right) \mathbb{E}[(\sum_{i=1}^n X_i')^{-b^*}] = \int_{\mathbb{R}_+^n} \frac{\prod_{i=1}^n x_i^{a_i + b_i - 1} e^{-x_i}}{(\sum_{i=1}^n x_i)^{b^*}} \ d\lambda^{\otimes n}(x).$$

Comme les deux membres de droite sont égaux, on a finalement

$$\mathbb{E}[\prod_{i=1}^n Y_i^{b_i}] = \frac{\prod_{i=1}^n \Gamma(a_i + b_i)}{\prod_{i=1}^n \Gamma(a_i)} \mathbb{E}[S'^{-b^*}].$$

Mais la loi de S' est connue : comme $\sum_i X_i'$ est la somme de variables aléatoires indépendantes suivant la loi $\Gamma(a_i + b_i, 1)$, $\sum_i X_i'$ suit la loi $\Gamma(a^* + b^*, 1)$. Les moments d'ordre $\alpha > -(a^* + b^*)$ ont été calculés

en 6.11.3.
En particulier, prenant $\alpha = -b^*$, on trouve $\mathbb{E}[S'^{-b^*}] = \frac{\Gamma(a^*)}{\Gamma(a^*+b^*)}$, d'où finalement

$$\mathbb{E}[\prod_{i=1}^{n} Y_i^{b_i}] = \frac{\tilde{B}(a+b)}{\tilde{B}(a)}.$$

6.12.2 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
$$\mathbb{E}f(\frac{S_n}{n}) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \mathbb{P}(S_n = k) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}$$
$$= B_n(x).$$
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

$$\begin{aligned} & B_n(x) - f(x) \\ = & \int_\Omega \left[f\left(\frac{S_n}{n}\right) - f(x) \right] \, d\mathbb{P} \\ = & \int_{|\frac{S_n}{n} - x| \le \varepsilon} \left[f\left(\frac{S_n}{n}\right) - f(x) \right] \, d\mathbb{P} + \int_{|\frac{S_n}{n} - x| > \varepsilon} \left[f\left(\frac{S_n}{n}\right) - f(x) \right] \, d\mathbb{P}. \end{aligned}$$

Ainsi, on a par l'inégalité triangulaire, pour tout $\varepsilon > 0$

$$\begin{aligned} & |B_n(x) - f(x)| \\ \le & \delta(\varepsilon)\mathbb{P}\left(\left|\frac{S_n}{n} - x\right| \le \varepsilon\right) + \int_{|\frac{S_n}{n} - x| > \varepsilon} \left| f\left(\frac{S_n}{n}\right) - f(x) \right| \, d\mathbb{P} \\ \le & \delta(\varepsilon) + 2\|f\|_\infty \mathbb{P}\left(\left|\frac{S_n}{n} - x\right| > \varepsilon\right). \end{aligned}$$

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

$$\forall \varepsilon > 0 \qquad \overline{\lim_{n \to +\infty}} \, \|B_n - f\|_\infty \le \delta(\varepsilon),$$

et enfin $\overline{\lim\limits_{n \to +\infty}} \, \|B_n - f\|_\infty \le 0$, ce qui donne le résultat voulu.

6.13 Exercices sur l'espérance

6.13.1 Exercices corrigés

Exercice 129. 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.

$\to$ indication $\to$ solution

Exercice 130. 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 $\mathbb{E}[(X-m)^2] = \inf_{x\in\mathbb{R}} \mathbb{E}[(X-x)^2]$.
2. On dit que λ est une médiane de X si
$$\mathbb{P}(X \geq \lambda) \geq 1/2 \text{ et } \mathbb{P}(X \leq \lambda) \geq 1/2.$$
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 $p \neq 1/2$.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 131. 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.
$\rightarrow$ indication $\rightarrow$ solution

Exercice 132. Soit X une variable aléatoire uniforme sur l'intervalle $[-1,1]$. Posons $Y = |4X^2-1|$. Calculer $\mathbb{E}(Y),\ \mathbb{E}(X+Y),\ \mathbb{E}[(Y+4X^2)^3]$.
$\rightarrow$ indication $\rightarrow$ solution

Exercice 133.
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 $\mathbb{E}\left(\min\{Y_1,\dots,Y_n\}\right)$.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 134. *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.

→ indication → solution

Exercice 135. 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 $x_1 + \cdots + x_8 = 40$ et A_i les éléments de A_0 avec $x_i \geq 10$, on pourra par exemple remarquer que $A_1 \cap A_2 \cap \cdots \cap A_i$ est en bijection avec l'ensemble des solutions de $y_1 + \cdots + y_8 = 40 - 10i$ dans $\mathbb{N}^8$.

→ indication → solution

Exercice 136.

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 $\{2, 4, 6, \dots, 2N\}$.

→ indication → solution

Exercice 137. *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 $a_1, \ldots, a_k$ entiers de somme n :

$$\mathbb{P}(N^1 = a_1, \ldots, N^k = a_k) = \binom{n}{a_1, a_2, \ldots, a_k} p_1^{a_1} \ldots p_k^{a_k},$$

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.

→ indication → solution

Exercice 138. 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.

→ indication → solution

Exercice 139. 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, $\mathbb{P}(A_k) \le x^{-2}\mathbb{E}(\mathbb{1}_{A_k} S_k^2)$. → indication → solution

Exercice 140. 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.

→ indication → solution

Exercice 141. 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 $\mathbb{E}(\min(X, Y))$.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 142. Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes de loi exponentielle de paramètre $a > 0$: $\mathbb{P}(X_k \geq x) = e^{-ax}$ pour $x \geq 0$. Posons $M = \inf\{X_1, \ldots, X_n\}$.

1. Calculer $\mathbb{P}(M \geq x)$ pour tout $x \in \mathbb{R}$. En déduire la loi de M.
2. Pour $t \geq 0$, on pose $N_t = \text{card}(k \in \{1, \ldots, n\} | X_k \geq t)$. 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).

$\rightarrow$ indication $\rightarrow$ solution

Exercice 143. 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 ? $\rightarrow$ indication $\rightarrow$ solution

Exercice 144. 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 145. 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. $\rightarrow$ indication $\rightarrow$ solution

Exercice 146. *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 $Y_1, \ldots, Y_n$ des variables aléatoires indépendantes suivant la loi $\Gamma(\frac{1}{p}, 1)$. On pose $X_k = Y_k^{1/p}$ et $S = X_1^p + \cdots + X_n^p$. 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 $T_n = \{(x_1, \ldots, x_n) \in \mathbb{R}_+^n ; x_1 + \cdots + x_n \leq 1\}$. Calculer $\displaystyle\int_{T_n} \frac{d\lambda^{\otimes n}(x)}{x_1 + \cdots + x_n}$.

$\to$ indication $\to$ solution

Exercice 147. *Triangles et loi de Dirichlet.*

1. Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes et identiquement distribuées suivant la loi exponentielle $\mathcal{E}(1)$. Calculer la loi du vecteur $(U_1, \ldots, U_{n-1}, S_n)$ défini par
$$S_n = \sum_{i=1}^n X_i \quad \text{et} \quad U_i = X_i/S_n.$$
En déduire la loi de $U = (U_1, \ldots, U_{n-1})$ puis celle du vecteur $V = (U_1, \ldots, U_n)$ où $U_n = X_n/S_n$.
2. Soit T un triangle équilatéral de sommets a, b, c. On choisit un point v au hasard dans T. Déterminer la loi de la surface du triangle de sommets a, b, v.

$\to$ indication $\to$ solution

$\to$ indication $\to$ solution

Exercice 148. Soit (X, Y) un couple de variables aléatoires de densité de probabilité :
$$g(x,y) = \begin{cases} \frac{1}{2}x\,y \ \text{ si } (x,y) \in D \\ \\ 0 \text{ si } (x,y) \notin D \end{cases}$$
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 $Z = R^2$.

$\to$ indication $\to$ solution

Exercice 149. 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.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 150. *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 $0 < \text{Re } z < 1$, on a
$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin \pi z}.$$
   C'est la formule des compléments.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 151. *Nombres de Bell et loi de Poisson.*
Le n-ième nombre de Bell, noté B_n, est le nombre de partitions de l'ensemble $\{1, \dots, n\}$.

1. À chaque fonction f de $\{1, \dots, n\}$ dans $\{1, \dots, p\}$, 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 $P_0 = 1$, $P_1 = X$, $P_i(X) = X(X-1)$, ..., $P_n(X) = X(X-1)\dots(X-i+1)$, on a l'identité entre polynômes
$$X^n = \sum_{\Pi \text{ Partition de } \{1,\dots,n\}} P_{N(\Pi)}(X).$$
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 $B_n \leq n! \frac{e^{n-1}}{(\log n)^n}$.

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 [6], page 102, ou Flajolet–Sedgewick [16], page 560.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 152. 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.

→ indication → solution

Exercice 153. 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. → indication → solution

Exercice 154. *Une preuve probabiliste d'un développement eulérien.*

1. Soient $s > 1$ et $(N_i)_{i\ge 1}$ une suite de variables aléatoires indépendantes telles que $1 + N_i \sim \mathcal{G}(1 - p_i^{-s})$, où $(p_i)_{i\ge 1}$ 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

$\rightarrow$ indication $\rightarrow$ solution

Exercice 155. *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. $\rightarrow$ indication $\rightarrow$ solution

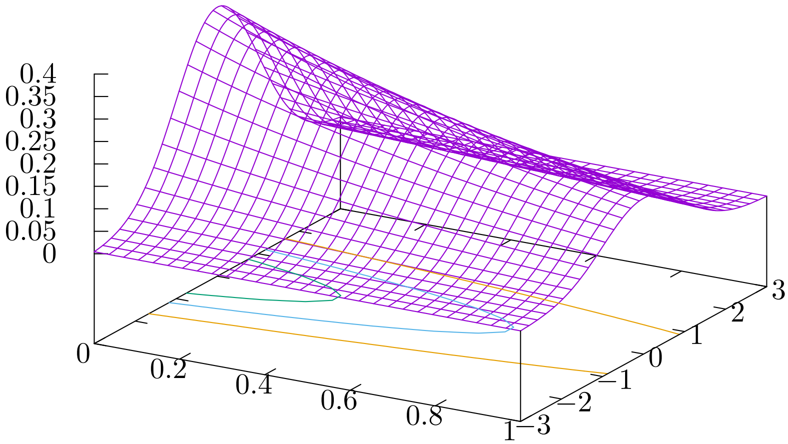
6.13.2 Exercices non corrigés

Exercice 156. 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 ?

$\rightarrow$ indication

Exercice 157. 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 ?



→ indication

Exercice 158. 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

$$\mathbb{P}(A\Delta B) = \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A \cap B).$$

→ indication

Exercice 159. Soient A, B deux éléments observables. Montrer que

$$|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \leq \sqrt{\mathbb{P}(A)\mathbb{P}(B)}.$$

→ indication

Exercice 160. 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 $\mathbb{P}(X = \frac{\pi}{6}) = \frac{1}{2}$.
→ indication

Exercice 161. *Une application des inégalités de Bonferroni.*

On place sept dames sur un échiquier torique 41×41 de telle manière qu'aucune dame ne puisse en prendre une autre. Soit ϕ une permutation des cases de l'échiquier.

Montrer qu'il existe x tel que x et $\phi(x)$ puissent chacun être pris par au moins une des dames.

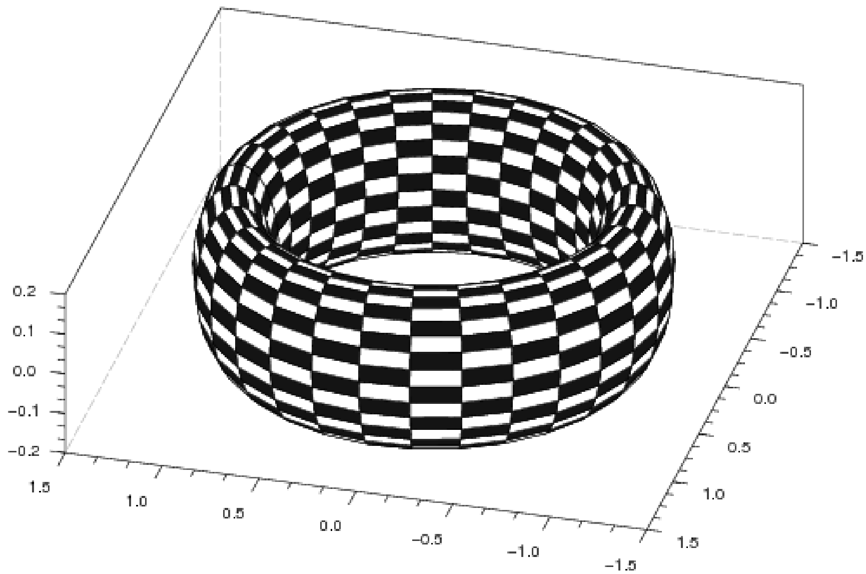


FIGURE 6.1 – Un échiquier torique

On rappelle que les dames peuvent prendre les pièces qui sont sur la même ligne, sur la même colonne, ou sur une même diagonale.

$\rightarrow$ indication

Exercice 162. 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

$$\frac{1}{n}\sum_{k=1}^{n}(a_k - m)^2 \leq \frac{(a_n - a_1)^2}{4}.$$

$\rightarrow$ indication

Exercice 163. 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 $\mathbb{E}(\sigma(X)) \geq \mathbb{E}X$.
$\to$ indication

Exercice 164. 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 $\mathrm{Var}\, X$.
$\to$ indication

Exercice 165. 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 ?
$\to$ indication

Exercice 166. Soit X une variable aléatoire suivant la loi uniforme sur $[0, 1]$. Déterminer la loi de $Z = -\log(1 - X)$. $\to$ indication

Exercice 167. Soit X une variable aléatoire de densité f. Montrer que la variable aléatoire $|X|$ admet comme densité

$$x \mapsto \mathbb{1}_{\mathbb{R}_+}(x)(f(x) + f(-x)).$$

$\to$ indication

Exercice 168. Soit X une variable aléatoire positive de densité f. Montrer que la variable aléatoire $X^{1/2}$ admet comme densité

$$x \mapsto \mathbb{1}_{\mathbb{R}_+}(x) 2x f(x^2).$$

$\to$ indication

Exercice 169. Soit X une variable aléatoire normale centrée réduite. Montrer que la variable aléatoire X^2 est à densité et la déterminer.
$\to$ indication

Exercice 170. *Simulation de la gaussienne par la méthode du rejet en utilisant la double exponentielle.*
On pose

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2) \text{ et } g(z) = \frac{1}{2} \exp(-|z|).$$

1. Montrer que g est bien une densité sur $\mathbb{R}$.
2. Déterminer une constante k satisfaisant $\forall z \in \mathbb{R},\ f(z) \leq k g(z)$.
3. Simuler une loi gaussienne centrée réduite par la méthode du rejet.
4. Comparer cette méthode à la méthode de Box-Muller.

$\to$ indication

Exercice 171. 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).

$\rightarrow$ indication

Exercice 172. 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 $\sum_{i=1}^{+\infty}(-b(i)) \leq \frac{b(0)}{2}$.

→ indication

Exercice 173. 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
$$\mathbb{E}X = \frac{\binom{n+1}{r+1}}{\binom{n}{r}} = \frac{n+1}{r+1}.$$

→ indication

Exercice 174. 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.

→ indication

Exercice 175. *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 $X_i = \max\left(\mathbb{1}_{\{Y_i\geq 1\}}, Z_i\right)$, $X = \sum\limits_{i=1}^{n} X_i$ et $Y = \sum\limits_{i=1}^{n} Y_i$.

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 $X_1, \ldots, X_n$ sont des variables aléatoires indépendantes suivant respectivement la loi de Bernoulli de paramètre p_i, on a alors pour tout $A \subset \mathbb{N}$:

$$|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| \leq \sum_{i=1}^{n} p_i^2.$$

$\rightarrow$ indication

Chapitre 7

Espaces $\mathcal{L}^p$ et L^p

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Pour $p \in [1, +\infty[$, on note $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ l'ensemble des applications mesurables de $(\Omega, \mathcal{A}, \mu)$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telles que

$$\int_\Omega |f(x)|^p \, d\mu(x) < +\infty.$$

On note $\mathcal{L}^\infty(\Omega, \mathcal{A}, \mu)$ l'ensemble des applications mesurables de $(\Omega, \mathcal{A}, \mu)$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telles que $\|f\|_{\infty,ess} < +\infty$, où

$$\|f\|_{\infty,\text{ess}} = \inf\{M \in \mathbb{R}; \mu(\{x \in \Omega; |f(x)| > M\}) = 0\}.$$

Comme

$$\{x \in \Omega; |f(x)| > \|f\|_{\infty,\text{ess}}\} = \bigcup_{n \geq 1} \{x \in \Omega; |f(x)| > \|f\|_{\infty,\text{ess}} + 1/n\},$$

et qu'une réunion dénombrable d'ensembles de mesure nulle est de mesure nulle, on a $|f| \leq \|f\|_{\infty,\text{ess}}$ presque partout.

On dit que des nombres p et q de $]1, +\infty[$ sont des exposants conjugués s'ils vérifient

$$\frac{1}{p} + \frac{1}{q} = 1.$$

On convient également que 1 et l'infini sont des exposants conjugués.

7.1 De $\mathcal{L}^p$ à L^p

7.1.1 Inégalité de Hölder

Théorème 7.1. *Soient p et q des exposants conjugués de $]1, +\infty[$, $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, $f \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$, $g \in \mathcal{L}^q(\Omega, \mathcal{A}, \mu)$. Alors $fg \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ et*

$$\left|\int_\Omega f(x)g(x)\ d\mu(x)\right| \leq \left(\int_\Omega |f(x)|^p\ d\mu(x)\right)^{1/p} \left(\int_\Omega |g(x)|^q\ d\mu(x)\right)^{1/q}.$$

Démonstration. Si f est nulle $\mu-$presque partout, alors l'inégalité est évidente (c'est en fait une égalité). Idem pour g. Dans le cas inverse, on a

$$\left(\int_\Omega |f(x)|^p\ d\mu(x)\right)^{1/p} > 0 \quad \text{et} \quad \left(\int_\Omega |g(x)|^q\ d\mu(x)\right)^{1/q} > 0.$$

Bien entendu, $|\int_\Omega fg\ d\mu| \leq \int_\Omega |f|.|g|\ d\mu$.
En remplaçant f par $|f|/\left(\int_\Omega |f(x)|^p\ d\mu(x)\right)^{1/p}$ et g par $|g|/\left(\int_\Omega |g(x)|^q\ d\mu(x)\right)^{1/q}$, on peut donc se ramener au cas où f et g sont positives avec

$$\int_\Omega f(x)^p\ d\mu(x) = \int_\Omega g(x)^q\ d\mu(x) = 1.$$

Or pour tous x, y dans $\overline{\mathbb{R}}_+$, on a l'inégalité [1]

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}.$$

Si x ou y est infini ou nul, c'est évident. Sinon, on peut écrire $x = e^{a/p}$, $y = e^{b/q}$ et utiliser la convexité de la fonction exponentielle. Ainsi, comme f et g sont positives, on a

$$f(x)g(x) \leq \frac{f(x)^p}{p} + \frac{g(x)^q}{q},$$

d'où

$$\int_\Omega f(x)g(x)\ d\mu(x) \leq \int_\Omega \frac{f(x)^p}{p}\ d\mu(x) + \int_\Omega \frac{g(x)^q}{q}\ d\mu(x) = \frac{1}{p} + \frac{1}{q} = 1.$$

□

Remarque 7.2. *On a μ presque partout $|fg| \leq |f|\|g\|_{\infty,ess}$, et en intégrant $\|fg\|_1 \leq \|f\|_1\|g\|_{\infty,ess}$.*

Exemple: Application à la fonction Gamma
Soient $x, y > 0$ et $\theta \in]0,1[$. En posant $p = \frac{1}{\theta}$ et $q = \frac{1}{1-\theta}$, l'inégalité de Hölder donne

$$\begin{aligned}\int_0^{+\infty} e^{-t}t^{\theta x+(1-\theta)y-1}\ d\lambda(t) &= \int_0^{+\infty} (e^{-t}t^{x-1})^\theta (e^{-t}t^{y-1})^{1-\theta}\ d\lambda(t)\\ &\leq \left(\int_0^{+\infty} e^{-t}t^{x-1}\ d\lambda(t)\right)^\theta \left(\int_0^{+\infty} e^{-t}t^{y-1}\ d\lambda(t)\right)^{1-\theta}\end{aligned}$$

1. Parfois appelée inégalité de Young.

soit $\Gamma(\theta x+(1-\theta)y) \leq \Gamma(x)^\theta \Gamma(y)^{1-\theta}$ ou encore

$$\log \Gamma(\theta x+(1-\theta)y) \leq \theta \log \Gamma(x)+(1-\theta)\log \Gamma(y).$$

Le logarithme de la fonction Gamma est donc une fonction convexe. On dit qu'elle est logarithmiquement convexe.

7.1.2 Inégalité triangulaire (ou inégalité de Minkowski)

Théorème 7.3. *Soient $p \in [1,+\infty]$, $(\Omega,\mathcal{A},\mu)$ un espace mesuré, f et g deux éléments de $\overline{\mathcal{V}}(\Omega,\mathcal{A},\mu)$. On a*

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

Démonstration. Dans le cas où $p=1$, il s'agit d'une conséquence immédiate de l'inégalité triangulaire sur $\mathbb{R}$ et de la positivité de l'intégrale. Si $p=+\infty$, on a alors $|f(x)| \leq \|f\|_\infty$ $\mu-$presque partout et $|g| \leq \|g\|_\infty$ $\mu-$presque partout, ce qui implique $|f+g| \leq \|f\|_\infty+\|g\|_\infty$ $\mu-$presque partout et le résultat en découle. Supposons donc $p \in]1,+\infty[$ et notons q l'exposant conjugué de p. Comme précédemment, on peut supposer que f et g ne sont pas nulles presque partout. Aussi, si $\int_\Omega |f(x)|^p\, d\mu(x)=+\infty$ ou si $\int_\Omega |g(x)|^p\, d\mu(x)=+\infty$, l'inégalité est évidente. On suppose donc que ces deux quantités sont finies. Comme $|f+g|^p \leq (|f|+|g|)^p$, on peut supposer sans perte de généralité que f et g sont positives. Maintenant, comme $\left(\frac{f+g}{2}\right)^p \leq \frac{f^p+g^p}{2}$ par convexité de la fonction $x \mapsto x^p$, il s'ensuit également $\int_\Omega |f(x)+g(x)|^p\, d\mu(x) < +\infty$. On écrit alors

$$(f+g)^p = f(f+g)^{p-1}+g(f+g)^{p-1}.$$

L'inégalité de Hölder donne

$$\int_\Omega f(f+g)^{p-1}\, d\mu \leq \left(\int_\Omega f^p\, d\mu\right)^{1/p}\left(\int_\Omega (f+g)^{(p-1)q}\, d\mu\right)^{1/q},$$

soit en remarquant que $p=(p-1)q$

$$\int_\Omega f(f+g)^{p-1}\, d\mu \leq \left(\int_\Omega f^p\, d\mu\right)^{1/p}\left(\int_\Omega (f+g)^p\, d\mu\right)^{1/q}.$$

De même, on obtient en échangeant les rôles de f et g

$$\int_\Omega g(f+g)^{p-1}\, d\mu \leq \left(\int_\Omega g^p\, d\mu\right)^{1/p}\left(\int_\Omega (f+g)^p\, d\mu\right)^{1/q}.$$

En additionnant ces inégalités, on trouve

$$\int_\Omega (f+g)^p\, d\mu \leq \left(\left(\int_\Omega f^p\, d\mu\right)^{1/p}+\left(\int_\Omega g^p\, d\mu\right)^{1/p}\right)\left(\int_\Omega (f+g)^p\, d\mu\right)^{1/q},$$

d'où
$$\left(\int_\Omega (f+g)^p \, d\mu\right)^{1/p} \le \left(\int_\Omega f^p \, d\mu\right)^{1/p} + \left(\int_\Omega g^p \, d\mu\right)^{1/p}.$$
$\square$

Il est maintenant simple de constater que si l'on pose
$$\|f\|_p = \left(\int_\Omega |f|^p \, d\mu\right)^{1/p},$$
on définit une semi-norme sur l'espace vectoriel $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$, de même que $\|\cdot\|_{\infty,\text{ess}}$ définit une semi-norme sur l'espace vectoriel $\mathcal{L}^\infty(\Omega, \mathcal{A}, \mu)$.

Remarquons bien qu'en général, l'application $\|.\|_p$ ne définit pas une norme sur $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ car l'axiome de séparation peut être pris en défaut.
En effet, sur $\mathcal{L}^p(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, on a bien $\|\mathbb{1}_\mathbb{Q}\|_p = 0$, mais bien sûr, $\mathbb{1}_\mathbb{Q} \neq 0$.

Notons $V = \{v \in \mathcal{L}^p ; \|v\|_p = 0\}$. D'après l'inégalité triangulaire, V est un sous-espace vectoriel de $\mathcal{L}^p$. Un raisonnement simple (à faire en exercice) permet en fait de montrer que $V = \{v \in \mathcal{L}^p ; v = 0 \quad \mu- \text{ p.p.}\}$.

Notons L^p le quotient de l'espace vectoriel $\mathcal{L}^p$ par son sous-espace vectoriel V.

Remarque 7.4.
1. *L'ensemble V dépend de μ, donc les classes d'équivalences correspondant à deux mesures μ et ν ne sont pas les mêmes.*
2. *Si Ω est fini et $\mu(\Omega) < +\infty$, alors tous les espaces L^p pour $1 \le p \le +\infty$ sont identiques.*
3. *Si Ω est dénombrable, si $\mathcal{A} = \mathcal{P}(\Omega)$ et si $\mu(\{x\}) > 0$ pour tout $x \in \Omega$, alors $L^p = \mathcal{L}^p$ pour tout $1 \le p \le +\infty$.*

Lemme 7.5. *Soit μ une mesure finie. Si $1 \le p \le r \le +\infty$, on a alors l'inclusion $L^r \subset L^p$.*

Démonstration. Supposons $r < +\infty$. Soit $f \in L^r$. Dans ce cas, on a toujours $|f|^p \le 1 + |f|^r$ (pour voir cela, il suffit de séparer les cas $|f| \le 1$ et $|f| > 1$). On a alors
$$\int_\Omega |f|^p \, d\mu \le \int_\Omega (1 + |f|^r) \, d\mu = \mu(\Omega) + \int_\Omega |f|^r \, d\mu < +\infty$$
car f est dans L^r et donc f est dans L^p.
Si $f \in L^\infty$, on a $|f(x)|^p \le \|f\|_\infty^p$ et donc $\int_\Omega |f|^p \, d\mu \le \|f\|_\infty^p \mu(\Omega) < +\infty$. $\square$

Remarque 7.6. *Ce résultat est FAUX si μ n'est pas finie. Par exemple, si $\mu = \lambda$, on remarque que la fonction identiquement égale à 1 est dans L^∞, alors qu'elle n'appartient à aucun L^p pour $1 \le p < +\infty$.*

Soient f et g deux éléments de la même classe : $k = f - g \in V$. D'après l'inégalité triangulaire $\|f\|_p \leq \|g\|_p + \|k\|_p = \|g\|_p$.
De même $\|g\|_p \leq \|f\|_p + \|k\|_p = \|f\|_p$, d'où $\|f\|_p = \|g\|_p$. La semi-norme passe donc au quotient : pour $f \in L^p$, on note $\|f\|_p = \|g\|_p$ où g est un représentant quelconque de la classe f. Évidemment, $f \mapsto \|f\|_p$ est encore une semi-norme sur L^p.
Mais en réalité, $f \mapsto \|f\|_p$ est une norme sur L^p. En effet, supposons $\|f\|_p = 0$. Soit g un représentant de f. On a $\|g\|_p = 0$, donc $g \in V$, ce qui signifie que g est dans la classe de 0, donc f est le zéro de L^p. On a donc prouvé le résultat suivant.

Théorème 7.7. *Soit $1 \leq p < +\infty$. Alors $(L^p, \|\cdot\|_p)$ est un espace vectoriel normé.*

Bien que $\mathcal{L}^p$ ne soit pas un espace vectoriel normé, on pourra lire fréquemment pour des fonctions $(f_n)_{n\geq 1}, f$ de $\mathcal{L}^p$: $(f_n)_{n\geq 1}$ converge dans $\mathcal{L}^p$ (ou parfois $(f_n)_{n\geq 1}$ converge dans L^p) vers f. Cela signifie que $\lim\limits_{n\to+\infty} \|f_n - f\|_p = 0$, ou de manière équivalente, que la suite des classes dans L^p des éléments de $(f_n)_{n\geq 1}$ converge dans L^p vers la classe de f dans L^p.

7.2 Complétude de L^p

Théorème 7.8. *Pour tout $p \in [1, +\infty[$, L^p est complet.*

En d'autres termes, il s'agit de montrer que $(L^p, \|\cdot\|_p)$ est un espace de Banach.

Afin de démontrer ce résultat, nous introduisons deux lemmes qui sont en fait les étapes de la preuve.

Lemme 7.9. *Soit (f_n) une suite d'éléments de L^p avec*

$$\sum_{n=1}^{+\infty} \|f_n\|_p < +\infty.$$

Alors la suite $\left(\sum\limits_{k=1}^{n} f_k\right)$ converge dans L^p quand n tend vers l'infini.

Démonstration. On note g_n un représentant de f_n. On va montrer qu'il existe une fonction g dans $\mathcal{L}^p$ telle que $\left\|\sum\limits_{k=1}^{n} g_k - g\right\|_p$ tend vers 0, ce qui donnera la convergence de la suite $\left(\sum\limits_{k=1}^{n} f_k\right)$ vers la classe de g.

Supposons d'abord que les g_k sont positives. Dans ce cas, la suite de fonctions $S_n = \sum_{k=1}^{n} g_k$ converge simplement vers une fonction g mesurable (éventuellement infinie en certains points). Cependant, d'après l'inégalité triangulaire

$$\int_\Omega S_n^p \, d\mu \leq \left(\sum_{k=1}^{n} \|g_k\|_p \right)^p,$$

et donc d'après le théorème de convergence monotone

$$\int_\Omega g^p \, d\mu \leq \left(\sum_{k=1}^{+\infty} \|g_k\|_p \right)^p < +\infty.$$

Ainsi g est dans $\mathcal{L}^p$. Soient n et n' des entiers tels que $n' \geq n$. On a

$$(S_{n'} - S_n)^p = \left(\sum_{k=n+1}^{n'} g_k \right)^p.$$

Faisons tendre n' vers $+\infty$: d'après le théorème de convergence dominée, on a

$$\int_\Omega (g - S_n)^p \, d\mu = \lim_{n' \to +\infty} \int_\Omega (S_{n'} - S_n)^p \, d\mu,$$

d'où

$$\|g - S_n\|_p = \lim_{n' \to +\infty} \|S_{n'} - S_n\|_p.$$

Cependant, d'après l'inégalité triangulaire

$$\|S_{n'} - S_n\|_p \quad \leq \quad \sum_{k=n+1}^{n'} \|g_k\|_p \leq \sum_{k=n+1}^{+\infty} \|g_k\|_p,$$

d'où

$$\|g - S_n\|_p \leq \sum_{k=n+1}^{+\infty} \|g_k\|_p.$$

Mais on reconnaît là le reste d'une série convergente, donc $\|g - S_n\|_p$ tend bien vers 0 lorsque n tend vers l'infini.

Dans le cas général, écrivons $g_k = g_k^+ - g_k^-$. On définit évidemment $g^+ = \sum g_k^+$ et $g^- = \sum g_k^-$, $S_n^+ = \sum_{k=1}^{n} g_k^+$, $S_n^- = \sum_{k=1}^{n} g_k^-$. La série de terme général $\|g_k^+\|_p$ est convergente car $\|g_k^+\|_p \leq \|g_k\|_p$.
On montre ainsi que $\|S_n^+ - g^+\|_p$ tend bien vers 0, de même que $\|g^- - S_n^-\|_p$ tend bien vers 0. Enfin, l'inégalité triangulaire permet de conclure que la quantité $\|g - S_n\|_p$ tend bien vers 0. □

Ainsi, on a montré que dans L^p, toute série absolument convergente est convergente. Pour conclure, il suffit de s'appuyer sur le résultat d'analyse suivant.

Lemme 7.10. *Un espace vectoriel normé où toute série absolument convergente converge est complet.*

Démonstration. Remarquons d'abord que si une suite de Cauchy admet une sous-suite convergente, elle converge. En effet supposons que la suite (x_n) est de Cauchy avec x_{n_k} qui converge vers ℓ. Soient k_0 tel que $\|x_{n_k} - \ell\| \leq \varepsilon/2$ pour $k \geq k_0$ et b_0 tel que $\|x_k - x_{k'}\| \leq \varepsilon/2$ lorsque k et k' dépassent b_0. Alors $\|x_n - \ell\| \leq \varepsilon$ dès que n dépasse $\max(b_0, n_{k_0})$.

Soit maintenant (x_n) une suite de Cauchy dans un espace où toute série absolument convergente converge. On pose $n_0 = 1$, puis pour $k \geq 1$

$$n_k = \inf\{n > n_{k-1} : i, i' \geq n \Longrightarrow \|x_i - x_{i'}\| \leq 2^{-k}\}.$$

Cette suite d'indices est strictement croissante et est bien définie car la suite (x_k) est de Cauchy. Par construction, $\|x_{n_k} - x_{n_{k+1}}\| \leq 2^{-k}$ pour $k \geq 1$, donc la série de terme général $x_{n_k} - x_{n_{k+1}}$ est absolument convergente. Comme on a fait l'hypothèse ici qu'une série absolument convergente est convergente, elle est donc convergente, ce qui veut dire que (x_{n_k}) est convergente. (x_n) est donc une suite de Cauchy qui admet une sous-suite convergente, elle est donc convergente. □

Théorème 7.11. *L'espace L^∞ est complet.*

Démonstration. Soit $(g_n)_{n\geq 1}$ une suite de Cauchy de L^∞. Pour tout $n \geq 1$, on note f_n une fonction de $\mathcal{L}^\infty$ qui est un représentant de la classe de g_n. On note

$$\begin{aligned} B = & \bigcup_{n\geq 1} \{\omega \in \Omega; |f_n(\omega)| > \|f_n\|_{\infty,\text{ess}}\} \\ & \cup \bigcup_{n\geq 1, p\geq 1} \{\omega \in \Omega; |f_n(\omega) - f_p(\omega)| > \|f_n - f_p\|_{\infty,\text{ess}}\} \end{aligned}$$

Comme B est la réunion dénombrable d'ensembles de mesure nulle sous μ, on a $\mu(B) = 0$. Posons $G = \Omega \backslash B$. La suite des fonctions $(f_n \mathbb{1}_G)_{n\geq 1}$ est une suite de Cauchy dans l'espace des fonctions bornées sur Ω muni de la norme infinie : comme cette suite de fonctions est à valeurs dans un espace complet, elle converge vers une fonction f. La fonction f est mesurable, comme limite ponctuelle de fonctions mesurables. Il est maintenant aisé de constater que la suite (g_n) converge dans L^∞ vers la classe de f (mettons g). Cela découle de l'inégalité

$$\|g_n - g\|_{\infty,\text{ess}} = \|f_n \mathbb{1}_G - f\|_{\infty,\text{ess}} \leq \|f_n \mathbb{1}_G - f\|_\infty.$$

□

Théorème 7.12. *Soit $1 \leq p \leq +\infty$. Soient $f, (f_n)_{n\geq 1}$ des fonctions dans $\mathcal{L}^p$ telles que $(f_n)_{n\geq 1}$ converge dans $\mathcal{L}^p$ vers f. Alors, il existe une suite strictement croissante d'indices $(n_k)_{k\geq 1}$ telle que $(f_{n_k})_{k\geq 1}$ converge presque partout vers f.*

Démonstration. On pose $g_n = |f - f_n|^p$. On sait que (g_n) converge dans $\mathcal{L}^1$ vers 0 et nous devons montrer l'existence d'une suite strictement croissante d'indices $(n_k)_{k\geq 1}$ telle que $(g_{n_k})_{k\geq 1}$ converge presque partout vers 0.
On pose $n_0 = 1$, puis pour $k \geq 1$:

$$n_k = \inf\{n > n_{k-1} : i, i' \geq n \Longrightarrow \|g_i - g_{i'}\|_1 \leq 2^{-k}\}.$$

Cette suite d'indices est strictement croissante et est bien définie car (g_k) est de Cauchy dans $\mathcal{L}^1$. Par construction, $\|g_{n_k} - g_{n_{k+1}}\|_1 \leq 2^{-k}$ pour $k \geq 1$, donc la série de terme général $\|g_{n_k} - g_{n_{k+1}}\|_1$ est convergente. Mais

$$\sum_{k=1}^{+\infty} \|g_{n_k} - g_{n_{k+1}}\|_1 = \sum_{k=1}^{+\infty} \int |g_{n_k} - g_{n_{k+1}}|\, d\mu = \int \sum_{k=1}^{+\infty} |g_{n_k} - g_{n_{k+1}}|\, d\mu.$$

La fonction positive

$$\sum_{k=1}^{+\infty} |g_{n_k} - g_{n_{k+1}}|$$

est intégrable, elle est donc en particulier finie presque partout. En un point x tel que

$$\sum_{k=1}^{+\infty} |g_{n_k}(x) - g_{n_{k+1}}(x)| < +\infty,$$

la suite $(g_{n_k}(x))_{k\geq 1}$ converge. Ainsi $(g_{n_k})_{k\geq 1}$ converge presque partout vers une fonction g^* positive ou nulle. Mais d'après le lemme de Fatou,

$$\int g^* d\mu = \int \underline{\lim}_{k\to+\infty} g_{n_k} d\mu \leq \underline{\lim}_{k\to+\infty} \int g_{n_k}\, d\mu = 0,$$

donc g^* est nulle $\mu-$presque partout, ce qui achève la preuve. $\square$

En revanche, la convergence dans L^p n'entraîne pas la convergence presque partout. La plupart des contre-exemples sont basés sur le modèle suivant, dit « phénomène de bosse glissante ». Pour tout $n \geq 1$, on se donne un recouvrement de $[-n, n]$ par des ensembles $A_{n,1}, \dots, A_{n,N_n}$, de telle manière que

$$\lim_{n\to+\infty} \max_{1\leq i\leq N_n} \lambda(A_{n,i}) = 0.$$

On peut prendre par exemple $N_n = n^2$ et $A_{n,k} = [\frac{k-1}{n}, \frac{k}{n}] \cup [-\frac{k}{n}, -\frac{k-1}{n}]$, ce qui nous donne $\max\limits_{1\leq i\leq N_n} \lambda(A_{n,i}) = \frac{2}{n}$.

Notons que $E = \{(n,i); n \in \mathbb{N}^* \text{ et } i \in \{1, \dots, N_n\}\}$ est dénombrable. On peut donc choisir une bijection ϕ de $\mathbb{N}$ dans E. Soit $x \in \mathbb{R}$. Pour tout $n \geq |x|$, on a $x \in [-n, n] \subset \bigcup\limits_{i=1}^{N_n} A_{n,i}$. Il existe donc $i \in \{1, \dots, N_n\}$ tel que $x \in A_{n,i}$.

Ainsi l'ensemble des $e \in E$ tels que $x \in A_e$ est infini, donc $x \in \overline{\lim}_{n\to+\infty} A_{\phi(n)}$.

Ainsi $\mathbb{R} = \overline{\lim_{n\to+\infty}} A_{\phi(n)}$. Soit $\varepsilon > 0$. Il existe N_0 tel que $\max_{1\le i\le N_n} \lambda(A_{n,i}) < \varepsilon$ pour $n > N_0$. On voit ainsi que l'ensemble des $e \in E$ tels que $\lambda(A_e) \geq \varepsilon$ est inclus dans l'ensemble fini $\{(n,i); n \in \{1,\ldots,N_0\}$ et $i \in \{1,\ldots,N_n\}\}$. Par suite, l'ensemble des n tels que $\lambda(A_{\phi(n)}) \geq \varepsilon$ est également fini, ce qui montre que $\lim_{n\to+\infty} \lambda(A_{\phi(n)}) = 0$.

On conclut que $f_n = \mathbb{1}_{A_{\phi(n)}}$ converge dans L^p vers 0, tandis que

$$\forall x \in \mathbb{R}, \qquad \overline{\lim_{n\to+\infty}} f_n(x) = 1.$$

7.3 Théorèmes d'approximation

Théorème 7.13. *Soit $\mathcal{S}$ l'ensemble des fonctions simples g sur $(\Omega,\mathcal{F})$ telles que*

$$\mu(\{x \in \Omega; g(x) \neq 0\}) < +\infty.$$

Pour tout $p \in [1,+\infty[$, $\mathcal{S}$ est dense dans $\mathcal{L}^p(\mu)$ (et donc les classes de ces fonctions sont denses dans $L^p(\mu)$).

Démonstration. Il est facile de voir que $\mathcal{S}$ est inclus dans $\mathcal{L}^p(\mu)$. Soit $f \in \mathcal{L}^p$. Supposons $f \geq 0$ et prenons f_n comme dans le lemme 4.7. On a

$$2^{-np}\, \mathbb{1}_{\{f_n>0\}} \leq f_n^p\, \mathbb{1}_{\{f_n>0\}} \leq f^p,$$

d'où

$$2^{-np}\, \mu(f_n > 0) \leq \int_\Omega f^p\, d\mu,$$

et donc $f_n \in \mathcal{S}$. Or $|f_n - f|^p \leq f^p$, donc d'après le théorème de convergence dominée, $\int_\Omega |f_n - f|^p\, d\mu$ tend vers 0, c'est-à-dire que f_n tend vers f dans $\mathcal{L}^p$. Le cas général s'ensuit en séparant partie positive et partie négative, comme dans la preuve du théorème 7.8. $\square$

Théorème 7.14. *Soit $p \in [1,+\infty[$. Les classes des fonctions continues à support compact forment une partie dense dans $L^p(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$.*

Démonstration. Notons A l'adhérence dans $\mathcal{L}^p$ de l'ensemble des fonctions continues à support compact. On commence par montrer que A contient les indicatrices des compacts de $\mathbb{R}^d$. Soit en effet K un compact de $\mathbb{R}^d$. On pose, pour $n \geq 1$, $f_n(x) = (1 - nd(x,K))^+$. Comme K est fermé, f_n converge simplement vers l'indicatrice de K. De plus, on a $0 \leq f_n(x) - \mathbb{1}_K \leq \mathbb{1}_{K+\overline{B}(0,1)}$ et $|f_n(x) - \mathbb{1}_K|^p \leq \mathbb{1}_{K+\overline{B}(0,1)}$, donc par convergence dominée, f_n converge dans $\mathcal{L}^p$ vers l'indicatrice de K. Soit alors B un borélien tel que $\lambda(B) < +\infty$. D'après le théorème B.13 en annexe, la mesure de Lebesgue est régulière ;

pour $\varepsilon > 0$ on peut trouver K compact avec $K \subset B$ et $\lambda(B) - \varepsilon \leq \lambda(K)$. On a ainsi $\|\mathbb{1}_K - \mathbb{1}_B\|_p^p = \lambda(B \backslash K) \leq \varepsilon$, ce qui montre que $\mathbb{1}_B \in A$. Comme A est un espace vectoriel, il contient les combinaisons linéaires des indicatrices des boréliens de mesure finie, donc $\mathcal{L}^p$ tout entier vu le théorème 7.13. □

7.4 Exercices sur les espaces L^p

7.4.1 Exercices corrigés

Exercice 176. Soient $p_0 > 1$ et X une variable aléatoire positive telle que $X \in L^{p_0}(\Omega, \mathcal{F}, \mathbb{P})$. Pour $p \in]0, p_0]$, on pose $N(p) = (\mathbb{E}[X^p])^{1/p}$.

1. En appliquant l'inégalité de Hölder aux fonctions X^p et 1 avec des exposants bien choisis, montrer que la fonction $p \mapsto N(p)$ est croissante sur $]0, p_0]$.
2. Montrer que $N(p)$ admet une limite réelle lorsque p tend vers 0. Déterminer cette limite dans le cas où X suit la loi uniforme sur $[0, 1]$.
3. Soit $p \in]0, p_0]$. On définit la fonction
$$\begin{array}{rcl} F : \mathbb{R}_+ \times \mathbb{R}_+ & \to & \mathbb{R}_+ \\ (x, t) & \mapsto & pt^{p-1}\mathbb{1}_{\{x \geq t\}}. \end{array}$$
Montrer que $F \in L^1(\mathbb{R}_+^2, \mathcal{B}(\mathbb{R}_+^2), \mathbb{P}_X \otimes \lambda)$, puis que
$$\mathbb{E}[X^p] = \int_{[0,+\infty[} pt^{p-1}\mathbb{P}(X \geq t)\, d\lambda(t).$$
Note : on rappelle que $\mathbb{1}_{\{x \geq t\}}$ vaut 1 si $x \geq t$, 0 sinon.
4. Soit X une variable aléatoire positive. On note
$$\|X\|_{\infty,\text{ess}} = \sup\{M > 0 : \mathbb{P}(X \geq M) > 0\}.$$
On suppose dorénavant que X est telle que $\|X\|_{\infty,\text{ess}} < +\infty$.
   (a) Soit $M > \|X\|_{\infty,\text{ess}}$. Montrer que pour tout $p > 1$, $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ avec $N(p) \leq M$.
   (b) Montrer que $\overline{\lim_{p \to +\infty}} N(p) \leq \|X\|_{\infty,\text{ess}}$.
   (c) Soit $M < \|X\|_{\infty,\text{ess}}$. Montrer que pour tout $p > 1$, on a
   $N(p) \geq M\mathbb{P}(X \geq M)^{1/p}$, puis que $\underline{\lim}_{p \to +\infty} N(p) \geq M$.
   (d) Montrer que $\lim_{p \to +\infty} N(p) = \|X\|_{\infty,\text{ess}}$.
5. Montrer que la limite de $N(p)$ lorsque p tend vers 0 vaut 0 si $\log X$ n'est pas intégrable, $\exp(\mathbb{E}[\log X])$ sinon.

→ indication → solution

Exercice 177. Soit $(x_n)_{n\geq 1}$ le terme général d'une suite positive de limite nulle, telle que la série de terme général (x_n) diverge.
On pose $s_n = x_1 + \cdots + x_n$ et, pour x réel, on note $\{x\}$ la partie fractionnaire de x ($\{x\} = x - \lfloor x \rfloor$)). On pose $A_n = [\{s_n\}, \{s_{n+1}\}]$ si $\{s_n\} \leq \{s_{n+1}\}$, $A_n = \varnothing$ sinon. Montrer que

$$\overline{\lim_{n\to+\infty}} \ A_n \supset]0,1[.$$

Montrer que la suite $f_n(x) = \mathbb{1}_{A_n}(x)$ converge dans L^1 vers 0, mais ne converge pas λ-presque partout vers 0.

Quel est l'ensemble des valeurs d'adhérence de la suite $(\{s_n\})_{n\geq 1}$? → indication → solution

Exercice 178. *Théorèmes d'Egoroff et de Lusin.*

1. Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré avec $\mu(\Omega) < +\infty$. Soit $(f_n)_{n\geq 1}$ une suite de fonctions $(\Omega, \mathcal{F})-(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurables convergeant μ-presque partout vers f. Montrer que pour tout $\varepsilon > 0$, il existe $A \in \mathcal{F}$ tel que $\mu(A) < \varepsilon$ et tel que f_n converge uniformément vers f sur $\Omega\backslash A$.
   Indication : poser $B_{k,n} = \bigcap_{i\geq n} \left\{x \in E : |f_i(x) - f(x)| > \frac{1}{k}\right\}$, puis montrer qu'il existe n_k tel que $\mu(B_{k,n_k}) \leq \frac{\varepsilon}{2^k}$ et poser $A = \bigcup_{k\geq 1} B_{k,n_k}$. Ce résultat est le théorème d'Egoroff.
2. Soient $[a,b]$ un intervalle compact de $\mathbb{R}$ et $f \in L^1([a,b])$. Montrer que pour tout $\varepsilon > 0$, il existe un ensemble $F \subset [a,b]$ tel que f est continue sur F et $\lambda([a,b]\backslash F) \leq \varepsilon$. Ce résultat est le théorème de Lusin.
   Indication : on rappelle que les fonctions continues sont denses dans $L^1([a,b])$.
3. Étendre le théorème de Lusin aux fonctions mesurables à valeurs réelles.

→ indication → solution

Exercice 179. *Étude des variations de la fonction Gamma.*
On rappelle que la relation $\Gamma(x+1) = x\Gamma(x)$ permet de prolonger la fonction Γ à $\mathbb{R}\backslash\{-n; n \in \mathbb{N}\}$.

1. Montrer que la fonction $\frac{\Gamma'}{\Gamma}$ est croissante sur $]0, +\infty[$.
2. En déduire que $\frac{\Gamma'}{\Gamma}$ est croissante sur chaque intervalle de la forme $]-(n+1), -n[$, avec $n \geq 0$.
3. Sur chaque intervalle $]-(n+1), -n[$, avec $n \geq 0$, étudier les variations de Γ (croissance, convexité, limites aux bords).

→ indication → solution

7.4.2 Exercices non corrigés

Exercice 180. Soit f une fonction de $\mathbb{R}$ dans $\mathbb{R}$ intégrable et soit $\tilde{f}$ la classe de f dans $L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$. Montrer que $\tilde{f}$ contient au plus une fonction continue.
$\rightarrow$ indication

Exercice 181. Étudier l'appartenance à $\mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ et à $\mathcal{L}^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ des fonction suivantes :

1. $f(t) = e^{-|t|}$.
2. $g(t) = \frac{\sin t}{t}$.
3. $h(t) = \frac{1}{\sqrt{|t|}(1+t^2)}$.

$\rightarrow$ indication

Exercice 182. Étudier dans $\mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ et dans $\mathcal{L}^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ la convergence des suites suivantes :

1. $f_n(t) = \sqrt{n}\exp(-n^2t^2)$.
2. $g_n(t) = \frac{n^2 \sin(nt)}{2\pi}\mathbb{1}_{[-\pi/n, \pi/n]}(t)$.
3. $h_n(t) = \frac{2}{\pi n^2}\sqrt{n^2 - t^2}\mathbb{1}_{[-/n, n]}(t)$.

$\rightarrow$ indication

Exercice 183. Soit $E = \{f \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)\ ;\ |f| \leq 1\ \lambda - p.p\}$. Montrer que E est un sous-ensemble fermé de $\mathcal{L}^1(\mathbb{R}, \lambda)$.
$\rightarrow$ indication

Exercice 184. Montrer que la fonction $f\ :\ x \mapsto \dfrac{1}{\sqrt{x}(1 + |\log(x)|)}$ est dans $\mathcal{L}^p(]0, +\infty[, \lambda)$ si et seulement si $p = 2$.
$\rightarrow$ indication

Exercice 185. Soient $p \geq 1$ et $C_{\mathbb{N}^*}$ la mesure de comptage sur $\mathbb{N}^*$. On note simplement $\ell^p(\mathbb{N}^*)$ pour $L^p(\mathbb{N}^*, \mathcal{P}(\mathbb{N}^*), C_{\mathbb{N}^*})$ Donner un exemple de fonction qui est dans $\ell^p(\mathbb{N}^*)$ pour tout $p > 1$, mais qui n'est pas dans $\ell^1(\mathbb{N}^*)$. $\rightarrow$ indication

Exercice 186. Montrer que si f et g appartiennent à $\mathcal{L}^1(X, \mu)$ alors $\sqrt{|f^2 + g^2|}$ appartient aussi à $\mathcal{L}^1(X, \mu)$.
$\rightarrow$ indication

Exercice 187. Soient $\alpha \in \mathbb{R}$ et $p \in [1,\ +\infty[$. On note f_α l'application de $]0,\ +\infty[$ dans $]0,\ +\infty[$ définie par $f_\alpha(x) = x^\alpha$.

1. Pour quelle(s) valeur(s) de α, la fonction f_α est-elle dans $\mathcal{L}^p(]0, 1], \lambda)$? Calculer alors les normes de f_α dans chacun de ces espaces.

2. Même question avec les espaces $\mathcal{L}^p([1,\infty[,\lambda)$.

$\rightarrow$ indication

Exercice 188. Donner un exemple de suite (f_n) dans $\mathcal{L}^1(X,\mu)$ telle que

1. (f_n) converge vers f presque partout mais (f_n) ne converge pas vers f dans $\mathcal{L}^1$;
2. (f_n) converge vers f dans $\mathcal{L}^1$ mais (f_n) ne converge pas vers f presque partout ;
3. (f_n) converge vers f presque partout, $(\int f_n d\mu)$ converge vers $\int f d\mu$, mais (f_n) ne converge pas vers f dans $\mathcal{L}^1$.

$\rightarrow$ indication

Exercice 189. Soient $(X,\ \mathcal{M},\ \mu)$ un espace mesuré avec $\mu(X) = 1$ et f, g des fonctions mesurables sur X à valeurs dans $[0,\ +\infty]$ telles que $fg \geq 1$. Montrer que l'on a $(\int_X f d\mu)(\int_X g d\mu) \ \geq\ 1$.

$\rightarrow$ indication

Exercice 190. Soient $f \in \mathcal{L}^p(X,\mu)$, $g \in \mathcal{L}^q(X,\mu)$ et r tel que $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$. Montrer que $fg \in \mathcal{L}^r(X,\mu)$ et que $||fg||_r \ \leq\ ||f||_p||g||_q$.

$\rightarrow$ indication

Exercice 191. *Inégalité de Hardy.*
Soit $p \in]1,+\infty[$. Pour f dans $\mathcal{L}^p(]0,+\infty[)$ et pour $x > 0$, on pose

$$T(f)(x) \ = \ \frac{1}{x}\int_{]0,x[} f d\lambda\ .$$

1. Montrer que $T(f)$ est bien définie sur $]0,+\infty[$.
2. On suppose dans cette question que f est positive continue à support compact.
   (a) Montrer que $T(f)$ est dérivable sur $]0,+\infty[$ et calculer sa dérivée.
   (b) Montrer que $T(f) \in \mathcal{L}^p(]0,+\infty[)$.
   (c) Montrer que $\displaystyle\int_{]0,\infty[} T(f)^p d\lambda \ = \ \frac{p}{p-1}\int_{]0,\infty[} T(f)^{p-1} f d\lambda$.
   (d) En déduire que $||T(f)||_p \ \leq\ \dfrac{p}{p-1}||f||_p$.
   (e) Montrer que cette inégalité reste vraie pour f de signe quelconque.
3. Soit $f \in \mathcal{L}^p(]0,+\infty[)$.

(a) Montrer que si (f_n) est une suite de fonctions continues à support compact qui converge vers f dans $\mathcal{L}^p(]0,+\infty[)$, alors $T(f_n)$ converge vers $T(f)$ $\lambda-$presque partout, puis que la suite $(T(f_n))$ est de Cauchy dans $L^p(]0,+\infty[)$ et enfin que $(T(f_n))$ converge vers $T(f)$ dans $L^p(]0,+\infty[)$.

(b) En déduire que $||T(f)||_p \;\leq\; \dfrac{p}{p-1}||f||_p$.

$\rightarrow$ indication

Exercice 192. Soit $(X,\mathcal{T},\mu)$ un espace mesuré tel que $\mu(X) < +\infty$. On se donne $p \in]1,+\infty[$ et $f : X \rightarrow \mathbb{R}$ une application mesurable. On suppose que pour toute fonction $g \in \mathcal{L}^p(X,\mu)$, la fonction fg est intégrable et il existe $C > 0$ telle que pour toute fonction $g \in \mathcal{L}^p(X,\mu)$ on ait $\left|\int fg\,d\mu\right| \;\leq\; C||g||_p$. Montrer que $f \in \mathcal{L}^q(X,\mu)$ où q est défini par $\frac{1}{p}+\frac{1}{q} \;=\; 1 \rightarrow$ indication

Chapitre 8

Convolution et transformation de Fourier

Dans ce chapitre, p est un réel appartenant à $[1, +\infty[$.

8.1 Produit de convolution

Tout d'abord, nous procédons à quelques remarques utiles pour la suite.

Si f_1, f_2 sont deux fonctions de $\mathcal{L}^1$ qui représentent le même élément de L^1, alors $\int f_1 \, d\mu$ et $\int f_2 \, d\mu$ sont égales. On peut donc se permettre d'écrire $\int f \, d\mu$ pour $f \in L^1$.

L'application $T_t : f \mapsto (x \mapsto f(x-t))$ passe au quotient dans $\mathcal{L}^1(\mathbb{R}^d)$, car si $f_1 = f_2$ presque partout, alors $f_1(.-t) = f_2(.-t)$ presque partout.

Théorème 8.1. *Pour toute fonction f dans L^p, l'application*

$$t \mapsto T_t f$$

est continue sur $\mathbb{R}^d$ pour $\| \cdot \|_p$.

En d'autres termes, la translation est continue dans L^p.

Démonstration. Comme $\|T_{t+h} f - T_t f\|_p = \|T_h(T_t f) - (T_t f)\|_p$, il suffit de montrer la continuité en 0. Commençons par le cas où f est une fonction continue à support compact. Comme f est continue, $T_h f$ tend simplement vers f. En utilisant le théorème de convergence dominée, on obtient alors la convergence dans $\mathcal{L}^p$ de $T_h f$ vers f.

Passons au cas général. D'après le théorème 7.14, on peut trouver g et h telles que $f = g + h$, g est continue à support compact et $\|h\|_p \leq \varepsilon$. On a

$$(T_t f - f) = (T_t g - g) + (T_t h - h),$$

d'où

$$\begin{aligned}\|T_t f - f\|_p &\leq \|T_t g - g\|_p + \|T_t h\|_p + \|h\|_p \\ &\leq \|T_t g - g\|_p + 2\|h\|_p.\end{aligned}$$

Cela entraîne, en faisant tendre t vers 0

$$\overline{\lim_{t\to 0}}\ \|T_t f - f\|_p \leq 2\varepsilon.$$

Comme cette inégalité est vraie pour tout ε, on en déduit que $\overline{\lim_{t\to 0}}\ \|T_t f - f\|_p = 0$, ce qui est le résultat voulu. □

8.1.1 Convolution dans $\mathcal{L}^1$

Soient f, g deux éléments de $\mathcal{L}^1(\lambda^d)$. On a alors

$$\begin{aligned}
& \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f(x-t)||g(t)|\ d\lambda^d(t) \right) d\lambda^d(x) \\
= & \int_{\mathbb{R}^d \times \mathbb{R}^d} |f(x-t)||g(t)|\ d\lambda^d \otimes \lambda^d(t,x) \quad \text{(Tonelli)} \\
= & \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f(x-t)||g(t)|\ d\lambda^d(x) \right) d\lambda^d(t) \\
= & \int_{\mathbb{R}^d} |g(t)| \left(\int_{\mathbb{R}^d} |f(x-t)|\ d\lambda^d(x) \right) d\lambda^d(t) \\
= & \int_{\mathbb{R}^d} |g(t)| \left(\int_{\mathbb{R}^d} |f(x)|\ d\lambda^d(x) \right) d\lambda^d(t) \\
= & \left(\int_{\mathbb{R}^d} |f(x)|\ d\lambda^d(x) \right) \left(\int_{\mathbb{R}^d} |g(t)|\ d\lambda^d(t) \right) < +\infty.
\end{aligned}$$

Ainsi, la fonction $f * g$ définie par

$$x \mapsto f * g(x) = \int_{\mathbb{R}^d} f(x-t) g(t)\ d\lambda^d(t)$$

est définie en presque tout point x et elle est dans $\mathcal{L}^1$: cette fonction est le produit de convolution de f par g.

Les arguments évoqués précédemment fonctionnent encore. Le produit de convolution "passe au quotient" et définit ainsi une application de $L^1 \times L^1$ dans L^1.

Au passage, nous avons démontré :

$$\|f * g\|_1 \leq \|f\|_1 \|g\|_1.$$

En reprenant le calcul précédant et en supposant que f et g sont dans $\mathcal{L}^1$, le théorème de Fubini permet alors d'écrire

$$\begin{aligned}
&\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} f(x-t)g(t)\; d\lambda^d(t)\right) d\lambda^d(x)\\
=\;& \int_{\mathbb{R}^d} g(t)\left(\int_{\mathbb{R}^d} f(x-t)\; d\lambda^d(x)\right) d\lambda^d(t)\\
=\;& \int_{\mathbb{R}^d} g(t)\left(\int_{\mathbb{R}^d} f(x)\; d\lambda^d(x)\right) d\lambda^d(t)\\
=\;& \left(\int_{\mathbb{R}^d} f(x)\; d\lambda^d(x)\right)\left(\int_{\mathbb{R}^d} g(t)\; d\lambda^d(t)\right),
\end{aligned}$$

soit

$$\int_{\mathbb{R}^d}(f*g)(x)\; d\lambda^d(x) = \left(\int_{\mathbb{R}^d} f(x)\; d\lambda^d(x)\right)\left(\int_{\mathbb{R}^d} g(t)\; d\lambda^d(t)\right). \tag{8.1}$$

8.1.2 Autres produits

Supposons maintenant que $g \in \mathcal{L}^1$ et que $f \in \mathcal{L}^p$. On a par l'inégalité de Hölder

$$\begin{aligned}
&\int |f(x-t)||g(t)|\; d\lambda^d(t)\\
=\;& \int |f(x-t)||g(t)|^{1/p}|g(t)|^{1/q}\; d\lambda^d(t)\\
\le\;& \left(\int |f(x-t)|^p|g(t)|\; d\lambda^d(t)\right)^{1/p}\left(\int |g(t)|\; d\lambda^d(t)\right)^{1/q}.
\end{aligned}$$

D'où

$$\int\left(\int |f(x-t)||g(t)|\; d\lambda^d(t)\right)^p d\lambda^d(x) \le \iint |f(x-t)|^p|g(t)|\; d\lambda^d(t)d\lambda^d(x)\|g\|_1^{p/q}.$$

Par ailleurs, on obtient par Fubini

$$\begin{aligned}
&\int\left(\int |f(x-t)|^p|g(t)|\; d\lambda^d(t)\right) d\lambda^d(x)\\
=\;& \int\left(\int |f(x-t)|^p|g(t)|\; d\lambda^d(x)\right) d\lambda^d(t)\\
=\;& \int\left(\int |f(x-t)|^p\; d\lambda^d(x)\right) |g(t)|d\lambda^d(t)\\
=\;& \int \|f\|_p^p|g(t)|d\lambda^d(t) = \|f\|_p^p\|g\|_1.
\end{aligned}$$

En résumé,

$$\int\left(\int |f(x-t)||g(t)|\; d\lambda^d(t)\right)^p\; d\lambda^d(x) \quad \le \|f\|_p^p\|g\|_1^{1+p/q}.$$

Ainsi, l'intégrale

$$\int f(x-t)g(t)\; d\lambda^d(t)$$

converge pour presque tout x et l'application

$$x \mapsto f * g(x) = \int f(x-t)g(t)\, d\lambda^d(t)$$

représente un élément de $\mathcal{L}^p$ avec

$$\int |f * g(t)|^p\, d\lambda^d(t) \leq \|f\|_p^p \|g\|_1^{1+p/q},$$

soit

$$\|f * g\|_p \leq \|f\|_p \|g\|_1.$$

Remarque 8.2 (importante)**.** *Quel que soit l'espace où l'on définit les fonctions, on a toujours*

$$\int f(x-t)g(t)\, d\lambda^d(t) = \int g(x-t)f(t)\, d\lambda^d(t)$$

pour les x tels que

$$\int |f(x-t)g(t)|\, d\lambda^d(t) < +\infty.$$

On en déduit ainsi que $f * g = g * f$ toutes les fois où cela a un sens.

8.1.3 Approximations de l'unité

Théorème 8.3. *Soit $(\phi_k)_{k\geq 1}$ une suite de fonctions mesurables positives telles que*

$$\forall k \geq 1 \quad \int_{\mathbb{R}^d} \phi_k(x)\, d\lambda^d(x) = 1$$

et pour tout $\delta > 0$,

$$\lim_{k\to+\infty} \int_{\mathbb{R}^d \setminus B(0,\delta)} \phi_k(x)\, d\lambda^d(x) = 0.$$

Alors

— *pour toute fonction f uniformément continue et bornée sur $\mathbb{R}^d$, la suite $(f * \phi_k)$ converge uniformément vers f.*

— *pour toute fonction f dans $\mathcal{L}^p$, la suite $(f * \phi_k)$ converge vers f dans $\mathcal{L}^p$.*

Démonstration. Soit $x \in \mathbb{R}^d$ et fixons $\delta > 0$ quelconque.

$$\begin{aligned} f * \phi_k(x) - f(x) &= \int_{\mathbb{R}^d} f(x-t)\phi_k(t)\, d\lambda^d(t) - f(x) \\ &= \int_{\mathbb{R}^d} (f(x-t) - f(x))\phi_k(t)\, d\lambda^d(t) \\ &= \int_{B(0,\delta)} (f(x-t) - f(x))\phi_k(t)\, d\lambda^d(t) \\ &\quad + \int_{\mathbb{R}^d \setminus B(0,\delta)} (f(x-t) - f(x))\phi_k(t)\, d\lambda^d(t). \end{aligned}$$

D'où

$$|f * \phi_k(x) - f(x)| \le \int_{B(0,\delta)} \omega_f(\delta)\phi_k(t)\, d\lambda^d(t) + \int_{\mathbb{R}^d\setminus B(0,\delta)} 2\|f\|_\infty \phi_k(t)\, d\lambda^d(t),$$

où ω_f désigne le module de continuité de f. En passant au supremum en x, on obtient

$$\|f * \phi_k - f\|_\infty \le \omega_f(\delta) + 2\|f\|_\infty \int_{\mathbb{R}^d\setminus B(0,\delta)} \phi_k(t)\, d\lambda^d(t).$$

Puis, faisant tendre k vers l'infini, on trouve

$$\overline{\lim_{k\to+\infty}}\ \|f * \phi_k - f\|_\infty \le \omega_f(\delta).$$

Comme f est uniformément continue, on obtient le premier résultat voulu en faisant tendre δ vers 0.

Prenons maintenant f dans $\mathcal{L}^p$. On a

$$\begin{aligned}
&|f * \phi_k(x) - f(x)| \\
&= |\int f(x-t) - f(x)\phi_k(t)^{\frac{1}{p}}\phi_k^{1-\frac{1}{p}}(t) d\, d\lambda^d(t)| \\
&\le \left(\int \phi_k(t) d\, d\lambda^d(t)|\right)^{1-1/p} \left(\int |f(x-t) - f(x)|^p \phi_k(t)\, d\lambda^d(t)\right)^{1/p} \\
&\le \left(|\int T_t f(x) - f(x)|^p \phi_k(t)\, d\lambda^d(t)\right)^{1/p}.
\end{aligned}$$

En élevant à la puissance p et en intégrant, le théorème de Tonelli nous donne

$$\|f * \phi_k - f\|_p^p \le \int_{\mathbb{R}^d} \|T_t f - f\|_p^p \phi_k(t)\, d\lambda^d(t).$$

Soit $\varepsilon > 0$. D'après le théorème 8.1, $\|T_t f - f\|_p^p$ tend vers 0 lorsque t tend vers 0. On peut donc se donner δ tel que $\|T_t f - f\|_p \le \varepsilon$ sur $B(0,\delta)$. Comme pour tout t, $\||T_t f - f\|_p^p \le (2\|f\|_p)^p$, en découpant comme dans la première partie, on a

$$\|f * \phi_k - f\|_p^p \le \varepsilon^p + (2\|f\|_p)^p \int_{\mathbb{R}^d\setminus B(0,\delta)} \phi_k(t)\, d\lambda^d(t).$$

En passant à la limite supérieure, on a $\overline{\lim_{k\to+\infty}}\ \|f * \phi_k - f\|_p^p \le \varepsilon^p$, soit $\overline{\lim_{k\to+\infty}}\ \|f * \phi_k - f\|_p \le \varepsilon$. On conclut en faisant tendre ε vers 0. □

Théorème 8.4. *Soit ϕ une fonction mesurable positive telle que*

$$\int_{\mathbb{R}^d} \phi(x)\, d\lambda^d(x) = 1.$$

Pour tout $k \geq 1$, posons $\phi_k(x) = k^d\phi(kx)$.
Alors, pour toute fonction f dans $\mathcal{L}^p$, la suite $(f * \phi_k)$ converge vers f dans $\mathcal{L}^p$.

Démonstration. Un changement de variable linéaire nous donne $\int_{\mathbb{R}^d\setminus B(0,\delta)} \phi_k(x)\ d\lambda(x) = \int_{\mathbb{R}^d} \mathbb{1}_{]k\delta,+\infty[}(\|x\|)\phi(x)$. En prenant $\delta = 0$, on voit que ϕ_k est d'intégrale 1. Le théorème de convergence dominée donne ensuite $\lim_{k\to+\infty} \int_{\mathbb{R}^d\setminus B(0,\delta)} \phi_k(x)\ d\lambda(x) = 0$, et on peut appliquer le théorème précédent. □

8.1.4 Régularisation

Théorème 8.5. *Soient $f \in \mathcal{L}^p(\mathbb{R}^d)$ et g une fonction de classe C^1 à support compact. Alors $f * g$ est de classe C^1 sur $\mathbb{R}^d$, avec*

$$D_x(f * g) = \int f(t) D_{x-t}g\ d\lambda^d(t).$$

Démonstration. Soit M tel que $g(x) = 0$ pour $\|x\| \geq M$. Soit $R > 0$. Par définition, on a

$$f * g(x) = \int f(x-t)g(t)\ d\lambda^d(t) = \int g(x-t)f(t)\ d\lambda^d(t).$$

Ici, c'est bien sûr la deuxième écriture qui va nous intéresser.
Supposons $\|x\| \leq R$. La différentielle de $g(x-t)f(t)$, vue comme une fonction de x, est $f(t)D_{x-t}g$. Bien entendu, on a

$$|f(t)D_{x-t}g| \leq |f(t)|\|Dg\|_\infty\ \mathbb{1}_{B(0,R+M)}(t).$$

Comme $f \in \mathcal{L}^p$ et $\|Dg\|_\infty\ \mathbb{1}_{B(0,R+M)} \in \mathcal{L}^q$, donc $|f|\|Dg\|_\infty\ \mathbb{1}_{B(0,R+M)}$ est dans $\mathcal{L}^1$. Le théorème de convergence dominée pour la différenciation sous le signe somme donne alors le résultat voulu. □

Corollaire 8.6. *Soient $f \in \mathcal{L}^p(\mathbb{R}^d)$ et g une fonction de classe C^k à support compact. Alors $f * g$ est de classe C^k sur $\mathbb{R}^d$, avec*

$$D_x^\alpha(f * g) = \int f(t) D_{x-t}^\alpha g\ d\lambda^d(t),$$

où on a supposé que le multi-indice $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$ vérifie
$\alpha_1 + \alpha_2 + \cdots + \alpha_d \leq k$.

On rappelle que $D_x^\alpha f$ est égale à l'évaluation de la différentielle partielle $\frac{\partial^{\alpha_1+\cdots+\alpha_d} f}{\partial^{\alpha_1}x_1 \dots \partial^{\alpha_d}x_d}$ au point $x = (x_1, \dots, x_d)$.

Démonstration. Par récurrence sur k. □

Corollaire 8.7. *Les fonctions C^∞ à support compact forment un ensemble dense dans $\mathcal{L}^p$.*

Démonstration. On peut toujours approcher une fonction quelconque de L^p par une fonction à support compact. Ensuite, cette dernière s'approche dans L^p par un polynôme ; cela provient immédiatement du Théorème 8.4 et du corollaire précédent. □

8.2 Transformée de Fourier

Définition. *Soit $f \in \mathcal{L}^1(\mathbb{R}^d)$. On appelle transformée de Fourier de f, et l'on note $\hat{f}$ la fonction à valeurs complexes définie sur $\mathbb{R}^d$ par*

$$\hat{f}(t) = \int e^{i\langle x,t\rangle} f(x) \, d\lambda^d(x).$$

Évidemment, $f \mapsto \hat{f}$ est linéaire, et comme $|e^{i\langle x,t\rangle} f(x)| \leq |f(x)|$, on a

$$\forall f \in \mathcal{L}^1, \quad \|\hat{f}\|_\infty \leq \|f\|_1.$$

Remarque 8.8. *Il existe de nombreuses définitions de la transformée de Fourier. Le lecteur trouvera par exemple*

$$\hat{f}(t) = \int e^{-i\langle x,t\rangle} f(x) \, d\lambda^d(x),$$

ou encore

$$\hat{f}(t) = \int e^{-2\pi i\langle x,t\rangle} f(x) \, d\lambda^d(x).$$

Il ne s'agit que d'une convention et il est naturel en probabilité d'utiliser la définition donnée ici, comme nous le verrons dans le chapitre suivant lors de la définition de la fonction caractéristique.

8.2.1 Propriétés élémentaires

Proposition 8.9. *Pour $f, g \in \mathcal{L}^1$, on a*

— $\widehat{f * g} = \hat{f}.\hat{g}$.
— $\widehat{T_x f}(t) = e^{i\langle x,t\rangle} \hat{f}(t)$.
— *Si* $g(t) = f(t/\alpha)$ *avec* $\alpha > 0$, *alors* $\hat{g}(t) = \alpha^d \hat{f}(\alpha t)$.
— *Si* $g(t) = f(t)e^{i\langle t,\theta\rangle}$, *alors* $\hat{g}(t) = \hat{f}(t + \theta)$.
— $\int f(x) \, d\lambda^d(x) = \hat{f}(0)$.

Démonstration. La première propriété mérite qu'on y consacre quelques lignes.

On a

$$\begin{aligned}
\widehat{f*g}(t) &= \int e^{i\langle x,t\rangle}\left(\int f(x-u)g(u)\,d\lambda^d(u)\right)\,d\lambda^d(x)\\
&= \int\left(\int e^{i\langle x,t\rangle}f(x-u)g(u)\,d\lambda^d(u)\right)\,d\lambda^d(x)\\
&= \int\left(\int f(x-u)e^{i\langle t,x-u\rangle}g(u)e^{i\langle t,u\rangle}\,d\lambda^d(u)\right)\,d\lambda^d(x)\\
&= \int (F*G)(x)\,d\lambda^d(x),
\end{aligned}$$

où $F(x)=f(x)e^{i\langle x,t\rangle}$ et $G(x)=g(x)e^{i\langle x,t\rangle}$. Mais d'après l'équation (8.1), on a

$$\int (F*G)(x)\,d\lambda^d(x)=\left(\int F(x)d\lambda^d(x)\right)\left(\int G(x)d\lambda^d(x)\right),$$

d'où le résultat voulu. □

Théorème 8.10 (Riemann–Lebesgue)**.** *Soit f une fonction intégrable sur $\mathbb{R}$, à valeurs réelles ou complexes. Alors*

$$\hat{f}(t)=\int_{\mathbb{R}} e^{itx}f(x)\,dx \underset{|t|\to+\infty}{\longrightarrow} 0.$$

Démonstration. Traitons d'abord le cas où f est C^∞ à support compact. Soit f une telle fonction. Comme f est à support compact (notons $[a,b]$ un intervalle contenant ce support), on a par intégration par parties

$$\int_{\mathbb{R}} e^{itx}f(x)\,dx=\int_a^b e^{itx}f(x)\,dx=\frac{i}{t}f(a)e^{iat}-\frac{i}{t}f(b)e^{ibt}+\frac{i}{t}\int_{\mathbb{R}} e^{itx}f'(x)\,dx.$$

Chacun des trois termes est de la forme $1/t$ multiplié par une quantité bornée, donc l'expression a bien une limite nulle. On utilise alors la densité des fonctions C^∞ à support compact dans $\mathcal{L}^1$ (corollaire 8.7). Soit $f\in\mathcal{L}^1$ et $\varepsilon>0$. Il existe f_ε dans $\mathcal{L}^1$, C^∞ à support compact, telle que $\int_{\mathbb{R}}|f-f_\varepsilon|\,d\lambda\le\varepsilon$. Pour tout t réel, on a

$$\begin{aligned}
\left|\int_{\mathbb{R}} e^{itx}f(x)\,dx\right| &\le \left|\int_{\mathbb{R}} e^{itx}f_\varepsilon(x)\,dx\right|+\left|\int_{\mathbb{R}} e^{itx}(f(x)-f_\varepsilon(x))\,dx\right|\\
&\le \left|\int_{\mathbb{R}} e^{itx}f_\varepsilon(x)\,dx\right|+\varepsilon
\end{aligned}$$

d'où $\overline{\lim\limits_{t\to+\infty}}\ |\int_{\mathbb{R}} e^{itx}f(x)\,dx|\le\varepsilon$. Comme $\varepsilon>0$ peut être aussi petit que l'on veut, on a $\overline{\lim\limits_{t\to+\infty}}\ |\int_{\mathbb{R}} e^{itx}f(x)\,dx|=0$, ce qui donne le résultat voulu. □

8.2.2 Théorème d'inversion

Théorème 8.11. *Soit $f \in \mathcal{L}^1(\mathbb{R}^d)$ telle que $\hat{f} \in \mathcal{L}^1(\mathbb{R}^d)$, alors on a*

$$f(t) = \frac{1}{(2\pi)^d} \int e^{-i\langle x,t\rangle} \hat{f}(x)\, d\lambda^d(x) \quad \textit{p.p.}$$

Pour montrer ce résultat, on a besoin du lemme suivant :

Lemme 8.12. *Soit $G(t) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{\langle t,t\rangle}{2}}$. Alors $\hat{G}(t) = e^{-\frac{\langle t,t\rangle}{2}} = (2\pi)^{d/2} G(x)$.*

Démonstration. En utilisant le théorème de Fubini, il est aisé de voir qu'il suffit de démontrer le résultat en dimension 1. Le résultat a déjà été démontré dans l'exercice corrigé 68 du chapitre 4. Cela sera redémontré dans ce chapitre sous forme d'un exercice non corrigé, ainsi que, par deux méthodes différentes, dans le prochain chapitre. □

Démonstration du théorème 8.11. Pour $k \geq 1$, posons $G_k(t) = k^d G(kt)$. On a

$$\widehat{G}_k(t) = k^d k^{-d} \widehat{G}(t/k) = (2\pi)^{d/2} G(t/k).$$

On recommence cette procédure, ce qui nous donne

$$\widehat{\widehat{G}}_k(t) = (2\pi)^{d/2} k^d \widehat{G}(kt) = (2\pi)^d k^d G(kt) = (2\pi)^d G_k(t),$$

et comme G_k est paire

$$\widehat{\widehat{G}}_k(-t) = (2\pi)^d G_k(t),$$

soit

$$\frac{1}{(2\pi)^d} \int e^{-i\langle x,t\rangle} \widehat{G}_k(x)\, d\lambda^d(x) \quad = \quad G_k(t).$$

On a donc

$$\begin{aligned} f * G_k(t) &= \frac{1}{(2\pi)^d} \int \int e^{-i\langle t-u,x\rangle} \widehat{G}_k(x) f(u)\, d\lambda^d(x)\, d\lambda^d(u) \\ &= \frac{1}{(2\pi)^d} \int e^{-i\langle x,t\rangle} \widehat{G}_k(x) \hat{f}(x)\, d\lambda^d(x). \end{aligned}$$

En utilisant le théorème de convergence dominée, on voit que le terme de droite tend vers $\frac{1}{(2\pi)^d} \int e^{-i\langle x,t\rangle} \hat{f}(x)\, d\lambda^d(x)$ lorsque k tend vers l'infini. Mais d'après le théorème 8.4, le membre de gauche converge dans $\mathcal{L}^1$ vers f. Comme la convergence dans $\mathcal{L}^1$ entraîne la convergence d'une sous-suite presque partout, l'unicité de la limite donne l'égalité voulue. □

8.3 Exercices sur la transformation de Fourier

8.3.1 Exercices corrigés

Exercice 193. Calculer le produit de convolution $f * g$ des fonctions suivantes définies sur $\mathbb{R}$ $(a > 0, b > 0)$:

1. $f(x) = \exp(-\frac{x^2}{2a^2})$ et $g(x) = \exp(-\frac{x^2}{2b^2})$.
   (On admettra que $\int_{\mathbb{R}} \exp(-\frac{x^2}{2})\, d\lambda(x) = \sqrt{2\pi}$.)
2. $f(x) = \mathbb{1}_{[-a,a]}(x)$ et $g(x) = \mathbb{1}_{[-b,b]}(x)$.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 194. Pour tout entier n, on définit la fonction

$$g_n(x) = (1 - x^2)^n \, \mathbb{1}_{[-1,1]}(x).$$

On pose $a_n = \displaystyle\int_{\mathbb{R}} g_n(x)\, dx$ et $k_n = a_n^{-1} g_n$.

1. Montrer que la suite (a_n) tend vers 0 et que $a_n \geq \frac{2}{n+1}$ pour tout entier $n \geq 0$.
2. Soit f une fonction uniformément continue sur $\mathbb{R}$ et bornée. Montrer que $f * k_n$ converge uniformément vers f.
3. Soit f une fonction continue à support dans $[-\frac{1}{2}, \frac{1}{2}]$. Montrer que la restriction de $f * k_n$ à $[-\frac{1}{2}, \frac{1}{2}]$ est un polynôme de degré $\leq 2n$.
4. En déduire le théorème de Weierstrass : toute fonction continue d'un intervalle $[a, b]$ dans $\mathbb{R}$ est limite uniforme sur $[a, b]$ d'une suite de polynômes.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 195. Soit $E \in \mathcal{B}(\mathbb{R})$ tel que $0 < \lambda(E) < +\infty$.

1. Montrer que $x \mapsto f(x) = (\mathbb{1}_E * \mathbb{1}_{-E})(x)$ est continue sur $\mathbb{R}$.
2. En déduire que $E - E = \{x - y \,/\, x \in E,\ y \in E\}$ est un voisinage de 0.
   Indication : remarquer que si $f(x) \neq 0$, alors $x \in E - E$.

$\rightarrow$ indication $\rightarrow$ solution

8.3.2 Exercices non corrigés

Exercice 196. Soit $f = \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}$.

1. Déterminer $f * f$ et $f * f * f$.
2. On note $f^{(*)1} = f$ et pour $n \geq 2$, $f^{(*)n} = f^{(*)(n-1)} * f$. Vérifier que pour tout $n \geq 1$, $f^{(*)n} \in \mathcal{L}^1(\mathbb{R})$ et que $\|f^{(*)n}\|_1 = 1$.

3. Montrer que pour tout $n \geq 2$, $f^{(*)n}$ est de classe $\mathcal{C}^{n-2}$.

→ indication

Exercice 197. Soient f et $g \in \mathcal{L}^1(\mathbb{R}^n)$. Montrer que si f (resp. g) est nulle presque partout en dehors d'un ensemble A (resp. B) alors $f * g$ est nulle presque partout en dehors de $A + B = \{a + b; a \in A, b \in B\}$.
→ indication

Exercice 198. Soit f la fonction de $\mathbb{R}$ dans lui-même définie par $f(x) = e^{-\frac{x^2}{2}}$.

1. Déterminer la transformée de Fourier de f en remarquant que $\widehat{f}$ est solution d'une équation différentielle linéaire.
2. Soit A une matrice carrée réelle symétrique d'ordre n définie positive. Déterminer la transformée de Fourier de la fonction de $\mathbb{R}^n$ dans $\mathbb{R}$ définie par $f(x) = e^{-\langle Ax,x\rangle}$ pour $x \in \mathbb{R}^n$.

→ indication

Exercice 199. 1. Soit $f \in L^1(\mathbb{R}^n)$ telle que $f * f = 0$. Montrer que $f = 0$.

2. Montrer que $\mathcal{L}^1(\mathbb{R})$ n'a pas d'unité pour la convolution.

→ indication

Exercice 200. Déterminer la transformée de Fourier de la fonction indicatrice d'un intervalle $[a, b]$. Montrer que $\mathbb{1}_{[-1,1]} * \mathbb{1}_{[-1,1]}$ est la transformée de Fourier d'une fonction de $L^1(\mathbb{R})$ qu'on déterminera. → indication

Exercice 201. Calculer la transformée de Fourier de la fonction f de $\mathbb{R}$ dans lui-même définie par $f(x) = e^{-a|x|}$, pour $x \in \mathbb{R}$ (où $a > 0$). En déduire la transformée de Fourier de la fonction $g : x \mapsto \dfrac{1}{a^2 + x^2}$.
→ indication

Exercice 202. *Lemme de Parseval.*
Soient μ et ν deux mesures de probabilité sur $\mathbb{R}^d$, de transformée de Fourier respective $\widehat{\mu}$ et $\widehat{\nu}$. Montrer que

$$\int_{\mathbb{R}^d} \widehat{\mu}(x)\, d\nu(x) = \int_{\mathbb{R}^d} \widehat{\nu}(x)\, d\mu(x).$$

→ indication

Chapitre 9

Fonction génératrice et fonction caractéristique, transformée de Laplace

L'objet du présent chapitre est de montrer comment certaines propriétés des lois des variables (ou des vecteurs) aléatoires, peuvent être étudiées à l'aide de fonctions auxiliaires définies à partir des lois. Trois transformations sont étudiées ici

— la fonction caractéristique (ou transformée de Fourier), qui est l'outil générique,
— la fonction génératrice, adaptée au cas des variables à valeurs entières,
— la transformée de Laplace, adaptée au cas des variables à valeurs réelles positives.

9.1 Fonction génératrice d'une variable aléatoire à valeurs dans $\mathbb{N}$

Définition. *On appelle fonction génératrice d'une variable aléatoire X à valeurs dans $\mathbb{N}$ la fonction*

$$z \mapsto G_X(z) := \mathbb{E}[z^X] = \sum_{k=0}^{+\infty} \mathbb{P}(X = k) z^k.$$

Usuellement, on définit cette fonction sur l'intervalle réel $[-1, 1]$, mais elle est, en fait, toujours définie sur la boule unité complexe fermée, notée $\bar{B}(0, 1)$.

Remarque 9.1. *Si la loi de X est à support fini, alors G_X est un polynôme.*

9.1.1 Fonction génératrice et indépendance

Théorème 9.2. *Si deux variables aléatoires X et Y sont indépendantes, on a $G_{X+Y} = G_X G_Y$.*

Démonstration. Soit $z \in \bar{B}(0,1)$. On a

$$G_{X+Y}(z) = \mathbb{E}[z^{X+Y}] = \mathbb{E}[z^X z^Y] = \mathbb{E}[z^X]\mathbb{E}[z^Y] = G_X(z)G_Y(z).$$

□

9.1.2 Calculs de fonctions génératrices

Loi de Bernoulli

La fonction génératrice d'une loi de Bernoulli de paramètre p est $z \mapsto (1-p) + pz$, car si X suit une telle loi, on a

$$\mathbb{E}[z^X] = \mathbb{P}(X=0)z^0 + \mathbb{P}(X=1)z^1 = (1-p) + pz.$$

Loi binomiale

Si $X_1, \ldots, X_n$ sont des variables aléatoires indépendantes suivant la loi de Bernoulli de paramètre p, alors $S_n = X_1 + \cdots + X_n$ suit la loi binomiale de paramètres n et p. Ainsi, on déduit du théorème 9.2 que la fonction génératrice d'une loi binomiale de paramètres n et p est

$$G_{S_n}(z) = G_{X_1}(z) \times \cdots \times G_{X_n}(z) = ((1-p) + pz)^n.$$

Loi géométrique de paramètre $p \in]0,1[$

Soient X suivant la loi géométrique de paramètre p et $z \in B(0,1)$. La fonction génératrice de X vaut alors, par le théorème de transfert :

$$\begin{aligned} G_X(z) &= \mathbb{E}[z^X] = \sum_{n=1}^{+\infty} p(1-p)^{n-1} z^n = \sum_{n=0}^{+\infty} p(1-p)^n z^{n+1} \\ &= pz \sum_{n=0}^{+\infty} ((1-p)z)^n = \frac{pz}{1-(1-p)z}. \end{aligned}$$

Loi de Poisson

Soit X une variable aléatoire suivant une loi de Poisson de paramètre $\lambda > 0$. On a

$$\begin{aligned} G_X(s) &= \sum_{k=0}^{+\infty} \mathbb{P}(X = k)s^k = \sum_{k=0}^{+\infty} e^{-\lambda} \frac{\lambda^k}{k!} s^k = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(\lambda s)^k}{k!} \\ &= e^{-\lambda} e^{\lambda s} = e^{-\lambda(1-s)}. \end{aligned}$$

9.1.3 Fonction génératrice et loi

Théorème 9.3. *Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. Sur $[0,1[$, la fonction $z \mapsto G_X(z)$ est infiniment dérivable et ses dérivées sont toutes positives, avec*

$$G_X^{(n)}(s) = \mathbb{E}[X(X-1)\dots(X-n+1)s^{X-n}].$$

En particulier

$$\mathbb{P}(X = n) = \frac{G_X^{(n)}(0)}{n!},$$

ce qui montre que la fonction génératrice caractérise la loi.

Démonstration. La fonction $z \mapsto G_X(z)$ est la somme d'une série entière de rayon de convergence au moins égal à 1. Ainsi, $z \mapsto G_X(z)$ est holomorphe sur le disque unité ouvert et y est infiniment dérivable, avec pour tout z dans le disque ouvert unité :

$$G_X^{(n)}(z) = \sum_{k=n}^{+\infty} k(k-1)\dots(k-n+1)\mathbb{P}(X = k)z^{k-n}.$$

Il suffit maintenant d'appliquer le théorème de transfert pour constater que le membre de droite est l'espérance de $X(X-1)\dots(X-n+1)z^{X-n}$.

En prenant $z = 0$, on obtient

$$\begin{aligned} G_X^{(n)}(0) &= \mathbb{E}\left[X(X-1)\dots(X-n+1)\, \mathbb{1}_{\{X-n=0\}}\right] \\ &= \mathbb{E}\left[n(n-1)\dots(n-n+1)\, \mathbb{1}_{\{X-n=0\}}\right] \\ &= n!\mathbb{P}(X = n). \end{aligned}$$

La restriction à un intervalle de $\mathbb{R}$ d'une fonction holomorphe est évidemment une fonction infiniment dérivable. Lorsque $s \in [0,1[$, on a pour tout $\omega \in \Omega$:

$$X(\omega)(X(\omega)-1)\dots(X(\omega)-n)(X(\omega)-n+1)s^{X(\omega)-n} \geq 0.$$

Comme l'espérance d'une variable aléatoire positive est positive, le résultat s'ensuit. □

9.1.4 Application : convolution de lois de Poisson

Théorème 9.4. *Soient X et Y deux variables aléatoires indépendantes, X suivant une loi de Poisson de paramètre λ et Y une loi de Poisson de paramètre μ. Alors $X+Y$ suit une loi de Poisson de paramètre $\lambda+\mu$.*

Démonstration. On a vu qu'une variable aléatoire réelle suivant une loi de Poisson de paramètre λ admet pour fonction génératrice $s \mapsto e^{-\lambda(1-s)}$, ceci quel que soit $\lambda > 0$.
En particulier, il s'ensuit que $G_X(s) = e^{-\lambda(1-s)}$ et $G_Y(s) = e^{-\mu(1-s)}$. Or, on a $G_{X+Y}(s) = G_X(s)G_Y(s) = e^{-\lambda(1-s)}e^{-\mu(1-s)} = e^{-(\lambda+\mu)(1-s)}$. Ainsi $X+Y$ a la même fonction génératrice qu'une loi de Poisson de paramètre $\lambda+\mu$. Mais d'après le théorème 9.3, la fonction génératrice détermine la loi, donc $X+Y$ suit la loi de Poisson de paramètre $\lambda+\mu$. □

9.1.5 Fonction génératrice et espérance

Théorème 9.5. *Soit X une variable aléatoire à valeurs dans $\mathbb{N}$.
Alors $\mathbb{E}X < +\infty$ si et seulement si G_X admet une dérivée à gauche en 1. Dans ce cas $G'_X(1) = \mathbb{E}X$.*

Démonstration. On note ν la loi de X. Pour $x \in [0,1[$, on a

$$\frac{G_X(1) - G_X(x)}{1-x} = \sum_{n=0}^{+\infty} \frac{1-x^n}{1-x}\nu(n).$$

Pour tout n, on a $\frac{1-x^n}{1-x} = 1+x+\dots+x^{n-1}$: c'est donc une fonction croissante de x. De plus $\lim\limits_{x\to 1^-} \frac{1-x^n}{1-x} = n$. D'après le théorème de convergence monotone (on intègre sur $\mathbb{N}$ par rapport à la mesure de comptage), on a donc

$$\lim_{x\to 1^-} \frac{G_X(1) - G_X(x)}{1-x} = \sum_{n=0}^{+\infty} n\nu(n) = \int x \, d\nu(x) = \mathbb{E}X.$$

□

9.2 Fonctions caractéristiques

9.2.1 Motivations

La fonction caractéristique est un outil analogue à la fonction génératrice, qui permet de généraliser les techniques des fonctions génératrices aux variables aléatoires à valeurs réelles, et même aux vecteurs aléatoires. Dans d'autres branches des mathématiques, la fonction caractéristique est appelée "transformée de Fourier", que nous avons déjà présentée au chapitre 8. Si X est une variable aléatoire à valeurs dans $\mathbb{R}^d$, on peut considérer, pour $t \in \mathbb{R}^d$, $\exp(i\langle t, X\rangle)$ comme une variable aléatoire à valeurs complexes, ce qui signifie que ses parties réelle $\cos(\langle t, X\rangle)$ et imaginaire $\sin(\langle t, X\rangle)$ sont des variables aléatoires réelles. Comme ces variables sont bornées (par 1), elles admettent une espérance, qu'il est naturel d'écrire

$$\mathbb{E}\exp(i\langle t, X\rangle) = \mathbb{E}\cos(\langle t, X\rangle) + i\mathbb{E}\sin(\langle t, X\rangle).$$

Définition. *On appelle fonction caractéristique d'une mesure de probabilité μ sur $\mathbb{R}^d$ la fonction complexe définie en tout point de $\mathbb{R}^d$ par*

$$\forall t = (t_1, \dots, t_d) \in \mathbb{R}^d,$$
$$\phi_\mu(t_1, \dots, t_d) = \int_{\mathbb{R}^d} \exp(i(t_1 x_1 + t_2 x_2 + \cdots + t_d x_d))\, d\mu(x).$$

Par extension, on appelle fonction caractéristique d'un vecteur aléatoire X et on note ϕ_X la fonction caractéristique de sa loi. Ainsi

$$\forall t \in \mathbb{R}^d \quad \phi_X(t) = \phi_{\mathbb{P}_X}(t) = \int_{\mathbb{R}^d} \exp(i\langle t, x\rangle)\, d\mathbb{P}_X(x) = \mathbb{E}e^{i\langle t, X\rangle}.$$

Remarque 9.6. *La fonction caractéristique ne dépend en fait que de la loi de X, notée $\mathbb{P}_X$. Il s'agit tout simplement de la transformée de Fourier de la loi $\mathbb{P}_X$.*

On va démontrer ici un résultat d'analyse important, qui justifie la dénomination de fonction caractéristique et rend cet outil pertinent.

Théorème 9.7. *Soient μ et ν deux mesures de probabilité sur $\mathbb{R}^d$. On a*

$$\forall t \in \mathbb{R}^d \quad \phi_\mu(t) = \phi_\nu(t) \iff \mu = \nu.$$

Démonstration. On va se concentrer sur le cas de la dimension 1. Les dimensions plus grandes compliquent en effet les écritures sans apporter d'idée nouvelle.

En fait, on va établir la formule d'inversion suivante :

$$\mu(]a, b[) + \frac{1}{2}(\mu(a) + \mu(b)) = \lim_{T \to +\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \phi_\mu(t)\, dt.$$

Posons

$$I_T(a, b) = \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \phi_\mu(t)\, dt.$$

Remplaçons ϕ_μ par sa définition :

$$I_T(a, b) = \frac{1}{2\pi} \int_{-T}^{T} \int \frac{e^{-ita} - e^{-itb}}{it} e^{itx}\, d\mu(x)\, dt.$$

On peut utiliser le théorème de Fubini, ce qui nous donne

$$\begin{aligned} I_T(a,b) &= \frac{1}{2\pi}\int\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}e^{itx}\,dt\,d\mu(x) \\ &= \frac{1}{2\pi}\int\left(\int_{-T}^{T}\frac{e^{it(x-a)}-e^{it(x-b)}}{it}\,dt\right)\,d\mu(x). \end{aligned}$$

Or la formule de Moivre et la parité du cosinus donnent

$$\int_{-T}^{T}\frac{e^{it(x-a)}-e^{it(x-b)}}{it}\,dt = \int_{-T}^{T}\frac{\sin(t(x-a))-\sin(t(x-b))}{t}\,dt$$

d'où

$$I_T(a,b) = \int\left(\frac{\mathrm{sign}(x-a)}{2\pi}\int_{-|x-a|T}^{|x-a|T}\frac{\sin t}{t}\,dt - \frac{\mathrm{sign}(x-b)}{2\pi}\int_{-|x-b|T}^{|x-b|T}\frac{\sin t}{t}\,dt\right)\,d\mu(x).$$

De plus, l'application

$$y \mapsto \int_{-y}^{y}\frac{\sin t}{t}\,dt$$

est une application continue qui admet comme limite π lorsque y tend vers l'infini. En particulier, sa norme est bornée par une constante M.

La quantité apparaissant sous l'intégrale est donc bornée par M/π. Lorsque T tend vers l'infini, elle converge vers la fonction

$$I_{a,b} = \begin{cases} 0, & \text{si } x < a \\ 1/2, & \text{si } x = a \\ 1, & \text{si } x \in]a,b[\\ 1/2, & \text{si } x = b \\ 0, & \text{si } x > b. \end{cases}$$

Ainsi, $I_T(a,b)$ converge vers $\int I_{a,b}\,d\mu$, ce qui donne la convergence vers la limite annoncée. Si $\mu(a)=\mu(b)=0$, alors $I_{a,b}$ est $\mu-$presque sûrement l'indicatrice de $]a,b[$, ce qui donne $\int I_{a,b}\,d\mu = \int \mathbb{1}_{]a,b[}\,d\mu = \mu(]a,b[)$.

Ainsi, si deux mesures μ et ν ont la même fonction caractéristique, on a $\mu(]a,b[) = \nu(]a,b[)$ quels que soient a et b dans

$$\mathbb{R}\backslash\{x\in\mathbb{R}; \mu(\{x\})>0 \text{ ou } \nu(\{x\})>0\}.$$

Mais ces ensembles forment un π-système qui engendre la tribu, donc les deux mesures coïncident. □

Donnons une conséquence frappante de ce théorème qui nous sera utile dans l'étude des vecteurs gaussiens.

Théorème 9.8. *Soient X et Y deux vecteurs aléatoires sur $\mathbb{R}^d$ tels que pour tout $a \in \mathbb{R}^d$, $\langle X, a\rangle$ et $\langle Y, a\rangle$ ont même loi. Alors X et Y ont même loi.*

Démonstration. On va montrer que X et Y ont même fonction caractéristique, ce qui assurera qu'ils ont même loi. Soit $a \in \mathbb{R}^d$. On pose $Z = \langle X, a\rangle$ et $T = \langle Y, a\rangle$. Comme Z et T ont même loi, on a $\mathbb{E}e^{iZ} = \mathbb{E}e^{iT}$. Mais on a aussi par définition de Z et T : $\mathbb{E}e^{iZ} = \mathbb{E}e^{i\langle X,a\rangle} = \phi_X(a)$ et $\mathbb{E}e^{iT} = \mathbb{E}e^{i\langle Y,a\rangle} = \phi_Y(a)$, donc $\phi_X(a) = \phi_Y(a)$. Ainsi $\phi_X = \phi_Y$, donc X et Y ont même loi. □

9.2.2 Propriétés des fonctions caractéristiques

Théorème 9.9. *Soit μ une mesure de probabilité sur $\mathbb{R}^d$. On a les propriétés suivantes :*

i) $\phi_\mu(0) = 1$.

ii) $|\phi_\mu| \le 1$.

iii) *ϕ est uniformément continue sur $\mathbb{R}^d$.*

iv) *Si $e_1, \dots, e_p$ sont des éléments de $\mathbb{R}^d$, la matrice A de taille $p \times p$ définie par $a_{k,l} = \phi_\mu(e_k - e_l)$ est hermitienne positive.*

Démonstration. i) $\phi_\mu(0) = \int_{\mathbb{R}^d} e^{i\langle x,0\rangle} d\mu(x) = \int_{\mathbb{R}^d} 1 \; d\mu = 1$.

ii) $|\phi_\mu(t)| = |\int_{\mathbb{R}^d} e^{i\langle x,t\rangle} d\mu(x)| \le \int_{\mathbb{R}^d} |e^{i\langle x,t\rangle}| d\mu(x) = \int_{\mathbb{R}^d} 1 \; d\mu = 1$.

iii) Soient $t, t' \in \mathbb{R}^d$. On a

$$\begin{aligned} |\phi_\mu(t) - \phi_\mu(t')| &= \left| \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} - e^{i\langle t',x\rangle} \; d\mu(x) \right| \\ &= \left| \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} (1 - e^{i\langle (t'-t),x\rangle}) \; d\mu(x) \right| \\ &\le \int_{\mathbb{R}^d} |1 - e^{i\langle (t'-t),x\rangle}| \; d\mu(x). \end{aligned}$$

Il suffit maintenant de voir que la fonction $u \mapsto \int_{\mathbb{R}^d} |1 - e^{i\langle u,x\rangle}| \; d\mu(x)$ admet une limite nulle en 0 pour conclure. Or ce dernier point est assuré par le théorème de convergence dominée de Lebesgue, car pour $a, y \in \mathbb{R}$, on a $|1 - e^{iay}| \le \min(|ay|, 2)$.

iv) On remarque que, pour tout $p-$uplet de nombres réels $x_1, \dots, x_p$, on a

$$\begin{aligned}
0 &\leq \int_{\mathbb{R}^d} \left| \sum_{k=1}^{p} x_k \exp(i\langle e_k, x\rangle) \right|^2 d\mu(x) \\
&= \int_{\mathbb{R}^d} \sum_{k=1}^{p} \sum_{l=1}^{p} x_k x_l \exp(i\langle e_k - e_l, x\rangle)\, d\mu(x) \\
&= \sum_{k=1}^{p} \sum_{l=1}^{p} x_k x_l \int_{\mathbb{R}^d} \exp(i\langle e_k - e_l, x\rangle)\, d\mu(x) \\
&= \sum_{k=1}^{p} \sum_{l=1}^{p} x_k x_l \phi_\mu(e_k - e_l).
\end{aligned}$$

□

Exemple: Pour $n = 1$, soit f une fonction positive 2π-périodique intégrable sur $[-\pi, \pi[$, et soit μ la mesure définie par

$$d\mu(x) := \frac{f(x)}{2\pi}\, \mathbb{1}_{[-\pi,\pi[}(x)\, d\lambda(x).$$

$\phi_{\mu(k)}$ est alors le $k-$ième coefficient de Fourier de f. On prend alors classiquement $e_k = k$.
Ainsi, si f est non nulle, la matrice A est en fait définie positive (c'est-à-dire que la forme quadratique définie par A est strictement positive pour $v \neq 0$: pour tout vecteur colonne $v \neq 0$, on a $v^t A v > 0$) car

$$\begin{aligned}
\sum_{k=1}^{p} \sum_{l=1}^{p} x_k x_l \phi_\mu(k-l) = 0 &\Longrightarrow \int_{[-\pi,\pi[} \left| \sum_{k=1}^{p} x_k \exp(ikx) \right|^2 d\mu(x) = 0 \\
&\Longrightarrow \tfrac{1}{2\pi} \int_{[-\pi,\pi[} \left| \sum_{k=1}^{p} x_k \exp(ikx) \right|^2 f(x)\, d\lambda(x) = 0 \\
&\Longrightarrow \left| \sum_{k=1}^{p} x_k \exp(ikx) \right|^2 f(x) = 0 \text{ p.p. sur } [-\pi, \pi].
\end{aligned}$$

Comme un polynôme trigonométrique non nul n'a qu'un nombre fini de zéros sur $[-\pi, \pi]$, on en déduit que f est presque partout nulle, ce qu'on avait exclu. En fait, un théorème difficile nous dit que les propriétés énoncées ci-dessus sont largement suffisantes pour permettre d'affirmer qu'une fonction donnée est une fonction caractéristique. Il s'agit du théorème de Bochner, que nous admettrons ici.

Proposition 9.10 (Bochner)**.** *Soit $\phi : \mathbb{R}^n \to \mathbb{C}$ une fonction continue en 0, vérifiant $\phi(0) = 1$ et de type positif, c'est-à-dire que si $e_1, \dots, e_p$ sont des éléments quelconques de $\mathbb{R}^n$, la matrice A de taille $p \times p$ définie par*
$a_{k,l} = \phi(e_k - e_l)$ est hermitienne positive.

Alors il existe une unique mesure de probabilité μ sur $\mathbb{R}^n$ telle que $\phi = \phi_\mu$.

Le théorème très simple ci-après est d'usage courant.

Théorème 9.11. *Soient X un vecteur aléatoire de $\mathbb{R}^d$, A une application linéaire de $\mathbb{R}^d$ dans $\mathbb{R}^n$ et $b \in \mathbb{R}^n$. On pose $Y = AX + b$. Alors*

$$\forall t \in \mathbb{R}^n \quad \phi_Y(t) = e^{i\langle b,t\rangle}\phi_X(A^*t).$$

Démonstration. On écrit

$$\begin{aligned}\phi_Y(t) &= \mathbb{E}e^{i\langle Y,t\rangle} = \mathbb{E}e^{i\langle AX+b,t\rangle} = \mathbb{E}e^{i\langle AX,t\rangle}e^{i\langle b,t\rangle} \\ &= \mathbb{E}e^{i\langle X,A^*t\rangle}e^{i\langle b,t\rangle} = e^{i\langle b,t\rangle}\phi_X(A^*t).\end{aligned}$$

□

9.2.3 Fonction caractéristique et indépendance

Théorème 9.12. *Soient X et Y deux vecteurs aléatoires indépendants à valeurs dans $\mathbb{R}^d$. Alors*

$$\forall t \in \mathbb{R}^d, \quad \phi_{X+Y}(t) = \phi_X(t)\phi_Y(t).$$

Démonstration. On utilise l'indépendance de X et Y :

$$\phi_{X+Y}(t) = \mathbb{E}e^{i\langle t,X+Y\rangle} = \mathbb{E}e^{i\langle t,X\rangle}e^{i\langle t,Y\rangle} = \mathbb{E}e^{i\langle t,X\rangle}\mathbb{E}e^{i\langle t,Y\rangle} = \phi_X(t)\phi_Y(t).$$

□

Théorème 9.13. *Soient X et Y deux vecteurs aléatoires indépendants, X étant à valeurs dans $\mathbb{R}^n$ et Y à valeurs dans $\mathbb{R}^p$. Alors le vecteur (X,Y) de dimension $n+p$ admet comme fonction caractéristique la fonction*

$$\forall (s,t) \in \mathbb{R}^n \times \mathbb{R}^p, \quad \phi_{(X,Y)}(s,t) = \phi_X(s)\phi_Y(t).$$

Démonstration. On écrit

$$\begin{aligned}\phi_{(X,Y)}(s,t) &= \mathbb{E}e^{i(\langle s,X\rangle + \langle t,Y\rangle)} = \mathbb{E}[e^{i\langle s,X\rangle}e^{i\langle t,Y\rangle}] \\ &= \mathbb{E}e^{i\langle s,X\rangle}\mathbb{E}e^{i\langle t,Y\rangle} = \phi_X(s)\phi_Y(t).\end{aligned}$$

□

Corollaire 9.14. *Si μ et ν sont des mesures de probabilité respectivement définies sur $\mathbb{R}^n$ et $\mathbb{R}^p$, alors*

$$\forall (s,t) \in \mathbb{R}^n \times \mathbb{R}^p, \quad \phi_{\mu\otimes\nu}(s,t) = \phi_\mu \otimes \phi_\nu(s,t) = \phi_\mu(s)\phi_\nu(t).$$

Démonstration. Il suffit de considérer un couple (X,Y) de variables aléatoires de loi $\mu\otimes\nu$. Comme X et Y sont indépendantes, on a

$$\phi_{\mu\otimes\nu}(s,t) = \phi_{(X,Y)}(s,t) = \phi_X(s)\phi_Y(t) = \phi_\mu(s)\phi_\nu(t).$$

□

Le théorème 9.13 admet une réciproque.

Théorème 9.15. *Soient X et Y deux vecteurs aléatoires, X étant à valeurs dans $\mathbb{R}^n$ et Y à valeurs dans $\mathbb{R}^p$. Si*

$$\forall (s,t) \in \mathbb{R}^n \times \mathbb{R}^p, \quad \phi_{(X,Y)}(s,t) = \phi_X(s)\phi_Y(t),$$

alors X et Y sont indépendants.

Démonstration. On suppose que $\phi_{\mathbb{P}_{(X,Y)}}(s,t) = \phi_{\mathbb{P}_X}(s)\phi_{\mathbb{P}_Y}(t)$. Mais d'après le corollaire précédent, on a $\phi_{\mathbb{P}_X}(s)\phi_{\mathbb{P}_Y}(t) = \phi_{\mathbb{P}_X \otimes \mathbb{P}_Y}(s,t)$.
Ainsi $\phi_{\mathbb{P}_{(X,Y)}}(s,t) = \phi_{\mathbb{P}_X \otimes \mathbb{P}_Y}(s,t)$. Comme la fonction caractéristique caractérise la loi, on a $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$, ce qui signifie que X et Y sont indépendants. □

En revanche, le théorème 9.12 n'admet pas de réciproque : on verra en exercice des variables aléatoires X et Y non-indépendantes telles que $\phi_{X+Y} = \phi_X \phi_Y$.

9.2.4 Fonction caractéristique et moments

Théorème 9.16. *Si X est une variable aléatoire à valeurs dans $\mathbb{R}^d$ de loi μ telle que $\|X\|$ admette un moment d'ordre N, alors la fonction caractéristique de X est de classe C^N.*

Si $k_1, \ldots, k_d$ sont des entiers naturels dont la somme $k = k_1 + \cdots + k_d$ ne dépasse pas N, on a alors

$$\forall u \in \mathbb{R}^d \quad \frac{\partial^k \phi_X}{\partial_1^{k_1} \ldots \partial_d^{k_d}}(u) = i^k \int_{\mathbb{R}^d} e^{i\langle u,x\rangle} \prod_{j=1}^d x_j^{k_j} \, d\mu(x). \tag{9.1}$$

Ainsi, on a en particulier

$$\frac{\partial^k \phi_X}{\partial_1^{k_1} \ldots \partial_d^{k_d}}(0) = i^k \mathbb{E}\left[\prod_{j=1}^d X_j^{k_j} \right]. \tag{9.2}$$

Démonstration. On montre d'abord le lemme suivant : si $P \in \mathbb{R}[X_1, \ldots, X_d]$ est le monôme de degré k : $P(X) = X_1^{\alpha_1} \ldots X_d^{\alpha_d}$ avec $\alpha_1 + \cdots + \alpha_d = k$ et $\mathbb{E}(\|X\|^{k+1}) < +\infty$, alors

$$\Psi : u \mapsto \mathbb{E}(e^{i\langle u,X\rangle} P(X))$$

est une application de classe C^1 de $\mathbb{R}^d$ dans $\mathbb{C}$ avec pour tout $j \in \{1, \ldots, d\}$:

$$\forall u \in \mathbb{R}^d \quad \frac{\partial \Psi}{\partial_j}(u) = \mathbb{E}(ie^{i\langle u,X\rangle} X_j P(X)).$$

Le lemme se montre facilement à l'aide du théorème de dérivation sous le signe intégrale et de la majoration $|P(x)| \leq \|x\|_\infty^k$.

La première formule s'en déduit simplement par récurrence sur k. Pour la deuxième formule, il suffit de prendre $u = 0$. □

Dans le cas des variables aléatoires réelles, le théorème s'énonce évidemment plus simplement.

Corollaire 9.17. *Si X est une variable aléatoire réelle admettant un moment d'ordre N, alors la fonction caractéristique de X est de classe C^N et on a*

$$\forall k \in \{1, \dots, N\}, \quad \phi_X^{(k)}(0) = i^k \mathbb{E}[X^k]. \tag{9.3}$$

En particulier, si X admet un moment d'ordre 2 et est centrée avec une variance σ^2, on a alors le développement limité en 0 :

$$\phi_X(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2).$$

Démonstration. Si X admet un moment d'ordre 2, alors ϕ_X est de classe C^2 et donc

$$\phi_X(t) = \phi_X(0) + \phi_X'(0)t + \frac{\phi_X''(0)}{2}t^2 + o(t^2).$$

On a de plus $\phi_X(0) = 1$, $\phi_X'(0) = i\mathbb{E}X = 0$ et $\phi_X''(0) = -\mathbb{E}X^2 = -\operatorname{Var} X$ car $\mathbb{E}X = 0$. Il suffit de substituer pour conclure. □

Réciproquement, la connaissance d'une certaine régularité de la fonction caractéristique permet d'en déduire l'existence de certains moments.

Théorème 9.18. *Soit $n \geq 0$ un entier naturel. Si la variable aléatoire réelle X est telle que ϕ_X est $2n$ fois dérivable, alors X admet un moment d'ordre $2n$.*

Démonstration. On montre le résultat par récurrence sur n. Pour $n = 0$, c'est évident. Supposons le résultat acquis jusqu'au rang n. Si ϕ_X est $2n + 2$ fois dérivable, la fonction $\psi(t) = \frac{\phi_X^{(2n)}(t)+\phi_X^{(2n)}(-t)}{2}$ est une fonction paire, 2 fois dérivable. D'après l'hypothèse de récurrence ϕ_X admet un moment d'ordre $2n$, d'où $\phi_X^{(2n)}(t) = (-1)^n \mathbb{E}(X^{2n} \exp(itX))$, ce qui entraîne que

$$\psi(t) = (-1)^n \mathbb{E}(X^{2n} \cos(tX)).$$

Comme la fonction ψ est paire et deux fois dérivable, il existe b réel tel que $\psi(t) = \psi(0) + bt^2 + o(t^2)$. Ainsi, $\frac{\psi(0)-\psi(t)}{t^2}$ admet une limite finie lorsque t tend vers 0, et il en est de même pour

$$(-1)^n \frac{\psi(0) - \psi(t)}{t^2} = \mathbb{E}\left(X^{2n} \frac{1 - \cos(tX)}{t^2}\right).$$

Le lemme de Fatou nous donne alors

$$\begin{aligned} \frac{1}{2}\mathbb{E}(X^{2n+2}) &= \mathbb{E}\left(\varliminf_{t\to 0} X^{2n} \frac{1 - \cos(tX)}{t^2}\right) \\ &\leq \varliminf_{t\to 0} \mathbb{E}\left(X^{2n} \frac{1 - \cos(tX)}{t^2}\right) < +\infty. \end{aligned}$$

□

Cette correspondance partielle entre la régularité de la fonction caractéristique et l'existence de moments peut évoquer au lecteur une autre correspondance, celle entre la taille des coefficients d'une série de Fourier et la régularité de la fonction somme (voir par exemple Zuily–Queffelec [31]).

Certains contre-exemples issus de la théorie des séries de Fourier peuvent être mis à profit pour construire des contre-exemples dans le contexte probabiliste qui nous occupe. Ainsi, il serait tentant de conjecturer que si ϕ_X est C^1, alors X est intégrable, mais on va voir que ce n'est pas le cas.

En effet, on sait que si les coefficients d'une série trigonométrique sont absolument convergents, alors la convergence est uniforme, mais on sait aussi (voir l'exercice 75) que la réciproque est fausse. Le théorème suivant fournit des contre-exemples : si $(\varepsilon_n)_{n\geq 1}$ est une suite décroissante de limite nulle, on sait que la série de terme général $\varepsilon_k \frac{\sin(kt)}{k}$ converge uniformément.

Ce résultat de convergence uniforme peut encore servir à construire des fonctions caractéristiques de classe C^1 associées à des lois non intégrables.

Posons en effet $S(t) = \sum_{n=1}^{+\infty} \varepsilon_n \frac{1-\cos(nt)}{n^2}$ et $S_N(t) = \sum_{n=1}^{N} \varepsilon_n \frac{1-\cos(nt)}{n^2}$. Comme $S_N(x) = \int_0^x S_N'(t)\,dt$ et que S_N' converge uniformément vers une fonction g continue, on obtient à la limite $S(x) = \int_0^x g(t)\,dt$. Il s'ensuit que S est C^1. Prenons $B \geq \sum_{n=1}^{+\infty} \frac{\varepsilon_n}{n^2}$. On a

$$B - S(x) = (B - \sum_{n=1}^{+\infty} \frac{\varepsilon_n}{n^2}) + \frac{1}{2} \sum_{n\neq 0} \frac{\varepsilon_n e^{inx}}{n^2},$$

de sorte que $\frac{B-S(x)}{B}$ est la fonction génératrice de la variable aléatoire dont la loi est donnée par $\mathbb{P}(X=0) = 1 - \frac{1}{B}\sum_{n=1}^{+\infty} \frac{\varepsilon_n}{n^2}$ et $\mathbb{P}(X=k) = \frac{\varepsilon_k}{2Bk^2}$ pour $k \in \mathbb{Z}\backslash\{0\}$. On a alors

$$\mathbb{E}|X| = \sum_{k\in\mathbb{Z}} |k|\mathbb{P}(X=k) = \frac{1}{B}\sum_{k=1}^{+\infty} \frac{\varepsilon_k}{k}.$$

Ainsi, si on choisit ε_k décroissante de limite nulle et telle que $\sum_{k=1}^{+\infty} \frac{\varepsilon_k}{k} = +\infty$, on a le contre-exemple voulu. La suite $\varepsilon_k = \frac{1}{1+\log k}$ convient encore.

9.2.5 Fonctions caractéristiques des variables aléatoires à valeurs dans $\mathbb{N}$

Pour une variable aléatoire à valeurs dans $\mathbb{N}$, le calcul de la fonction caractéristique est équivalent à celui de la fonction génératrice. En effet, on a la formule :

$$\forall t \in \mathbb{R}, \quad \phi_X(t) = \mathbb{E}[e^{itX}] = \mathbb{E}[(e^{it})^X] = G_X(e^{it}).$$

On laisse au lecteur le soin de calculer la fonction caractéristique de la loi de Bernoulli, de la loi binomiale, de la loi de Poisson et de la loi géométrique.

9.2.6 Quelques fonctions caractéristiques de mesures à densité

Pour une mesure à densité, le calcul de la fonction caractéristique est en fait le calcul de la transformée de Fourier de la densité. Lorsque l'on sort des cas simples où une primitive peut facilement être trouvée, ces intégrales sont souvent calculées en utilisant des techniques issues de la théorie de l'analyse complexe, par exemple en utilisant la méthode des résidus ou encore en appliquant un théorème de prolongement analytique. D'autres méthodes peuvent être utiles, par exemple reconnaître la transformée de Fourier d'une fonction connue, ou utiliser un théorème d'inversion, ou expliciter une équation différentielle satisfaite par la fonction caractéristique, puis la résoudre. Nous traiterons d'abord la méthode de la théorie de la variable complexe qui, nous ne le cachons pas, a notre préférence, mais donnerons également un calcul de pure analyse réelle, pour ne pas pénaliser le lecteur qui n'est pas familier de cette théorie.

Loi uniforme sur $[a,b]$

On commence par calculer la fonction caractéristique de la loi uniforme sur $[-1,1]$. Pour $t \neq 0$, on a

$$\phi_X(t) = \mathbb{E}e^{itX} = \int_{-1}^{1} \frac{1}{2} e^{itx}\, dx = \left[\frac{e^{itx}}{2it}\right]_{-1}^{1} = \frac{e^{it} - e^{-it}}{2it} = \frac{\sin t}{t}.$$

La formule se prolonge par continuité pour $t = 0$ (avec $\phi_X(0) = 1$).

Maintenant, si on pose $Y = \frac{a+b}{2} + \frac{b-a}{2}X$, alors Y suit la loi uniforme sur $[a,b]$ et on a

$$\phi_Y(t) = e^{i\frac{a+b}{2}t}\phi_X\left(\frac{b-a}{2}t\right) = e^{i\frac{a+b}{2}t}\frac{\sin\frac{(b-a)t}{2}}{\frac{(b-a)t}{2}}.$$

Loi exponentielle de paramètre λ

On commence par calculer la fonction caractéristique de la loi exponentielle de paramètre 1. Pour tout réel t, on a

$$\phi_X(t) = \mathbb{E}e^{itX} = \int_0^{+\infty} e^{-x}e^{itx}\, dx = \left[\frac{e^{(-1+it)x}}{-1+it}\right]_0^{+\infty} = \frac{0-1}{-1+it} = \frac{1}{1-it}.$$

Maintenant, si on pose $Y = \frac{1}{\lambda}X$, alors Y suit la loi $\mathcal{E}(\lambda)$ et on a

$$\phi_Y(t) = \phi_X\left(\frac{1}{\lambda}t\right) = \frac{1}{1-i\frac{t}{\lambda}} = \frac{\lambda}{\lambda - it}.$$

Variable aléatoire gaussienne

Théorème 9.19. *La fonction caractéristique de la loi normale $\mathcal{N}(m, \sigma^2)$ est*

$$t \mapsto \exp(imt) \exp\left\{-\frac{1}{2}\sigma^2 t^2\right\}.$$

Démonstration. On va d'abord déterminer la fonction caratéristique de la loi $\mathcal{N}(0, 1)$. Nous devons calculer

$$\phi_X(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-x^2/2) \exp(itx)\ d\lambda(x).$$

Ainsi, cela achèvera la preuve du lemme 8.12.
Méthode 1 : utilisation de la théorie des fonctions holomorphes.
Pour $z \in \mathbb{C}$, on pose

$$f_z(x) = \exp(-x^2/2) \exp(xz).$$

Pour tout $x \in \mathbb{R}$, la fonction $z \mapsto f_z(x)$ est holomorphe sur $\mathbb{C}$. D'autre part, pour $z \in D(0, R)$ la fonction $x \mapsto f_z(x)$ est dominée par la fonction intégrable $x \mapsto \exp(-x^2/2) \exp(R|x|)$, car $\forall x \in \mathbb{R}, \quad \forall z \in D(0, R)$, on a

$$|\exp(-x^2/2) \exp(xz)| = \exp(-x^2/2) \exp(x\text{Re } z) \le \exp(-x^2/2) \exp(R|x|).$$

Il s'ensuit que $z \mapsto \int_{\mathbb{R}} f_z(x) d\lambda(x)$ est holomorphe.

Pour z réel, on a

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-x^2/2) \exp(zx)\ d\lambda(x) \\
= \quad &\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-((x-z)^2 - z^2)/2)\ d\lambda(x) \\
= \quad &\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-(x-z)^2/2)\ d\lambda(x) \exp\left(\frac{z^2}{2}\right) \\
= \quad &\exp\left(\frac{z^2}{2}\right),
\end{aligned}$$

car l'expression intégrée n'est autre que la densité de la loi $\mathcal{N}(z, 1)$. Mais si deux fonctions holomorphes coïncident sur $\mathbb{R}$, elles coïncident sur $\mathbb{C}$. On a donc

$$\forall z \in \mathbb{C}, \quad \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-x^2/2) \exp(zx)\ d\lambda(x) = \exp\left(\frac{z^2}{2}\right).$$

On particularise alors z en it, t étant réel, et on obtient

$$\phi_X(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-x^2/2) \exp(itx)\ dx = \exp\left(-\frac{t^2}{2}\right).$$

Méthode 2 : utilisation d'une équation différentielle.
On pose $g_t(x) = e^{itx}$. On a $\phi_X(t) = \mathbb{E}[g_t(X)]$.

Avec l'aide du théorème de dérivation sous le signe intégrale, on montre facilement que

$$\phi'_X(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-x^2/2) ix \exp(itx)\, dx = i\mathbb{E}[X g_t(X)].$$

On peut alternativement employer un argument plus général : comme X est intégrable, la formule $\phi'_X(t) = i\mathbb{E}[X g_t(X)]$ n'est autre que l'équation (9.1) pour $N = 1$ et en dimension 1.

Mais d'après la formule d'intégration par parties de la variable gaussienne (vue au lemme 6.51), on a $\mathbb{E}[X g_t(X)] = \mathbb{E}[g'_t(X)]$. Cependant, pour tout x réel, on a $g'_t(x) = itg_t(x)$, d'où

$$\phi'_X(t) = i\mathbb{E}[X g_t(X)] = i\mathbb{E}[g'_t(X)] = i\mathbb{E}[itg_t(X)] = -t\mathbb{E}[g_t(X)] = -t\phi_X(t).$$

L'équation différentielle se résout classiquement.
On pose $F(t) = \exp(t^2/2)\phi_X(t)$. On a $F(0) = 1$ et $F'(t) = 0$, donc F est constante égale à un, ce qui donne $\phi_X(t) = \exp(-t^2/2)$.

Pour passer au cas général, on pose $Y = \sigma X + m$; on a $Y \sim \mathcal{N}(m, \sigma^2)$, et alors

$$\begin{aligned} \phi_Y(t) &= \mathbb{E}(e^{itY}) = e^{it(\sigma X + m)} \\ &= e^{imt}\mathbb{E}(e^{it\sigma X}) = e^{imt}\phi_X(\sigma t) = e^{imt} e^{-\frac{\sigma^2 t^2}{2}}. \end{aligned}$$

□

Loi de Cauchy

Théorème 9.20. *La fonction caractéristique de la loi de Cauchy $\mathcal{C}(a, b)$ est*

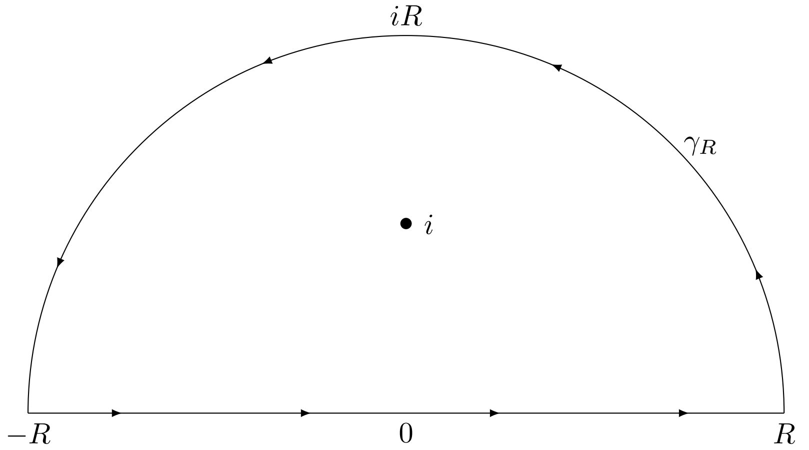
$$t \mapsto \phi(t) = e^{iat} e^{-b|t|}.$$

Démonstration. Rappelons que la loi de Cauchy $\mathcal{C}(a, b)$ admet comme densité par rapport à la mesure de Lebesgue :

$$x \mapsto \frac{1}{\pi} \frac{b}{(x-a)^2 + b^2}.$$

Comme dans les exemples précédents, on a une famille de lois qui est stable par les transformations affines, donc on va d'abord calculer la fonction caractéristique de $\mathcal{C}(0, 1)$ et on s'y ramènera.

Méthode 1 : utilisation de la théorie des fonctions holomorphes.
On suppose d'abord que $t > 0$. Pour $R > 1$, on intègre la fonction $f(z) = \frac{e^{itz}}{1+z^2}$ sur le contour γ_R.



Le seul pôle de f à l'intérieur de la courbe est i, donc pour $R > 1$, on a

$$\frac{1}{2i\pi}\int_{\gamma_R} f(z)\,dz = \mathbf{Res}_i f(z).$$

Or $\frac{1}{1+(i+h)^2} = \frac{1}{h}\frac{1}{2i+h}$, donc $\text{Res}_i f(z) = \frac{e^{-t}}{2i}$. Ainsi,

$$\forall R > 1, \quad \frac{1}{\pi}\int_{\gamma_R} f(z)\,dz = e^{-t}.$$

Par ailleurs,

$$\int_{\gamma_R} f(z)\,dz = \int_{-R}^{R} f(t)\,dt + iR\int_0^{\pi} f(Re^{i\theta})e^{i\theta}\,d\theta.$$

Mais, lorsque $z = Re^{i\theta}$ et $\sin\theta \geq 0$, on a

$$|f(Re^{i\theta})e^{i\theta}| = \frac{e^{-Rt\sin\theta}}{|1+z^2|} \leq \frac{1}{R^2-1}.$$

(C'est ici qu'on utilise l'hypothèse $t > 0$.) Ainsi,

$$\lim_{R\to+\infty} iR\int_0^{\pi} f(Re^{i\theta})e^{i\theta}\,d\theta = 0.$$

On en déduit que

$$e^{-t} = \frac{1}{\pi}\int_{-\infty}^{+\infty} f(t)\,dt.$$

Donc si $X \sim \mathcal{C}(0,1)$, on a alors

$$\forall t > 0 \quad \phi_X(t) = e^{-t}.$$

Comme la loi de X est symétrique ($\mathbb{P}_X = \mathbb{P}_{-X}$) et que $\phi(0) = 1$, on a

$$\forall t \in \mathbb{R} \quad \phi_X(t) = e^{-|t|}.$$

Méthode 2 : utilisation de l'inversion de Fourier.
Cette méthode n'est pas naturelle, mais elle est classique. Elle suppose de "connaître le résultat à l'avance". Posons $f(x) = e^{-|x|}$. Un calcul simple donne $\hat{f}(t) = \frac{2}{1+t^2}$. On remarque que $\hat{f} \in \mathcal{L}^1$, donc d'après le théorème d'inversion, on a pour $\lambda-$presque tout x

$$e^{-|x|} = \int_{\mathbb{R}} \frac{1}{2\pi} e^{-ixt} \frac{2}{1+t^2}\, d\lambda(t) = \int_{\mathbb{R}} e^{ixt} \frac{1}{\pi} \frac{1}{1+t^2}\, d\lambda(t) = \phi_X(x)$$

par symétrie. L'ensemble des $x \in \mathbb{R}$ tels que $\phi_X(x) \neq e^{-|x|}$ est un ouvert, car $x \mapsto \phi_X(x) - e^{-|x|}$ est continue. Comme il est de mesure nulle, il est vide, ce qui nous donne $\phi_X(t) = e^{-|t|}$ pour tout t dans $\mathbb{R}$.
Pour conclure, on remarque enfin que si on pose $Y = bX + a$, on a
$Y \sim \mathcal{C}(a, b)$, et alors

$$\phi_Y(t) = \mathbb{E}e^{itY} = e^{it(bX+a)} = e^{iat}\mathbb{E}e^{itbX} = e^{iat}\phi_X(bt) = e^{iat}e^{-b|t|}.$$

□

9.3 Transformée de Laplace

On note $\mathbb{H}_+ = \{a + ib; (a, b) \in \mathbb{R}_+^* \times \mathbb{R}\}$ le demi-plan positif ouvert et $\overline{\mathbb{H}}_+ = \{a + ib; (a, b) \in \mathbb{R}_+ \times \mathbb{R}\}$ le demi-plan positif fermé. À toute mesure de probabilité μ sur $\mathbb{R}_+$, on associe une fonction $\mathcal{L}_\mu$ définie sur le demi-plan positif fermé par

$$\forall z \in \overline{\mathbb{H}}_+ \quad \mathcal{L}_\mu(z) = \int_{\mathbb{R}_+} e^{-zt}\, d\mu(t).$$

C'est la transformée de Laplace de μ. On définit également la fonction génératrice des moments M_μ par

$$M_\mu(z) = \mathcal{L}_\mu(-z) = \int_{\mathbb{R}_+} e^{zt}\, d\mu(t).$$

$M_\mu(z)$ peut éventuellement valoir $+\infty$. On définit la transformée de Laplace d'une variable aléatoire positive X par

$$\forall z \in \overline{\mathbb{H}}_+ \quad \mathcal{L}_X(z) = \mathbb{E}[e^{-zX}] = \int_{\mathbb{R}_+} e^{-zt}\, d\mathbb{P}_X(t).$$

Comme on l'a fait pour les fonctions génératrices ou caractéristiques, il est facile de voir que si X et Y sont des variables aléatoires indépendantes, on a

$$\forall z \in \overline{\mathbb{H}}_+ \quad \mathcal{L}_{X+Y}(z) = \mathcal{L}_X(z)\mathcal{L}_Y(z).$$

Théorème 9.21. *$\mathcal{L}_\mu$ est continue sur $\overline{\mathbb{H}}_+$, holomorphe sur $\mathbb{H}_+$, avec*

$$\forall n \in \mathbb{N} \quad \forall z \in \mathbb{H}_+, \quad \mathcal{L}_\mu^{(n)}(z) = (-1)^n \int_{\mathbb{R}_+} t^n e^{-zt}\, d\mu(t).$$

La restriction de $\mathcal{L}_\mu$ au demi-axe réel positif est telle que M_μ est absolument monotone, c'est-à-dire que toutes ses dérivées sont positives (pour tout $n \geq 0$, $z > 0$, on a $(-1)^n \mathcal{L}_\mu^{(n)}(z) \geq 0$). Cette restriction suffit à caractériser μ.

Démonstration. La continuité découle aisément du théorème 4.18 de continuité sous le signe intégrale, et de la majoration $|e^{zx}| \leq 1$. L'holomorphie et l'expression des dérivées découlent de cette même majoration et du théorème 4.21. La positivité des dérivées est évidente sur l'expression trouvée. Si $\mathcal{L}_\mu$ et $\mathcal{L}_\nu$ coïncident sur le demi-axe réel positif, elles coïncident sur le demi-espace positif ouvert, par le principe de prolongement analytique. Par continuité, elles coïncident sur l'axe des imaginaires purs, donc les fonctions caractéristiques coïncident, ce qui entraîne l'égalité des mesures. □

De même que le théorème de Bochner décrivait les fonction caractéristiques, le théorème suivant, dû à Bernstein, permet de caractériser les transformées de Laplace des mesures de probabilité. La preuve qui suit, inspirée de Dubourdieu [9], est cependant plus simple que les preuves du théorème de Bochner.

Théorème 9.22 (de Bernstein)**.** *Soit f une fonction continue sur $\mathbb{R}_-$, infiniment dérivable sur $\mathbb{R}_-^*$ et absolument monotone, c'est-à-dire que*

$$\forall x < 0, \quad \forall n \geq 0, \quad f^{(n)}(x) \geq 0$$

et $f(0) = 1$. Alors, il existe une unique mesure de probabilité μ telle que

$$\forall x \geq 0, \quad f(-x) = \mathcal{L}_\mu(x) = \int_{\mathbb{R}_+} e^{-xt} \, d\mu(t).$$

Démonstration. On va commencer par montrer que quels que soient les réels x et N avec $-N < x < 0$, on a le développement en série

$$f(x) = \sum_{k=0}^{+\infty} \frac{f^{(k)}(-N)}{k!}(x+N)^k.$$

Par la formule de Taylor avec reste intégral, on a

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(-N)}{k!}(x+N)^k + R_n, \text{ avec } R_n = \int_{-N}^{x} \frac{(x-t)^n}{n!} f^{(n+1)}(t) \, dt.$$

Il faut donc montrer que le reste tend vers 0. Si on applique encore la formule de Taylor avec reste intégral à f' entre t et x avec $t \leq x$, on a

$$f'(x) = \sum_{k=0}^{n} \frac{f^{(k+1)}(t)}{k!}(x-t)^k + \int_{t}^{x} \frac{(x-u)^n}{n!} f^{(n+2)}(u) \, du.$$

Comme le reste est positif, on a $f'(x) \geq \sum\limits_{k=0}^{n} \frac{f^{(k+1)}(t)}{k!}(x-t)^k$, ce qui montre

la convergence de la série de terme général positif $\frac{f^{(k+1)}(t)}{k!}(x-t)^k$.
En particulier, on a pour tout $k \geq 0 : 0 \leq \frac{f^{(k+1)}(t)}{k!}(x-t)^k \leq f'(x)$ et $\lim_{k\to+\infty} \frac{f^{(k+1)}(t)}{k!}(x-t)^k = 0$. Il suffit alors d'appliquer le théorème de convergence dominée pour obtenir la convergence de R_n vers 0.

Pour $N > -x$, on a donc un développement en série de $f(x)$, que l'on peut réécrire

$$\begin{aligned} f(x) &= \sum_{k=0}^{+\infty} \left(1+\frac{x}{N}\right)^k \frac{N^k}{k!} f^{(k)}(-N) \\ &= \sum_{k=0}^{+\infty} g_N^x\left(\frac{k}{N}\right) \frac{N^k}{k!} f^{(k)}(-N), \end{aligned}$$

avec $g_N^x(\lambda) = \left(1+\frac{x}{N}\right)^{N\lambda}$. Cela peut encore se réécrire

$$f(x) = \int_{\mathbb{R}} g_N^x(\lambda)\, d\mu_N(\lambda), \text{ où } \mu_N = \sum_{k=0}^{+\infty} \frac{N^k}{k!} f^{(k)}(-N)\delta_{k/N}.$$

Par le théorème de convergence monotone, on a

$$\lim_{x\to 0^-} f(x) = \lim_{x\to 0^-} \sum_{k=0}^{+\infty} \left(1+\frac{x}{N}\right)^k \frac{N^k}{k!} f^{(k)}(-N) = \sum_{k=0}^{+\infty} \frac{N^k}{k!} f^{(k)}(-N) = \mu_N(\mathbb{R}).$$

Ainsi, $\mu_N(\mathbb{R}) = 1$ et μ_N est une mesure de probabilité à support dans $\mathbb{R}_+$. Comme g_N^x est “presque” la fonction exponentielle, on peut penser que si μ_N converge (en un sens à préciser), la limite fournira une solution au problème. On peut déjà remarquer que si l'on considère le point $-N + N\exp(x/N)$, on a $-N \leq -N + N\exp(x/N) < 0$, donc

$$f(-N + N\exp(x/N)) = \int_{\mathbb{R}} g_N^{-N+N\exp(x/N)}(\lambda)\, d\mu_N(\lambda) = \int_{\mathbb{R}} e^{\lambda x}\, d\mu_N(\lambda). \tag{9.4}$$

À l'aide du théorème de Fubini, on montre que pour toute mesure ν sur $\mathbb{R}_+$, on a

$$\forall \alpha > 0 \quad \int_0^{+\infty} \nu([0,u])\alpha e^{-\alpha u}\, du = \int_{\mathbb{R}_+} e^{-\alpha t}\, d\nu(t).$$

En particulier, si on note F_N la fonction de répartition associée à μ_N, on a

$$\int_0^{+\infty} F_N(u)\alpha e^{-\alpha u}\, du = \int_{\mathbb{R}_+} e^{-\alpha t}\, d\mu_N(t) = f(-N + N\exp(-\alpha/N)).$$

D'après le théorème 5.6 de Helly, on peut trouver une suite (n_k) et une fonction càdlàg (continue à droite et limitée à gauche) F telle que $F_{n_k}(x)$ tend vers $F(x)$ en tout point de continuité de F.

Comme F_{n_k} tend presque partout vers F, il vient, en faisant tendre k vers

l'infini
$$\forall \alpha > 0, \quad \int_0^{+\infty} F(u)\alpha e^{-\alpha u}\, du = f(-\alpha).$$
Montrons que F est une fonction de répartition : comme F est càdlàg, nulle sur $\mathbb{R}_-^*$, il suffit de montrer que $\lim\limits_{u\to+\infty} F(u) = 1$. F est croissante et majorée par 1 : notons ℓ sa limite. On a pour tout $\alpha > 0$:
$$1 \geq \ell = \int_0^{+\infty} \ell\alpha e^{-\alpha u}\, du \geq \int_0^{+\infty} F(u)\alpha e^{-\alpha u}\, du = f(-\alpha).$$
En faisant tendre α vers 0, on obtient $\ell = 1$. Ainsi, F est bien une fonction de répartition, correspondant à une mesure μ, et on a pour tout $\alpha > 0$
$$f(-\alpha) = \int_0^{+\infty} F(u)\alpha e^{-\alpha u}\, du = \int_{\mathbb{R}_+} e^{-\alpha t}\, d\mu(t),$$
ce qui est bien la représentation annoncée. L'unicité a été établie au théorème précédent. □

9.4 Application aux marches aléatoires

Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes identiquement distribuées à valeurs dans $\mathbb{Z}^d$. On note μ la loi de X_1 et ϕ sa fonction caractéristique.

La suite des sommes partielles $(S_n)_{n\geq 0}$, définie par $S_0 = 0$ et pour $n \geq 1$:
$$S_n = \sum_{k=1}^{n} X_k,$$
forme ce que l'on appelle une marche aléatoire sur $\mathbb{Z}^d$ issue de 0.

On cherche a déterminer si la marche aléatoire $(S_n)_{n\geq 0}$ passe une infinité de fois en zéro.

Le cas le plus simple est celui de la marche aléatoire simple en dimension 1 : $\mu = \frac{1}{2}(\delta_1 + \delta_{-1})$. Dans ce cas, on donnera plusieurs solutions au problème.

9.4.1 Transience et récurrence

Théorème 9.23. *Notons $S = \sum\limits_{n=0}^{+\infty} \mathbb{1}_{\{S_n=0\}}$ le nombre de passages de la marche en zéro. On a*
$$\mathbb{P}(S = +\infty) = \begin{cases} 0 & \text{si } \sum\limits_{k=0}^{+\infty} \mathbb{P}(S_k = 0) < +\infty, \\ 1 & \text{si } \sum\limits_{k=0}^{+\infty} \mathbb{P}(S_k = 0) = +\infty. \end{cases}$$

Dans le premier cas, on dit que la marche est transiente ; dans le deuxième cas, on dit qu'elle est récurrente.

Démonstration. D'après Tonelli, on a $\mathbb{E}[S] = \sum_{n=0}^{+\infty} \mathbb{P}(S_n = 0)$. Ainsi, si la série converge, S est intégrable. Alors S est fini presque sûrement, ce qui nous donne le premier cas [1]. Posons $B = \{S < +\infty\}$. On peut découper B suivant l'instant T de dernier retour en 0 : $\{T = n\} = \{S_n = 0\} \cap \{\forall k > n; S_k \neq 0\}$

$$\begin{aligned}
B &= \bigcup_{n=0}^{+\infty} \{T = n\} \\
&= \bigcup_{n=0}^{+\infty} \{S_n = 0\} \cap \{\forall k > n; S_k \neq 0\} \\
&= \bigcup_{n=0}^{+\infty} \{S_n = 0\} \cap \{\forall k > n; S_k - S_n \neq 0\}.
\end{aligned}$$

C'est ici le point important : on a réécrit l'intersection comme intersection de deux événements indépendants. Comme la réunion est disjointe, on a

$$\begin{aligned}
\mathbb{P}(B) &= \sum_{n=0}^{+\infty} \mathbb{P}(\{S_n = 0\} \cap \{\forall k > n; S_k - S_n \neq 0\}) \\
&= \sum_{n=0}^{+\infty} \mathbb{P}(S_n = 0)\mathbb{P}(\forall k > n; S_k - S_n \neq 0\}) \\
&= \sum_{n=0}^{+\infty} \mathbb{P}(S_n = 0)\mathbb{P}(\forall k > 0; S_k - S_0 \neq 0\}) \\
&= \left(\sum_{n=0}^{+\infty} \mathbb{P}(S_n = 0) \right) \mathbb{P}(T = 0).
\end{aligned}$$

Ainsi, si la série diverge, alors $\mathbb{P}(B) = 0$ si $\mathbb{P}(T = 0) = 0$ et $\mathbb{P}(B) = +\infty$ si $\mathbb{P}(T = 0) > 0$. Mais une probabilité ne peut être infinie, donc $\mathbb{P}(B) = 0$ et $\mathbb{P}(S = +\infty) = 1$. $\square$

Dans le cas de la marche symétrique sur $\mathbb{Z}$ (c'est-à-dire $\mu = \frac{1}{2}(\delta_1 + \delta_{-1})$), le calcul de $\mathbb{P}(S_n = 0)$ est très simple : cette probabilité est nulle aux instants impairs (parité), et aux temps pairs, on a

$$\mathbb{P}(S_{2n} = 0) = \frac{1}{2^{2n}} \binom{2n}{n}.$$

Une manière élégante de justifier l'égalité peut être de remarquer que comme $\frac{1+X_1}{2}$ suit la loi de Bernoulli de paramètre $1/2$, $B_n = \frac{n+S_n}{2}$ suit la loi Bino-

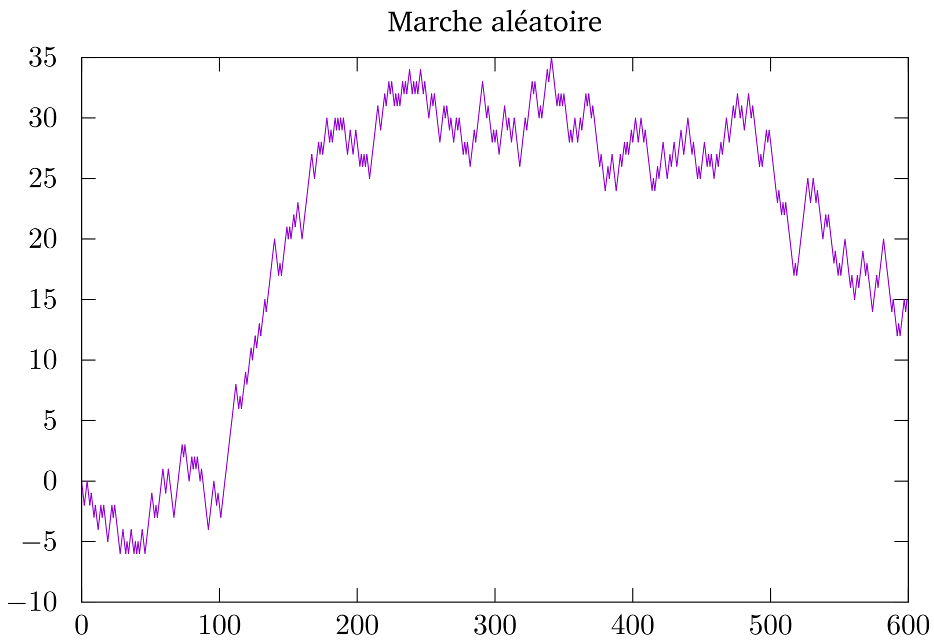
1. Le lecteur initié aura reconnu que nous avons redémontré ici le 1er lemme de Borel-Cantelli. En revanche, pour le deuxième cas, on ne peut appliquer le 2ème lemme de Borel-Cantelli car les événements $\{S_k = 0\}$ ne sont pas indépendants. En réalité, ce théorème est un cas particulier d'un critère de récurrence des chaînes de Markov (voir [18]).

miale de paramètres n et $1/2$: ainsi $\mathbb{P}(S_{2n} = 0) = \mathbb{P}(B_{2n} = n)$.

On peut alors donner un équivalent de $\mathbb{P}(S_{2n} = 0)$ grâce à la formule de Stirling : $n! \sim \sqrt{2\pi n}\,\left(\frac{n}{e}\right)^n$, d'où l'on tire $\mathbb{P}(S_{2n} = 0) \sim \frac{1}{\sqrt{\pi n}}$, ce qui donne la divergence de la série.

On peut remarquer qu'en fait, on n'a pas vraiment besoin de la formule de Stirling, mais seulement de la partie "équivalent de Wallis" de la preuve.

Ci-après, une réalisation de la marche aléatoire symétrique sur $\mathbb{Z}$ (avec en abscisse le temps) :



Pour étudier des cas plus généraux, on va commencer par donner une formule explicite permettant de retrouver la loi d'une variable aléatoire à valeurs dans $\mathbb{Z}^d$ à partir de sa fonction caractéristique. On montre dans l'exercice 207 comment cette approche peut être appliquée plus simplement en dimension un pour montrer la divergence de la série.

Lemme 9.24. *Soit X une variable aléatoire à valeurs dans $\mathbb{Z}^d$, de fonction caractéristique ϕ_X. On a*

$$\forall n \in \mathbb{Z}^d \quad \mathbb{P}(X = n) = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \phi_X(t) e^{-i\langle n,t\rangle}\, d\lambda^{\otimes d}(t).$$

Démonstration. Soit $n \in \mathbb{Z}^d$. Notons m la loi uniforme sur $[-\pi, \pi]^d$.
La fonction $(\theta, x) \mapsto e^{i\langle \theta, x-n\rangle}$ est bornée, donc intégrable par rapport à la

mesure $m \otimes \mathbb{P}_X$. Par le théorème de Fubini, on a, d'une part

$$\begin{aligned}\int_{[-\pi,\pi]^d \times \mathbb{R}^d} e^{i\langle \theta, x-n\rangle}\, d(m\otimes \mathbb{P}_X)(\theta,x) &= \int_{[-\pi,\pi]^d} \left(\int_{\mathbb{R}^d} e^{i\langle \theta, x-n\rangle}\, d\mathbb{P}_X(x)\right)\, dm(\theta)\\ &= \int_{[-\pi,\pi]^d} \mathbb{E}[e^{i\langle \theta, X-n\rangle}]\, dm(\theta),\end{aligned}$$

et d'autre part

$$\int_{[-\pi,\pi]^d \times \mathbb{R}^d} e^{i\langle \theta, x-n\rangle}\, d(m\otimes \mathbb{P}_X)(\theta,x) = \int_{\mathbb{R}^d} \left(\int_{[-\pi,\pi]^d} e^{i\langle \theta, x-n\rangle}\, dm(\theta)\right)\, d\mathbb{P}_X(x).$$

Or, pour $k \in \mathbb{Z}^d$, $k = (k_1, \dots, k_d)$ et $n = (n_1, \dots, n_d)$, on a

$$\int_{[-\pi,\pi]^d} e^{i\langle \theta, k-n\rangle}\, dm(\theta) = \prod_{p=1}^{d} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\theta(k_p-n_p)}\, d\theta = \prod_{p=1}^{d} \delta_{n_p,k_p} = \mathbb{1}_{\{n\}}(k).$$

Comme $\mathbb{P}_X$-presque tout réel est entier, on a alors

$$\begin{aligned}\int_{[-\pi,\pi]^d \times \mathbb{R}^d} e^{i\langle \theta, x-n\rangle}\, d(m\otimes \mathbb{P}_X)(\theta,x) &= \int_{\mathbb{R}^d} \mathbb{1}_{\{n\}}(x)\, d\mathbb{P}_X(x)\\ &= \mathbb{P}_X(\{n\}) = \mathbb{P}(X=n).\end{aligned}$$

□

Théorème 9.25 (Chung et Fuchs). *La marche aléatoire est transiente si et seulement si la fonction caractéristique ϕ de X_1 vérifie*

$$\lim_{r\to 1^-} \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} Re\left(\frac{1}{1-r\phi(t)}\right)\, d\lambda^{\otimes d}(t) < +\infty.$$

Démonstration. Posons, pour $r<1$, $\mathcal{S}_r = \sum_{n=0}^{+\infty} \mathbb{1}_{\{S_n=0\}} r^n$ et $S = \sum_{n=0}^{+\infty} \mathbb{1}_{\{S_n=0\}}$. On a vu que $\mathbb{E}[S] = \sum_{n=0}^{+\infty} \mathbb{P}(S_n=0)$, et, comme $S = \lim_{r\to 1^-} \mathcal{S}_r$, le théorème de convergence monotone nous dit que $\mathbb{E}[S] = \lim_{r\to 1^-} \mathbb{E}[\mathcal{S}_r]$.

Soit N un entier naturel. Comme $\mathcal{S}_r = \lim_{N\to+\infty} \sum_{n=0}^{N} \mathbb{1}_{\{S_n=0\}} r^n$, le théorème de convergence monotone implique que

$$\mathbb{E}[\mathcal{S}_r] = \lim_{N\to+\infty} \mathbb{E}\left[\sum_{n=0}^{N} \mathbb{1}_{\{S_n=0\}} r^n\right] = \lim_{N\to+\infty} \sum_{n=0}^{N} r^n \mathbb{P}(S_n=0).$$

Or

$$\begin{aligned}\mathbb{P}(S_n=0) &= \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \phi_{S_n}(t)\, d\lambda^{\otimes d}(t)\\ &= \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \phi(t)^n\, d\lambda^{\otimes d}(t), \qquad (9.5)\end{aligned}$$

donc on obtient

$$\begin{aligned}\sum_{n=0}^{N} r^n\mathbb{P}(S_n=0) &= \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d}\sum_{n=0}^{N}(r\phi(t))^n \, d\lambda^{\otimes d}(t)\\ &= \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d}\frac{1-(r\phi(t))^{N+1}}{1-r\phi(t)}\, d\lambda^{\otimes d}(t).\end{aligned}$$

En appliquant le théorème de convergence dominée, on obtient alors

$$\mathbb{E}[\mathcal{S}_r] = \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d}\frac{1}{1-r\phi(t)}\, d\lambda^{\otimes d}(t).$$

Comme $\mathcal{S}_r$ est réel, $\mathbb{E}[\mathcal{S}_r]$ est réel, et donc

$$\begin{aligned}\mathbb{E}[\mathcal{S}_r] &= \mathrm{Re}\,\frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d}\frac{1}{1-r\phi(t)}\, d\lambda^{\otimes d}(t)\\ &= \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d}\mathrm{Re}\,\frac{1}{1-r\phi(t)}\, d\lambda^{\otimes d}(t).\end{aligned}$$

On conclut en faisant tendre r vers 1^- et en utilisant le théorème 9.23. □

Corollaire 9.26. *Si*

$$\int_{[-\pi,\pi]^d}\mathit{Re}\,\frac{1}{1-\phi(t)}\, d\lambda^{\otimes d}(t) = +\infty,$$

alors la marche est récurrente.

Démonstration. En effet, grâce au lemme de Fatou, on a

$$+\infty = \int_{[-\pi,\pi]^d}\mathrm{Re}\,\frac{1}{1-\phi(t)}\, d\lambda^{\otimes d}(t) \le \varliminf_{r\to 1^-}\int_{[-\pi,\pi]^d}\mathrm{Re}\,\frac{1}{1-r\phi(t)}\, d\lambda^{\otimes d}(t),$$

et on applique le critère de Chung et Fuchs. □

Théorème 9.27. *Si $\frac{1}{1-\phi}$ est intégrable sur $[-\pi,\pi]^d$, alors la marche aléatoire est transiente.*

Démonstration. Grâce à (9.5), on a

$$\begin{aligned}\sum_{n=0}^{N}\mathbb{P}(S_n=0) &= \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d}\sum_{n=0}^{N}\phi(t)^n\, d\lambda^{\otimes d}(t)\\ &= \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d}\frac{1-\phi(t)^{N+1}}{1-\phi(t)}\, d\lambda^{\otimes d}(t).\end{aligned}$$

Comme $\left|\frac{1-\phi(t)^{N+1}}{1-\phi(t)}\right| \le \frac{2}{|1-\phi(t)|}$, on a l'inégalité

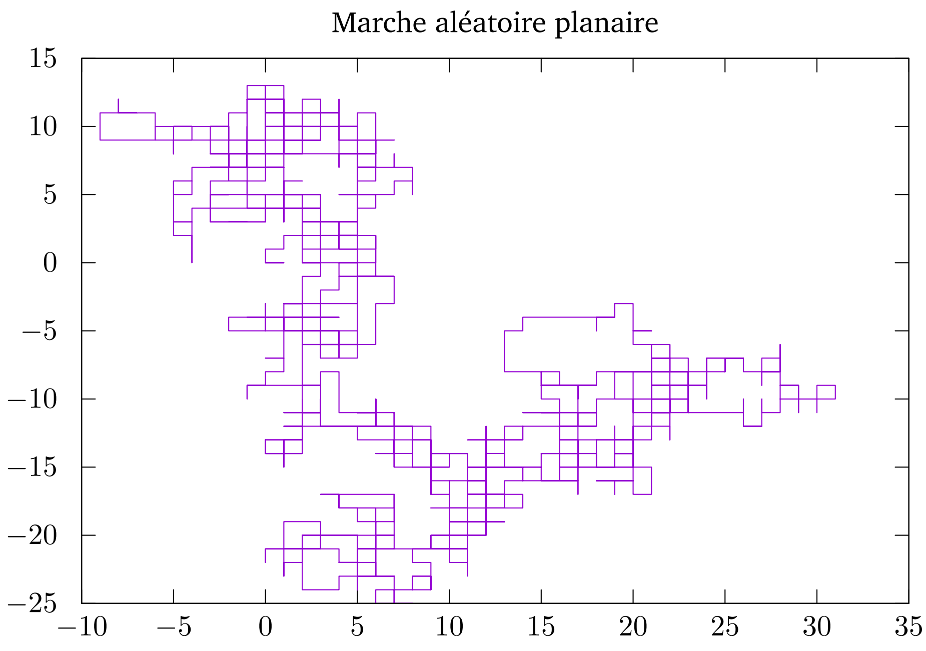
$$\sum_{n=0}^{N}\mathbb{P}(S_n=0) \le \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d}\frac{2}{|1-\phi(t)|}\, d\lambda^{\otimes d}(t).$$

Ceci entraîne la convergence de la série de terme général positif $\mathbb{P}(S_n = 0)$ et on applique le théorème 9.23. □

9.4.2 Marche aléatoire simple sur $\mathbb{Z}^d$

L'exemple le plus classique est celui où μ est la mesure uniforme sur $\{x \in \mathbb{Z}^d; \|x\|_1 = 1\}$. On l'appelle la marche aléatoire simple sur $\mathbb{Z}^d$.

Ci-après, une simulation des 1500 premiers pas d'une marche aléatoire symétrique aux plus proches voisins en dimension 2 :



Théorème 9.28 (Pólya, 1921)**.** *La marche aléatoire simple sur $\mathbb{Z}^d$ est récurrente pour $d = 1, 2$, et transiente si $d \geq 3$.*

Démonstration. On a aisément $\phi(\theta) = \frac{1}{d} \sum_{k=1}^{d} \cos\theta_k$ et un développement limité montre que

$$1 - \phi(\theta_1, \ldots, \theta_d) = \frac{1}{2d}\|\theta\|_2^2 + o(\|\theta\|_2^2).$$

On obtient ainsi que l'intégrale $\int_{[-\pi,\pi[^d} \frac{1}{1-\phi(\theta)}\, d\lambda^{\otimes d}(\theta)$ a même nature que $\int_{[-\pi,\pi[^d} \frac{1}{\|\theta\|_2^2}\, d\theta_1 \ldots d\theta_d$ ou que $\int_{B(0,1)} \frac{1}{\|\theta\|_2^2}\, d\theta_1 \ldots d\theta_d$.

Il reste à voir la nature de l'intégrale $\int_{B(0,1)} \frac{1}{\|\theta\|_2^2} \, d\theta_1 \dots d\theta_d$, qui est aisément déterminée grâce au théorème 4.48 d'intégration des fonctions radiales. □

Remarque 9.29. *Il est possible, dans ce cas particulier, d'étudier la nature de la série de terme général $\mathbb{P}(S_n = 0)$ en utilisant seulement des estimées sur des coefficients binomiaux, voir par exemple l'ouvrage de Lesigne [25].*

9.5 Exercices sur les fonctions génératrices et les fonctions caractéristiques

9.5.1 Exercices corrigés

Exercice 203. Pour tout $r > 0$, on pose $\Gamma(r) = \int_0^\infty e^{-t} t^{r-1} dt$. On rappelle que la loi $\Gamma(r, \lambda)$ est la probabilité de densité

$$\gamma_{r,\lambda}(t) = \frac{\lambda^r}{\Gamma(r)} e^{-\lambda t} t^{r-1} \, \mathbb{1}_{\mathbb{R}^+}(t).$$

1. Calculer la transformée de Laplace de $\Gamma(r, \lambda)$.
2. Montrer que si X et Y sont des variables indépendantes de loi respectives $\Gamma(r, \lambda)$ et $\Gamma(s, \lambda)$ alors $X + Y$ suit une loi $\Gamma(r + s, \lambda)$.
3. Soient $X_1, \dots, X_n$ des variables indépendantes de loi exponentielle de paramètre λ. Calculer la loi de $X_1 + \dots + X_n$.
4. Soient $X_1, \dots, X_n$ des variables indépendantes de loi $\mathcal{N}(0, 1)$. Montrer que X_1^2 suit une loi Γ et en déduire la loi de $X_1^2 + \dots + X_n^2$ (appelée loi du χ^2).

→ indication → solution

Exercice 204. Soient (X_n) une suite de variables aléatoires indépendantes de même loi non dégénérée à valeurs dans $\mathbb{N}$ et T une variable aléatoire à valeurs dans $\mathbb{N}^*$ indépendante des précédentes. On définit, pour tout $n \in \mathbb{N}^*$, la variable $S_n = \sum_{j=1}^{n} X_j$, puis $S(\omega) = S_{T(\omega)}(\omega)$ pour tout $\omega \in \Omega$.

1. Si G_T et G_{X_1} désignent les fonctions génératrices de T et X_1, montrer que la fonction génératrice de S est donnée par $G_S = G_T \circ G_{X_1}$.
2. *Formule de Wald.*
   Si X_1 et T admettent les moyennes (espérances) m et t, montrer que $\mathbb{E}[S] = mt$.

→ indication → solution

Exercice 205. *Loi de Laplace.*
Soient X et ε deux variables aléatoires indépendantes, où X suit la loi exponentielle de paramètre 1 et ε la loi de Rademacher : $\mathbb{P}_\varepsilon = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$.
On appelle *loi de Laplace* la loi de εX.

1. Montrer que la loi de Laplace est une loi à densité.
2. Calculer la fonction caractéristique de la loi de Laplace.

→ indication → solution

Exercice 206. On considère quatre variables aléatoires indépendantes identiquement distribuées $X_{1,1}$,$X_{1,2}$, $X_{2,1}$, $X_{2,2}$ de loi $\mathcal{N}(0,1)$.

1. On note $U = X_{1,1}X_{2,2}$ et $V = X_{1,2}X_{2,1}$. Déterminer la fonction caractéristique de U et de V.
2. Montrer que le déterminant $\begin{vmatrix} X_{1,1} & X_{1,2} \\ X_{2,1} & X_{2,2} \end{vmatrix}$ suit la loi de Laplace définie dans l'exercice précédent.

→ indication → solution

Exercice 207. *Probabilité de retour en zéro au temps n d'une marche aléatoire symétrique.*

1. Soit X une variable aléatoire à valeurs dans $\mathbb{Z}$. Montrer que
$$\mathbb{P}(X = 0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathbb{E}[e^{i\theta X}]\, d\theta.$$
2. Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes identiquement distribuées, avec $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$. On considère la somme partielle $S_n = X_1 + \cdots + X_n$.
   (a) Montrer que $\mathbb{P}(S_{2n} = 0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos^{2n}(\theta)\, d\theta$. En déduire la divergence de la série de terme générale $\mathbb{P}(S_{2n} = 0)$.
   (b) Montrer l'équivalent à l'infini $\mathbb{P}(S_{2n} = 0) \sim \dfrac{1}{\sqrt{\pi n}}$.

→ indication → solution

Exercice 208. Soit μ une mesure de probabilité sur $\mathbb{R}$, de fonction caractéristique ϕ.

1. Montrer que pour tout réel a et pour tout $T > 0$,
$$\frac{1}{2T}\int_{[-T,T]} e^{-ita}\phi(t)\, d\lambda(t) = \mu(\{a\}) + \int_{\mathbb{R}\setminus\{a\}} \frac{\sin(T(x-a))}{T(x-a)}\, d\mu(x)$$

2. En déduire que $\mu(\{a\}) = \lim_{T \to +\infty} \frac{1}{2T} \int_{[-T,T]} e^{-ita} \phi(t)\, d\lambda(t)$.
3. Montrer enfin qu'une mesure de probabilité μ sur $\mathbb{R}$ dont la fonction caractéristique a une limite nulle en $+\infty$ et en $-\infty$ est sans atome.
   Rappel : un atome de μ est un singleton de mesure strictement positive.

$\to$ indication $\to$ solution

9.5.2 Exercices non corrigés

Exercice 209. 1. On suppose que X et Y sont des variables aléatoires à valeurs entières, de fonctions génératrices G_X et G_Y. Soit A un événement indépendant de X et de Y, avec $\mathbb{P}(A) = p$. On note Z la variable aléatoire définie par $Z(\omega) = X(\omega)\, \mathbb{1}_A(\omega) + Y(\omega)\, \mathbb{1}_{A^c}(\omega)$, i.e.

$$Z(\omega) = \begin{cases} X(\omega) \text{ si } \omega \in A \\ Y(\omega) \text{ si } \omega \notin A. \end{cases}$$

Montrer que la fonction génératrice de Z est $pG_X + (1-p)G_Y$.

2. On lance trois fois de suite un dé à six faces. À chaque série de trois lancers est associé un score. Le score se calcule comme suit. Si le troisième lancer est un “1”, le score est le nombre de nombres pairs apparus lors des deux premiers lancers. Sinon, le score est le nombre de “6” apparus lors des deux premiers lancers.

   Quelle est la loi du score ?

   Exemples :
   - $2-4-1$ rapporte 2 points
   - $6-1-2$ rapporte 1 point
   - $1-4-2$ rapporte 0 point
   - $5-2-3$ rapporte 0 point.

$\to$ indication

Exercice 210. Soient N une variable aléatoire suivant une loi de Poisson de paramètre λ et $(X_i)_{i\geq 1}$ une suite de variables aléatoires réelles indépendantes identiquement distribuées de loi de Bernoulli de paramètre p, cette suite étant aussi indépendante de N. Montrer que $S = X_1 + X_2 + \cdots + X_N$ suit la loi de Poisson de paramètre λp.

$\to$ indication

Exercice 211. 1. Soit K un vecteur aléatoire n-dimensionnel dont les composantes sont indépendantes et suivent la loi de Poisson de paramètre λ. Soit L un vecteur aléatoire n-dimensionnel dont les composantes sont indépendantes et suivent la loi de Bernoulli de paramètre p. On suppose que K et L sont indépendants. Déterminer la fonction génératrice de $\langle K, L\rangle$.

2. On suppose que, pour tout $k \geq 0$, X_k suit une loi de Poisson de paramètre $k\lambda$. Soit T une variable aléatoire indépendante de la suite $(X_k)_{k\geq 0}$ et suivant la loi binomiale $\mathcal{B}(n,p)$. Déterminer la fonction génératrice de X_T.

3. En déduire que $\langle K, L \rangle$ suit la même loi que X_T.

→ indication

Exercice 212. Un joueur joue à pile ou face avec $n \geq 2$ pièces équilibrées qu'il lance simultanément. S'il n'obtient aucun pile, son gain est nul et la partie s'arrête. S'il obtient au moins un pile, il relance la première pièce autant de fois qu'il a obtenu de piles à la première phase du jeu et gagne autant d'unités que le nombre de piles obtenus lors de cette deuxième série de lancers. On note X_1 le nombre de piles obtenus à la première étape, et X_2 le gain du joueur.

1. Déterminer la fonction génératrice de X_2.
2. En déduire que

$$\forall k \in \{0, \dots, n\}, \quad \mathbb{P}(X_2 = k) = \binom{n}{k} \frac{1}{3^k} \left(\frac{3}{4}\right)^n.$$

$\rightarrow$ indication

Exercice 213. Montrer que la convolée de deux lois de Cauchy est une loi de Cauchy.
$\rightarrow$ indication

Exercice 214. Donner un exemple de variable aléatoire X dont la fonction caractéristique ϕ_X vérifie $(\phi_X)^2 = \phi_{2X}$. En déduire que la propriété

$$\forall t \in \mathbb{R}, \quad \phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$$

n'implique PAS que X et Y sont indépendantes.
$\rightarrow$ indication

Exercice 215. Soit $X \sim \mathcal{N}(0,1)$. Montrer que $\forall n \geq 0, \quad \mathbb{E}[X^{2n}] = \frac{(2n)!}{n! 2^n}$.
$\rightarrow$ indication

Exercice 216. Soient X et Y deux variables aléatoires admettant chacune un moment d'ordre 3. On pose $\phi(x,y) = \mathbb{E}\exp(i(xX + yY))$.

1. Montrer que X^2Y est intégrable.
2. Quelle est la régularité de ϕ ?
3. Exprimer $\mathbb{E}[X^2Y]$ en fonction de ϕ.

$\rightarrow$ indication

Exercice 217. Soit X une variable aléatoire dont la fonction caractéristique est de module constant. On se propose de montrer que X est constante.

1. Soit t non nul. Posons $A_t = \frac{\theta_t}{t} + \frac{2\pi}{t}\mathbb{Z}$. Montrer qu'il existe θ_t tel que $\phi(t) = e^{i\theta_t}$ et $\mathbb{P}(X \in A_t) = 1$.
2. Soient t et t' non nuls, $x, y \in A_t \cap A_{t'}$. Montrer que $x - y \in \frac{2\pi\mathbb{Z}}{t} \cap \frac{2\pi\mathbb{Z}}{t'}$. Conclure.

→ indication

Exercice 218. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes identiquement distribuées à valeurs dans $\mathbb{Z}$. On suppose que X_1 est centrée, avec un moment d'ordre deux. Montrer que la marche aléatoire $(S_n)_{n\geq 1}$ associée, définie par $S_n = X_1 + \cdots + X_n$ est récurrente. → indication

Exercice 219. Soient (X,Y) et (X',Y') deux vecteurs aléatoires. On suppose que, pour tout disque euclidien ouvert D du plan, on a

$$\mathbb{P}((X,Y)\in D) = \mathbb{P}((X',Y')\in D).$$

1. Montrer que pour tout demi-plan ouvert H du plan, on a
$$\mathbb{P}((X,Y)\in H) = \mathbb{P}((X',Y')\in H).$$
2. Soient a,b réels. Montrer que $aX+bY$ et $aX'+bY'$ ont même loi.
3. Montrer que les vecteurs (X,Y) et (X',Y') ont même loi.

→ indication

Exercice 220. Soit Y une variable aléatoire à valeurs dans $\mathbb{N}^*$. On note ϕ_Y sa fonction caractéristique. On fabrique une variable aléatoire Z à partir de Y : Z est choisi au hasard de manière uniforme entre 0 et $Y-1$. Ainsi, on suppose que

$$\forall n\in\mathbb{N}^* \quad \forall k\in\{0,\dots,n-1\} \quad \mathbb{P}(Z=k|Y=n) = \frac{1}{n}.$$

1. Montrer que
$$\forall t\in\mathbb{R}\backslash 2\pi\mathbb{Z} \quad \mathbb{E}\Big(e^{itZ}\mathbb{1}_{\{Y=n\}}\Big) = \frac{1}{n}\frac{1-e^{int}}{1-e^{it}}\mathbb{P}(Y=n).$$
2. On note ϕ_Z la fonction caractéristique de Z. Montrer que
$$\forall t\in\mathbb{R}\backslash 2\pi\mathbb{Z} \quad \phi_Z(t) = \frac{i}{e^{it}-1}\int_0^t \phi_Y(x)\,dx.$$
3. On suppose que Y admet un moment d'ordre 1. Montrer que Z admet un moment d'ordre 1, puis que $\mathbb{E}Z = \frac{\mathbb{E}Y-1}{2}$.
4. Soient X_1 et X_2 deux variables aléatoires indépendantes suivant la loi géométrique de paramètre p, avec $p\in]0,1[$. On pose $Y = X_1+X_2-1$. Montrer que la fonction caractéristique de Y vérifie
$$\forall t\in\mathbb{R}\backslash 2\pi\mathbb{Z} \quad \phi_Y(t) = \frac{p^2e^{it}}{(1-(1-p)e^{it})^2}.$$
5. Montrer que dans ce cas, $Z+1$ suit la loi géométrique de paramètre p.

→ indication

Chapitre 10

Convergences, lois des grands nombres

10.1 Convergence presque sûre

Définition. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. On dit qu'une suite de variables (ou de vecteurs) aléatoires $(X_n)_{n\geq 0}$ définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ converge presque sûrement vers une variable (ou un vecteur) aléatoire X si*

$$\mathbb{P}(\{\omega \in \Omega; X_n(\omega) \to X(\omega)\}) = 1.$$

On écrit alors $X_n \xrightarrow{\text{p.s.}} X$.

La convergence presque sûre n'est autre que la convergence presque partout relativement à une mesure de probabilité. On a alors les résultats classiques suivants. Si $X_n \xrightarrow{\text{p.s.}} X$ et $Y_n \xrightarrow{\text{p.s.}} Y$, (avec X et Y dans $\mathbb{R}^d$, $d \geq 1$) alors

— $\forall a \in \mathbb{R}, \quad aX_n \xrightarrow{\text{p.s.}} aX$.
— $X_n + Y_n \xrightarrow{\text{p.s.}} X + Y$.
— $\langle X_n, Y_n\rangle \xrightarrow{\text{p.s.}} \langle X, Y\rangle$.

Plus généralement, si $(X_i)_i$ et X sont à valeurs dans une partie E de $\mathbb{R}^d$ et si $X_n \xrightarrow{\text{p.s.}} X$, alors pour toute fonction f continue définie sur E, on a $f(X_n) \xrightarrow{\text{p.s.}} f(X)$.

Remarque 10.1. *Il est intéressant de remarquer que la convergence presque sûre d'une suite de vecteurs aléatoires est équivalente à la convergence presque sûre de chacune de ses composantes.*

On se souvient qu'on a étudié au chapitre 7 les liens entre convergence dans L^p et convergence presque partout. Ainsi, d'après le théorème 7.12, la convergence dans L^p entraîne la convergence presque partout (appelée ici convergence presque sûre) d'une sous-suite. Cependant, même dans le cadre d'une mesure de probabilités, la convergence dans L^p n'entraine toujours pas la convergence presque sûre : on peut par exemple se reporter à l'exercice corrigé 177.

10.1.1 Rappels d'analyse

En probabilités, le retour aux ε est très fréquent. Si l'on ne veut pas que cela devienne trop compliqué, il importe de bien connaître les outils d'analyse permettant de simplifier les choses. Rappelons quelques propriétés des limites supérieures.

— Pour $x \in \mathbb{R}$, on a

$$\lim_{n\to+\infty} x_n = x \iff \overline{\lim_{n\to+\infty}} |x_n - x| = 0.$$

— On a les équivalences

$$\overline{\lim_{n\to+\infty}} x_n \le M \iff \forall\varepsilon > 0 \quad \{n : x_n \ge M + \varepsilon\} \text{ est fini} \tag{10.1}$$

et

$$\overline{\lim_{n\to+\infty}} x_n \ge M \iff \forall\varepsilon > 0 \quad \{n : x_n \ge M - \varepsilon\} \text{ est infini.} \tag{10.2}$$

10.1.2 Limites supérieures, inférieures d'ensembles

Dans la pratique, comment montre-t-on que $\overline{\lim_{n\to+\infty}} X_n = M$ presque sûrement ? Comme vous l'avez deviné, on montre que $\overline{\lim_{n\to+\infty}} X_n \ge M$ presque sûrement, puis que $\overline{\lim_{n\to+\infty}} X_n \le M$ presque sûrement.

L'équivalence 10.2 dit exactement que

$$\left\{ \overline{\lim_{n\to+\infty}} X_n \ge M \right\} = \bigcap_{\varepsilon>0} \overline{\lim_{n\to+\infty}} \{X_n \ge M - \varepsilon\}.$$

Comme la suite $\overline{\lim_{n\to+\infty}} \{X_n \ge M - \varepsilon\}$ est monotone en ε, on a

$$\left\{ \overline{\lim_{n\to+\infty}} X_n \ge M \right\} = \bigcap_{\varepsilon\in\mathbb{Q}_+^*} \overline{\lim_{n\to+\infty}} \{X_n \ge M - \varepsilon\}. \tag{10.3}$$

L'avantage est que l'intersection est maintenant dénombrable. Or, on a le résultat classique très utile suivant.

Théorème 10.2. *L'intersection d'une famille dénombrable d'événements est de probabilité* 1 *si et seulement si chacun des événements est de probabilité* 1.

Démonstration. Soit D un ensemble d'indices dénombrable. Soit $(A_n)_{n\in D}$ une famille d'événements indicée par D. On pose $A = \bigcap_{n\in D} A_n$. Pour tout n, $A \subset A_n$, donc $\mathbb{P}(A) \le \mathbb{P}(A_n)$. Ainsi si $\mathbb{P}(A) = 1$, on a pour tout $n \in D$ $\mathbb{P}(A_n) = 1$. Réciproquement, on a

$$\mathbb{P}(A^c) = \mathbb{P}(\bigcup_{n\in D} A_n^c) \le \sum_{n\in D} \mathbb{P}(A_n^c) \le \sum_{n\in D} 0 = 0.$$

Donc $\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - 0 = 1$. □

Pour prouver que $\overline{\lim}_{n\to+\infty} X_n \ge M$ presque sûrement, il suffit donc de prouver que $\forall a < M, \quad \mathbb{P}\left(\overline{\lim}_{n\to+\infty} \{X_n \ge a\}\right) = 1$.

De la même manière, on voit que pour avoir $\overline{\lim}_{n\to+\infty} X_n \le M$ presque sûrement, il suffit donc de prouver que $\forall a > M, \quad \mathbb{P}\left(\underline{\lim}_{n\to+\infty} \{X_n < a\}\right) = 1$ ou, de manière équivalente, que $\forall a > M, \quad \mathbb{P}\left(\overline{\lim}_{n\to+\infty} \{X_n \ge a\}\right) = 0$.

On peut donc énoncer le théorème suivant.

Théorème 10.3. *Soient* X_n *une suite de variables aléatoires et* M *un réel. On suppose que*

1. $\forall a < M, \quad \mathbb{P}\left(\overline{\lim}_{n\to+\infty} \{X_n \ge a\}\right) = 1$,
2. $\forall a > M, \quad \mathbb{P}\left(\overline{\lim}_{n\to+\infty} \{X_n \ge a\}\right) = 0$.

Alors, $\overline{\lim}_{n\to+\infty} X_n = M$ *presque sûrement.*

Le théorème suivant très important en est une application directe

Théorème 10.4 (Critère fondamental de convergence presque-sûre)**.** *La suite de variables aléatoires* (X_n) *converge presque sûrement vers la variable aléatoire* X *si et seulement si*

$$\forall \varepsilon > 0 \quad \mathbb{P}\left(\overline{\lim}_{n\to+\infty} \{|X_n - X| \ge \varepsilon\}\right) = 0.$$

Démonstration. Il suffit d'appliquer le lemme précédent à la suite de variables aléatoires $(|X_n - X|)_{n\geq 0}$, avec $M = 0$ et $a = \varepsilon$. □

10.2 Convergence en probabilité

Définition. *On dit que la suite $(X_n)_{n\geq 1}$ converge en probabilité vers X si*

$$\forall \varepsilon > 0 \qquad \lim_{n\to+\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0.$$

On écrit alors $X_n \xrightarrow{\mathbb{P}} X$.

On peut remarquer que si X_n tend en probabilité vers X et Y_n en probabilité vers Y, alors

— le couple (X_n, Y_n) tend en probabilité vers (X, Y).

— $X_n + Y_n$ tend en probabilité vers $X + Y$.

En effet, on a les inégalités

$$\mathbb{P}((\|(X_n, Y_n) - (X, Y)\|_\infty \geq \varepsilon) \leq \mathbb{P}(|X_n - X| \geq \varepsilon) + \mathbb{P}(|Y_n - Y| \geq \varepsilon)$$

et $\mathbb{P}(|(X_n + Y_n) - (X + Y)| \geq \varepsilon) \leq \mathbb{P}(|X_n - X| \geq \varepsilon/2) + \mathbb{P}(|Y_n - Y| \geq \varepsilon/2)$.

10.2.1 Comparaison avec les autres modes de convergence

Convergence dans L^p et convergence en probabilité

Théorème 10.5. *La convergence dans L^p ($p \geq 1$) implique la convergence en probabilité.*

Démonstration. On a

$$\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{P}(|X_n - X|^p \geq \varepsilon^p) \leq \frac{\mathbb{E}|X_n - X|^p}{\varepsilon^p}.$$

□

La réciproque est fausse : la convergence en probabilité n'entraîne pas la convergence dans L^p (voir par exemple l'exercice corrigé 222).

Convergence presque sûre et convergence en probabilité

Théorème 10.6. *La convergence presque sûre implique la convergence en probabilité.*

Démonstration. Soit $\varepsilon > 0$. D'après le théorème 10.4, on a

$$\mathbb{P}\left(\overline{\lim_{n\to+\infty}} \{|X_n - X| \geq \varepsilon\}\right) = 0.$$

Or, d'après le théorème de continuité séquentielle décroissante, on a

$$\mathbb{P}\left(\overline{\lim_{n\to+\infty}} \{|X_n - X| \geq \varepsilon\}\right) = \lim_{n\to+\infty} \mathbb{P}\left(\bigcup_{k=n}^{+\infty} \{|X_k - X| \geq \varepsilon\}\right).$$

Comme

$$0 \leq \mathbb{P}(|X_n - X| \geq \varepsilon) \leq \mathbb{P}\left(\bigcup_{k=n}^{+\infty} \{|X_k - X| \geq \varepsilon\}\right),$$

on en déduit que

$$\lim_{n\to+\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0.$$

Comme ε est quelconque, on conclut que X_n converge en probabilité vers X. □

En revanche, la convergence en probabilité n'entraîne pas la convergence presque sûre. Un exemple sera traité un peu plus loin dans ce chapitre.

Remarque 10.7. *Il y a une différence fondamentale entre la convergence presque sûre et les autres modes de convergence évoqués ici : pour une suite $(X_n)_{n\geq 1}$, la convergence en probabilité ou dans L^p ne met en jeu que la suite des distributions des variables individuelles, alors que la convergence presque sûre met en jeu les distributions conjointes.*

10.2.2 Loi faible des grands nombres

On va commencer par un résultat qui n'est optimal, ni dans ses hypothèses, ni dans sa conclusion, mais dont la preuve, très classique, est à mémoriser. On pourra alors passer à des énoncés plus raffinés.

Théorème 10.8. *Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires de même loi, admettant un moment d'ordre 2 et deux à deux non corrélées.*
On pose

$$S_n = \sum_{k=1}^n X_k \quad \text{et} \quad M_n = \frac{1}{n} S_n.$$

Alors $M_n \xrightarrow{L^2} \mathbb{E}X_0$. On dit que M_n converge en moyenne quadratique vers $\mathbb{E}X_0$. On a également $M_n \xrightarrow{\mathbb{P}} \mathbb{E}X_0$.

Démonstration. On a $\mathbb{E}M_n = \frac{1}{n}\mathbb{E}S_n = \frac{1}{n} \sum_{k=1}^n \mathbb{E}X_k = \frac{1}{n} n\mathbb{E}X_0 = \mathbb{E}X_0$. Par conséquent $\mathbb{E}|M_n - \mathbb{E}X_0|^2 = \operatorname{Var} M_n = \frac{1}{n^2} \operatorname{Var} S_n$. Comme les X_k sont deux à deux non corrélées, on a

$$\operatorname{Var} S_n = \sum_{k=1}^n \operatorname{Var} X_k = n \operatorname{Var} X_1.$$

On a donc
$$\mathbb{E}\left(|M_n - \mathbb{E}X_0|^2\right) = \frac{\operatorname{Var} X_1}{n}, \tag{10.4}$$
qui tend bien vers zéro. □

Le théorème qui précède est la loi faible des grands nombres classique, qu'il faut retenir en première approche. Elle a toutefois le défaut de requérir un moment d'ordre deux, ce qui est un peu dommage. Nous vous proposons donc un critère de convergence en probabilité plus général, qui s'accommode d'hypothèses d'intégrabilité plus faibles. Ce sera l'occasion de découvrir les arguments de troncature.

Théorème 10.9. *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires deux à deux indépendantes ; on note S_n les sommes partielles. Soit $(b_n)_{n\geq 1}$ une suite croissante de réels strictement positifs tendant vers l'infini. On pose également, pour $n \geq 1$ et $k \in \{1, \dots, n\}$:*
$$Y_{k,n} = X_k \mathbb{1}_{\{|X_k|\leq b_n\}}, \quad S'_n = \sum_{k=1}^n Y_{k,n}, \text{ et } \mu_n = \mathbb{E}[S'_n].$$
Si
$$\lim_{n\to+\infty} \sum_{k=1}^n \mathbb{P}(|X_k| > b_n) = 0 \text{ et } \lim_{n\to+\infty} \frac{1}{b_n^2}\sum_{k=1}^n \operatorname{Var}(Y_{k,n}) = 0,$$
alors $\frac{S_n-\mu_n}{b_n} \xrightarrow{\mathbb{P}} 0$.

Remarque 10.10. *Dans le cas où les variables aléatoires sont supposées globalement indépendantes, la réciproque est vraie, et peut être établie avec les outils du chapitre 14, voir par exemple Etemadi [12].*

Démonstration. Soit $\varepsilon > 0$.
$$\begin{aligned}\mathbb{P}\left(|\frac{S_n-\mu_n}{b_n}| \geq \varepsilon\right) &\leq \mathbb{P}\left(|\frac{S'_n-\mu_n}{b_n}| \geq \varepsilon\right) + \mathbb{P}\left(S_n \neq S'_n\right)\\ &\leq \frac{1}{b_n^2}\frac{1}{\varepsilon^2}\operatorname{Var} S'_n + \sum_{k=1}^n \mathbb{P}\left(Y_{k,n} \neq X_k\right)\\ &\leq \frac{1}{b_n^2}\frac{1}{\varepsilon^2}\sum_{k=1}^n \operatorname{Var} Y_{k,n} + \sum_{k=1}^n \mathbb{P}\left(|X_k| > b_n\right)\end{aligned}$$
ce qui donne le résultat voulu. □

Corollaire 10.11. *Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires deux à deux indépendantes de même loi, admettant un moment d'ordre 1.*
On pose
$$S_n = \sum_{k=1}^n X_k \quad \text{et} \quad M_n = \frac{1}{n}S_n.$$

Alors $M_n \xrightarrow{\mathbb{P}} \mathbb{E}X_0$.

Démonstration. On va appliquer le théorème 10.9 avec $b_n = n$. On a

$$\sum_{k=1}^{n} \mathbb{P}(|X_k| > b_n) = n\mathbb{P}(|X_1| > n) \leq \mathbb{E}[|X_1|\mathbb{1}_{\{|X_1|\geq n\}}],$$

qui tend vers 0 d'après le théorème de convergence dominée. De même

$$\frac{1}{b_n^2}\sum_{k=1}^{n} \mathrm{Var}(Y_{k,n}) = \frac{1}{n}\mathrm{Var}\, Y_{1,n} \leq \frac{1}{n}\mathbb{E}[X_1^2\mathbb{1}_{\{|X_1|\leq n\}}] = \mathbb{E}\left[|X_1|\frac{|X_1|\mathbb{1}_{\{|X_1|\leq n\}}}{n}\right],$$

qui tend vers zéro, encore grâce au théorème de convergence dominée. □

Le théorème 10.9 met aussi en lumière le fait que la convergence en probabilité est possible sans que les variables soient intégrables.

Exemple: Soit X une variable aléatoire symétrique dont la loi est la suivante. Pour $t \in [0, e[$, on a $\mathbb{P}(|X| > t) = \frac{1}{2}$. Pour $t \geq e$, on définit $\mathbb{P}(|X| > t) = \frac{e}{2t\log t}$. On a $\mathbb{E}[|X|] = \int_{[0,+\infty[} \mathbb{P}(|X| > t)\, d\lambda(t) = \frac{e}{2} + \int_e^{+\infty} \frac{e\ dt}{2t\log t} = +\infty$, donc X n'est pas intégrable (et a fortiori pas de carré intégrable). On considère $(X_n)_n$, une suite de variables aléatoires indépendantes, toutes de même loi que X. On veut montrer que $\frac{X_1+\cdots+X_n}{n}$ tend en probabilité vers 0. L'idée est d'appliquer le théorème avec $b_n = n$: on a alors

1. $\sum\limits_{k=1}^{n} \mathbb{P}(|X_k| > b_n) = n\mathbb{P}(|X| > n) \to 0$,
2. $\mathbb{E}(X_k\mathbb{1}_{\{|X_k|\leq n\}}) = 0$ car la loi de X est symétrique, donc $\mu_n = 0$,
3. $\frac{1}{n^2}\sum\limits_{k=1}^{n} \mathrm{Var}(X\mathbb{1}_{\{|X|\leq n\}}) = \frac{1}{n}\mathbb{E}[X^2\mathbb{1}_{\{|X|\leq n\}}]$. Il suffit donc de montrer que $\mathbb{E}[X^2\mathbb{1}_{\{|X|<n\}}] = o(n)$. Or,

$$\begin{aligned}\mathbb{E}[X^2\mathbb{1}_{\{|X|<n\}}] &= \int_{\mathbb{R}_+} \mathbb{P}(X^2\mathbb{1}_{\{|X|<n\}} > t)\, d\lambda(t)\\ &= \int_{[0,n]} \mathbb{P}(X^2\mathbb{1}_{\{|X|<n\}} > t)\, d\lambda(t) \leq \int_{[0,n^2]} \mathbb{P}(X > \sqrt{t})\, d\lambda(t)\\ &\leq e^2 + \int_{e^2}^{n^2} \frac{e\ dt}{2\sqrt{t}\log\sqrt{t}}\end{aligned}$$

   Lorsque t tend vers l'infini, on a $\frac{1}{\sqrt{t}\log\sqrt{t}} = o(\frac{1}{\sqrt{t}})$: comme l'intégrale $\int_{e^2}^{+\infty} \frac{d\lambda(t)}{\sqrt{t}}$ vaut $+\infty$, on en déduit que

$$\mathbb{E}[X^2\mathbb{1}_{\{|X|<n\}}] = o\left(\int_{e^2}^{n^2} \frac{dt}{\sqrt{t}}\right) = o(n),$$

   ce qui donne le résultat voulu.

10.3 Lemmes de Borel-Cantelli

10.3.1 Premier lemme de Borel–Cantelli

Théorème 10.12. *Soit $(A_n)_{n\geq 1}$ une suite d'événements. Si la série de terme général $\mathbb{P}(A_n)$ converge, alors $\mathbb{P}\left(\overline{\lim_{n\to+\infty}} A_n\right) = 0$.*

Démonstration. On pose $B_n = \bigcup_{k=n}^{+\infty} A_k$. La suite (B_n) est décroissante, et l'intersection des (B_n) est, par définition, $\overline{\lim_{n\to+\infty}} A_n$. D'après le théorème de continuité séquentielle décroissante, on a donc

$$0 \leq \mathbb{P}\left(\overline{\lim_{n\to+\infty}} A_n\right) = \lim_{n\to+\infty} \mathbb{P}(B_n).$$

Or on a de plus

$$\mathbb{P}(B_n) = \mathbb{P}\left(\bigcup_{k=n}^{+\infty} A_k\right) \leq \sum_{k=n}^{+\infty} \mathbb{P}(A_k) = r_n.$$

Comme r_n est le reste d'ordre n d'une série convergente, r_n est de limite nulle, et donc, par comparaison $\mathbb{P}\left(\overline{\lim_{n\to+\infty}} A_n\right) = 0$. □

Remarque 10.13. — *Dans un contexte probabiliste, on écrit parfois $\{A_n$ infiniment souvent$\}$ ou $\{A_n$ i.s.$\}$ pour $\overline{\lim_{n\to+\infty}} A_n$.*

La propriété "$\mathbb{P}\left(\overline{\lim_{n\to+\infty}} A_n\right) = 0$" signifie que presque sûrement, seul un nombre fini de A_n se réalisent.

— *On dit parfois que la suite $(X_n)_{n\geq 1}$ converge presque complètement vers X si quel que soit $\varepsilon > 0$, la série de terme général $\mathbb{P}(|X_n - X| > \varepsilon)$ converge. À l'aide du lemme de Borel-Cantelli et du critère fondamental de convergence presque sûre, il n'est pas difficile de voir que la convergence presque complète entraîne la convergence presque sûre. Ce petit raisonnement se retrouvera fréquemment dans les exercices. On peut noter que, comme la convergence en probabilité ou la convergence dans L^p, la convergence presque complète ne met en jeu que la suite des distributions des variables individuelles.*

10.3.2 Deuxième lemme de Borel-Cantelli

Le deuxième lemme de Borel-Cantelli est une sorte de réciproque du premier, dans le cas où les événements considérés sont indépendants. Ici, on choisit de présenter d'emblée une généralisation du deuxième lemme de Borel-Cantelli, due à Erdös et Renyi (1959). Le théorème classique s'en déduira aisément.

Théorème 10.14 (Erdös-Renyi). *Soit* $(A_n)_{n\geq 1}$ *une suite d'événements. On pose*

$$N_n = \sum_{k=1}^{n} \mathbb{1}_{A_k}, \quad N = \sum_{k=1}^{+\infty} \mathbb{1}_{A_k} \quad \textit{et} \quad m_n = \sum_{k=1}^{n} \mathbb{P}(A_k) = \mathbb{E}N_n.$$

Si

$$\lim_{n\to+\infty} m_n = +\infty \quad \textit{et} \quad \lim_{n\to+\infty} \frac{\operatorname{Var} N_n}{m_n^2} = 0,$$

alors

$$\mathbb{P}\left(\varlimsup_{n\to\infty} A_n\right) = 1.$$

Remarque 10.15. *On a* $\varlimsup_{n\to\infty} A_n = \{N = +\infty\}$.

Démonstration. Pour $m_n > a$, on a

$$\mathbb{P}(N \leq a) \leq \mathbb{P}(N_n \leq a) \leq \mathbb{P}(|N_n - m_n| \geq m_n - a) \leq \frac{\operatorname{Var} N_n}{(m_n - a)^2}.$$

En faisant tendre n vers $+\infty$, on en déduit que

$$\forall a \in \mathbb{N} \quad \mathbb{P}(N \leq a) = 0.$$

Ainsi

$$\mathbb{P}(N < +\infty) = \mathbb{P}(\lim_{a\to+\infty}\uparrow \{N \leq a\}) = \lim_{a\to+\infty} \mathbb{P}(N \leq a) = \lim_{a\to+\infty} 0 = 0.$$

□

Théorème 10.16 (Second lemme de Borel-Cantelli). *Soit* $(A_n)_{n\geq 1}$ *une suite d'événements indépendants. Si la série de terme général* $\mathbb{P}(A_n)$ *diverge, alors* $\mathbb{P}\left(\varlimsup_{n\to+\infty} A_n\right) = 1$.

Démonstration. On va appliquer le théorème précédent. Comme les $(A_k)_{k\geq 1}$ sont indépendants, leurs indicatrices sont des variables aléatoires indépendantes, et donc

$$\operatorname{Var} N_n = \sum_{k=1}^{n} \operatorname{Var} \mathbb{1}_{A_k} = \sum_{k=1}^{n} \mathbb{P}(A_k)\,(1 - \mathbb{P}(A_k)) \leq \sum_{k=1}^{n} \mathbb{P}(A_k) = m_n.$$

Ainsi $\frac{\operatorname{Var} N_n}{m_n^2} \leq \frac{1}{m_n}$. Comme $\lim\limits_{n\to+\infty} m_n = \sum\limits_{k=1}^{+\infty} \mathbb{P}(A_k) = +\infty$, le résultat s'ensuit. □

Remarque 10.17. *La propriété "$\mathbb{P}\left(\overline{\lim\limits_{n\to+\infty}} A_n\right) = 1$" signifie que presque sûrement, A_n est réalisé pour une infinité de valeurs de n.*

Scolie : La conclusion du second lemme de Borel-Cantelli reste-t-elle vraie si l'on suppose seulement que les $(A_k)_{k\geq 1}$ sont deux à deux indépendants ? Que les $(A_k)_{k\geq 1}$ sont négativement corrélés [1] ?

Corollaire 10.18 (loi 0–1 de Borel). *Soit $(A_n)_{n\geq 1}$ une suite d'événements indépendants. La probabilité de l'événement $\mathbb{P}\left(\overline{\lim\limits_{n\to+\infty}} A_n\right)$ ne peut valoir que 0 ou 1. Elle vaut 1 si et seulement si la série de terme général $\mathbb{P}(A_n)$ diverge.*

Démonstration. C'est une conséquence immédiate des deux lemmes de Borel-Cantelli. □

Théorème 10.19. *Soit $(X_n)_{n\geq 0}$ une suite convergeant en probabilité vers X. Alors, il existe une sous-suite $(X_{n_k})_{k\geq 1}$ telle que $X_{n_k} \xrightarrow[k\to\infty]{p.s.} X$.*

Démonstration. On pose $n_0 = 0$, puis pour $k \geq 1$:

$$n_k = \inf\left\{n > n_{k-1}; \mathbb{P}\left(|X_n - X| \geq \frac{1}{k}\right) \leq \frac{1}{2^k}\right\}.$$

À k fixé, $\mathbb{P}\left(|X_n - X| \geq \frac{1}{k}\right)$ tend vers 0 quand n tend vers l'infini, donc on a bien, pour tout k, $n_k < +\infty$.

De plus, on a pour tout $k \geq 1$,

$$\mathbb{P}\left(|X_{n_k} - X| \geq \frac{1}{k}\right) \leq \frac{1}{2^k}.$$

Comme la série de terme général $\frac{1}{2^k}$ converge, le premier lemme de Borel-Cantelli nous permet d'affirmer que

$$\mathbb{P}\left(\overline{\lim_{k\to+\infty}} \{|X_{n_k} - X| \geq \frac{1}{k}\}\right) = 0,$$

ce qui est équivalent à

$$\mathbb{P}\left(\underline{\lim_{k\to+\infty}} \{|X_{n_k} - X| < \frac{1}{k}\}\right) = 1.$$

1. C'est-à-dire que $\mathbb{P}(A_i \cap A_j) \leq \mathbb{P}(A_i)\mathbb{P}(A_j)$ pour $i \neq j$.

Cela veut dire que pour presque tout ω, il existe un $k_0(\omega)$ tel que

$$k \geq k_0(\omega) \Longrightarrow |X_{n_k}(\omega) - X(\omega)| < \frac{1}{k},$$

ce qui implique bien sûr que $X_{n_k}(\omega)$ tend vers $X(\omega)$ pour $\mathbb{P}$-presque tout ω. □

Corollaire 10.20. *Si (X_n) est une suite à valeurs dans $E \subset \mathbb{R}^d$ convergeant en probabilité vers X et si f est une fonction continue sur E, alors $f(X_n)$ converge en probabilité vers $f(X)$.*

Démonstration. Soit $\varepsilon > 0$. On doit montrer que la suite $u_n = \mathbb{P}(|f(X_n) - f(X)| \geq \varepsilon)$ converge vers 0. Soit $(n_k)_{k\geq 1}$ une suite d'entiers strictement croissante telle que $\lim_{k\to+\infty} u_{n_k} = \overline{\lim_{n\to+\infty}} u_n$. On pose $Y_k = X_{n_k}$. Y_k tend en probabilité vers X, donc il existe une suite $(m_k)_{k\geq 1}$ d'entiers strictement croissante telle que Y_{m_k} converge presque sûrement vers X. Ceci entraîne que $f(Y_{m_k})$ converge presque sûrement vers $f(X)$, et donc que $f(Y_{m_k})$ converge en probabilité vers $f(X)$. Une suite convergente a la même limite que chacune de ses sous-suites :

$$\lim_{k\to+\infty} \mathbb{P}(|f(Y_k) - f(X)| \geq \varepsilon) = \lim_{k\to+\infty} \mathbb{P}(|f(Y_{m_k}) - f(X)| \geq \varepsilon) = 0,$$

soit $\lim_{k\to+\infty} u_{n_k} = \overline{\lim_{n\to+\infty}} u_n = 0.$ □

10.4 Lois fortes des grands nombres

Il existe de nombreuses lois fortes des grands nombres, c'est-à-dire des théorèmes de convergence presque sûre pour les moyennes de suite de variables aléatoires. La plus ancienne est dûe à Émile Borel : elle concerne la répartition des chiffres du dévoloppement en base deux d'un réel de $[0, 1]$.

On va ici présenter deux théorèmes : le premier concerne des variables non-corrélées, qui n'ont pas nécessairement la même loi. Sa preuve est assez courte. Le second théorème présenté , dû à Etemadi, est l'aboutissement d'une longue suite d'améliorations successives : c'est ce résultat que nous appellerons « la loi forte des grands nombres ».

10.4.1 Deux lois fortes des grands nombres

Théorème 10.21. *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires de carré intégrables, deux à deux non corrélées, et telles que $\sup_{n\geq 1} \operatorname{Var} X_n < +\infty$. Alors, si* *l'on pose $S_n = X_1 + \cdots + X_n$, on a*

$$\frac{S_n - \mathbb{E}[S_n]}{n} \xrightarrow{p.s.} 0.$$

Démonstration. On pose $C = \sup_{n\geq 1} \operatorname{Var} X_n < +\infty$. Quitte à remplacer X_i par $X_i - \mathbb{E}[X_i]$, on peut supposer sans perte de généralité que les X_i sont centrés

$$\operatorname{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\sum_{k=1}^{n} \operatorname{Var} X_k \leq \frac{C}{n}.$$

Avec l'inégalité de Tchebitchev, cela donne

$$\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}[S_n]}{n}\right| \geq \varepsilon\right) \leq \varepsilon^{-2} \operatorname{Var} \frac{S_n}{n} \leq \frac{C}{n\varepsilon^2},$$

d'où

$$\mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right| \geq \varepsilon\right) \leq \frac{C}{n^2\varepsilon^2},$$

ce qui, avec le lemme de Borel-Cantelli, donne la convergence presque sûre de $\frac{S_{n^2}}{n^2}$ vers 0. Soit maintenant $n \geq 1$ et notons $p = p(n) = \lfloor\sqrt{n}\rfloor^2$. On a $n + 1 - 2\sqrt{n} = (\sqrt{n} - 1)^2 \leq p(n) \leq n$, donc $n - p(n) = O(\sqrt{n})$, et en particulier $\lim \frac{p(n)}{n} = 1$. Notons $D_{n,p} = X_{p+1} + \cdots + X_n$. On a

$$\operatorname{Var}\left(\frac{D_{n,p}}{n}\right) = \frac{1}{n^2} \operatorname{Var} D_{n,p} \leq \frac{C(n - p(n))}{n^2} = \frac{O(\sqrt{n})}{n^2} = O(\frac{1}{n^{3/2}}).$$

Par les mêmes arguments que précédemment, $\frac{D_{n,p}}{n}$ tend presque sûrement vers 0. Comme $\lim p(n) = +\infty$ et que $\frac{S_{p(n)}}{p(n)}$ a les mêmes termes que $\frac{S_{n^2}}{n^2}$, $\frac{S_{p(n)}}{p(n)}$ tend également presque sûrement vers 0, d'où on déduit que $\frac{S_n}{n}$ tend presque sûrement vers 0 avec l'identité

$$\frac{S_n}{n} = \frac{D_{n,p}}{n} + \frac{p(n)}{n}\frac{S_{p(n)}}{p(n)}.$$

□

Corollaire 10.22. *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires deux à deux indépendantes, de même loi μ. On suppose que μ admet un moment d'ordre 2. Alors*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{p.s.} \mathbb{E}X_1.$$

Démonstration. Des variables indépendantes sont non corrélées ; on applique le théorème précédant en notant qu'ici $\mathbb{E}[S_n] = n\mathbb{E}[X_1]$. □

En réalité, l'existence d'un moment d'ordre 2 n'est pas nécessaire, comme le montre le résultat suivant, dû à Etemadi :

Théorème 10.23 (Etemadi). *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires deux à deux indépendantes, de même loi μ. On suppose que μ admet un moment d'ordre 1. Alors*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{p.s.} \mathbb{E}X_1.$$

C'est ce dernier théorème que nous demandons au lecteur de retenir en première approche. Il sera démontré un peu plus loin sous la forme d'un exercice corrigé.

Notons que ce résultat avait été démontré par Kolmogorov en 1929 sous l'hypothèse plus forte d'une indépendance globale.

Remarque 10.24. *On remarque que les hypothèses de la loi forte des grands nombres sont, en un certain sens, minimales. En effet, on verra en exercice que si des variables aléatoires indépendantes identiquement distribuées $(X_n)_{n\geq 1}$ ne sont pas intégrables, alors $\sup_{n\geq 1} \frac{|S_n|}{n} = +\infty$ presque sûrement, alors qu'il serait presque sûrement fini si S_n/n convergeait presque sûrement.*

En particulier, si l'on reprend l'exemple précédent de la suite $(X_n)_{n\geq 1}$ de variables aléatoires symétriques identiquement distribuées vérifiant $\mathbb{P}(|X_k| > t) = \frac{1}{2}$ pour $t \in [0, e[$ et $\mathbb{P}(|X_k| > t) = \frac{e}{2t\log t}$ pour $t \geq e$, cette suite vérifie une loi faible des grands nombres puisque S_n/n tend en probabilité vers 0, mais pas une loi forte, puisque, on l'a vu, X_1 n'est pas intégrable. Ceci fournit un exemple de suite convergeant en probabilité, mais pas presque sûrement, ainsi qu'il avait été annoncé précédemment.

Par la loi faible des grands nombres, on obtient que $\frac{X_1+\cdots+X_n}{n}$ fluctue de moins en moins et converge vers l'espérance. Cela se traduit par : pour tout $\varepsilon > 0$, la moyenne empirique $\frac{X_1+\cdots+X_n}{n}$ est comprise entre $[\mathbb{E}(X_1) - \varepsilon]$ et $[\mathbb{E}(X_1) + \varepsilon]$ avec une probabilité tendant vers 1 lorsque n tend vers l'infini. La loi forte des grands nombres dit beaucoup plus. Avec probabilité 1, pour tout $\varepsilon > 0$, on peut trouver n_ε tel que $\left|\frac{X_1+\cdots+X_n}{n} - \mathbb{E}X_1\right| < \varepsilon$ si $n \geq n_\varepsilon$ et ces n_ε ne dépendent que de ε et du tirage de X_i. En particulier, il suffit d'effectuer une seule suite infinie de répétitions d'une expérience aléatoire pour découvrir empiriquement la valeur de $\mathbb{P}(A)$, qui est la fréquence d'apparition d'un événement A.

10.4.2 Probabilités et fréquences asymptotiques

Théorème 10.25. *Soit $(A_n)_{n\geq 0}$ une suite d'événements indépendants de même probabilité p sur $(\Omega, \mathcal{F}, \mathbb{P})$. Pour $\omega \in \Omega$, on note $N_n(\omega)$ le nombre d'événements qui sont réalisés parmi $A_1, \ldots, A_n$. Ainsi, on a*

$$N_n = \sum_{k=1}^{n} \mathbb{1}_{A_k} \text{ et } F_n = \frac{1}{n} N_n.$$

Alors $\mathbb{P}(F_n \to p) = 1$.

Démonstration. On pose $X_k = \mathbb{1}_{A_k}$ et on applique le théorème 10.23. X_k admet bien un moment d'ordre 1 car $0 \leq X_k \leq 1$ et l'on a $\mathbb{E}X_1 = \mathbb{P}(A_1) = p$. □

10.4.3 Méthode de Monte-Carlo

Nous allons développer ici la méthode de Monte-Carlo qui consiste à utiliser des expériences répétées pour évaluer une quantité ou résoudre un système déterministe. Considérons par exemple le problème de l'intégration numérique. Soit $g : \mathbb{R}^d \to \mathbb{R}$ une fonction intégrable par rapport à la mesure de Lebesgue. On veut approcher la valeur de l'intégrale

$$I = \int_{[0,1]} g(x)\, dx.$$

La méthode de Monte-Carlo consiste alors à écrire cette intégrale sous la forme $I = \mathbb{E}[g(U)]$ où U est une variable aléatoire de loi uniforme sur $[0,1]$, que l'on sait simuler (voir la section 5.6.8). Ainsi, si g^2 est intégrable et si $(U_i)_{i\geq 1}$ est une suite de variables aléatoires indépendante et identiquement distribuées, de loi uniforme sur $[0,1]$ alors la loi forte des grands nombres stipule que

$$\frac{1}{n} \sum_{i=1}^{n} g(U_i) \underset{n\to+\infty}{\overset{p.s.}{\to}} \mathbb{E}[g(U)].$$

Dit autrement, si on tire aléatoirement des nombres $u_1, \ldots, u_n$ compris entre 0 et 1, alors $\frac{1}{n}\left(g(u_1) + \ldots + g(u_n)\right)$ est une approximation de I. Le résultat général est le suivant :

Théorème 10.26. *Soit* $g : [0,1]^d \to \mathbb{R}$ *une fonction intégrable telle que* g^2 *est intégrable. Soit* $(X_i)_{i\geq 1}$ *une suite de variables aléatoires indépendantes et identiquement distribuées de loi uniforme sur* $[0,1]^d$.
Alors la suite $\left(\frac{1}{n} \sum_{i=1}^{n} g(X_i)\right)_{n\geq 1}$ *converge presque sûrement (et dans* L^1*) vers* $I = \int_{[0,1]^d} g(x)\, dx$.

Démonstration. Comme les variables X_i sont indépendantes, il en de même pour $g(X_i)$. De plus, les variables $g(X_i)$ sont de même loi. Par le théorème de transfert et l'hypothèse d'intégrabilité de g, les variables $g(X_i)$ sont intégrables. Ainsi, g étant de carré intégrable, la loi forte des grands nombres s'applique à la suite $g(X_i)$ et implique que la suite $\left(\frac{1}{n} \sum_{i=1}^{n} g(X_i)\right)_{n\geq 1}$ converge presque sûrement (et dans L^1) vers $\mathbb{E}[g(X_1)]$. Il ne reste qu'à déterminer cette dernière espérance. La variable X_1 suit la loi uniforme sur $[0,1]^d$ et donc le théorème de transfert donne

$$\mathbb{E}[g(X_1)] = \int_{\mathbb{R}^d} g(x)\, d\mathbb{P}_{X_1}(x) = \int_{\mathbb{R}^d} g(x) \mathbb{1}_{[0,1]^d}(x)\, dx = I.$$

□

Remarque 10.27. *Il est maintenant aisé de voir que le raisonnement ci-dessus est inchangé si, au lieu de prendre des variables aléatoires de loi uniforme sur $[0,1]^d$, on considère des variables de densité f que l'on sait simuler et telle que $\int_{\mathbb{R}^d} \frac{g^2(x)}{f(x)}\, dx < +\infty$.*

Nous connaissons donc ainsi une valeur approchée de l'intégrale I. Une question naturelle est de connaître sa vitesse de convergence et d'estimer l'erreur. Nous pouvons déjà donner une estimation très grossière de l'erreur d'approximation via la loi faible des grands nombres.

Proposition 10.28. *Soient $(X_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes et identiquement distribuées et $g : \mathbb{R}^d \to \mathbb{R}$ telle que $\mathbb{E}[g(X_1)^2] < +\infty$. Alors pour tout $\varepsilon > 0$, on a*

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n g(X_i) - \mathbb{E}[g(X_1)]\right| \geq \varepsilon\right) \leq \frac{\mathrm{Var}(X_1)}{n\varepsilon^2}.$$

Démonstration. Soit $\varepsilon > 0$. Par l'inégalité de Markov, on obtient

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n g(X_i) - \mathbb{E}[g(X_1)]\right| \geq \varepsilon\right) &= \mathbb{P}\left(\left(\frac{1}{n} \sum_{i=1}^n g(X_i) - \mathbb{E}[g(X_1)]\right)^2 \geq \varepsilon^2\right) \\ &= \mathbb{P}\left(\left(\sum_{i=1}^n (g(X_i) - \mathbb{E}[g(X_1)])\right)^2 \geq n^2\varepsilon^2\right) \\ &\leq \frac{1}{n^2\varepsilon^2}\, \mathbb{E}[\left(\sum_{i=1}^n g(X_i) - n\mathbb{E}[g(X_1)]\right)^2].\end{aligned}$$

Comme les variables X_i sont indépendantes, on a

$$\mathbb{E}[\left(\sum_{i=1}^n g(X_i) - n\mathbb{E}[g(X_1)]\right)^2] = \mathrm{Var}(\sum_{i=1}^n g(X_i)) = n\, \mathrm{Var}(g(X_1)).$$

On en déduit le résultat. □

Nous verrons des résultats beaucoup plus précis dans la section 13.2.

10.4.4 Exercice : une preuve de la loi forte des grands nombres

La preuve, désormais classique, présentée ici sous forme d'exercice, est due à Etemadi [11]. On admirera l'habileté avec laquelle sont combinés deux types d'arguments classiques : les arguments de troncature, et l'utilisation de sous-suites pour lesquelles il y a convergence presque complète.

La belle méthode d'Etemadi a d'ailleurs fait florès : on pourra par exemple voir dans le traité de Bulinski et Shashkin [4] comment elle peut être utilisée pour généraliser le théorème 10.21 à certaines familles de variables aléatoires dépendantes.

Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires positives de même loi deux à deux indépendantes, admettant un moment d'ordre un. On pose

$$S_n = \sum_{i=1}^{n} X_i \text{ et } Q_n = \frac{1}{n}S_n.$$

On considère également les variables aléatoires tronquées : $X_i^* = X_i\mathbb{1}_{\{X_i\leq i\}}$ et les sommes et quotients associés : $S_n^* = \sum_{i=1}^{n} X_i^*$ et $Q_n^* = S_n^*/n$.

1. Montrer que
$$\operatorname{Var} S_n^* \leq \sum_{i=1}^{n} \mathbb{E}\left((X_i^*)^2\right).$$
En déduire que
$$\operatorname{Var} S_n^* \leq n\mathbb{E}[X_1^2\mathbb{1}_{X_1\leq n}].$$
2. Soit $\beta > 1$. On note u_n l'entier le plus proche de β^n. Montrer que $u_n \sim \beta^n$. En déduire que
$$\sum_{n=N}^{+\infty} \frac{1}{u_n} = O(\frac{1}{u_N}).$$
3. Montrer qu'il existe une constante $C > 0$ telle que
$$\forall N \geq 1 \quad \sum_{n=N}^{+\infty} \frac{1}{u_n} \leq C\frac{1}{u_N}.$$
4. Montrer que
$$\sum_{n=1}^{+\infty} \operatorname{Var} Q_{u_n}^* \leq \mathbb{E}\left(X_1^2 \sum_{n=1}^{+\infty} \frac{1}{u_n}\mathbb{1}_{X_1\leq u_n}\right),$$
puis
$$\sum_{n=1}^{+\infty} \operatorname{Var} Q_{u_n}^* \leq \mathbb{E}\left(X_1^2 \sum_{n:u_n\geq X_1} \frac{1}{u_n}\right).$$
5. En déduire que
$$\sum_{n=1}^{+\infty} \operatorname{Var} Q_{u_n}^* < +\infty.$$

6. Montrer que
$$Q^*_{u_n} - \mathbb{E}Q^*_{u_n} \to 0 \text{ p.s.}$$
7. Montrer que
$$\lim_{n\to+\infty} \mathbb{E}[X^*_n] = \mathbb{E}[X_1].$$
8. Montrer que $\lim \mathbb{E}[Q^*_n] = \mathbb{E}[X_1]$. (Indication : on pourra utiliser le théorème de Cesàro [2])
9. En déduire que
$$Q^*_{u_n} \to \mathbb{E}[X_1] \text{ p.s.}$$
10. Montrer que la série de terme général $\mathbb{P}(X_n \neq X^*_n)$ converge.
11. A l'aide du lemme de Borel-Cantelli, montrer que pour presque tout ω, il existe un $n_0(\omega)$ tel que les suites $(X_n(\omega))$ et $(X^*_n(\omega))$ coïncident à partir du rang $n_0(\omega)$.
12. Montrer que
$$Q_{u_n} \to \mathbb{E}[X_1] \text{ p.s.}$$
13. Si $u_n \le k \le u_{n+1}$, montrer que
$$\frac{u_n}{u_{n+1}} Q_{u_n} \le Q_k \le \frac{u_{n+1}}{u_n} Q_{u_{n+1}}.$$
14. En déduire que
$$\frac{1}{\beta}\mathbb{E}[X_1] \le \underline{\lim}_{k\to+\infty} Q_k \le \overline{\lim}_{k\to+\infty} Q_k \le \beta\mathbb{E}[X_1] \text{ p.s.}$$
15. On note
$$\Omega_\beta = \left\{\omega \in \Omega; \frac{1}{\beta}\mathbb{E}[X_1] \le \underline{\lim}_{k\to+\infty} Q_k(\omega) \le \overline{\lim}_{k\to+\infty} Q_k(\omega) \le \beta\mathbb{E}[X_1]\right\}.$$
Montrer que
$$\mathbb{P}\left(\bigcap_{n=1}^{+\infty} \Omega_{1+1/n}\right) = 1.$$
16. Montrer que
$$\lim_{k\to+\infty} Q_k = \mathbb{E}[X_1] \text{ p.s.}$$
17. Montrer que le résultat de la dernière question demeure vrai si l'on ne suppose plus que les (X_n) sont des variables positives.

2. Ernesto Cesàro, mathématicien italien (1859–1906). Ses travaux portent essentiellement sur la géométrie différentielle.

Solution

1. Les variables aléatoires $(X_i^*)_{i\geq 1}$ sont deux à deux indépendantes, car elles sont fabriquées à partir des variables aléatoires $(X_i)_{i\geq 1}$ qui sont elles mêmes deux à deux indépendantes. On a donc
$$\begin{aligned}\operatorname{Var} S_n^* = \sum_{i=1}^n \operatorname{Var} X_i^* &\leq \sum_{i=1}^n \mathbb{E}\left((X_i^*)^2\right)\\ &\leq \sum_{i=1}^n \mathbb{E}[X_i^2 \mathbb{1}_{\{X_i\leq i\}}]\\ &\leq \sum_{i=1}^n \mathbb{E}[X_i^2 \mathbb{1}_{\{X_i\leq n\}}] = n\mathbb{E}[X_1^2 \mathbb{1}_{\{X_1\leq n\}}],\end{aligned}$$
où la dernière égalité découle de l'identique distribution des X_i.
2. On a pour tout n $|u_n - \beta^n| \leq 1$. Comme β^n tend vers l'infini, on a $u_n - \beta^n = o(\beta^n)$, soit $u_n \sim \beta^n$. $\frac{1}{u_n} \sim \beta^{-n}$ et que la série à termes positifs β^{-n} est convergence, on a l'équivalence des restes :
$$\sum_{n=N}^{+\infty} \frac{1}{u_n} \sim \sum_{n=N}^{+\infty} \beta^{-n} = (1-\beta^{-1})^{-1}\beta^{-N} \sim (1-\beta^{-1})^{-1}\frac{1}{u_N}.$$
3. La suite $u_N \sum_{n=N}^{+\infty} \frac{1}{u_n}$ converge lorsque N tend vers l'infini, elle est donc bornée par une constante C.
4. On a
$$\begin{aligned}\sum_{n=1}^{+\infty} \operatorname{Var} Q_{u_n}^* = \sum_{n=1}^{+\infty} \frac{1}{u_n^2} \operatorname{Var} S_{u_n}^* &\leq \sum_{n=1}^{+\infty} \frac{1}{u_n^2} u_n \mathbb{E} X_1^2 \mathbb{1}_{\{X_1\leq u_n\}}\\ &\leq \mathbb{E}\left(X_1^2 \sum_{n=1}^{+\infty} \frac{1}{u_n} \mathbb{1}_{\{X_1\leq u_n\}}\right)\end{aligned}$$
Mais on peut remarquer que l'on a
$$\mathbb{E}\left(X_1^2 \sum_{n=1}^{+\infty} \frac{1}{u_n} \mathbb{1}_{\{X_1\leq u_n\}}\right) = \mathbb{E}\left(X_1^2 \sum_{n:u_n\geq X_1} \frac{1}{u_n}\right).$$
5. On en déduit que
$$\sum_{n=1}^{+\infty} \operatorname{Var} Q_{u_n}^* \leq \left(\mathbb{E} X_1^2 \frac{C}{u_{\inf\{n\in\mathbb{N};u_n\geq X_1\}}}\right) \leq \mathbb{E}\left[X_1^2 \frac{C}{X_1}\right] = C\mathbb{E} X_1 < +\infty.$$
6. Soit $\varepsilon > 0$ quelconque. Comme $\mathbb{P}(|Q_{u_n}^* - \mathbb{E} Q_{u_n}^*| \geq \varepsilon) \leq \frac{\operatorname{Var} Q_{u_n}^*}{\varepsilon^2}$ (par Chebychev), la question précédente permet d'affirmer que la série de

terme général $\mathbb{P}(|Q^*_{u_n} - \mathbb{E}Q^*_{u_n}| \geq \varepsilon)$ est convergente. D'après le lemme de Borel-Cantelli, on peut donc dire que

$$\mathbb{P}\left(\overline{\lim_{n\to+\infty}} \{|Q^*_{u_n} - \mathbb{E}Q^*_{u_n}| \geq \varepsilon\}\right) = 0,$$

ceci pour ε strictement positif quelconque. D'après le critère fondamental de convergence presque sûre, on peut alors affirmer que

$$Q^*_{u_n} - \mathbb{E}Q^*_{u_n} \to 0 \text{ p.s.}$$

7. $\mathbb{E}X^*_n = \mathbb{E}X_n 1\!\!1_{\{X_n \leq n\}} = \mathbb{E}X_1 1\!\!1_{\{X_1 \leq n\}}$, car X_1 et X_n ont même loi. Comme la suite $(X_1 1\!\!1_{\{X_1 \leq n\}})_{n \geq 1}$ converge en croissant vers X_1, le théorème de convergence monotone permet d'affirmer que la suite $(\mathbb{E}X^*_n)$ converge vers $\mathbb{E}X_1$.
8. $\mathbb{E}Q^*_n = \frac{1}{n}\left(\sum_{k=1}^{n} \mathbb{E}X^*_n\right)$. D'après la question précédente $\mathbb{E}X^*_n$ converge vers $\mathbb{E}X_1$, et avec le théorème de Cesàro, les moyennes convergent également vers $\mathbb{E}X_1$.
9. On a $Q^*_{u_n} - \mathbb{E}X_1 = \left(Q^*_{u_n} - \mathbb{E}Q^*_{u_n}\right) + \left(\mathbb{E}Q^*_{u_n} - \mathbb{E}X_1\right)$. $(\mathbb{E}Q^*_{u_n})$ est une sous-suite d'une suite (déterministe) qui converge vers $\mathbb{E}X_1$, donc elle converge aussi vers $\mathbb{E}X_1$. Comme $Q^*_{u_n} - \mathbb{E}Q^*_{u_n}$ tend vers 0 presque sûrement et que la convergence presque sûre est compatible avec la somme, on en déduit que $Q^*_{u_n}$ tend presque sûrement vers $\mathbb{E}X_1$.
10. $\mathbb{P}(X_n \neq X^*_n) = \mathbb{P}(X_n > n) = \mathbb{P}(X_1 > n)$. Pour tout $n \geq 0$, on a $\mathbb{P}(X_1 > n) \leq \int_{]n-1,n]} \mathbb{P}(X_1 > t)\, d\lambda(t)$. Ainsi
$$\sum_{n=1}^{+\infty} \mathbb{P}(X_n \neq X^*_n) = \sum_{n=1}^{+\infty} \mathbb{P}(X_1 > n) \leq \int_0^{+\infty} \mathbb{P}(X_1 > t)\, dt \leq \mathbb{E}X_1 < +\infty.$$
11. Ainsi, avec le lemme de Borel-Cantelli, $\mathbb{P}\left(\overline{\lim_{n\to+\infty}} \{X_n \neq X^*_n)\}\right) = 0$, ce qui équivaut à $\mathbb{P}\left(\underline{\lim}_{n\to+\infty} \{X_n = X^*_n\}\right) = 1$: pour presque tout ω, on a $\omega \in \{X_n = X^*_n)\}$ dès que $n \geq n_0(\omega)$, c'est-à-dire que les suites $(X_n(\omega))$ et $(X^*_n(\omega))$ coïncident à partir du rang $n_0(\omega)$.
12. Pour presque tout ω, on peut écrire dès que $u_n \geq n_0(\omega)$
$$Q_{u_n} - \mathbb{E}X_1 = \frac{S_{u_{n_0-1}} - S^*_{u_{n_0-1}}}{u_n} + \left(Q^*_{u_n} - \mathbb{E}X^*_1\right).$$
Bien sûr, $\frac{S_{u_{n_0-1}} - S^*_{u_{n_0-1}}}{u_n}$ tend vers 0 lorsque n tend vers l'infini. Quant au deuxième terme, il tend presque sûrement vers zéro, d'après la question 9.

Détaillons le raisonnement :
si $\omega \in \left(\underline{\lim}_{n\to+\infty} \{X_n = X_n^*)\}\right) \cap \{Q^*_{u_n} \to \mathbb{E}X_1\}$,
alors $\omega \in \{Q_{u_n} \to \mathbb{E}X_1\}$. Comme l'intersection de deux ensembles de mesure 1 est un ensemble de mesure 1, on en déduit que Q_{u_n} converge presque sûrement vers $\mathbb{E}X_1$.

13. Comme on fait des sommes de termes positifs, on a $S_{u_n} \leq S_k \leq S_{u_{n+1}}$. On en déduit que
$$\frac{u_n}{u_{n+1}} Q_{u_n} \leq \frac{u_n}{k} Q_{u_n} \leq Q_k \leq \frac{u_{n+1}}{k} Q_{u_{n+1}} \leq \frac{u_{n+1}}{u_n} Q_{u_{n+1}}.$$

14. Si n_k désigne la partie entière du logarithme de n en base β, on a
$$\frac{u_{n_k}}{u_{n_k+1}} Q_{u_{n_k}} \leq Q_k \leq \frac{u_{n_k+1}}{u_{n_k}} Q_{u_{n_k+1}}.$$
Comme $\lim_{k\to+\infty} \frac{u_{n_k}}{u_{n_k+1}} = \frac{1}{\beta} > 0$ et $\lim_{k\to+\infty} \frac{u_{n_k+1}}{u_{n_k}} = \beta > 0$, on en déduit
$$\frac{1}{\beta}\mathbb{E}X_1 \leq \underline{\lim}_{k\to+\infty} Q_k \leq \overline{\lim}_{k\to+\infty} Q_k \leq \beta\mathbb{E}X_1 \text{ p.s.}$$

15. On a montré à la question précédente que pour tout $\beta > 1$, $\mathbb{P}(\Omega_\beta) = 1$. En particulier $\forall n \geq 1 \quad \mathbb{P}(\Omega_{1+1/n}) = 1$. Comme l'intersection d'une famille dénombrable d'événements de probabilité un est de probabilité un, on a bien
$$\mathbb{P}\left(\bigcap_{n=1}^{+\infty} \Omega_{1+1/n} \right) = 1.$$

16. Soit $\omega \in \bigcap_{n=1}^{+\infty} \Omega_{1+1/n}$. On a pour tout $n \geq 1$
$$\frac{1}{1+\frac{1}{n}}\mathbb{E}X_1 \leq \underline{\lim}_{k\to+\infty} Q_k(\omega) \leq \overline{\lim}_{k\to+\infty} Q_k(\omega) \leq \left(1+\frac{1}{n}\right)\mathbb{E}X_1.$$
En faisant tendre n vers l'infini, on obtient
$$\mathbb{E}X_1 \leq \underline{\lim}_{k\to+\infty} Q_k(\omega) \leq \overline{\lim}_{k\to+\infty} Q_k(\omega) \leq \mathbb{E}X_1,$$
soit $\lim_{k\to+\infty} Q_k(\omega) = \mathbb{E}X_1$. Comme $\mathbb{P}\left(\bigcap_{n=1}^{+\infty} \Omega_{1+1/n} \right) = 1$, on a bien
$$\lim_{k\to+\infty} Q_k = \mathbb{E}X_1 \text{ p.s.}$$

17. C'est un raisonnement classique : si $(X_n)_{n\geq 1}$ est une suite de variables aléatoires indépendantes identiquement distribuées, on va alors poser

$X_n^+ = \max(0, X_n)$ et $X_n^- = \max(0, -X_n)$: on a $X_n = X_n^+ - X_n^-$. Ainsi

$$\frac{1}{n}\left(\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\left(\sum_{k=1}^{n} X_k^+\right) - \frac{1}{n}\left(\sum_{k=1}^{n} X_k^-\right).$$

Le théorème démontré à la question précédente (loi des grands nombres pour des suites de variables aléatoires positives intégrables deux à deux indépendantes identiquement distribuées) s'applique aux deux termes de la différence, de sorte que $\frac{1}{n}\left(\sum_{k=1}^{n} X_k\right)$ converge presque sûrement vers $\mathbb{E}X_1^+ - \mathbb{E}X_1^- = \mathbb{E}X_1$.

10.5 Exercices sur la convergence presque sûre

10.5.1 Exercices corrigés

Exercice 221. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes telles que pour tout $n \geq 1$, on ait $\mathbb{P}(X_n = \sqrt{n}) = \mathbb{P}(X_n = -\sqrt{n}) = \frac{1}{2}$. On pose $S_n = \sum_{i=1}^{n} X_i$.

1. Montrer que $\frac{S_n}{n^{3/2}} \xrightarrow{\mathbb{P}} 0$.
2. Montrer que $\frac{S_{n^2}}{n^3} \xrightarrow{\text{p.s.}} 0$.

   Indication : montrer que $\forall \varepsilon > 0$, on a $\sum_{n=1}^{+\infty} \mathbb{P}\left(\frac{|S_{n^2}|}{n^3} > \varepsilon\right) < \infty$ et appliquer le lemme de Borel-Cantelli.
3. En s'inspirant de la preuve du théorème 10.21, montrer que $S_n/n^{3/2}$ converge presque sûrement vers 0.

→ indication → solution

Exercice 222. Soit X une variable aléatoire de densité $f(x) = e^{-x}\mathbb{1}_{\mathbb{R}_+}(x)$. Posons $Y_n = X\ \mathbb{1}_{[0,n[}(X) + e^{2n}\ \mathbb{1}_{[n,+\infty[}(X)$.

1. Vérifier que Y_n converge vers X en probabilité et presque sûrement.
2. Calculer $\mathbb{E}(Y_n)$.
3. Est-il vrai que Y_n converge vers X dans L^1 ?

→ indication → solution

Exercice 223. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires telle que

$$\mathbb{P}(X_n = -n) = \mathbb{P}(X_n = n) = \frac{1}{2n^2} \text{ et } \mathbb{P}(X_n = 0) = 1 - \frac{1}{n^2}.$$

La suite $(X_n)_{n\geq 1}$ satisfait-elle la loi forte des grands nombres ? → indication → solution

Exercice 224. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes identiquement distribuées, $\mathbb{P}(X_1 = 1) = p$, $\mathbb{P}(X_1 = 0) = 1 - p$. Soit Y_k une variable aléatoire telle que $Y_k = 0$ si $X_k = X_{k+1}$ et $Y_k = 1$ si $X_k \neq X_{k+1}$. Posons $S_n = Y_1 + \cdots + Y_n$.

1. Calculer la moyenne et la variance de S_n.
2. Montrer que $\frac{S_n}{n}$ converge dans L^2 vers $2p(1-p)$.
3. Étudier la convergence presque sûre.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 225. On fait une infinité de lancers d'une pièce de monnaie équilibrée. Quelle est la probabilité de l'événement “on obtient une infinité de fois deux “face” consécutifs” ?

$\rightarrow$ indication $\rightarrow$ solution

Exercice 226. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes identiquement distribuées avec $\mathbb{E}|X_1| = +\infty$.

1. Soit $a > 0$. Montrer que $\overline{\lim\limits_{n\to+\infty}} \frac{|X_n|}{n} \geq a$ p.s.
2. On pose $S_n = \sum\limits_{k=1}^{n} X_k$. Montrer que $\sup\limits_{n\geq 1} \frac{|S_n|}{n} = +\infty$ p.s.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 227. *Variables M-dépendantes. Application au singe dactylographe.* Soit M un entier. On dit que des variables aléatoires $(X_n)_{n\geq 1}$ sont deux-à-deux M-dépendantes si X_i et X_j sont indépendantes dès que $|i-j| \geq M$.

1. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires de même loi intégrable, que l'on suppose deux à deux M-dépendantes.
   (a) Montrer que $\frac{X_n}{n}$ converge presque sûrement vers 0.
   Indication : on pourra utiliser le lemme de Borel-Cantelli.
   (b) Montrer que la suite $(X_n)_{n\geq 1}$ vérifie la loi forte des grands nombres.
   Indication : remarquer que $\frac{1}{Mn}\sum_{k=1}^{Mn} X_k = \frac{1}{M}\sum_{i=1}^{M} \frac{1}{i}\sum_{j=0}^{i-1} X_{Mj+i}$.
2. Soit (X_n) une suite de variables aléatoires indépendantes identiquement distribuées à valeurs dans D fini ou dénombrable.
   On note $\mathbb{P}(X_1 = i) = p_i$ pour $i \in D$.
   Montrer que pour tout $\ell \geq 1$ et pour tout $(a_1, \ldots, a_\ell) \in D^\ell$, si on pose
$$S_n = \sum_{k=0}^{n-1} \mathbb{1}_{\{X_{k+1}=a_1,\ldots,X_{k+\ell}=a_\ell\}}, \text{ alors on a } \frac{S_n}{n} \xrightarrow{\text{p.s.}} \prod_{i=1}^{\ell} p_{a_i}.$$

3. Commenter la citation suivante : “Concevons qu’on ait dressé un million de singes à frapper au hasard sur les touches d’une machine à écrire et que, sous la surveillance de contremaîtres illettrés, ces singes dactylographes travaillent avec ardeur dix heures par jour avec un million de machines à écrire de types variés. Les contremaîtres illettrés rassembleraient les feuilles noircies et les relieraient en volumes. Et au bout d’un an, ces volumes se trouveraient renfermer la copie exacte des livres de toute nature et de toutes langues conservés dans les plus riches bibliothèques du monde. Telle est la probabilité pour qu’il se produise pendant un instant très court, dans un espace de quelque étendue, un écart notable de ce que la mécanique statistique considère comme le phénomène le plus probable.” Émile Borel, *Mécanique Statistique et Irréversibilité*, J. Phys. 5e série, vol. 3, 1913, pp.189-196.

$\to$ indication $\to$ solution

Exercice 228. *Représentation g-adique.*
On sait démontrer que pour un entier $g \geq 2$ fixé, chaque réel $\omega \in [0,1[$ admet une unique représentation sous la forme

$$\omega = \sum_{i=1}^{\infty} a_i g^{-i} \quad \text{où} \quad a_i \in \{0,1,\dots,g-1\}$$

telle que la suite a_i ne soit pas constamment égale à $g-1$ à partir d’un certain rang. Cette écriture est appelée la représentation g-adique. Lorsque $g=2$, on parle de représentation dyadique.

Après avoir établi l’existence et l’unicité de cette représentation, on se propose d’étudier la répartition des a_i et on montrera que pour λ-presque tout $\omega \in [0,1[$, les a_i sont uniformément répartis sur $\{0,1,\dots,g-1\}$. Plus précisément, on dit que ω est g-normal si pour tout $\ell \geq 1$ et quels que soient $b_1,\dots,b_\ell$ compris entre 0 et $g-1$, on a

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{a_i=b_1, a_{i+1}=b_2,\dots,a_{i+\ell-1}=b_\ell\}} = \frac{1}{g^\ell} \text{ p.s.}$$

(c’est-à-dire que la fréquence asymptotique de chaque séquence de chiffres de longueur ℓ est $1/g^\ell$).

On montrera que presque tous les nombres dans $[0,1[$ sont g-normaux.

On considère l’espace probabilisé $(\Omega,\mathcal{F},\mathbb{P}) = ([0,1[,\mathcal{B}([0,1[),\lambda_{|[0,1[})$. Soit $g \geq 2$ un entier. On pose $X_0^g(\omega)=\omega$.

1. On définit les variables A_i^g et X_i^g par les récurrences $X_i^g = \{gX_{i-1}^g\}$ et $A_i^g = \lfloor gX_i \rfloor$. Montrer que pour tout $\omega \in [0,1[$, on a

$$\omega = X_0(\omega) = \sum_{i=0}^{+\infty} \frac{A_i^g(\omega)}{g^{i+1}} \text{ avec } A_i^g \in \{0,1,\dots,g-1\}$$

et que la suite $A_i^g(\omega)$ contient une infinité de termes différents de $g-1$: c'est le développement g-adique de ω. Vérifier l'unicité d'une telle décomposition.

2. On note G l'application de $[0,1[$ dans $\{0,\dots,g-1\}\times[0,1[$ qui à x associe $G(x)=(\lfloor gx\rfloor,\{gx\})$. Montrer que la loi image de $\mathbb{P}$ par G est $\mathcal{U}(\{0,\dots,g-1\})\otimes\mathbb{P}$.

   Indication : on commencera par calculer $\mathbb{P}_G(\{k\}\times[0,x])$.
3. Montrer que les X_i^g suivent la loi uniforme sur $[0,1]$ et que les A_i^g suivent la loi uniforme sur $\{0,\dots,g-1\}$.
4. Montrer que les $(A_i^g)_{i\geq 0}$ sont indépendants.
5. En utilisant les résultats de l'exercice 227, montrer que $\mathbb{P}$-presque tout nombre dans $[0,1[$ est g-normal.
6. Un nombre $\omega\in[0,1[$ s'appelle *absolument normal* s'il est g-normal pour tout $g=2,3,\dots$. Montrer finalement que $\mathbb{P}$-presque tout nombre dans $[0,1[$ est absolument normal.

Nous ne connaissons pourtant que très peu d'exemples concrets de nombres g-normaux et nous ignorons tout sur la normalité des nombres $\pi,\sqrt{2},\log 2$ ou encore e. À titre d'exemple, notons que le nombre

$$0,12345678910111213\dots$$

est 10-normal. En revanche, et bien que, comme nous l'avons vu, presque tout réel $\omega\in[0,1[$ ait cette propriété, on ne sait pas exhiber de nombre qui soit g-normal quel que soit $g\geq 2$.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 229. *Existence de variables aléatoires indépendantes suivant une loi quelconque.*
Grâce à l'exercice précédent, nous remarquons qu'on sait construire une suite de variables aléatoires indépendantes identiquement distribuées sur $(\Omega,\mathcal{F},\mathbb{P})$ suivant la loi $\mathrm{Ber}(1/2)$. Soit $(X_n)_{n\geq 1}$ une telle suite. On pose $S_n=\sum\limits_{k=1}^{n}\frac{X_k}{2^k}$ et $S=\sum\limits_{k=1}^{+\infty}\frac{X_k}{2^k}$.

1. Soient $n\geq 1$ et $k\in\{0,\dots,2^n-1\}$. Montrer que
$$\left\{S<\frac{k}{2^n}\right\}\Delta\left\{S_n<\frac{k}{2^n}\right\}\subset\bigcap_{k=n+1}^{+\infty}\{X_k=1\}.$$
   En déduire que $\mathbb{P}\left(S<\frac{k}{2^n}\right)=\frac{k}{2^n}$.
2. Déterminer la loi de S.

3. Que dire de la loi de la suite de variables aléatoires $(U_i)_{i\geq 1}$ définies par $U_i = \sum_{k=1}^{+\infty} \frac{X_{2^k(2i+1)}}{2^k}$?
4. Soit (μ_i) une suite de mesures de probabilités sur $\mathbb{R}$. En utilisant le théorème 5.36 et l'exercice précédent, montrer qu'on peut construire sur l'espace probabilisé $([0,1[,\mathcal{B}([0,1[),\lambda_{|[0,1[})$ une famille $(Z_i)_{i\geq 1}$ de variables indépendantes telles que pour tout i, Z_i suit la loi μ_i.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 230. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant une loi exponentielle de paramètre 1. Calculer $\overline{\lim_{n\to+\infty}} \frac{X_n}{\log n}$. $\rightarrow$ indication $\rightarrow$ solution

Exercice 231. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes telle que pour tout n, X_n suit une loi binomiale $\mathcal{B}(n, \frac{1}{n^{1,01789}})$. Montrer que l'ensemble des valeurs d'adhérence de $(X_n)_{n\geq 1}$ est presque sûrement égal à un ensemble déterministe que l'on précisera. $\rightarrow$ indication $\rightarrow$ solution

Exercice 232. *Lemme de Kochen–Stone [24].*

1. *Inégalité de Paley-Zygmund.*
   Soit X une variable aléatoire de carré intégrable et d'espérance strictement positive. Montrer que pour tout $\lambda \in]0,1[$,
   $$\mathbb{P}(X > \lambda\mathbb{E}[X]) \geq (1-\lambda)^2 \frac{(\mathbb{E}X)^2}{\mathbb{E}[X^2]}.$$
   Indication : majorer et minorer $\mathbb{E}[X\mathbb{1}_{\{X\leq\lambda\}}]$.
2. Soit $(B_n)_n$ une famille d'événements. Montrer que
   $$\mathbb{P}\left(\overline{\lim_{n\to+\infty}} B_n\right) \geq \overline{\lim_{n\to+\infty}} \mathbb{P}(B_n).$$
3. Soit $(B_n)_{n\geq 1}$ une famille d'événements telle que $\sum_{n=1}^{+\infty} \mathbb{P}(B_n) = +\infty$.
   On pose $N_n = \sum_{k=1}^{n} \mathbb{1}_{B_k}$ et on rappelle que
   $\{\overline{\lim_{n\to+\infty}} B_n\} = \{\lim_{n\to+\infty} N_n = +\infty\}$. Montrer que
   $$\mathbb{P}\left(\overline{\lim_{n\to+\infty}} B_n\right) \geq \overline{\lim_{n\to+\infty}} \frac{(\mathbb{E}N_n)^2}{\mathbb{E}[N_n^2]}.$$
   Ce résultat est le lemme de Kochen–Stone.

4. Que se passe-t-il lorsque les événements B_n sont deux-à-deux indépendants ?

$\to$ indication $\to$ solution

Exercice 233. *Une application du lemme de Kochen–Stone.*
Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes identiquement distribuées, avec $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$. On construit la marche aléatoire simple $(S_n)_{n\geq 1}$ en posant $S_n = X_1 + \cdots + X_n$. On rappelle les résultats suivants qui ont été vu précédemment.

— On a l'équivalent en l'infini $\mathbb{P}(S_{2n} = 0) \sim \frac{1}{\sqrt{\pi n}}$ (voir l'exercice 207).

— Pour tout $A \in \mathcal{Q} = \bigcap\limits_{n\geq 1} \sigma((S_i)_{i\geq n})$, on a $\mathbb{P}(A) = 0$ ou $\mathbb{P}(A) = 1$ (c'est le corollaire 5.42 obtenu comme conséquence de la loi du 0–1 de Hewitt et Savage).

À la lumière de ces résultats et du lemme de Kochen–Stone vu à l'exercice précédent, on veut montrer que

$$\mathbb{P}(S_{2n^2} = 0 \text{ pour une infinité de valeurs de } n) = 1.$$

1. On pose $B_n = \{S_{2n^2} = 0\}$ et $N_n = \sum\limits_{k=1}^{n} \mathbb{1}_{B_k}$. Montrer que

$$\mathbb{E}[N_n^2] = \mathbb{E}[N_n] + 2\sum_{k=2}^{n}\sum_{p=1}^{k-1} \mathbb{P}(S_{2p^2} = 0)\mathbb{P}(S_{2(k^2-p^2)} = 0),$$

   puis qu'il existe une constante C telle que

$$\forall n \geq 1, \quad \mathbb{E}[N_n(N_n-1)] \leq C\sum_{k=2}^{n}\sum_{p=1}^{k-1} c_{p,k} \text{ avec } c_{p,k} = \frac{1}{p\sqrt{k^2-p^2}}.$$

2. On pose $s_k = \sum\limits_{p=1}^{k-1} c_{p,k}$. Montrer qu'il existe une constante D telle que

$$\forall k \geq 2, \quad s_k \leq D\frac{\log k}{k}.$$

   Indication : on pourra remarquer que pour $p \leq k/2$, on a $c_{p,k} \geq c_{k-p,k}$.

3. Montrer que

$$\overline{\lim_{n\to+\infty}} \frac{\mathbb{E}[N_n^2]}{(\mathbb{E}[N_n])^2} < +\infty.$$

4. Conclure.

$\to$ indication $\to$ solution

Exercice 234. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires de même loi non dégénérée (c'est-à-dire que leur loi n'est pas une mesure de Dirac). Soit

$(a_n)_{n\geq 1}$ une suite réelle. On suppose que $(a_nX_n)_{n\geq 1}$ tend en probabilité vers zéro. Montrer que $(a_n)_{n\geq 1}$ tend vers zéro. Montrer que la réciproque est aussi vraie. $\to$ indication $\to$ solution

Exercice 235. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant les lois de Poisson de paramètres respectifs λ_n. On pose $Y_n = \prod\limits_{i=1}^{n} X_i$.

1. Montrer que $\mathbb{P}(X_n \neq 1) \geq 1 - 1/e$.
   En déduire que $\mathbb{P}(X_n \neq 1$ pour une infinité de $n) = 1$.
2. On pose $p = \prod_{n=1}^{+\infty}(1 - e^{-\lambda_n})$. Montrer que

$$Y_n \xrightarrow{\text{p.s.}} \begin{cases} 0 & \text{avec probabilité } 1-p \\ +\infty & \text{avec probabilité } p \end{cases}$$

3. Calculer cette probabilité quand
   (a) $\lambda_n = o(\log(n))$,
   (b) $\lambda_n = 2\log(n+1)$.

$\to$ indication $\to$ solution

Exercice 236. *Limites supérieures de variables de Poisson*.
Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires telle que pour tout n, X_n suit la loi de Poisson de paramètre λ_n.

1. On suppose que $\sum\limits_{n=1}^{+\infty} \lambda_n^2 < +\infty$. Montrer que

$$\varlimsup_{n\to+\infty} X_n \leq 1 \text{ p.s.}$$

   Indication : on pourra remarquer que $\mathbb{1}_{\{X_n\geq 2\}} \leq \frac{1}{2}X_n(X_n-1)$.
2. Plus généralement, que dire si $\sum\limits_{n=1}^{+\infty} \lambda_n^k < +\infty$, avec k un entier positif ?
3. On suppose désormais que les variables $(X_n)_{n\geq 1}$ sont indépendantes et que $\lambda_n = \frac{1}{n^\alpha}$, où α est un réel strictement positif. Caractériser à l'aide de α l'ensemble aléatoire formé des valeurs d'adhérence de $(X_n)_{n\geq 1}$.

$\to$ indication $\to$ solution

Exercice 237. 1. On désigne par $\mathbb{F}_q$ un corps fini de cardinal q. Soit E un $\mathbb{F}_q$-espace vectoriel de dimension n. Montrer que pour $r \leq n$, E possède

$$\prod_{i=0}^{r-1}(q^n - q^i)$$

familles libres de cardinal r. On note alors $\text{Gr}_r(E)$ l'ensemble des sous-espaces vectoriels de E de dimension r. Montrer que

$$|\text{Gr}_r(E)| = \frac{Q(n)}{Q(r)Q(n-r)},$$

où $Q(n) = \prod_{i=1}^{n} (q^i - 1)$.

2. Soit $(X_{i,j})_{i,j\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi uniforme sur $\mathbb{F}_q$. On note M_n la matrice des $(X_{i,j})_{1\leq i,j\leq n}$ et R_n son rang.
   (a) Montrer que R_n/n tend en probabilité vers 1, puis que $\mathbb{E}(R_n) \sim n$.
   (b) Montrer qu'il existe c tel que presque sûrement, pour n assez grand $R_n \geq n - c\sqrt{\log n}$.

$\to$ indication $\to$ solution

10.5.2 Exercices non corrigés

Exercice 238. Soient $s \in]0,1[$ et (x_n) une suite de limite ℓ. On pose, pour $n \geq 1$: $m_n = \sum_{j=0}^{n} \binom{n}{j} s^j (1-s)^{n-j} x_j$. Montrer que $\lim\limits_{n\to+\infty} m_n = \ell$. $\to$ indication

Exercice 239. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi $\mathcal{N}(m, \sigma^2)$. Montrer que la suite $\frac{X_1^2+\cdots+X_n^2}{n}$ converge presque sûrement et déterminer sa limite. $\to$ indication

Exercice 240. On considère une suite infinie de lancers de “pile ou face” avec une pièce équilibrée. On note N_n le nombre de “pile” observés parmi les n premiers lancers. Montrer que presque sûrement, il existe n tel que $N_n \geq n/3$.
$\to$ indication

Exercice 241. On dit que la suite $(X_n)_{n\geq 1}$ converge presque complètement vers X si quel que soit $\varepsilon > 0$, la série de terme général $\mathbb{P}(|X_n - X| > \varepsilon)$ converge. Montrer que les deux énoncés suivant sont équivalents :

— $(X_n)_{n\geq 1}$ converge presque complètement vers 0.

— Si une suite (X'_n) de variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$ est telle que pour tout n, X_n et X'_n ont même loi, alors $(X'_n)_{n\geq 1}$ converge presque sûrement vers 0.

$\to$ indication

Exercice 242. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires identiquement distribuées telle qu'il existe $\alpha > 0$ avec $\mathbb{E}\exp(\alpha|X_1|) < +\infty$. Montrer que

$$\lim_{n\to+\infty} \frac{X_n}{(\log n)^{\frac{3}{2}}} = 0.$$

$\to$ indication

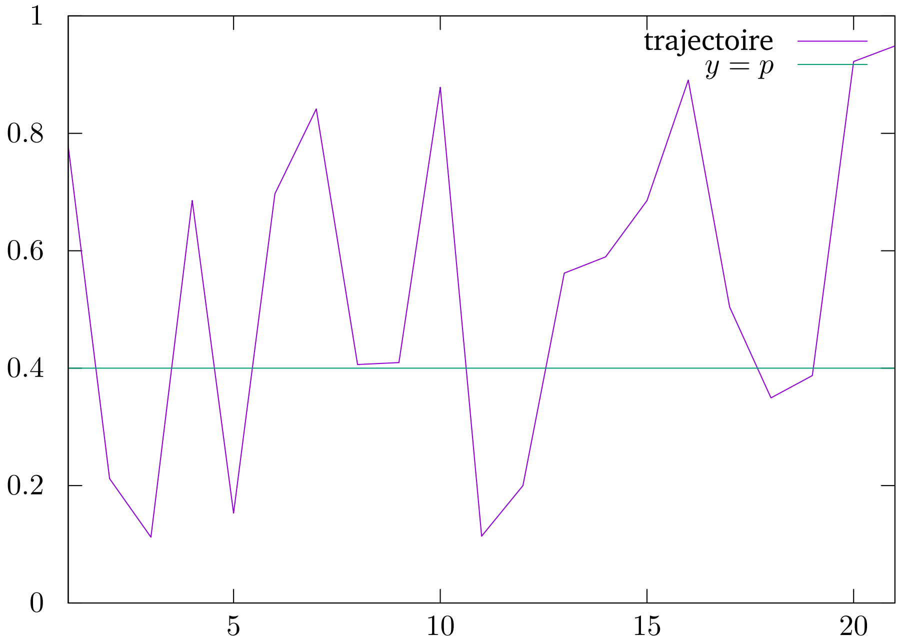
Exercice 243. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi $\mathcal{N}(0,1)$.

1. Calculer $\overline{\lim}_{n\to+\infty} \frac{X_n}{\sqrt{2\log n}}$.
2. On se donne maintenant une deuxième suite $(Y_n)_{n\geq 1}$ de variables aléatoires indépendantes suivant la loi $\mathcal{N}(0,1)$. Cette deuxième suite est supposée indépendante de la première. Comparer

$$\overline{\lim_{n\to+\infty}} \frac{X_n}{\sqrt{2\log n}} + \overline{\lim_{n\to+\infty}} \frac{Y_n}{\sqrt{2\log n}} \quad \text{et} \quad \overline{\lim_{n\to+\infty}} \frac{(X_n+Y_n)}{\sqrt{2\log n}}.$$

$\to$ indication

Exercice 244. Soient $p \in [0,1]$ et $(U_n)_{n\geq 0}$ une suite de variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On note T_n le nombre de fois où la ligne polygonale reliant successivement $(0,U_0),(1,U_1),(2,U_2),\dots$ coupe la droite d'équation $y=p$ avant le temps n. Dans notre exemple, $p=0.4$ et $T_{20}=8$.



Montrer que $\frac{T_n}{n}$ converge presque sûrement et déterminer sa limite. $\to$ indication

Exercice 245. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires de Bernoulli indépendantes de paramètre $1/2$. On pose

$$M_n = \begin{pmatrix} 2+X_n & 1 \\ 1 & 2+X_n \end{pmatrix}$$

et

$$A_n = M_n \times M_{n-1} \times \cdots \times M_2 \times M_1.$$

1. Montrer que la suite $(\det A_n)^{1/n}$ converge presque sûrement et déterminer sa limite.
2. Soit $(x,y) \in \mathbb{R}^2 \setminus \{(0,0)\}$. On pose

   $$Y_n = A_n \begin{pmatrix} x \\ y \end{pmatrix}.$$

   Montrer que

   $$\|Y_n\|^{1/n} \xrightarrow{\text{p.s.}} \begin{cases} \sqrt{2} & \text{si } x+y=0 \\ \sqrt{12} & \text{si } x+y \neq 0. \end{cases}$$

→ indication

Exercice 246. *Ensemble triadique de Cantor, le retour.*
Posons $K_0 = [0,1]$, puis $K_{n+1} = f_1(K_n) \cup f_2(K_n)$ où $f_1(x) = \frac{x}{3}$ et $f_2(x) = \frac{x+2}{3}$. On définit la mesure μ_n sur $[0,1]$ par $\mu_n(A) = \lambda(A|K_n) = \frac{\lambda(A\cap K_n)}{\lambda(K_n)}$ et F_n la fonction de répartition associée à μ_n. On pose $K = \bigcap_{n=1}^{+\infty} K_n$.

1. Montrer que K_n est décroissante et que

   $$K_{n+1} = ([0,1/3] \cap f_1(K_n)) \cup ([2/3,1] \cap f_2(K_n)).$$

2. Montrer que la suite (F_n) vérifie la récurrence $F_{n+1} = TF_n$ avec

   $$Tf(x) = \begin{cases} \frac{1}{2} f(3x) & \text{si } x \in [0,1/3] \\ \frac{1}{2} & \text{si } x \in [1/3,2/3] \\ \frac{1}{2}(1+f(3x-2)) & \text{si } x \in [2/3,1] \end{cases}.$$

   En déduire que F_n converge vers une fonction continue F qui est la fonction de répartition d'une loi à support dans $[0,1]$. On note $\mathbb{K}$ cette loi.
3. On reprend les notations de l'exercice 228, avec $g=3$. Montrer que la loi image de $\mathbb{K}$ par G est $\mathcal{U}(\{0,2\}) \otimes \mathbb{K}$. En déduire que sous $\mathbb{K}$, les $(A_i^g)_{i\geq 0}$ sont indépendantes suivant la loi uniforme sur $\{0,2\}$.
4. Montrer que K est de mesure 1 sous $\mathbb{K}$, mais de mesure nulle sous $\mathbb{P} = \lambda|_{[0,1[}$.

5. On va aller un peu plus loin dans l'étude de la régularité de F.
   Pour $\alpha \in]0,1]$, on note E_α l'ensemble des fonctions de $[0,1]$ dans $\mathbb{R}$ qui vérifient $f(0)=0$, $f(1)=1$ et
   $$\|f\|_\alpha = \sup_{0\le s<t\le 1} \frac{|f(s)-f(t)|}{|s-t|^\alpha} < +\infty.$$
   Montrer que $(E_\alpha, \|.\|_\alpha)$ est un espace métrique complet. En déduire que $F \in E_\alpha$ pour $\alpha < \frac{\log 2}{\log 3}$. Inversement, montrer que
   $$\overline{\lim_{y\to\omega}} \frac{|F(\omega)-F(y)|}{|\omega - y|^{\frac{\log 2}{\log 3}}} > 0 \quad \mathbb{K}\text{-p.s.}$$
   En particulier, on notera que pour $\mathbb{K}$ presque tout ω, F n'est pas dérivable en ω, tandis que, comme on l'a vu au chapitre 5, F est dérivable en λ presque tout ω.

Le $\frac{\log 2}{\log 3}$ qui apparaît ici comme valeur critique de la régularité Hölder est la dimension fractale de l'ensemble de Cantor. Il y a en réalité plusieurs notions de dimension fractale, qui coïncident dans cet exemple particulier. Pour en savoir plus, on pourra se référer au cours très abordable de Queffelec [30], ou encore aux ouvrages spécialisés de Falconer [13, 14] ou Mattila [27].

$\rightarrow$ indication

Chapitre 11

Convergence en loi

11.1 Convergence en loi

11.1.1 Définition

On dit qu'une suite $(\mu_n)_{n\geq 1}$ de mesures de probabilité sur $\mathbb{R}^d$ converge faiblement vers la mesure de probabilité μ lorsque pour toute fonction f continue bornée de $\mathbb{R}^d$ dans $\mathbb{R}$, on a [1]

$$\lim_{n\to+\infty} \int_{\mathbb{R}^d} f\ d\mu_n = \int_{\mathbb{R}^d} f\ d\mu. \tag{11.1}$$

Par extension, on dit qu'une suite de variables aléatoires $(X_n)_n$ converge en loi vers la variable aléatoire X (ou vers la loi μ) si la suite de mesures $(\mathbb{P}_{X_n})$ converge faiblement vers $\mathbb{P}_X$ (ou vers la loi μ).

Ainsi, dire que X_n converge en loi vers X signifie que pour toute fonction continue bornée, $\mathbb{E}f(X_n)$ converge vers $\mathbb{E}f(X)$.
On note cette propriété $X_n \Longrightarrow X$.

Rappelons que si μ et ν sont deux mesures qui chargent finiment les compacts de $\mathbb{R}^d$ et telles que pour toute fonction f continue bornée de $\mathbb{R}^d$ dans $\mathbb{R}$, on a $\int_{\mathbb{R}^d} f\ d\mu = \int_{\mathbb{R}^d} f\ d\nu$, alors $\mu = \nu$.

On en déduit immédiatement l'unicité de la limite pour la convergence en loi.

1. Il est possible de définir plusieurs notions de convergence pour des mesures finies $(\mu_n)_{n\geq 1}$ et μ. On parle alors de convergence étroite lorsque (11.1) est vérifiée pour les fonctions continues bornées, et de convergence vague lorsque (11.1) est vérifiée pour les fonctions continues à support compact. Ces deux notions coïncident lorsque les $(\mu_n)_{n\geq 1}$ et μ sont des mesures de probabilité. Ne travaillant qu'avec des mesures de probabilité, nous ne parlerons ici que de convergence faible.

Théorème 11.1. *Soit g une fonction continue définie sur $\mathbb{R}^d$. Si la suite $(X_n)_{n\geq 1}$ converge en loi vers X, alors la suite $(Y_n)_{n\geq 1}$ définie par $Y_n = g(X_n)$ converge en loi vers $g(X)$.*

Démonstration. Soit f une fonction continue bornée. On a par définition de Y_n : $\mathbb{E}f(Y_n) = \mathbb{E}f(g(X_n)) = \mathbb{E}(f \circ g)(X_n)$. Comme f et g sont continues, $f \circ g$ est continue. Comme f est bornée, $f \circ g$ est aussi bornée. Ainsi, $f \circ g$ est continue, bornée et $(X_n)_{n\geq 1}$ converge en loi vers X, donc $\mathbb{E}(f \circ g)(X_n)$ converge vers $\mathbb{E}(f \circ g)(X) = \mathbb{E}f(g(X))$, ce qui achève la preuve. □

Corollaire 11.2. *Soient $(X_n)_{n\geq 1}$ et $(Y_n)_{n\geq 1}$ deux suites de vecteurs aléatoires à valeurs dans $\mathbb{R}^d$. Si $(X_n, Y_n)_{n\geq 1}$ converge en loi vers (X, Y), alors*

1. *$X_n + Y_n$ converge en loi vers $X + Y$.*
2. *$\langle X_n, Y_n\rangle$ converge en loi vers $\langle X, Y\rangle$.*

Démonstration. Il suffit d'appliquer le théorème précédent à la fonction continue $(x, y) \mapsto x + y$ et à la fonction continue $(x, y) \mapsto \langle x, y\rangle$. □

11.1.2 Premiers exemples

Un critère de convergence en loi

Théorème 11.3 (Lemme de Scheffé)**.**
Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré; $f, (f_n)_{n\geq 1} : \Omega \to \mathbb{R}_+$ des applications positives intégrables par rapport à μ telles que

a) $f_n \to f$ *μ–p.p.*

b) $\lim\limits_{n\to\infty} \int_\Omega f_n \, d\mu(x) = \int_\Omega f \, d\mu(x)$.

Alors $f_n \xrightarrow{L^1} f$.

Démonstration. On a

$$\begin{array}{rcl} f + f_n & = & \max(f, f_n) + \min(f, f_n) \\ \text{et} \quad |f - f_n| & = & \max(f, f_n) - \min(f, f_n). \end{array}$$

On en déduit immédiatement que

$$f + f_n - |f - f_n| = 2\min(f, f_n),$$

d'où

$$|f - f_n| = f + f_n - 2\min(f, f_n) = -f + f_n + 2(f - \min(f, f_n)).$$

Ainsi, on obtient

$$\|f_n - f\|_{L^1(\mu)} = -\int_\Omega f(x) d\mu(x) + \int_\Omega f_n(x) d\mu(x) + 2\int_\Omega (f - \min(f, f_n))(x) d\mu(x).$$

D'après la deuxième hypothèse, $-\int_\Omega f(x)d\mu(x) + \int_\Omega f_n(x)d\mu(x)$ tend vers 0 lorsque n tend vers $+\infty$. D'autre part, on a $0 \le f - \min(f, f_n) \le f$ car f est à valeurs positives ou nulles, et $f - \min(f, f_n) \to 0 \quad \mu-$p.p. D'après le théorème de convergence dominée, on a donc $\int_\Omega (f - \min(f, f_n)) \ d\mu \to 0$ p.p., d'où le résultat. □

Corollaire 11.4. *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Soient ν et $(\nu_n)_{n\ge1}$ des mesures de probabilité sur $(\Omega, \mathcal{A}, \mu)$ admettant les densités f et $(f_n)_{n\ge1}$ par rapport à μ. On suppose que $f_n \to f$ $\mu-$p.p. Alors (ν_n) converge faiblement vers ν.*

Démonstration. Soit g une fonction continue bornée sur Ω. On a

$$\begin{aligned}
\left|\int g(x)\ d\nu_n(x) - \int g(x)\ d\nu(x)\right| &= \left|\int g(x)f(x)\ d\mu(x) - \int g(x)f_n(x)\ d\mu(x)\right| \\
&\le \int |g(x)(f(x) - f_n(x))|\ d\mu(x) \\
&\le \|g\|_\infty \int |f(x) - f_n(x)|\ d\mu(x),
\end{aligned}$$

qui tend vers 0 d'après le théorème de Scheffé. Comme cette convergence a lieu pour toute fonction continue bornée g, on trouve que ν_n converge en loi vers ν. □

Corollaire 11.5. *Soient X et $(X_n)_{n\ge1}$ des variables aléatoires discrètes à valeurs dans un ensemble dénombrable D. On suppose que*

$$\forall k \in D, \qquad \lim_{n\to+\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k).$$

Alors $(X_n)_{n\ge1}$ converge en loi vers X.

Démonstration. Il suffit de remarquer que $\mathbb{P}_X$ et $\mathbb{P}_{X_n}$ pour $n \ge 1$ ont une densité par rapport à la mesure de comptage sur D et appliquer le corollaire précédent. □

Convergence de la loi binomiale vers la loi de Poisson

Théorème 11.6. *Soit, pour $n \ge 1$, une variable aléatoire X_n suivant la loi binomiale de paramètres n et p_n. On suppose que*

$$\lim_{n\to\infty} np_n = \lambda > 0.$$

Alors X_n converge en loi vers la loi de Poisson de paramètre λ.

Démonstration. D'après le corollaire 11.5, il suffit de montrer que pour tout entier $k \geq 0$, $\mathbb{P}(X_n = k)$ converge vers $e^{-\lambda}\frac{\lambda^k}{k!}$. On a

$$\mathbb{P}(X_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} = \binom{n}{k} p_n^k (1-p_n)^n (1-p_n)^{-k}.$$

On a les équivalents quand n tend vers l'infini $\binom{n}{k} = \frac{n(n-1)\ldots(n-k+1)}{k!} \sim \frac{n^k}{k!}$ et $p_n^k \sim (\lambda/n)^k = \lambda^k n^{-k}$. D'autre part, on a l'équivalent à l'infini

$$\log(1-p_n)^n = n\log(1-p_n) = n(-p_n + o(p_n)) \sim -np_n \sim -\lambda.$$

Ainsi, $\log(1-p_n)^n$ converge vers $-\lambda$ donc $(1-p_n)^n$ converge vers $e^{-\lambda}$. En mettant ensemble les équivalents, on obtient le résultat souhaité. □

Application pratique. Si n est "grand" et np "pas trop grand", on peut remplacer la loi binomiale par une loi de Poisson. D'après une grand-mère statisticienne, n est grand à partir de 30 et np n'est pas trop grand jusqu'à 10. Ce théorème peut être interprété de la manière suivante : la loi de Poisson est une bonne modélisation pour le nombre de fois où un événement rare survient (par exemple, un tremblement de terre).

Remarque 11.7. *La loi de Poisson peut intervenir lorsque l'on compte des événements rares, même lorsque l'on compte des événements dépendants, pourvu que cette dépendance soit limitée. Par exemple soit σ_n une permutation aléatoire, suivant la loi uniforme sur $\mathfrak{S}_n$. Si l'on note D_n le nombre de points fixes de σ_n, en relisant les résultats de l'exercice 42 sur le problème des dérangements, on voit que $\mathbb{P}(D_n = k) = \binom{n}{k}\frac{d_{n-k}}{n!} = \frac{1}{k!}\frac{d_{n-k}}{(n-k)!}$. Comme $d_n \sim \frac{1}{e}n!$, on conclut que D_n converge en loi vers la loi de Poisson de paramètre 1.*

Convergence de la loi hypergéométrique vers la loi binomiale

Théorème 11.8. *Soit, pour $j \geq 1$, une variable aléatoire X_j suivant la loi hypergéométrique de paramètres (N_j, n_j, k). On suppose que N_j tend vers l'infini et que l'on a*

$$\lim_{j\to\infty} \frac{n_j}{N_j} = p.$$

Alors X_j converge en loi vers une variable de loi binomiale $\mathcal{B}(k,p)$.

Démonstration. D'après le corollaire 11.5, il suffit de montrer que pour tout entier $i \geq 0$, $\mathbb{P}(X_j = i)$ converge vers $\binom{k}{i}p^i(1-p)^{k-i}$. On a, à i fixé, l'équivalent à l'infini (en j)

$$\mathbb{P}(X_j = i) = \frac{\binom{n_j}{i}\binom{N_j - i}{k-i}}{\binom{N_j}{k}} \sim \frac{\frac{n_j^i}{i!}\frac{(N_j-n_j)^{k-i}}{(k-i)!}}{\frac{N_j^k}{k!}}.$$

Or

$$\frac{\frac{n_j^i}{i!}\frac{(N_j-n_j)^{k-i}}{(k-i)!}}{\frac{N_j^k}{k!}} = \frac{k!}{i!(k-i)!}\left(\frac{n_j}{N_j}\right)^i\left(1-\frac{n_j}{N_j}\right)^{k-i}.$$

Comme cette dernière quantité converge vers $\binom{k}{i}p^i(1-p)^{k-i}$ lorsque j tend vers l'infini, cela achève la preuve. □

11.1.3 Théorème de Portmanteau

On va maintenant énoncer le théorème de Portmanteau [2] dans $\mathbb{R}^d$. Exceptée la propriété (6), qui n'a pas de sens hors de $\mathbb{R}^d$, le reste de l'équivalence demeure vrai dans un cadre beaucoup plus général que celui dans lequel nous nous sommes placés ici.

Théorème 11.9. *Les propositions suivantes sont équivalentes :*

1. *μ_n converge faiblement vers μ.*
2. *Pour toute fonction f uniformément continue bornée de $\mathbb{R}^d$ dans $\mathbb{R}$, on a*
$$\lim_{n\to+\infty}\int_{\mathbb{R}^d} f\, d\mu_n = \int_{\mathbb{R}^d} f\, d\mu.$$
3. *Pour tout fermé F, $\mu(F) \geq \overline{\lim_{n\to+\infty}}\, \mu_n(F)$.*
4. *Pour tout ouvert O, $\mu(O) \leq \underline{\lim_{n\to+\infty}}\, \mu_n(O)$.*
5. *Pour tout borélien A dont la frontière ∂A vérifie $\mu(\partial A) = 0$,*
on a $\lim_{n\to+\infty} \mu_n(A) = \mu(A)$.
6. *Pour tout pavé $A = \prod_{i=1}^d\,]a_i, b_i]$ dont la frontière ∂A vérifie $\mu(\partial A) = 0$,*
on a $\lim_{n\to+\infty} \mu_n(A) = \mu(A)$.

Démonstration. On va prouver successivement $(3) \iff (4)$ puis les implication $(1) \Longrightarrow (2) \Longrightarrow (3) \Longrightarrow (4) \Longrightarrow (5) \Longrightarrow (6) \Longrightarrow (1)$.

2. Les exégètes divergent quant à l'origine du nom. Faut-il retenir que l'on peut y accrocher n'importe quoi, ou, dans le sens anglais du mot, qu'il permet d'emporter ce que l'on veut ? En tout cas, une chose est sûre, c'est à Patrick Billingsley que revient l'honneur d'avoir, dans la 2e édition de *Convergence of Probability Measures*, rendu justice au travail de Portmanteau [29].

— Pour voir que $(3) \iff (4)$, il suffit de remarquer que

$$\begin{aligned}
& \sup_{O \textbf{ ouvert}} \left(\mu(O) - \varliminf_{n\to+\infty} \mu_n(O)\right) \\
= & \sup_{F \textbf{ fermé}} \left(\mu(F^c) - \varliminf_{n\to+\infty} \mu_n(F^c)\right) \\
= & \sup_{F \textbf{ fermé}} \left(1-\mu(F) - \varliminf_{n\to+\infty} (1-\mu_n(F))\right) \\
= & \sup_{F \textbf{ fermé}} \left(-\mu(F) - \varliminf_{n\to+\infty} -\mu_n(F)\right) \\
= & \sup_{F \textbf{ fermé}} \left(-\mu(F) + \varlimsup_{n\to+\infty} \mu_n(F)\right).
\end{aligned}$$

Ainsi, si l'un des supremum est négatif, alors l'autre l'est aussi.

— Preuve de $(2) \Longrightarrow (3)$. Soit F un fermé de $\mathbb{R}^d$. Pour tout $x \in \mathbb{R}^d$, on pose $d_F(x) = d(x,F) = \inf(\|y-x\| : y \in F)$ et, pour $\varepsilon > 0$, G_ε est la fonction continue définie sur $\mathbb{R}$ par $G_\varepsilon(x) = \left(1 - \frac{|x|}{\varepsilon}\right)^+$. On a $G_\varepsilon \circ d_F \geq \mathbb{1}_F$, donc

$$\varlimsup_{n\to+\infty} \int G_\varepsilon \circ d_F \, d\mu_n \geq \varlimsup_{n\to+\infty} \int \mathbb{1}_F \, d\mu_n.$$

Comme $G_\varepsilon \circ d_F$ est uniformément continue (c'est la composée d'une application 1-lipschitzienne et d'une application $\frac{1}{\varepsilon}$-lipschitzienne), on a

$$\varlimsup_{n\to+\infty} \int G_\varepsilon \circ d_F \, d\mu_n = \int G_\varepsilon \circ d_F \, d\mu,$$

d'où

$$\forall \varepsilon > 0, \quad \int G_\varepsilon \circ d_F \, d\mu \geq \varlimsup_{n\to+\infty} \mu_n(F).$$

Or par définition de la mesure image, on a $\int G_\varepsilon \circ d_F \, d\mu = \int G_\varepsilon \, d\mu_{d_F}$. Lorsque ε tend vers 0, G_ε converge vers l'indicatrice de 0 et donc par convergence dominée, $\int G_\varepsilon \circ d_F \, d\mu = \int G_\varepsilon \, d\mu_{d_F}$ converge vers

$$\int \mathbb{1}_{\{0\}} \, d\mu_{d_F} = \mu_{d_F}(\{0\}) = \mu(d_F = 0) = \mu(F).$$

— Preuve de $(3,4) \Longrightarrow 5$. On a $\subset A \subset \overline{A}$, d'où

$$\begin{aligned}
& \varlimsup_{n\to+\infty} \mu_n(A) \leq \varlimsup_{n\to+\infty} \mu_n(\overline{A}) \leq \mu(\overline{A}), \\
\text{et} \quad & \varliminf_{n\to+\infty} \mu_n(A) \geq \varliminf_{n\to+\infty} \mu_n() \geq \mu().
\end{aligned}$$

Par ces deux inégalités, on obtient

$$\mu() \leq \underline{\lim}_{n\to+\infty} \mu_n(A) \leq \overline{\lim}_{n\to+\infty} \mu_n(A) \leq \mu(\overline{A}).$$

Comme $\mu(\overline{A}) - \mu() = \mu(\partial A) = 0$, la suite $(\mu_n(A))_{n\geq 1}$ admet une limite supérieure qui coïncide avec sa limite inférieure. Elle converge donc vers $\mu(\overline{A}) = \mu()$, c'est-à-dire vers $\mu(A)$, car $\mu() \leq \mu(A) \leq \mu(\overline{A})$.

— $(5) \Longrightarrow (6)$ est évident

— Preuve de $(6) \Longrightarrow (1)$. L'idée est d'approcher la fonction f par une somme d'indicatrices de pavés dont la frontière est de mesure nulle. On va commencer par exhiber un grand ensemble de mesure nulle.
Soit T une base de $\mathbb{R}$ vu comme un $\mathbb{Q}$-espace vectoriel telle que $1 \in T$. On peut toujours se ramener au cas où tous les éléments de la base sont dans $[0,1]$. Pour $t \in T$ et $k \in \{1,\dots,d\}$, notons

$$\mathcal{P}_t^k = \{x \in \mathbb{R}^d ; x_k - t \in \mathbb{Q}\}.$$

$\mathcal{P}_t^k$ est une réunion d'hyperplans orthogonaux au k-ième vecteur de la base de $\mathbb{R}^d$. À k fixé, les ensembles $(\mathcal{P}_t^k)_{t\in[0,1[\cap T}$ sont disjoints. Comme μ est une probabilité, l'ensemble des $t \in [0,1[\cap T$ tels que $\mu(\mathcal{P}_t^k) > 0$ est au plus dénombrable. Comme T n'est pas dénombrable [3], il existe t_k tel que $\mu(\mathcal{P}_{t_k}^k) = 0$. Pour un entier p, on pose $\bar{B}_p = \{x \in \mathbb{R}^d ; \|x-t\|_\infty \leq p\}$, où $t = (t_1,\dots,t_d)$. Soit $\varepsilon > 0$. On peut trouver p tel que $\mu(\bar{B}_p) \geq 1 - \frac{\varepsilon}{2}$. D'après ce qui précède, $\mu(\partial \bar{B}_p) = 0$. On peut donc trouver n_0 tel que $n \geq n_0$ entraîne $\mu_n(\bar{B}_p) \geq 1 - \varepsilon$. Or $\bar{B}_p$ est compact. On note ω_f le module de continuité de la restriction de f à $\bar{B}_p$, défini pour $\eta > 0$ par

$$\omega_f(\eta) = \sup(|f(x) - f(y)| : \ \|x-y\|_\infty \leq \eta, x, y \in \bar{B}_p).$$

Pour $N \geq 1$, on pose

$$f_N(x) = \mathbb{1}_{\bar{B}_p}(x)\, f\left(t_1 + \frac{\lceil N(x-t_1)\rceil}{N}, \dots, t_d + \frac{\lceil N(x-t_d)\rceil}{N}\right).$$

On a

$$|f_N(x) - f(x)| \leq \omega_f(1/N) + 2\|f\|_\infty \mathbb{1}_{\bar{B}_p^c}(x).$$

Ainsi

$$\forall N \in \mathbb{N}, \quad \int_{\mathbb{R}^d} |f_N(x) - f(x)| d\mu(x) \leq \omega_f(1/N) + 2\|f\|_\infty \frac{\varepsilon}{2},$$

$$\forall N \in \mathbb{N}, \quad \overline{\lim}_{n\to\infty} \int_{\mathbb{R}^d} |f_N(x) - f(x)| d\mu_n(x) \leq \omega_f(1/N) + 2\|f\|_\infty \varepsilon.$$

3. En effet, si on avait $T = \{t^i ; i \in \mathbb{N}^*\}$, on aurait $\mathbb{R} = \bigcup_{n=1}^{+\infty} \{ \sum_{i=1}^{n} \lambda_i t^i ; (\lambda_1,\dots,\lambda_n) \in \mathbb{Q}^n\}$, ce qui contredirait le fait que $\mathbb{R}$ n'est pas dénombrable.

Soit n_0 un entier tel que $\omega_f(1/n_0) \leq \varepsilon$. On a

$$\int_{\mathbb{R}^d} |f_{n_0}(x) - f(x)| \; d\mu(x) \leq \varepsilon + 2\|f\|_\infty \frac{\varepsilon}{2}$$

et

$$\overline{\lim_{n\to\infty}} \int_{\mathbb{R}^d} |f_{n_0}(x) - f(x)| \; d\mu_n(x) \leq \varepsilon + 2\|f\|_\infty \varepsilon.$$

On a de plus,

$$f_{n_0}(x) = \sum_{y \in \left(t + \frac{1}{n_0}\mathbb{Z}^d\right) \cap \bar{B}_p} f(y) \; \mathbb{1}_{\left\{y + \left]-\frac{1}{n_0}, 0\right]^d\right\}}(x).$$

Ainsi, f_N s'écrit comme une combinaison linéaire finie d'indicatrices de pavés de $\mathbb{R}^d$ dont la frontière est de μ-mesure nulle. On a donc

$$\lim_{n\to+\infty} \int f_N \; d\mu_n = \int f_N \; d\mu.$$

Comme

$$\begin{aligned}\int f \; d\mu - \int f \; d\mu_n &= \int f \; d\mu - \int f_N \; d\mu + \int f_N \; d\mu - \int f_N \; d\mu_n \\ &\quad + \int f_N \; d\mu_n - \int f \; d\mu_n,\end{aligned}$$

on a

$$\begin{aligned} &\overline{\lim_{n\to\infty}} \left| \int_{\mathbb{R}^d} f(x) \; d\mu(x) - \int_{\mathbb{R}^d} f(x) \; d\mu_n(x) \right| \\ \leq \; & \int_{\mathbb{R}^d} |f(x) - f_{n_0}(x)| \; d\mu(x) \\ & + \overline{\lim_{n\to\infty}} \left| \int_{\mathbb{R}^d} f_{n_0}(x) \; d\mu(x) - \int_{\mathbb{R}^d} f_{n_0}(x) \; d\mu_n(x) \right| \\ & + \overline{\lim_{n\to\infty}} \int_{\mathbb{R}^d} |f_{n_0}(x) - f(x)| \; d\mu_n(x) \\ \leq \; & (\varepsilon + \|f\|_\infty \varepsilon) + 0 + (\varepsilon + 2\|f\|_\infty \varepsilon) \\ \leq \; & \varepsilon(2 + 3\|f\|_\infty). \end{aligned}$$

Comme ε est arbitraire, on en déduit que $\int_{\mathbb{R}^d} f(x) \; d\mu_n(x)$ converge vers $\int_{\mathbb{R}^d} f(x) \; d\mu(x)$, et cela pour toute fonction continue bornée f. La condition (1) est donc vérifiée.

□

Corollaire 11.10. *Soient (X_n) une suite de variables aléatoires réelles, et X une variable aléatoire de fonction de répartition F_X. On a équivalence entre*

— *(X_n) converge en loi vers X.*

— *Pour tout point x où la fonction de répartition de X est continue, $F_{X_n}(x)$ tend vers $F_X(x)$ lorsque n tend vers l'infini.*

Démonstration. Sens direct : posons $A =]-\infty, x]$. A est un borélien dont la frontière $\{x\}$ est telle que $\mathbb{P}_X(\{x\}) = 0$, car $\{x\}$ est un point de continuité de F_X. Comme $\mathbb{P}_{X_n}(A) = F_{X_n}(x)$ et $\mathbb{P}_X(A) = F_X(x)$, il suffit donc d'appliquer le théorème de Portmanteau pour conclure. Réciproquement, soit $]a, b]$ un intervalle de $\mathbb{R}$, avec $\mathbb{P}_X(\partial]a, b]) = \mathbb{P}_X(\{a; b\}) = 0$. a et b étant des points de continuité de F_X, on a par hypothèse $F_X(a) = \lim\limits_{n\to+\infty} F_{X_n}(a)$ et

$F_X(b) = \lim\limits_{n\to+\infty} F_{X_n}(b)$, donc

$$\mathbb{P}_X(]a, b]) = F_X(b) - F_X(a) = \lim_{n\to+\infty} F_{X_n}(b) - F_n(a) = \lim_{n\to+\infty} \mathbb{P}_{X_n}(]a, b]).$$

On a vérifié la propriété (6) du Théorème de Portmanteau, et par conséquent X_n converge en loi vers X. □

Remarque 11.11. *Cette dernière conséquence de la convergence en loi est très utile, par exemple en statistique (voir le chapitre 13).*

Le théorème suivant est également très utile.

Théorème 11.12. *Si des mesures de probabilités $(\mu_n)_{n\geq 1}$ et μ sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ sont telles que pour toute fonction f continue positive à support compact de $\mathbb{R}^d$ dans $\mathbb{R}$, on a*

$$\lim_{n\to+\infty} \int_{\mathbb{R}^d} f\ d\mu_n = \int_{\mathbb{R}^d} f\ d\mu,$$

alors μ_n converge faiblement vers μ.

Démonstration. Soit $\varepsilon > 0$. Notons g_A la fonction continue, affine sur $[A/2, A]$, valant 1 sur $]-\infty, A/2]$, 0 sur $[A, +\infty[$, et $h_A = g_A \circ \|.\|$. On a

$$\int_{\mathbb{R}^d} h_A\ d\mu \geq \mu(B(0, A/2)),$$

de sorte que si l'on prend A suffisamment grand, on a

$$\int_{\mathbb{R}^d} h_A\ d\mu \geq 1 - \varepsilon.$$

Comme h_A est continue positive à support compact, $\int_{\mathbb{R}^d} h_A\ d\mu_n$ converge vers $\int_{\mathbb{R}^d} h_A\ d\mu$. Ainsi, il existe n_0 tel que

$$n \geq n_0 \Longrightarrow \int_{\mathbb{R}^d} h_A\ d\mu_n \geq 1 - 2\varepsilon \quad \text{et} \quad \mu_n(B(0, A)) \geq 1 - 2\varepsilon.$$

Soit maintenant f une fonction continue bornée positive. On a

$$\begin{aligned}\int f\,d\mu_n - \int f\,d\mu &= \left(\int fh_{2A}\,d\mu_n - \int fh_{2A}\,d\mu\right)\\ &\quad + \left(\int f(1-h_{2A})\,d\mu_n - \int f(1-h_{2A})\,d\mu\right).\end{aligned}$$

Ainsi pour $n \geq n_0$, on a

$$\left|\int f\,d\mu_n - \int f\,d\mu\right| \leq \left|\int fh_{2A}\,d\mu_n - \int fh_{2A}\,d\mu\right| + 3\varepsilon\|f\|_\infty.$$

Comme fh_{2A} est une fonction continue positive à support compact, la suite des intégrales $\int fh_{2A}\,d\mu_n$ converge vers $\int fh_{2A}\,d\mu$, d'où

$$\overline{\lim_{n\to+\infty}} \left|\int f\,d\mu_n - \int f\,d\mu\right| \leq 4\varepsilon\|f\|_\infty.$$

Comme ε est arbitraire, on en déduit que $\int f\,d\mu_n$ converge vers $\int f\,d\mu$.

Le passage à une fonction continue bornée de signe quelconque ne pose pas de problème, car $f = \max(f,0) - \max(-f,0)$ et le résultat s'ensuit par linéarité. □

11.1.4 Lien avec les autres modes de convergence

Théorème 11.13. *Soient $(X_n)_{n\geq 1}$ une suite de variables aléatoires et X une variable aléatoire.*

1. *Si X_n converge en probabilité vers X, alors X_n converge en loi vers X.*
2. *Si X_n converge en loi vers une constante a (ou de manière équivalente vers une masse de Dirac δ_a), alors X_n converge en probabilité vers a.*

Démonstration. 1. Soit f une fonction continue bornée. Soit $x_n = \mathbb{E}f(X_n)$. La suite $(x_n)_{n\geq 1}$ est bornée. Soit a une valeur d'adhérence de $(x_n)_{n\geq 1}$, avec $\lim_{k\to+\infty} x_{n_k} = a$. Comme X_{n_k} converge en probabilité vers X, on peut (d'après le théorème 10.19) en extraire une sous-suite $(X_{n_{m_k}})$ telle que $\lim_{k\to+\infty} X_{n_{m_k}} = X$ presque sûrement. Par continuité, on trouve que $\lim_{k\to+\infty} f(X_{n_{m_k}}) = f(X)$ presque sûrement. Ainsi, le théorème de convergence dominée donne $\lim_{k\to+\infty} \mathbb{E}f(X_{n_{m_k}}) = \mathbb{E}f(X)$, c'est-à-dire $\lim_{k\to+\infty} x_{n_{m_k}} = \mathbb{E}f(X)$. Or $(x_{n_{m_k}})$ est une sous-suite de (x_{n_k}) qui converge elle-même vers a, donc $a = \mathbb{E}f(X)$. Comme $(x_n)_{n\geq 1}$ est une suite bornée dont $\mathbb{E}f(X)$ est l'unique valeur d'adhérence, elle converge vers $\mathbb{E}f(X)$. De plus, pour toute fonction continue bornée f, $\mathbb{E}f(X_n)$ converge vers $\mathbb{E}f(X)$, donc $(X_n)_{n\geq 1}$ converge en loi vers X.

2. Soit $\varepsilon > 0$. Posons $F = \{x \in \mathbb{R}^d ; \|a - x\| \geq \varepsilon\}$.
   On a $\mathbb{P}(\|X_n - a\| \geq \varepsilon) = \mathbb{P}_{X_n}(F)$. Or F est fermé et $\mathbb{P}_{X_n}$ converge faiblement vers δ_a, donc, d'après le théorème de Portmanteau,
$$\varlimsup_{n\to\infty} \mathbb{P}_{X_n}(F) \leq \delta_a(F) = 0.$$
   Ainsi, pour tout $\varepsilon > 0$, $\mathbb{P}(\|X_n - a\| \geq \varepsilon)$ tend vers 0 quand n tend vers l'infini : $(X_n)_{n\geq 1}$ converge en probabilité vers a.

□

Le résultat suivant est connu sous le nom de théorème de Slutsky (parfois lemme de Slutsky) et est très utile dans la pratique.

Théorème 11.14. *Soient $(X_n)_{n\geq 1}$ et $(Y_n)_{n\geq 1}$ deux suites de vecteurs aléatoires à valeurs dans $\mathbb{R}^d$. Si $(X_n)_{n\geq 1}$ converge en loi vers un vecteur aléatoire X et $(Y_n)_{n\geq 1}$ converge en loi vers un vecteur constant c, alors*
- *$X_n + Y_n$ converge en loi vers $X + c$.*
- *$\langle X_n, Y_n\rangle$ converge en loi vers $\langle X, c\rangle$.*

Démonstration. Si on montre que $(X_n, Y_n)_{n\geq 1}$ converge en loi vers (X, c), alors il suffira d'appliquer le théorème 11.1 pour obtenir le résultat. Pour ce faire, nous allons utiliser le théorème de Portmanteau.

Tout d'abord, prouvons que $(X_n, c)_{n\geq 1}$ converge en loi vers (X, c). Ce résultat est vrai si on sait montrer que pour toute fonction continue bornée f, $\mathbb{E}f(X_n, c)$ converge vers $\mathbb{E}f(X, c)$. Pour ce faire, posons $g(x) = f(x, c)$. Bien entendu, g hérite des propriétés de continuité et bornitude de f. Comme la suite (X_n) converge en loi vers X, le théorème de Portmanteau implique que $\mathbb{E}g(X_n)$ converge vers $\mathbb{E}g(X)$. Cette dernière convergence équivaut à la convergence de $\mathbb{E}f(X_n, c)$ vers $\mathbb{E}f(X, c)$.

Comme (Y_n) converge en loi vers une constante, il s'agit en fait d'une convergence en probabilité. Ainsi, $\|(X_n, Y_n) - (X_n, c)\| = \|Y_n - c\|$ converge en probabilité vers 0. On a donc que $\|(X_n, Y_n) - (X_n, c)\|$ converge en probabilité vers 0 et (X_n, c) converge en loi vers (X, c).

Montrons que cela suffit à obtenir la convergence en loi de (X_n, Y_n) vers (X, c). En effet, si $U_n - V_n$ converge en probabilité vers 0 et U_n converge en loi vers U, alors V_n converge en loi vers U. Nous avons une fois de plus besoin du théorème de Portmanteau. Soit f une fonction uniformément continue, bornée par M. Soit $\varepsilon > 0$. Il existe $\delta > 0$ tel que $\|u - v\| \leq \delta$ implique

$|f(u)-f(v)| \leq \varepsilon$. On a alors

$$\begin{aligned}
|\mathbb{E}f(U_n)-\mathbb{E}f(V_n)| &\leq \mathbb{E}|f(U_n)-f(V_n)| \\
&\leq \varepsilon\mathbb{E}(\mathbb{1}_{\|U_n-V_n\|<\delta}) + 2M\mathbb{P}(\|U_n-V_n\| \geq \delta) \\
&\leq \varepsilon\mathbb{P}(\|U_n-V_n\|<\delta) + 2M\mathbb{P}(\|U_n-V_n\| \geq \delta) \\
&\leq \varepsilon + 2M\mathbb{P}(\|U_n-V_n\| \geq \delta).
\end{aligned}$$

À partir de là, on utilise l'inégalité triangulaire

$$\begin{aligned}
|\mathbb{E}f(V_n)-\mathbb{E}f(U)| &\leq |\mathbb{E}f(U_n)-\mathbb{E}f(V_n)| + |\mathbb{E}f(U_n)-\mathbb{E}f(U)| \\
&\leq \varepsilon + 2M\mathbb{P}(\|U_n-V_n\| \geq \delta) + |\mathbb{E}f(U_n)-\mathbb{E}f(U)|.
\end{aligned}$$

Comme les deux derniers termes de cette inégalité tendent vers 0, on conclut que $\overline{\lim\limits_{n\to+\infty}} |\mathbb{E}f(V_n)-\mathbb{E}f(U)| \leq \varepsilon$. Comme ε est quelconque, on en déduit le résultat cherché. □

11.2 Convergence en loi sur $\mathbb{R}^d$ grâce aux fonctions caractéristiques

11.2.1 Critère de convergence

Théorème 11.15 (premier théorème de Lévy). *Soit $(\mu_n)_{n\geq 1}$ une suite de mesures de probabilité et μ une mesure de probabilité donnée sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Alors la suite $(\mu_n)_{n\geq 1}$ converge faiblement vers μ si et seulement si*

$$\forall t \in \mathbb{R}^d, \quad \lim_{n\to+\infty} \phi_{\mu_n}(t) = \phi_\mu(t).$$

La preuve de ce théorème est reportée à la fin du chapitre.

11.2.2 Théorème de continuité de Lévy

Théorème 11.16 (continuité de Lévy). *Soient $(\mu_n)_{n\geq 1}$ une suite de mesures de probabilité et ϕ une fonction donnée. Si*

$$\forall t \in \mathbb{R}^d \quad \lim_{n\to+\infty} \phi_{\mu_n}(t) = \phi(t)$$

et si ϕ est continue en 0, alors il existe une unique mesure de probabilité μ telle que $\phi = \phi_\mu$ et la suite $(\mu_n)_{n\geq 1}$ converge faiblement vers μ.

La preuve de ce théorème est reportée à la fin du chapitre.

Ce dernier théorème est intéressant si la loi limite est une loi nouvelle, inconnue. L'appliquer lorsque la loi est une loi bien connue est assez maladroit.

11.2.3 Une application du théorème de Lévy

Le résultat qui suit peut être démontré sans l'aide du théorème de Lévy, mais ce dernier théorème en rend la preuve particulièrement simple.

Théorème 11.17. *Si μ_n tend faiblement vers μ et ν_n tend faiblement vers ν, alors la suite $(\mu_n \otimes \nu_n)_{n\geq 1}$ converge faiblement vers $\mu \otimes \nu$.*

Démonstration. Soient $s, t \in \mathbb{R}^d$. On a $\phi_{\mu_n \otimes \nu_n}(s,t) = \phi_{\mu_n}(s)\phi_{\nu_n}(t)$. Comme μ_n converge faiblement vers μ, $\phi_{\mu_n}(s)$ converge vers $\phi_\mu(s)$. De même, $\phi_{\nu_n}(t)$ converge vers $\phi_\nu(t)$. Ainsi $\phi_{\mu_n \otimes \nu_n}(s,t)$ converge vers $\phi_\mu(s)\phi_\nu(t) = \phi_{\mu\otimes\nu}(s,t)$, donc d'après le théorème de Lévy, la suite $(\mu_n \otimes \nu_n)_{n\geq 1}$ tend faiblement vers $\mu \otimes \nu$. □

Théorème 11.18. *Si μ_n tend faiblement vers μ et ν_n tend faiblement vers ν, alors la suite $(\mu_n * \nu_n)_{n\geq 1}$ tend faiblement vers $\mu * \nu$.*

Démonstration. Soient (X_n, Y_n) de loi $\mu_n \otimes \nu_n$ et (X, Y) de loi $\mu \otimes \nu$. D'après le théorème 11.17, (X_n, Y_n) converge en loi vers (X, Y), et donc d'après le corollaire 11.2, $X_n + Y_n$ converge en loi vers $X + Y$. Mais la loi de $X_n + Y_n$ est $\mu_n * \nu_n$ et la loi de $X + Y$ est $\mu * \nu$, donc le résultat est démontré. □

11.3 Théorème central limite en dimension 1

En dimension 1, le théorème s'énonce comme suit.

Théorème 11.19. *Soit $(X_n)_n$ une suite de variables aléatoires réelles indépendantes identiquement distribuées admettant un moment d'ordre 2. On note m l'espérance et σ^2 la variance communes à ces variables. Alors*

$$\frac{(X_1 + \cdots + X_n) - nm}{\sqrt{n}} \Longrightarrow \mathcal{N}(0, \sigma^2).$$

Démonstration. On pose $S_n = (X_1 + \cdots + X_n) - nm = \sum_{k=1}^{n} (X_k - m)$. Notons ϕ la fonction caractéristique de $X_1 - m$.
Comme les variables aléatoires $X_1 - m, \dots, X_n - m$ sont indépendantes et de même loi, la fonction caractéristique de $S_n/\sqrt{n}$ vaut

$$\phi_{S_n/\sqrt{n}}(t) = \phi_{S_n}\left(\frac{t}{\sqrt{n}}\right) = \prod_{k=1}^{n} \phi_{X_k - m}\left(\frac{t}{\sqrt{n}}\right) = \phi\left(\frac{t}{\sqrt{n}}\right)^n.$$

D'après le théorème de Lévy, pour montrer que $S_n/\sqrt{n}$ converge en loi vers $\mathcal{N}(0, \sigma^2)$, il suffit de montrer que

$$\forall t \in \mathbb{R} \qquad \lim_{n\to+\infty} \phi\left(\frac{t}{\sqrt{n}}\right)^n = \exp\left(-\frac{\sigma^2}{2}t^2\right),$$

car $t \mapsto \exp(-\frac{\sigma^2}{2}t^2)$ est la fonction caractéristique de la loi $\mathcal{N}(0, \sigma^2)$.

Pour ce faire, on utilisera le développement limité établi au corollaire 9.17 :

$$\phi(x) = 1 - \frac{\sigma^2}{2}x^2 + o(x^2). \tag{11.2}$$

L'introduction du logarithme complexe peut être évitée en remarquant que pour des nombres complexes z et u de module inférieur ou égal à 1, on a

$$\forall n \in \mathbb{N}^*, \quad |z^n - u^n| = \left|(z-u)\left(\sum_{k=0}^{n-1} z^k u^{n-1-k}\right)\right| \leq n|z-u|.$$

Il s'ensuit que

$$\begin{aligned}\left|\phi\left(\frac{t}{\sqrt{n}}\right)^n - \exp\left(-\frac{\sigma^2}{2}t^2\right)\right| &= \left|\phi\left(\frac{t}{\sqrt{n}}\right)^n - \exp\left(-\frac{\sigma^2}{2n}t^2\right)^n\right| \\ &\leq n\left|\phi\left(\frac{t}{\sqrt{n}}\right) - \exp\left(-\frac{\sigma^2}{2n}t^2\right)\right|.\end{aligned}$$

On a d'une part $\exp(-\frac{\sigma^2}{2n}t^2) = 1 - \frac{\sigma^2}{2n}t^2 + o(1/n)$, et d'autre part, d'après l'équation (11.2), $\phi(\frac{t}{\sqrt{n}}) = 1 - \frac{\sigma^2}{2n}t^2 + o(1/n)$. On obtient ainsi le résultat cherché, à savoir $n\left|\phi\left(\frac{t}{\sqrt{n}}\right) - \exp\left(-\frac{\sigma^2}{2n}t^2\right)\right| = o(1)$, ce qui achève la preuve. □

Une application importante est l'étude des fluctuations des fréquences de réussite dans une suite d'épreuves indépendantes de même probabilité, ou de manière équivalente, l'approximation d'une loi binomiale par une loi gaussienne.

Théorème 11.20. *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace probabilisé sur lequel est définie une suite d'événements indépendants* $(A_n)_{n\geq 1}$ *de même probabilité* p.
Pour $\omega \in \Omega$*, on note* $N_n(\omega)$ *le nombre d'événements qui sont réalisés parmi* $A_1, \dots, A_n$*. Ainsi, on a*

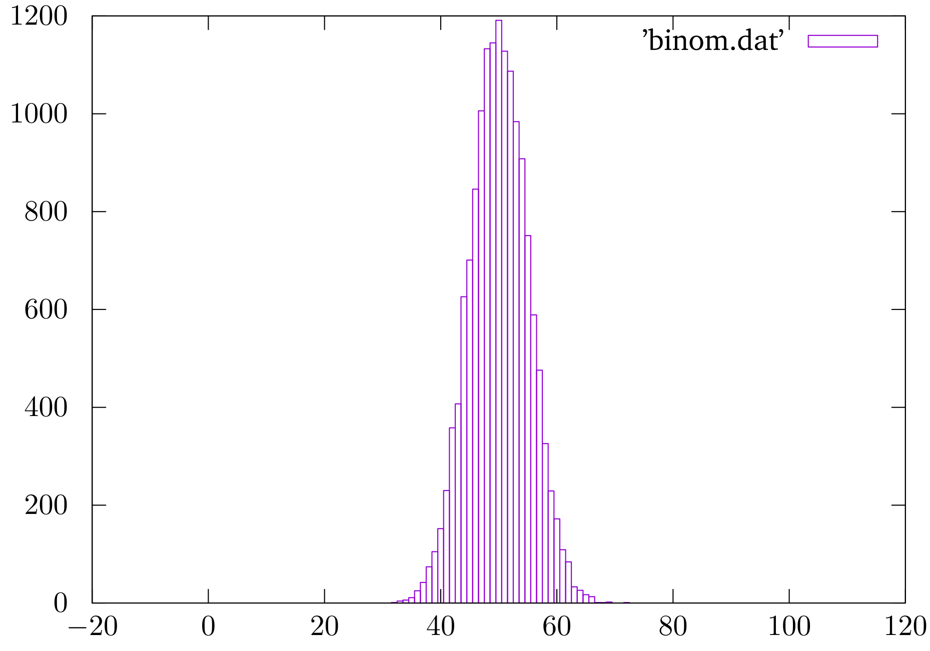
$$N_n = \sum_{k=1}^{n} \mathbb{1}_{A_k}$$

et N_n *suit la loi binomiale* $\mathcal{B}(n,p)$*. Alors on a*

$$\frac{N_n - np}{\sqrt{n}} \Longrightarrow \mathcal{N}(0, p(1-p)).$$

Démonstration. Il suffit d'appliquer le théorème précédent à la suite $(\mathbb{1}_{A_n})_{n\geq 1}$, qui est une suite de variables de Bernoulli indépendantes, d'espérance p et de variance $p(1-p)$. □

L'histogramme ci-dessous représente le nombre d'observations de chaque entier compris entre 0 et 100 pour une simulation de 15000 variables aléatoires indépendantes suivant la loi $\mathcal{B}(100, \frac{1}{2})$.



11.4 Preuve des théorèmes de Lévy

11.4.1 Tension

Définition. *On dit qu'une famille $\mathcal{M}$ de mesures de probabilités sur $\mathbb{R}^d$ est tendue si pour tout $\varepsilon > 0$, il existe un compact K de $\mathbb{R}^d$ tel que pour tout $\mu \in \mathcal{M}$, on a $\mu(K^c) \leq \varepsilon$.*

Exemple: Une famille constituée d'une unique mesure de probabilité μ sur $\mathbb{R}^d$ est tendue. En effet, considérons la suite d'ensemble $A_n = B(0, n)$. La suite (A_n^c) est décroissante et son intersection est vide, donc d'après le théorème de continuité séquentielle décroissante, $\mu(A_n^c)$ tend vers 0, ce qui montre que, pour n assez grand, $\mu(A_n^c)$ ne dépasse pas ε.

Lemme 11.21. *La réunion de deux familles tendues est une famille tendue.*

Démonstration. Soient $\mathcal{M}_1$ et $\mathcal{M}_2$ deux familles tendues. Soit $\varepsilon > 0$. Comme $\mathcal{M}_1$ est tendue, il existe un compact K_1 tel que $\forall \mu \in \mathcal{M}_1$, $\mu(K_1^c) \leq \varepsilon$. De même, il existe un compact K_2 tel que $\forall \mu \in \mathcal{M}_2$, $\mu(K_2^c) \leq \varepsilon$. Maintenant, si l'on pose $K = K_1 \cup K_2$, K est compact et il est facile de voir que

$$\forall \mu \in \mathcal{M}_1 \cup \mathcal{M}_2, \quad \mu(K^c) \leq \varepsilon.$$

Ainsi $\mathcal{M}_1 \cup \mathcal{M}_2$ est tendue. □

Corollaire 11.22. *Toute famille finie de mesures sur $\mathbb{R}^d$ est tendue.*

Le lemme suivant peut également être utile.

Lemme 11.23. *Soit $x \in \mathbb{R}^d$ et notons, pour $1 \leq k \leq d$, $\pi_k(x) = x_k$ la k-ième composante de x. Soit $\mathcal{M}$ une famille de mesures sur $\mathbb{R}^d$. $\mathcal{M}$ est tendue si et seulement si pour tout $1 \leq k \leq d$, la famille $\{\mu_{\pi_k}; \mu \in \mathcal{M}\}$ est tendue.*

Démonstration.
— Supposons que $\mathcal{M}$ est tendue. Soient $1 \leq k \leq d$ et $\varepsilon > 0$. Il existe un compact K tel que pour tout $\mu \in \mathcal{M}$, $\mu(K) \geq 1 - \varepsilon$. Alors, il est clair que $\pi_k(K)$ est compact et que pour tout $\mu \in \mathcal{M}$, on a

$$\mu_{\pi_k}(\pi_k(K)) = \mu(x \in \mathbb{R}^d : \pi_k(x) \in \pi_k(K)) \geq \mu(K) \geq 1 - \varepsilon.$$

— Réciproquement, supposons que $\{\mu_{\pi_k}; \mu \in \mathcal{M}\}$ est tendue. Soit $\varepsilon > 0$. Pour tout $1 \leq k \leq d$, il existe un compact K_k tel que pour tout $\mu \in \mathcal{M}$ on a $\mu_{\pi_k}(K_k) \geq 1 - \varepsilon/d$. Posons $K = K_1 \times K_2 \times \cdots \times K_d$. On a $K^c = \bigcup_{i=1}^{d} \pi_k^{-1}(K_k^c)$, donc

$$\mu(K^c) \leq \sum_{i=1}^{d} \mu(\pi_k^{-1}(K_k^c)) = \sum_{i=1}^{d} \mu_{\pi_k}(K_k^c) \leq \sum_{i=1}^{d} \varepsilon/d = \varepsilon.$$

□

Définition. *On dit qu'une famille $\mathcal{M}$ de mesures de probabilités est relativement compacte si et seulement si toute suite d'éléments de $\mathcal{M}$ admet une sous-suite qui converge en loi vers une mesure de probabilité.*

Théorème 11.24 (Prohorov)**.** *Toute famille tendue est relativement compacte.*

Démonstration. On donne la preuve uniquement dans le cas où $d = 1$ [4].

Soit $(\mu_n)_n$ une suite de mesures appartenant à une famille $\mathcal{M}$ tendue. On note F_n la fonction de répartition de μ_n. D'après le théorème 5.6 de Helly, on peut trouver une suite strictement croissante (n_k) et une fonction F continue à droite et croissante telle que $F_{n_k}(x)$ converge vers $F(x)$ en chaque point de continuité de F. On va montrer que F admet respectivement les limites

4. En dimension plus grande, des complications apparaissent : la première étape d'extraction ne pose pas de difficulté, mais la reconstruction d'une mesure à partir de la fonction de répartition est plus complexe. Voir par exemple Billingsley [1].

0 et 1 en $-\infty$ et en $+\infty$. Soit $\varepsilon > 0$. La famille $\mathcal{M}$ étant tendue, il existe un compact K tel que pour tout $\mu \in \mathcal{M}$, on ait $\mu(K) \geq 1 - \varepsilon$. Considérons $M = \sup\{|x|; x \in K\}$. Comme l'ensemble des points de discontinuité d'une fonction croissante est au plus dénombrable, il existe $a < -M$ et $b > M$ tels que F soit continue en a et en b. Ainsi, pour tout $k \geq 1$, on a

$$F_{n_k}(b) - F_{n_k}(a) = \mu_{n_k}(]a, b]) \geq \mu_{n_k}([-M, M]) \geq \mu_{n_k}(K) \geq 1 - \varepsilon.$$

Or a et b sont des points de continuité de F. On obtient, en faisant tendre k vers $+\infty$, que $F(b) - F(a) \geq 1 - \varepsilon$, soit $F(a) + (1 - F(b)) \leq \varepsilon$. Comme F est à valeurs dans $[0, 1]$, on a $F(a) \leq \varepsilon$ et $F(b) \geq 1 - \varepsilon$. Avec la monotonie de F, cela donne les limites voulues. Ainsi, d'après le corollaire 5.37, il existe une mesure de probabilité μ telle que pour tous $a < b$, $F(b) - F(a) = \mu(]a, b])$. Comme $F_{n_k}(x)$ converge vers $F(x)$ en chaque point de continuité de F, on conclut avec le corollaire 11.10 que $(\mu_{n_k})_k$ converge en loi vers μ. □

Corollaire 11.25. *Soit $(\mu_n)_{n\geq 1}$ une suite de mesures de probabilité telle que*

i) $\{\mu_n; n \geq 1\}$ *est tendue,*

ii) toute sous-suite de $(\mu_n)_{n\geq 1}$ qui converge en loi converge vers μ.

Alors $(\mu_n)_{n\geq 1}$ converge en loi vers μ.

Démonstration. Soit f une fonction continue bornée. On doit montrer que la suite $(\int f(x)\, d\mu_n(x))_{n\geq 1}$ converge vers $\int f(x)\, d\mu(x)$. Soit $(n_k)_{k\geq 1}$ une suite quelconque strictement croissante d'entiers. La suite $(\mu_{n_k})_{k\geq 1}$ est une suite d'éléments d'une famille tendue. On peut donc en extraire une sous-suite convergente $(\mu_{n_{k_i}})_{i\geq 1}$. Mais d'après la deuxième hypothèse, la limite ne peut être que μ. Il s'ensuit que

$$\lim_{i\to+\infty} \int f(x)\, d\mu_{n_{k_i}}(x) = \int f(x)\, d\mu(x).$$

Ainsi, de chaque sous-suite de la suite $(\int f(x)\, d\mu_n(x))_{n\geq 1}$, on peut extraire une sous-suite qui tend vers $\int f(x)\, d\mu(x)$.
Cela signifie que la suite $(\int f(x)\, d\mu_n(x))_{n\geq 1}$ converge vers $\int f(x)\, d\mu(x)$, ce qui achève la preuve. □

11.4.2 Théorèmes de Lévy

Théorème 11.26. *Soit (μ_n) une suite de mesures de probabilité sur $\mathbb{R}^d$. On a les résultats suivants :*

— *Si μ_n converge en loi vers une probabilité μ, alors la suite (ϕ_{μ_n}) converge ponctuellement vers ϕ_μ.*

— *Si la suite (ϕ_{μ_n}) converge ponctuellement vers ϕ_μ, où ϕ_μ est la fonction caractéristique de la loi de la probabilité μ, alors μ_n converge en loi vers μ.*

— *Si la suite* (ϕ_{μ_n}) *converge ponctuellement vers une fonction* ϕ *continue en l'origine, alors il existe une mesure de probabilité* μ *dont la fonction caractéristique est* ϕ_μ *et* μ_n *converge en loi vers* μ.

Démonstration. Le premier point résulte de la définition de la convergence en loi et du fait que pour tout t, la fonction $x \mapsto e^{i\langle t,x\rangle}$ est continue bornée.

Le deuxième point est une conséquence du troisième, en utilisant le fait que la fonction caractéristique ϕ_μ est bien continue en l'origine.

Montrons donc le troisième point (sans utiliser le deuxième). Ce qui nous manque, c'est la tension de la suite $(\mu_n)_{n\geq 1}$, car si $(\mu_n)_{n\geq 1}$ est tendue, elle admet une valeur d'adhérence μ par Prohorov et il ressort alors clairement du premier point que $\phi_\mu = \phi$. De plus, si une mesure μ_* est valeur d'adhérence de la suite $(\mu_n)_{n\geq 1}$, alors $\phi_{\mu_*} = \phi$, donc $\mu_* = \mu$. Grâce au corollaire précédent, cela implique alors la convergence en loi de μ_n vers μ. Il reste donc à montrer que $(\mu_n)_{n\geq 1}$ est tendue. Pour cela, il suffit de montrer que, pour tout $1 \leq k \leq d$, la suite de mesures marginales $(\pi_k^{-1}\mu_n)_{n\geq 1}$ est tendue. Notons qu'on a simplement $\phi_{\pi_k^{-1}\mu_n}(x) = \phi_{\mu_n}(0,\dots,0,x,0,\dots,0)$ et la suite de fonctions $(\phi_{\pi_k^{-1}\mu_n}(x))_{n\geq 1}$ converge simplement vers la fonction $x \mapsto \phi(0,\dots,0,x,0,\dots,0)$, qui est continue en 0. Ainsi, on est ramené à démontrer qu'une famille $(\nu_n)_{n\geq 1}$ de mesures de probabilités sur $\mathbb{R}$ dont les fonctions caractéristiques ψ_n convergent simplement vers une fonction ψ continue en 0 est tendue. Soient $n \geq 1$ et $a > 0$. On a

$$\begin{aligned}
\frac{1}{2a}\int_{-a}^{a}(1-\psi_n(t))\,dt &= \frac{1}{2a}\int_{-a}^{a}\int(1-e^{itx})\,d\nu_n(x)\,dt \\
&= \int \frac{1}{2a}\int_{-a}^{a}(1-e^{itx})\,dt\,d\nu_n(x) \\
&= \int\left(1-\frac{\sin ax}{ax}\right)\,d\nu_n(x).
\end{aligned}$$

La fonction $\theta \mapsto 1-\frac{\sin\theta}{\theta}$ est positive et pour $|\theta| \geq \pi/2$, on a $\left|\frac{\sin\theta}{\theta}\right| \leq \frac{2}{\pi}$, d'où

$$\begin{aligned}
\frac{1}{2a}\int_{-a}^{a}(1-\psi_n(t))\,dt &= \int\left(1-\frac{\sin ax}{ax}\right)\,d\nu_n(x) \\
&\geq \int \mathbb{1}_{\{|ax|\geq\pi/2\}}\left(1-\frac{2}{\pi}\right)\,d\nu_n(x) \\
&= \left(1-\frac{2}{\pi}\right)\nu_n\left(\{x\in\mathbb{R} : |x| \geq \frac{\pi}{2a}\}\right).
\end{aligned}$$

Soit maintenant $\varepsilon > 0$. On va montrer qu'il existe un compact qui porte à ε près la charge de chaque ν_n. Comme ψ est continue en 0, il existe $a > 0$ tel que

$$\sup_{t\in[-a,a]}|1-\psi(t)| \leq \frac{\varepsilon}{2}\left(1-\frac{2}{\pi}\right),$$

ce qui entraîne
$$\frac{1}{2a}\int_{-a}^{a}(1-\psi(t))\,dt \le \frac{\varepsilon}{2}\left(1-\frac{2}{\pi}\right).$$
D'autre part, d'après le théorème de convergence dominée, on sait que la quantité $\frac{1}{2a}\int_{-a}^{a}(1-\psi_n(t))\,dt$ converge vers $\frac{1}{2a}\int_{-a}^{a}(1-\psi(t))\,dt$, donc il existe n_0 tel que pour $n>n_0$, on ait
$$\frac{1}{2a}\int_{-a}^{a}(1-\psi_n(t))\,dt \le \varepsilon\left(1-\frac{2}{\pi}\right).$$
Cela entraîne $\nu_n(\{x\in\mathbb{R} : |x|\ge \frac{\pi}{2a}\})\le\varepsilon$. De plus, comme la famille $(\nu_n)_{n\le n_0}$ est finie, elle est tendue : il existe un compact K tel que pour tout $n\le n_0$, on ait $\nu_n(K)\ge 1-\varepsilon$. Enfin $K'=K\cup[-\frac{\pi}{2a},\frac{\pi}{2a}]$ est compact et pour tout $n\ge 1$, on a $\nu_n(K')\ge 1-\varepsilon$, ce qui achève la preuve. □

11.5 Exercices sur la convergence en loi

11.5.1 Exercices corrigés

Exercice 247. Soit $(U_n)_{n\ge1}$ une suite de variables aléatoires indépendantes de même loi $\mathcal{N}(0,1)$. Pour tout $\theta\in\mathbb{R}$, on définit la suite $(X_n)_{n\ge0}$ par la relation de récurrence $X_n=\theta X_{n-1}+U_n$ pour tout $n\ge1$ et $X_0=0$ presque sûrement.

1. Déterminer les lois des variables aléatoires X_n.
2. Étudier la convergence en loi de la suite de variables aléatoires $(X_n)_{n\ge0}$.

→ indication → solution

Exercice 248. Soit X_n suivant la loi uniforme sur $\{0,\dots,n-1\}$. Montrer que X_n/n converge en loi vers la loi uniforme sur $[0,1]$.
→ indication → solution

Exercice 249. On suppose que pour tout n, X_n suit la loi Gamma $\Gamma(2,n)$.

1. Montrer que X_n converge en loi vers 0.
2. En déduire un équivalent de $\int_0^{+\infty}|\sin t|e^{-nt}\,dt$.

→ indication → solution

Exercice 250. Soit $(X_n)_n$ une suite de variables aléatoires indépendantes identiquement distribuées de loi uniforme sur $[0,1]$.

1. Déterminer la loi de la variable aléatoire $M_n=\max_{1\le j\le n} X_j$.
2. Démontrer que M_n converge vers 1 en loi et presque sûrement.

3. Démontrer que la suite de variables aléatoires $(n(1-M_n))_n$ converge en loi et trouver la loi limite.

→ indication → solution

Exercice 251. *Une preuve probabiliste de la formule de Stirling.*
Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires indépendantes suivant la loi exponentielle de paramètre 1. On pose

$$S_n = \sum_{k=0}^{n} X_k.$$

1. Soient $(Y_n)_{n\geq 1}$ une suite de variables aléatoires convergeant en loi vers une variable aléatoire Y, $(a_n)_{n\geq 1}$ et $(b_n)_{n\geq 1}$ des suites de réels convergeant respectivement vers les réels a et b. Montrer que la suite de variables aléatoires $(a_nY_n+b_n)_{n\geq 1}$ converge en loi vers la variable aléatoire $aY+b$.
2. Montrer que $\frac{S_n-n}{\sqrt{n}}$ converge en loi vers la loi $\mathcal{N}(0,1)$.
3. Montrer que S_n suit la loi $\Gamma(n+1,1)$. En déduire que la densité de $\frac{S_n-n}{\sqrt{n}}$ s'écrit $g_n(x)=a_nh_n(x)$, avec
$$a_n = \frac{n^{n+1/2}e^{-n}\sqrt{2\pi}}{\Gamma(n+1)}$$
et
$$h_n(x) = \frac{1}{\sqrt{2\pi}}e^{-\sqrt{n}x}\left(1+\frac{x}{\sqrt{n}}\right)^n \mathbb{1}_{[-\sqrt{n},+\infty[}.$$
4. Montrer que
$$\lim_{n\to+\infty}\int_0^1 g_n(x)\,dx = \int_0^1 \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx.$$
5. Montrer que
$$\lim_{n\to+\infty}\int_0^1 h_n(x)\,dx = \int_0^1 \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx.$$
6. En déduire la formule de Stirling :
$$n! \sim \sqrt{2\pi}n^{n+1/2}e^{-n}.$$

→ indication → solution

Exercice 252. *Existence des lois stables d'indice α.*
Soit $(Y_{n,k})$ une suite de variables aléatoires indépendantes suivant la loi de Poisson $\mathcal{P}(cn^\alpha/|k|^{1+\alpha})$, où c est une constante positive et α un réel vérifiant $0<\alpha<2$. On pose

$$Z_n = \frac{1}{n}\sum_{k=-n^2}^{n^2} kY_{n,k}.$$

1. Montrer que la fonction caractéristique de Z_n s'écrit sous la forme $\phi_{Z_n}(\theta) = \exp(-2c\theta^\alpha u_n(\theta))$, avec
$$u_n(\theta) = \int_0^{n|\theta|} f\left(\frac{\theta}{n}\lfloor\frac{nx}{\theta}\rfloor\right)\, dx,$$
où $f(x) = \frac{1-\cos x}{x^{1+\alpha}}$.
Indication : on reconnaîtra une "somme de Riemann".
2. Montrer que, pour un choix approprié de c, la suite de fonctions (ϕ_{Z_n}) converge vers $\phi(\theta) = \exp(-|\theta|^\alpha)$. (Pour ce faire, on pourra remarquer que $|f(x)| \le \min\left(\frac{2}{x^{1+\alpha}}, \frac{1}{x^\alpha}, \frac{1}{2}\frac{1}{x^{\alpha-1}}\right)$.)
3. Soit α tel que $0 < \alpha \le 2$. Montrer qu'il existe une mesure m_α dont la fonction caractéristique est la fonction $\theta \mapsto \exp(-|\theta|^\alpha)$.
4. Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes identiquement distribuées suivant la loi m_α. Montrer qu'il existe un réel λ_n tel que $\lambda_n(X_1 + X_2 + \cdots + X_n)$ suive la loi m_α.

$\to$ indication $\to$ solution

Exercice 253. Soit (X_n) une suite de variables aléatoires convergeant en loi vers une variable aléatoire X.

1. Montrer que la famille $(X_n)_{n\ge1}$ est tendue.
2. On pose $\phi_n = \phi_{X_n}$. Montrer que la famille $(\phi_n)_{n\ge1}$ est uniformément équicontinue, c'est-à-dire que
$$\forall\varepsilon > 0\, \exists\eta > 0\, \forall n \ge 1,\ \forall(x,y) \in \mathbb{R}^2,\ |x-y| \le \eta \Longrightarrow |\phi_n(x)-\phi_n(y)| \le \varepsilon.$$

$\to$ indication $\to$ solution

Exercice 254. *Renouvellement.*
Soit $(X_n)_{n\ge1}$ une suite de variables aléatoires positives, indépendantes et de carré intégrable, avec $\mathbb{E}(X_n) = 1$.
On pose $S_n = X_1 + \cdots + X_n$ et $N_t = \inf\{n \ge 1; S_n \ge t\}$.
Montrer que $\frac{N_t - t}{\sqrt{t}}$ converge en loi vers $\mathcal{N}(0,1)$ quand t tend vers l'infini (c'est-à-dire que $\frac{N_{t_n}-t_n}{\sqrt{t_n}}$ converge en loi vers $\mathcal{N}(0,1)$ pour toute suite t_n tendant vers l'infini $+\infty$). $\to$ indication $\to$ solution

Exercice 255. 1. Interpréter la quantité $e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}$ comme une probabilité. À l'aide du théorème central de la limite, démontrer la relation
$$\lim_{n\to+\infty} e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!} = \frac{1}{2}.$$

2. On tire avec remise dans une urne contenant n boules jusqu'à ce qu'une boule ait été tirée deux fois. Quelle est l'espérance du nombre X_n de couleurs différentes qui ont été tirées ? Donner un équivalent lorsque n tend vers l'infini.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 256. Soient $(X_n)_{n\geq 1}$ des variables indépendantes de même loi avec $\mathbb{P}(X_n = \pm 1) = 1/2$. Alors $Z_n = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)$ converge en loi vers une variable aléatoire $Z \sim \mathcal{N}(0,1)$. Montrer qu'en revanche, il n'y a pas convergence en probabilité. $\rightarrow$ indication $\rightarrow$ solution

Exercice 257. Soient $(X_n)_{n\geq 1}$ des variables indépendantes de même loi avec $\mathbb{P}(X_n = \pm 1) = 1/2$. On pose $S_n = X_1 + \cdots + X_n$. Déterminer une constante A telle que $\mathbb{E}(|S_n|) \sim A\sqrt{n}$. $\rightarrow$ indication $\rightarrow$ solution

Exercice 258. *Convergence vers la loi Zêta. Application à la densité naturelle des couples d'entiers premiers entre eux et des entiers sans facteur carré.*

1. Soient $X, (X_n)_{n\geq 1}$ des variables aléatoires à valeurs dans $\mathbb{N}^*$. On suppose que :
   — (X_n) est tendue,
   — pour tout $N \geq 1$, $\mathbb{P}(N|X_n) \rightarrow \mathbb{P}(N|X)$.
   Le but de la question est de montrer que (X_n) converge en loi vers X. Pour p premier et x entier naturel non nul, on note $\nu_p(x)$ l'exposant de p dans la décomposition de x en produit de facteurs premiers (c'est la valuation p-adique de x). Pour $N \geq 1$, on note encore
   $$\psi_N(x) = \prod_{i=1}^{N} p_i^{\nu_{p_i}(x)}.$$
   (a) Montrer que pour tout entier naturel N, la suite de vecteurs aléatoires $(\nu_{p_1}(X_n), \ldots, \nu_{p_N}(X_n))_{n\geq 1}$ converge en loi vers le vecteur $(\nu_{p_1}(X), \ldots, \nu_{p_N}(X))$. En déduire que $\psi_N(X_n)$ converge en loi vers $\psi_N(X)$.
   (b) Soit $\varepsilon > 0$. Montrer qu'il existe N tel que $\mathbb{P}(X > p_N) \leq \varepsilon/3$ et pour tout $n \geq 1$, $\mathbb{P}(X_n > p_N) \leq \varepsilon/3$, puis que pour toute variable aléatoire Y à valeurs dans $\mathbb{N}^*$ et tout entier naturel i, on a
   $$|\mathbb{P}(Y = i) - \mathbb{P}(\psi_N(Y) = i)| \leq \mathbb{P}(Y > p_N).$$
   (c) Conclure.
2. Soient X_n, Y_n des variables aléatoires indépendantes suivant la loi uniforme sur $\{1, \ldots, n\}$. On note $Z_n = X_n \wedge Y_n$ et $W_n = r(X_n)$, où $r(n)$ est le plus grand entier a tel que a^2 divise n.
   Montrer que W_n et Z_n convergent en loi vers la loi Zêta de paramètre 2. Que valent $\lim_{n\to+\infty} \mathbb{P}(Z_n = 1)$ et $\lim_{n\to+\infty} \mathbb{P}(W_n = 1)$? Interpréter.

Remarque : on montre de la même manière le résultat suivant, dû à Ernesto Cesàro. Soient $X_n^1, \dots, X_n^m$ des variables aléatoires indépendantes suivant la loi uniforme sur $\{1, \dots, n\}$. On note $Z_n = X_n^1 \wedge \dots \wedge X_n^m$ et $W_n = r_m(X_n^1)$, où $r_m(n)$ est le plus grand entier a tel que a^m divise n. Alors W_n et Z_n convergent en loi vers la même loi Zêta de paramètre m.

On pourra également trouver dans [17] une extension des résultats de cet exercice à l'anneau $\mathbb{Z}[i]$.

$\rightarrow$ indication $\rightarrow$ solution

11.5.2 Exercices non corrigés

Exercice 259. Soit $(X_\lambda)_{\lambda>0}$ une famille de variables aléatoires telle que pour tout $\lambda > 0$, X_λ suive une loi de Poisson $\mathcal{P}(\lambda)$. Montrer que la suite de terme général

$$\frac{X_\lambda - \lambda}{\sqrt{\lambda}}$$

converge faiblement (en loi) vers la loi $\mathcal{N}(0, 1)$ lorsque λ tend vers l'infini. $\rightarrow$ indication

Exercice 260.

1. Soit $(U_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi uniforme sur $[0, 1]$. Pour $x > 0$, on pose

$$X_n^x = \sum_{k=1}^{n} \mathbb{1}_{\{U_k \leq \frac{x}{n}\}}.$$

Montrer que $(X_n^x)_{n\geq 1}$ converge en loi et déterminer la loi limite.

2. Soit $(u_n)_{n\geq 0}$ une suite croissante, de limite $\ell \in \mathbb{R}$. Pour $x > 0$, on pose

$$f(x) = \sum_{k=0}^{+\infty} e^{-x} \frac{x^k}{k!} u_k.$$

Montrer que f est une fonction croissante et déterminer sa limite en $+\infty$.

$\rightarrow$ indication

Exercice 261. Jean joue au jeu suivant. Sur chaque case d'un plateau carré de taille $n \times n$, il dispose une pièce de 1 €, les côtés visibles étant choisis au hasard (c'est-à-dire avec équiprobabilité), de manière indépendante. Ensuite, il tire au hasard un nombre X compris entre 1 et n. Deux possibilités s'offrent alors à lui :

— soit retourner tous les pions de la colonne X,

— soit retourner tous les pions de la ligne X.

Son but est de maximiser le nombre de faces. On suppose que Jean agit intelligemment. On note alors F_n (resp. P_n) le nombre de faces (resp. de piles) dans la configuration ainsi obtenue. On pose $D_n = F_n - P_n$.

Montrer que

$$\mathbb{E}D_n \underset{n\to+\infty}{\sim} \sqrt{\frac{4n}{\pi}},$$

mais que

$$\lim_{n\to+\infty} P(D_n \geq 0) = \frac{1}{2}.$$

$\to$ indication

Exercice 262. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires positives convergeant en loi vers X. Montrer que

$$\mathbb{E}[X] \leq \varliminf_{n\to+\infty} \mathbb{E}[X_n].$$

Indication : on pourra utiliser le fait que $\mathbb{E}[X_n] = \int_{\mathbb{R}^+} \mathbb{P}(X_n > t)\, d\lambda(t)$. $\to$ indication

Exercice 263. Soit $n \geq 1$. Montrer qu'on peut choisir un réel λ_n de telle sorte que la mesure μ_n dont le support est exactement $\{-n, -(n-1), \dots, n-1, n\}$ et vérifiant

$$\forall k \in \{-n, -(n-1), \dots, n-1, n\}, \quad \mu_n(k) = \lambda_n(2n+1-2|k|)$$

soit une mesure de probabilité. Soit maintenant X_n suivant la loi μ_n. Montrer que X_n/n converge en loi.$\to$ indication

Chapitre 12

Vecteurs gaussiens

Certaines notions déjà définies pour les variables aléatoires réelles se transposent aux vecteurs aléatoires. En probabilité et en statistiques, les variables gaussiennes et les vecteurs gaussiens jouent un rôle particulier.

Définition. *On dit qu'un vecteur aléatoire $X \in \mathbb{R}^d$ est gaussien si pour tout $a \in \mathbb{R}^d$ la variable aléatoire $\langle X, a \rangle$ est gaussienne.*

Rappelons qu'une variable aléatoire est gaussienne si elle est constante où si elle admet une densité de la forme $x \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-m)^2}{2\sigma^2}$ par rapport à la mesure de Lebesgue.

Nous verrons dans ce chapitre une généralisation (au cas des vecteurs aléatoires) du théorème central limite (vu au chapitre précédent). De plus, les vecteurs gaussiens sont très importants en statistique. En effet, dans le "monde gaussien", certaines choses sont simplifiées, comme la notion d'indépendance (voir la section 12.4), et l'un des tests statistiques les plus importants est basé sur l'approximation par des gaussiennes, le test du χ^2 (introduit section 12.8). En résumé, ce chapitre a deux principaux objectifs : à la fois généraliser le théorème central limite, et préparer le chapitre de statistiques.

12.1 Image affine d'un vecteur gaussien

Théorème 12.1. *L'image d'un vecteur gaussien X d'espérance m_X et de matrice de covariance C_X par une application affine $x \mapsto Ax + b$ est un vecteur gaussien d'espérance $m_Y = Am_X + b$ et de matrice de covariance $C_Y = AC_XA^*$.*

Démonstration. On pose $Y = AX + b$. Soit $a \in \mathbb{R}^d$.
On a $\langle Y, a \rangle = \langle AX, a \rangle + \langle b, a \rangle = \langle X, A^*a \rangle + \langle b, a \rangle$. Comme X est un vecteur gaussien $\langle X, A^*a \rangle$ est une variable aléatoire gaussienne. Quand on ajoute une

constante à une variable aléatoire gaussienne, on obtient une variable aléatoire gaussienne. Ainsi, pour tout a, $\langle Y, a\rangle$ est une variable aléatoire gaussienne, donc Y est un vecteur gaussien. L'expression de l'espérance et de la covariance est une conséquence du théorème 6.35. □

Corollaire 12.2. *Si $X = (X_1, \dots, X_d)$ est un vecteur gaussien, alors pour tout $I \subset \{1, \dots, d\}$, le vecteur $(X_i)_{i\in I}$ est gaussien.*

Démonstration. $(X_i)_{i\in I}$ est l'image de $X = (X_1, \dots, X_d)$ par l'application linéaire

$$\mathbb{R}^d \to \mathbb{R}^I, \quad (x_i)_{i\in\{1,\dots,d\}} \mapsto (x_i)_{i\in I}.$$

□

12.2 Exemple fondamental

Théorème 12.3. *Soient $X_1, \dots, X_d$ des variables aléatoires gaussiennes indépendantes. Alors $X = (X_1, \dots, X_d)$ est un vecteur gaussien.*

Démonstration. Soit $a \in \mathbb{R}^d$. Pour $k \in \{1, \dots, d\}$, on pose $S_k = \sum_{i=1}^{k} a_i X_i$. On montre par récurrence sur k que S_k est une variable aléatoire gaussienne. Pour $k = 1$, $S_1 = a_1 X_1$: quand on multiplie une variable gaussienne par une constante, on a toujours une variable aléatoire gaussienne. Supposons acquis que S_k est gaussienne. On a $S_{k+1} = S_k + a_{k+1}X_{k+1}$. Or $a_{k+1}X_{k+1}$ est une variable gaussienne indépendante de S_k, car S_k est $\sigma(X_1, \dots, X_k)$-mesurable. Or d'après le théorème 6.44, la somme de deux variables aléatoires gaussiennes indépendantes est une variable aléatoire gaussienne, donc S_{k+1} est une variable aléatoire gaussienne.

Comme $\langle X, a\rangle = S_d$ quel que soit a, on en déduit que X est un vecteur gaussien. □

12.3 Loi gaussienne

Théorème 12.4. *Soient C une matrice symétrique positive $d \times d$ et $m \in \mathbb{R}^d$. Alors, on peut construire un vecteur gaussien admettant m comme espérance et C comme matrice de covariance.*

Démonstration. Comme C est une matrice symétrique, on peut la diagonaliser avec une matrice de passage orthogonale. Comme elle est positive, les valeurs propres sont positives. Ainsi, on peut trouver une matrice O orthogonale et des réels positifs $\lambda_1, \dots, \lambda_d$ tels que

$$C = O \begin{pmatrix} \lambda_1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \lambda_2 & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \lambda_{d-1} & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \lambda_d \end{pmatrix} O^*.$$

Posons

$$A = O \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \sqrt{\lambda_{d-1}} & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \sqrt{\lambda_d} \end{pmatrix} O^*$$

et soient $(X_1, \dots, X_d)$ des variables aléatoires indépendantes suivant la loi $\mathcal{N}(0,1)$. L'espérance de X est $m_X = 0$ sa matrice de covariance $C_X = \mathrm{Id}_d$. D'après le théorème 12.3, X est un vecteur gaussien ; maintenant d'après le théorème 6.35, $Y = AX + m$ est un vecteur gaussien d'espérance $A.0 + m = m$ et de covariance $AC_X A^* = AA^* = C$. □

Théorème 12.5. *Soient X et Y deux vecteurs gaussiens ayant même espérance et même matrice de covariance. Alors X et Y ont même loi.*

Démonstration. Soit $a \in \mathbb{R}^d$. On pose $V = \langle X, a\rangle$ et $W = \langle Y, a\rangle$. L'espérance de V est $\langle m_X, a\rangle$ et la covariance de V est $\mathrm{Covar}(\langle X, a\rangle, \langle X, a\rangle) = \langle C_X a, a\rangle$. Comme X est gaussien, V est gaussienne, donc $V \sim \mathcal{N}(\langle m_X, a\rangle, \langle C_X a, a\rangle)$. De même $V \sim \mathcal{N}(\langle m_Y, a\rangle, \langle C_Y a, a\rangle)$. Comme $m_X = m_Y$ et $C_X = C_Y$, on en déduit que $V = \langle X, a\rangle$ et $W = \langle Y, a\rangle$ ont même loi. Comme cela est vrai quel que soit a, on déduit du théorème 9.8 que X et Y ont même loi. □

Notations. Soient $m \in \mathbb{R}^d$ et C une matrice $d \times d$ symétrique positive. On note $\mathcal{N}(m, C)$ la loi commune à tous les vecteurs gaussiens admettant m comme espérance et C comme matrice de covariance. La pertinence de cette définition est assurée par les théorèmes 12.4 et 12.5.

12.4 Loi gaussienne et indépendance

Théorème 12.6. *Soient $d_1, \ldots, d_n$ des entiers positifs de somme d. Soient $C_1, \ldots, C_n$ des matrices symétriques positives, et $m_1, \ldots, m_n$ des vecteurs, de tailles respectives $d_1, \ldots, d_n$. Alors*

$$\mathcal{N}(m_1, C_1) \otimes \mathcal{N}(m_2, C_2) \otimes \cdots \otimes \mathcal{N}(m_n, C_n) = \mathcal{N}(m, C),$$

avec

$$m = \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix} \quad \text{et} \quad C = \begin{pmatrix} C_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & C_n \end{pmatrix}.$$

Démonstration. Soient $Y_1, \ldots, Y_n$ des vecteurs gaussiens indépendants, avec $Y_k \sim \mathcal{N}(m_k, C_k)$. On va d'abord montrer que $X = (Y_1, \ldots, Y_n)$ est gaussien. Soit $a \in \mathbb{R}^n$. On peut écrire $a = (a_1, \ldots, a_n)$, avec a_k de taille d_k. On a

$$\langle X, a \rangle = \sum_{k=1}^{n} \langle Y_k, a_k \rangle.$$

Comme Y_k est gaussien, chaque variable aléatoire $\langle Y_k, a_k \rangle$ est une variable aléatoire gaussienne. Comme les Y_k sont indépendants, les variables aléatoires $\langle Y_k, a_k \rangle$ sont indépendantes. Or on sait qu'une somme de variables aléatoires gaussiennes indépendantes est une variable aléatoire gaussienne, donc $\langle X, a \rangle$ est une variable aléatoire gaussienne. Il s'ensuit que X est un vecteur gaussien. L'expression de l'espérance et de la matrice de covariance ne pose pas de problème, puisque des variables aléatoires indépendantes ne sont pas corrélées. □

Théorème 12.7. *Soient $d_1, \ldots, d_n$ des entiers positifs de somme d. On suppose que $X = (X_1, \ldots, X_n)$ est un vecteur gaussien dont la matrice de covariance est diagonale par blocs :*

$$C_X = \begin{pmatrix} C_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & C_n \end{pmatrix}.$$

Si l'on pose $Y_1 = (X_1, \ldots, X_{d_1})$, $Y_2 = (X_{d_1+1}, \ldots, X_{d_1+d_2})$ et $Y_n = (X_{d_1+d_2+\cdots+d_{n-1}+1}, \ldots, X_{d_1+d_2+\cdots+d_n})$, alors les vecteurs $Y_1, \ldots, Y_n$ sont des vecteurs gaussiens indépendants.

Démonstration. X a même espérance et même matrice de covariance que le vecteur aléatoire considéré au théorème précédent. Comme X et ce vecteur aléatoire sont tous deux gaussiens, ils ont tous deux la même loi. Ainsi, la loi de X n'est autre que $\mathcal{N}(m_1, C_1) \otimes \mathcal{N}(m_2, C_2) \otimes \cdots \otimes \mathcal{N}(m_n, C_n)$, ce qui signifie que les Y_i sont indépendants et que pour tout i, on a $Y_i \sim \mathcal{N}(m_i, C_i)$. □

En particulier, on a le corollaire suivant.

Corollaire 12.8. *Si le vecteur gaussien $X = (X_1, \dots, X_d)$ a une matrice de covariance dont tous les termes non-diagonaux sont nuls, alors $X_1, \dots, X_d$ sont des variables aléatoires indépendantes.*

12.5 Loi gaussienne à densité

Théorème 12.9. *Soient C une matrice symétrique définie positive et $m \in \mathbb{R}^d$. La loi (sur $\mathbb{R}^d$) $\mathcal{N}(m, C)$ admet comme densité par rapport à la mesure de Lebesgue la fonction*

$$f_{m,C}(y) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det C}} \exp\left(-\frac{1}{2} \langle C^{-1}(y-m), y-m \rangle \right).$$

Démonstration. On reprend les notations de la preuve du théorème 12.4. On a $X \sim \mathcal{N}(0, \mathrm{Id}_d)$ et $Y \sim \mathcal{N}(m, C)$. Comme C est définie positive, les λ_i sont ici strictement positifs, et donc que A est inversible. Comme X est composé de n variables aléatoires indépendantes à densité, la densité de X est le produit des densités, soit

$$f_X(x) = \prod_{k=1}^{d} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_k^2}{2} \right) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2} \langle x, x \rangle \right).$$

D'après le théorème 6.42, $Y = AX + m$ admet comme densité

$$\frac{1}{\det A} f_X(A^{-1}(y-m)) = \frac{1}{\det A} \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2} \langle A^{-1}(y-m), A^{-1}(y-m) \rangle \right).$$

Or on a

$$\begin{aligned}
\langle A^{-1}(y-m), A^{-1}(y-m) \rangle &= \langle (A^{-1})^* A^{-1}(y-m), y-m \rangle \\
&= \langle (A^{-1})^* A^{-1}(y-m), y-m \rangle \\
&= \langle (AA^*)^{-1}(y-m), y-m \rangle \\
&= \langle C^{-1}(y-m), y-m \rangle.
\end{aligned}$$

D'autre part, $\det C = \det AA^* = (\det A)^2$, donc $\det A = \sqrt{\det C}$. On en déduit que la densité de Y (c'est-à-dire de la loi de Y) est

$$f_{m,C}(y) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det C}} \exp\left(-\frac{1}{2} \langle C^{-1}(y-m), y-m \rangle \right).$$

□

Remarque 12.10. *On peut se demander ce qui se passe si C n'est pas inversible. Supposons que $X \sim \mathcal{N}(m, C)$, avec C de rang r. On peut diagonaliser C dans une base orthogonale et écrire C sous la forme*

$$C = O^* \mathit{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0) O,$$

avec O matrice orthogonale. Posons $Y = OX$. La matrice de covariance de Y est $OCO^* =$ diag$(\lambda_1, \dots, \lambda_r, 0, \dots, 0)$. Supposons que $m = 0$: dans ce cas les variables $Y_{r+1}, Y_{r+2}, \dots, Y_n$ sont identiquement nulles, ce qui signifie que Y est à valeurs dans Vect$(e_1, \dots, e_r)$, qui est un espace vectoriel de dimension r. Par suite, $X = O^*Y$ est à valeurs dans Vect$(O^*e_1, \dots, O^*e_r)$, qui est encore un espace vectoriel de dimension r. Dans le cas général, X prend ses valeurs dans l'espace affine de dimension r : $m +$ Vect$(O^*e_1, \dots, O^*e_r)$.

12.6 Fonction caractéristique des vecteurs gaussiens

Théorème 12.11. *En dimension quelconque, la fonction caractéristique de la loi normale $\mathcal{N}(m, C)$ est*

$$x \mapsto \exp(i\langle x, m\rangle) \exp\left(-\frac{1}{2}\langle Cx, x\rangle\right).$$

Démonstration. On a

$$\mathbb{E}\exp(i\langle X, t\rangle) = \mathbb{E}\exp(iY) = \phi_Y(1),$$

où $Y = \langle X, t\rangle$. Comme X est gaussien de covariance C et d'espérance m, Y est gaussien de covariance $\langle Ct, t\rangle$ et d'espérance $\langle t, m\rangle$. On déduit donc le résultat de la formule précédente et du théorème 9.19. □

12.7 Théorème central limite en dimension d

Théorème 12.12. *Soit $(X_n)_{n\geq 1}$ une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^d$, indépendants et identiquement distribués.*
On suppose que $\mathbb{E}\|X_1\|^2 < +\infty$. On note m l'espérance et C la matrice de covariance. Alors

$$\frac{(X_1 + \cdots + X_n) - nm}{\sqrt{n}} \Longrightarrow \mathcal{N}(0, C).$$

En fait, le théorème 12.12 peut être vu comme une conséquence du théorème 14.2. Afin de prouver ce résultat, on va énoncer un lemme comparable dans son esprit au theorème 9.8.

Lemme 12.13. *Soit $(X_n)_{n\geq 1}$ une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^d$ et X un vecteur aléatoire à valeurs dans $\mathbb{R}^d$. Si, pour tout $a \in \mathbb{R}^d$, $\langle X_n, a\rangle$ converge en loi vers $\langle X, a\rangle$, alors X_n converge en loi vers X.*

Démonstration. Soit $a \in \mathbb{R}^d$. On pose $Y_n = \langle X_n, a\rangle$. On a

$$\phi_{X_n}(a) = \mathbb{E}e^{i\langle X_n, a\rangle} = \mathbb{E}e^{iY_n} = \phi_{Y_n}(1).$$

Par hypothèse, Y_n converge en loi vers $\langle X, a\rangle$, donc $\phi_{Y_n}(1)$ converge vers $\phi_{\langle X,a\rangle}(1) = \mathbb{E}e^{i\langle X,a\rangle} = \phi_X(a)$. Donc $\phi_{X_n}(a)$ converge vers $\phi_X(a)$. Comme c'est vrai pour tout $a \in \mathbb{R}^d$, le théorème de Lévy implique que $(X_n)_{n\geq 1}$ converge en loi vers X. □

Démonstration du théorème 12.12. Soit X un vecteur aléatoire suivant la loi $\mathcal{N}(0, C)$. On pose

$$S_n = (X_1 + \cdots + X_n) - nm = \sum_{k=1}^{n}(X_k - m).$$

D'après le lemme précédent, il nous suffit de montrer que pour tout $a \in \mathbb{R}^d$ $\langle \frac{S_n}{\sqrt{n}}, a\rangle$ converge en loi vers $\langle X, a\rangle$.

Fixons $a \in \mathbb{R}^d$ et posons $Y_n = \langle X_n - m, a\rangle = a^*(X_n - m)$. Les Y_n sont des variables aléatoires indépendantes qui suivent toutes la même loi, la loi image de $\mathbb{P}_{X_1}$ par $x \mapsto \langle x - m, a\rangle = a^*(x - m)$. D'après le théorème 6.35, leur espérance est 0 et leur variance $a^*Ca = \langle Ca, a\rangle$. Ainsi, d'après le théorème 14.2, la suite $\frac{\sum_{k=1}^{n} Y_k}{\sqrt{n}}$ converge en loi vers $\mathcal{N}(0, \langle Ca, a\rangle)$. Mais on a

$$\frac{\sum_{k=1}^{n} Y_k}{\sqrt{n}} = \left\langle \frac{S_n}{\sqrt{n}}, a \right\rangle,$$

et, encore d'après le théorème 6.35, la loi de $\langle X, a\rangle$ est précisément la loi $\mathcal{N}(0, \langle Ca, a\rangle)$. Donc $\langle \frac{S_n}{\sqrt{n}}, a\rangle$ converge en loi vers $\langle X, a\rangle$, ce qui achève la démonstration. □

12.8 Convergence vers la loi du χ^2

12.8.1 Préliminaires

Définition. *Un n-échantillon $X_1, \ldots, X_n$ de loi μ est un ensemble de n vecteurs aléatoires indépendants identiquement distribués, suivant la loi μ.*

Proposition 12.14. *Soit X un vecteur gaussien dans $\mathbb{R}^n$ de loi $\mathcal{N}(0, Id_n)$. Alors la loi de $\|X\|_2^2$ est une loi à densité. On l'appelle loi du chi-deux à n degrés de liberté et on la note $\chi^2(n)$.*

Démonstration. Nous allons montrer un résultat plus fort par récurrence sur n. Notons f la densité de la loi gaussienne réelle centrée réduite. Nous allons prouver ici que la loi $\chi^2(n)$ admet la densité $f_\chi * f_\chi * \ldots * f_\chi = f_\chi^{*n}$, où $f_\chi(x) = \frac{1}{\sqrt{2\pi x}}e^{-x/2}\,\mathbb{1}_{]0,+\infty[}(x)$.

Si $n = 1$, alors $\|X\|_2^2 = X_1^2$, où X_1 est une variable gaussienne standard. On a alors pour tout $x > 0$, $\mathbb{P}(X_1^2 \leq x) = \mathbb{P}(|X_1| \leq \sqrt{x})$ et donc la densité de X_1^2 est

$$\begin{aligned} f_{X_1^2}(x) &= \left(\frac{1}{2\sqrt{x}} f(\sqrt{x}) + \frac{1}{2\sqrt{x}} f(-\sqrt{x})\right) \mathbb{1}_{]0,+\infty[}(x) \\ &= \frac{1}{\sqrt{2\pi x}} e^{-x/2} \mathbb{1}_{]0,+\infty[}(x). \end{aligned}$$

Notons $X = (X_1, \ldots, X_n)$. Les coordonnées de X sont toutes des gaussiennes centrées réduites indépendantes. Posons $Y = (X_1, \ldots, X_{n-1})$. Par hypothèse de récurrence, on sait que $\|Y\|_2^2$ suit la loi $\chi^2(n-1)$ qui admet la densité $f_\chi^{*(n-1)}$. De plus, on sait que X_n est indépendante de Y et $\|X\|_2^2 = \|Y\|_2^2 + X_n^2$. La loi de la somme de deux variables indépendantes de loi à densités est la loi dont la densité est donnée par le produit de convolution des densités, soit ici $f_{\|X\|_2^2} = f_{\|Y\|_2^2} * f_{X_n^2}$. Or $f_{\|X_n\|^2} = f_\chi$ et $f_{\|Y\|_2^2} = f_\chi^{*(n-1)}$, ce qui donne le résultat cherché. On conclut que la loi du χ^2 est bien une loi à densité. $\square$

Proposition 12.15. *On a* $\chi^2(n) = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$.

Démonstration. En effet, chaque variable X_k^2 suit la loi $\Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$. On a vu précédemment, au §6.8.3, que la somme de deux variables indépendantes de loi $\Gamma(a, \gamma)$ et $\Gamma(b, \gamma)$ est une variable de loi $\Gamma(a + b, \gamma)$. Une autre technique, que celle présentée précédemment, pour montrer ce résultat est la suivante. La transformée de Laplace de la variable Z de loi $\Gamma\left(a, \frac{1}{2}\right)$ est pour $u > 0$, en posant $y = (u + 1/2)x$

$$\begin{aligned} \phi_Z(u) &= \mathbb{E}e^{-uZ} = \frac{1}{2^a\Gamma(a)} \int_0^{+\infty} e^{-ux} x^{a-1} e^{-x/2}\, dx \\ &= \frac{1}{(1+2u)^a\Gamma(a)} \int_0^{+\infty} e^{-y} y^{a-1}\, dy = \frac{1}{(1+2u)^a}. \end{aligned}$$

Comme les variables X_k^2 sont toutes indépendantes, on trouve que la transformée de Laplace de $\|X\|_2^2$ est $\frac{1}{(1+2u)^{n/2}}$. La transformée de Laplace caractérise la loi et donc $\|X\|_2^2$ suit la loi $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$. $\square$

Théorème 12.16 (de Cochran). *Soit $X = (X_1, \ldots, X_n)$ un vecteur gaussien centré réduit. Pour F un sous-espace vectoriel de $\mathbb{R}^n$ de dimension p, on note p_F (respectivement $p_{F^\perp}$) la projection orthogonale sur F (respectivement $F^\perp$).*

Alors les vecteurs aléatoires $p_F X$ et $p_{F^\perp} X$ sont gaussiens indépendants, $p_F X$ suit la loi $\mathcal{N}(0, p_F)$ et $p_{F^\perp} X$ suit la loi $\mathcal{N}(0, p_{F^\perp})$. De plus, les variables aléatoires $\|p_F X\|_2^2$ et $\|p_{F^\perp} X\|_2^2$ sont indépendantes, $\|p_F X\|_2^2$ suit la loi $\chi^2(p)$ et $\|p_{F^\perp} X\|_2^2$ suit la loi $\chi^2(n-p)$.

Démonstration. Soient $f_1, \dots, f_p$ une base orthonormale de F et $f_{p+1}, \dots, f_n$ une base orthonormale de $F^\perp$. Pour tout i compris entre 1 et n, on pose $Y_i = \langle X, f_i \rangle$. Le vecteur Y est gaussien, en tant qu'image d'un vecteur gaussien par une transformation linéaire. On peut noter que $Y_1, \dots, Y_p$ sont les coordonnées de $p_F X$ dans la base $f_1, \dots, f_p$ tandis que $Y_{p+1}, \dots, Y_n$ sont les coordonnées de $p_{F^\perp} X$ dans la base $f_{p+1}, \dots, f_n$. L'application de $\mathbb{R}^n$ dans lui-même qui à x associe $(\langle x, f_1 \rangle, \dots, \langle x, f_n \rangle)$ est une isométrie. Sa matrice O est orthogonale et la matrice de covariance de Y est $O\mathrm{Id}_n O^* = \mathrm{Id}_n$: $Y_1, \dots, Y_n$ sont des variables indépendantes suivant la loi $\mathcal{N}(0,1)$. Si l'on note encore p_F la matrice de p_F dans la base canonique, la matrice de covariance de $p_F X$ est $p_F \mathrm{Id}_n p_F^* = p_F^2 = p_F$ car p_F est un projecteur orthogonal. Ainsi $p_F X \sim \mathcal{N}(0, p_F)$. De même $p_{F^\perp} X \sim \mathcal{N}(0, p_{F^\perp})$. L'indépendance de $p_F X$ et de $p_{F^\perp} X$ découle des identités

$$p_F X = \sum_{i=1}^{p} Y_i f_i \text{ et } p_{F^\perp} X = \sum_{i=p+1}^{n} Y_i f_i.$$

Comme les f_i sont orthonormés, on a aussi

$$\|p_F X\|_2^2 = \sum_{i=1}^{p} Y_i^2 \text{ et } \|p_{F^\perp} X\|_2^2 = \sum_{i=p+1}^{n} Y_i^2,$$

ce qui donne les lois du chi-deux annoncées. □

12.8.2 Le théorème de convergence

Nous allons ici présenter un théorème de convergence en loi qui prépare à un test statistique connu,le test du chi-deux, dont nous détaillerons l'usage dans le chapitre suivant.

On considère un n-échantillon $X_1, \dots, X_n$ de loi ν sur un ensemble fini E que l'on identifie à $\{1, \dots, q\}$.

On note N_j^n le nombre de variables aléatoires X_i parmi $X_1, \dots, X_n$ qui sont égales à j (on a donc $N_1^n + \dots + N_q^n = n$) et on pose

$$T_n = \sum_{j=1}^{q} \frac{1}{n\nu_j} (N_j^n - n\nu_j)^2.$$

L'idée du test du χ^2 d'adéquation est la suivante : comme $N_j^n = \sum_{i=1}^{n} \mathbb{1}_{\{X_i = j\}}$, on a alors convergence presque sûre de N_j^n/n vers ν_j par la loi des grands nombres. On peut donc espérer que T_n ne deviendra pas trop grand pour $n \to +\infty$ (on va voir en fait que T_n converge en loi vers une variable aléatoire finie).

Théorème 12.17. *Soit ν une loi sur $\{1,\dots,q\}$, $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes de loi ν. Alors les T_n convergent en loi vers une variable aléatoire de loi χ^2_{q-1} à $q-1$ degrés de liberté.*

Démonstration. Notons $(e_1,\dots,e_q)$ la base canonique de $\mathbb{R}^q$. On définit la variable q-dimensionnelle dont la $i-$ième composante vaut 1 si $X_j = i$ et 0 sinon. Autrement dit, on a $Y_j = e_{X_j}$. Bien sûr, les variables Y_j sont indépendantes et de même loi, avec $\mathbb{P}(Y_j = e_i) = \nu_i$. Posons $S_n = Y_1 + \dots + Y_n$. Soient m le vecteur espérance de Y_1 et K sa matrice de covariance. Par le théorème central limite, les variables $V_n = (S_n - nm)/\sqrt{n}$ convergent en loi vers un vecteur gaussien centré de covariance K. Si $V_{n,i}$ désigne la $i-$ième composante de V_n, comme $m = \sum\limits_{i=1}^{n} \nu_i e_i$, on a aisément

$$T_n = \sum_{i=1}^{q} \frac{V_{n,i}^2}{\nu_i}.$$

Donc, si $U = (U_1,\dots,U_q)$ est un vecteur gaussien centré de covariance K, alors les variables T_n convergent en loi vers

$$T = \sum_{i=1}^{q} \frac{U_i^2}{\nu_i}.$$

On peut noter que $T = \|\mathrm{diag}(\frac{1}{\sqrt{\nu_1}},\dots,\frac{1}{\sqrt{\nu_q}})U\|_2^2$. Calculons la matrice de covariance K. On a

$$\begin{aligned}
K_{i,j} &= \mathrm{Covar}(\langle Y_1, e_i\rangle, \langle Y_1, e_j\rangle) \\
&= \mathrm{Covar}(\mathbb{1}_{\{Z_1=i\}}, \mathbb{1}_{\{Z_1=j\}}) \\
&= \delta_{i,j}\nu_i - \nu_i\nu_j \\
&= \sqrt{\nu_i}(\delta_{i,j} - \sqrt{\nu_i}\sqrt{\nu_j})\sqrt{\nu_j}.
\end{aligned}$$

On en déduit que la matrice de covariance de $\mathrm{diag}\left(\frac{1}{\sqrt{\nu_1}},\dots,\frac{1}{\sqrt{\nu_q}}\right)U$ a comme matrice la matrice C donnée par $c_{i,j} = \delta_{i,j} - \sqrt{\nu_i}\sqrt{\nu_j}$. Or on remarque que

$$\langle (\mathrm{Id}_q - C)e_i, e_j\rangle = (\mathrm{Id}_q - C)_{i,j} = \sqrt{\nu_i}\sqrt{\nu_j} = \langle\langle v, e_i\rangle v, e_j\rangle,$$

où v est le vecteur unitaire $(\sqrt{\nu_1},\dots,\sqrt{\nu_q})$: $\mathrm{Id}_q - C$ est le projecteur orthogonal sur l'espace vectoriel engendré par v, et donc C le projecteur orthogonal sur $v^\perp$. Ainsi, d'après le thèorème de Cochran, $T = \|\mathrm{diag}(\frac{1}{\sqrt{\nu_1}},\dots,\frac{1}{\sqrt{\nu_q}})U\|_2^2$ suit la loi du χ^2 à $q-1$ degrés de libertés. □

12.9 Exercices sur les vecteurs gaussiens

12.9.1 Exercices corrigés

Exercice 264. *Loi uniforme sur la sphère.*
Soit X un vecteur aléatoire suivant la loi $\mathcal{N}(0, \mathrm{Id}_n)$. Pour $r > 0$, on note H_r^{n-1} la loi de $r\frac{X}{\|X\|_2}$. Soit O une matrice orthogonale. Montrer que H_r^{n-1} est invariante par O. On admettra dans la suite que H_r^{n-1} est l'unique mesure à support sur la sphère de rayon r et invariante par l'action du groupe orthogonal. $\rightarrow$ indication $\rightarrow$ solution

Exercice 265. *Lemme de Poincaré.*
Soit $S_a^{n-1} = \{x \in \mathbb{R}^n \ : ||x|| = a\}$ la sphère de rayon a plongée dans $\mathbb{R}^n$ où

$$||x|| = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

Soient μ_n la probabilité uniforme sur $S_{\sqrt{n}}^{n-1}$ (autrement dit $\mu_n = H_{\sqrt{n}}^{n-1}$), $\pi_n : \mathbb{R}^n \rightarrow \mathbb{R}$ la projection définie par $\pi_n(x) = x_1$ et γ_n la mesure image de μ_n par π_n. Ainsi, pour tout $a < b \in \mathbb{R}$,

$$\gamma_n([a,b]) = \mu_n(\pi_n^{-1}([a,b])) = \mu_n\left(\{x \in S_{\sqrt{n}}^{n-1} : a \leq x_1 \leq b\}\right).$$

À l'aide de l'exercice 264, démontrer le résultat suivant, appelé lemme de Poincaré. Pour toute fonction $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée,

$$\lim_{n\to+\infty} \int \Phi(x)\gamma_n(dx) = \int \Phi(x)\gamma(x)\ dx = \int \Phi(x)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\ dx.$$

$\rightarrow$ indication $\rightarrow$ solution

Exercice 266. Soit $(X,Y) \sim \mathcal{N}(0,\ \mathrm{Id}_2)$. Montrer que XY a même loi que la variable $\frac{1}{2}(X^2 - Y^2)$.
Indication : penser à la formule de polarisation

$$XY = \frac{1}{4}\left((X+Y)^2 - (X-Y)^2\right).$$

$\rightarrow$ indication $\rightarrow$ solution

Exercice 267. *Comparaisons stochastiques.*

1. Soient μ et ν deux lois sur $\mathbb{R}$. On dit que μ domine stochastiquement ν et on écrit $\mu \succeq \nu$ si pour tout fonction f mesurable positive croissante, on a $\mathbb{E}f(X) \leq \mathbb{E}f(Y)$. Montrer que les propriétés suivantes sont équivalentes :

(a) On a un couplage ordonné de μ et ν, c'est-à-dire qu'il existe un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ sur lequel vivent des variables aléatoires X et Y avec $\mathbb{P}_X = \mu$, $\mathbb{P}_Y = \nu$, et $X \geq Y$.

(b) $\mu \succeq \nu$.

(c) Pour tout t réel, $\mu(]t, +\infty[) \geq \nu(]t, +\infty[)$.

2. Montrer que si les mesures de probabilité $\mu_1, \mu_2, \nu_1, \nu_2$ vérifient $\mu_1 \succeq \mu_2$ et $\nu_1 \succeq \nu_2$, alors $\mu_1 * \mu_2 \succeq \nu_1 * \nu_2$.
3. Soient n et m des entiers naturels avec $m \leq n$, $0 \leq q \leq p \leq 1$. Montrer qu'on a
   — $\mathcal{B}(n,p) \succeq \mathcal{B}(m,p) \succeq \mathcal{B}(m,q)$
   — $\chi^2(n) \succeq \chi^2(m)$.
4. (a) Soient $(a_1, \dots, a_n)$ des réels et $X_1, \dots, X_n$ des variables gaussiennes centrées réduites indépendantes.

   Montrer que la loi de $\sum\limits_{i=1}^{n} (X_i + a_i)^2$ dépend uniquement de $\lambda = \sum\limits_{i=1}^{n} a_i^2$. On appellera cette loi la loi du χ^2 décentré et on la notera $\chi^2(n, \lambda)$.

   (b) Montrer que si $\lambda \geq \lambda'$, alors $\chi^2(n, \lambda) \succeq \chi^2(n, \lambda')$.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 268. La loi uniforme sur la sphère unité de dimension m est la loi de $\frac{X}{\|X\|_2}$, où $X \sim \mathcal{N}(0, I_m)$. Soit F un sous-espace de dimension n de $\mathbb{R}^m$, P_F la projection orthogonale sur F. Montrer que si M suit la loi uniforme sur la sphère de dimension m, alors $\|P_F(M)\|_2^2$ suit la loi Bêta de paramètres $(\frac{m}{2}, \frac{n-m}{2})$. $\rightarrow$ indication $\rightarrow$ solution

12.9.2 Exercices non corrigés

Exercice 269. Soient U et V deux variables aléatoires gaussiennes centrées et f une fonction croissante positive. Montrer que

$$\mathbb{E}[\|U\|^2] < \mathbb{E}[\|V\|^2] \Longrightarrow \mathbb{E}f(\|U\|) \leq \mathbb{E}f(\|V\|).$$

$\rightarrow$ indication

Exercice 270. Soient X et Y deux variables aléatoires indépendantes de loi $\mathcal{N}(0,1)$. On pose

$$U = \begin{pmatrix} 2X \\ 2X \end{pmatrix} \text{ et } V = \begin{pmatrix} 2X \\ \sqrt{5}Y \end{pmatrix}.$$

Montrer que U et V sont des vecteurs gaussiens centrés, puis que l'on a

$$\mathbb{E}\|U\|_2^2 < \mathbb{E}\|V\|_2^2,$$

tandis que
$$\mathbb{E}\|U\|_2^4 > \mathbb{E}\|V\|_2^4.$$
Comparer avec le résultat de l'exercice précédent. $\to$ indication

Exercice 271. Soit X un vecteur gaussien de matrice de covariance C. Montrer que
$$\mathbb{E}\|X\|_2^2 = \text{Tr } C.$$
$\to$ indication

Exercice 272.

1. Montrer qu'il existe un vecteur gaussien X dont la matrice de covariance est
$$C = \begin{pmatrix} 5 & 0 & 3 \\ 0 & 6 & 0 \\ 3 & 0 & 2 \end{pmatrix}.$$
2. Déterminer
$$\sup_{a\in\mathbb{R}^3} \frac{\mathbb{E}|\langle X, a\rangle|}{\|a\|_2}.$$

$\to$ indication

Exercice 273. Montrer qu'on peut trouver des variables aléatoires X_1, X_2, X_3 telles que pour tout $i \in \{1,2,3\}$, X_i suit la loi $\mathcal{N}(0,3)$ et
$$\mathrm{Covar}(X_1, X_2) = \mathrm{Covar}(X_1, X_3) = \mathrm{Covar}(X_3, X_2) = 1.$$
Calculer la variance de $X_1 + 2X_2 + 3X_3$. $\to$ indication

Exercice 274.

1. Soient X et Y deux variables aléatoires indépendantes suivant la loi $\mathcal{N}(0,1)$. Calculer la matrice de covariance du couple $(X, \alpha X + \beta Y)$.
2. Soit (Z,T) un vecteur gaussien centré, avec $\mathrm{Var}\, Z = 1$. Montrer qu'il existe α et β tels que la matrice de covariance de (Z,T) soit
$$\begin{pmatrix} 1 & \alpha \\ \alpha & \alpha^2+\beta^2 \end{pmatrix}.$$
En déduire que $\mathbb{E}[Z^2T^2] = 3\alpha^2 + \beta^2 = \mathrm{Var}\, T + 2\,\mathrm{Covar}(Z,T)^2$.
3. Soit (Z,T) un vecteur gaussien centré quelconque (on ne suppose plus que $\mathrm{Var}\, Z = 1$). Montrer que
$$\mathbb{E}[Z^2T^2] = \mathrm{Var}\, Z\, \mathrm{Var}\, T + 2\,\mathrm{Covar}(Z,T)^2.$$
4. Soit (Z,T) un vecteur gaussien de matrice de covariance
$$C = \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}.$$

Calculer $\mathbb{E}[Z^2T^2]$.
$\rightarrow$ indication

Exercice 275. Soit X un vecteur gaussien de matrice de covariance C. Montrer que

$$\mathbb{E}\|X\|_2^4 = 2\text{Tr }(C^*C) + (\text{Tr }C)^2.$$

$\rightarrow$ indication

Exercice 276. Soit X un vecteur gaussien dans $\mathbb{R}^n$ de loi $\mathcal{N}(0, C)$. On note $\lambda_1, \dots, \lambda_n$ les valeurs propres de C comptées avec leur ordre de multiplicité, ainsi que $\rho(C) = \max(|\lambda_i|; 1 \leq i \leq n\}$ le rayon spectral de C.
Soit α un réel vérifiant $\alpha < \rho(C)^{-1}$. Montrer que

$$\mathbb{E}\exp\left(\frac{\alpha}{2}\|X\|_2^2\right) = \prod_{i=1}^n (1 - \alpha\lambda_i)^{-1/2}.$$

$\rightarrow$ indication

Exercice 277. Soient X un vecteur gaussien dans $\mathbb{R}^n$ de loi $\mathcal{N}(0, \text{Id}_n)$ et O une matrice orthogonale. Montrer que $Y = OX$ a la même loi que X. $\rightarrow$ indication

Exercice 278. Soient A la matrice d'un projecteur orthogonal de $\mathbb{R}^n$ de rang r, $m \in \mathbb{R}^n$, X et Y deux variables aléatoires indépendantes suivant la loi $\mathcal{N}(m, A)$. Montrer que $\frac{\|X-Y\|_2^2}{2}$ suit la loi $\chi^2(r)$. $\rightarrow$ indication

Exercice 279. Soit X un vecteur gaussien de dimension d de loi $\mathcal{N}(0, K)$. Montrer que si le rang de K vaut d, alors $\langle K^{-1}X, X\rangle$ suit la loi $\chi^2(d)$. $\rightarrow$ indication

Chapitre 13

Statistique

En statistique, on observe un certain nombre de variables aléatoires, dont les lois sont (partiellement ou totalement) inconnues. À partir de ces observations, on cherche à obtenir le plus possible d'informations sur ces lois.

Par exemple, on fabrique des pièces sur une machine. Chaque pièce fabriquée est défectueuse avec une probabilité p inconnue. La valeur de p dépend du réglage de la machine et en particulier, plus p est proche de 0, meilleur est le réglage (mais il ne peut pas être parfait, bien entendu). Avant de lancer la fabrication, on veut s'assurer que la machine est "bien réglée" , c'est-à-dire que p est proche de 0 (même si p ne peut pas valoir 0). Pour ce faire, on fabrique un nombre n de pièces qui servent à tester le réglage. L'observation consiste à compter le nombre X de pièces défectueuses parmi les n pièces fabriquées. On se pose alors deux questions naturelles :

— trouver "la" valeur de p. Cela s'appelle estimer le paramètre p. Dans cet exemple, il est naturel de prendre comme estimateur la proportion de pièces défectueuses, soit $\hat{p}_n = X/n$,

— s'assurer que la vraie valeur de p ne dépasse pas un seuil critique p_0 fixé à l'avance (sinon, il faut refaire le réglage). On teste le fait que $p \leq p_0$.

Ces deux problèmes sont de nature mathématique différente. Leur point commun est qu'on ne peut pas arriver à une conclusion certaine. En effet, si n est assez grand, alors la valeur exacte de p sera proche de l'estimation $\hat{p}_n$, mais elle ne lui sera vraisemblablement pas égale. De même, on peut décider que $p \leq p_0$ si X/n est suffisamment petite, mais on ne sera jamais certain que la vraie valeur de p soit effectivement inférieure au seuil p_0.

On peut toujours représenter un problème de statistique de la manière suivante.

Définition. *On appelle modèle statistique la donnée de :*

1. *un espace d'états Ω (l'ensemble de tous les résultats possibles de l'expérience), que l'on munit de la tribu des événements $\mathcal{F}$,*
2. *une famille $(\mathbb{P}_\theta, \theta \in \Theta)$ de probabilités sur $(\Omega, \mathcal{F})$.*

Une variable aléatoire sur cet espace est appelée statistique.

On cherche toujours, à partir de la connaissance de l'observation (aléatoire) ω, à obtenir des renseignements sur la valeur inconnue (et non aléatoire) du paramètre θ :

1. soit on veut trouver la valeur de θ (ou d'une fonction de θ) et on parle alors d'estimation ponctuelle,
2. soit on veut savoir si θ se trouve dans une partie Θ_0 de l'ensemble Θ (ou dans son complémentaire) et on parle alors de test statistique.

Exemple:

1. Reprenons l'exemple de la fabrication de pièces. On observe le nombre X de pièces défectueuses, donc $\Omega = \{0, \dots, n\}$ avec la tribu $\mathcal{F} = \mathcal{P}(\Omega)$ et $X(\omega) = \omega$. L'ensemble des paramètres est $\Theta =]0, 1[$ et pour $\theta \in \Theta$, la probabilité $\mathbb{P}_\theta$ est la loi binomiale $\mathcal{B}(n, \theta)$.
2. On veut mesurer une longueur inconnue l et pour ce faire, on prend n mesures successives, dont les résultats sont $X_1, \dots, X_n$. Le modèle est constitué de $\Omega =]0, +\infty[^n$ muni de la tribu borélienne et des variables aléatoires $X_i(x_1, \dots, x_n) = x_i$ si $\omega = (x_1, \dots, x_n) \in \Omega$. Il est naturel de supposer que les X_i sont indépendantes, de même loi μ admettant pour moyenne la quantité l à mesurer. En terme de modèle statistique, Θ est l'ensemble de toutes les probabilités sur $]0, +\infty[$ et $\mathbb{P}_\theta$ est l'unique probabilité sur Ω pour laquelle les X_i sont indépendantes et de loi θ. L'espace Θ est donc très gros, mais on ne s'intéresse qu'à la fonction $f(\theta) = \int x \, d\theta(x)$, qui est la moyenne de la loi θ et supposée être égale à la quantité l cherchée.

On voit sur ces exemples deux sortes de modèles statistiques. Dans un cas, l'espace Θ est une partie de $\mathbb{R}$ (ou $\mathbb{R}^d$) : on a un problème paramétrique. Dans l'autre cas, Θ est l'espace de toutes les probabilités sur un ensemble donné : on a un problème non-paramétrique. Ici, on ne considérera que des problèmes paramétriques.

Les problèmes statistiques se posent très souvent dans le cadre des échantillons.

Définition. *Un modèle statistique $(\Omega, \mathcal{F}, (\mathbb{P}_\theta, \theta \in \Theta))$ étant fixé, on appelle échantillon une famille $(X_i)_{i \in I}$ de variables aléatoires à valeurs dans l'espace E (en général, $\mathbb{Z}, \mathbb{R}$ ou $\mathbb{R}^d$) telle que pour tout $\theta \in \Theta$, les variables aléatoires*

$(X_i)_{i\in I}$ forment une famille de variables aléatoires indépendantes identiquement distribuées. En particulier, on parlera de n-échantillon un ensemble $(X_1,\ldots,X_n)$ de n variables aléatoires sur $(\Omega,\mathcal{F})$, telles que pour tout $\theta\in\Theta$, les variables aléatoires $X_1,\ldots,X_n$ forment une famille de variables aléatoires indépendantes identiquement distribuées.

Remarque 13.1. *L'exemple de mesure de longueur précédent est un exemple de modèle basé sur un n-échantillon. Dans le cas des pièces défectueuses, il s'agit d'un 1-échantillon. On peut cependant le voir comme un modèle à n échantillons en notant X_i la variable aléatoire de Bernoulli qui vaut 1 si la pièce i est défectueuse. Dans cette version du modèle, on a $\Omega=\{0,1\}^n$ et aussi pour $\omega=(i_1,\ldots,i_n)\in\Omega$ et $X(\omega)=S_n(\omega)=X_1(\omega)+\cdots+X_n(\omega)=i_1+\cdots+i_n$, on a $\mathbb{P}_\theta(\omega)=\theta^{X(\omega)}(1-\theta)^{n-X(\omega)}$ pour $\theta\in[0,1]$.*

13.1 Estimateurs

On suppose donné le modèle statistique $(\Omega,\mathcal{F},(\mathbb{P}_\theta)_{\theta\in\Theta})$. Soit f une fonction connue sur Θ, supposée à valeurs réelles pour simplifier, $f:\Theta\to\mathbb{R}$. On veut estimer la quantité inconnue $f(\theta)$.

Estimer $f(\theta)$ signifie qu'au vu de l'observation ω, on "décide" que la valeur $f(\theta)$ vaut un certain nombre, noté $T(\omega)$ qui dépend de ω. On choisit donc une variable aléatoire réelle, appelée statistique, T. Dans le cadre de l'estimation, T est appelée estimateur de $f(\theta)$.

13.1.1 Lois empiriques

Si $X_1,\ldots,X_n$ est un n-échantillon, on peut lui associer une mesure de probabilité appelée loi empirique ou distribution empirique : il s'agit de

$$\mu_n=\frac{1}{n}\sum_{i=1}^n\delta_{X_i}.$$

Des estimateurs très classiques sont associés à la distribution empirique :

1. La fonction de répartition empirique est la fonction de répartition associée à la loi μ_n : on a
$$F_n(x)=\mu_n(]-\infty,x])=\frac{1}{n}\sum_{i=1}^n\mathbb{1}_{\{X_i\le x\}}.$$
C'est un estimateur de la fonction de répartition.
2. la moyenne empirique, qui est la moyenne de la distribution empirique

μ_n, est un estimateur de la moyenne

$$\bar{X}_n = \int_{\mathbb{R}} x \, d\mu_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

3. la variance empirique, qui est la variance de la distribution empirique μ_n, est un estimateur de la variance

$$S_n^2 = \int_{\mathbb{R}} (x - \bar{X}_n)^2 \, d\mu_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

Théorème 13.2. *Soit $X_1, \dots, X_n, \dots$ un échantillon infini d'une loi de carré intégrable. On a alors*

1. *la fonction de répartition empirique $F_n(x)$ converge (lorsque $n \to +\infty$) presque sûrement vers $F(x)$,*
2. *la moyenne empirique $\bar{X}_n$ converge presque sûrement vers $\mathbb{E}X_1$,*
3. *la variance empirique S_n^2 converge presque sûrement vers Var(X_1).*

Démonstration. 1. $F_n(x)$ est la somme de variables aléatoires indépendantes suivant la même loi de Bernoulli dont le paramètre est l'espérance $\mathbb{E}(\mathbb{1}_{\{X_1 \leq x\}}) = \mathbb{P}(X_1 \leq x) = F(x)$. De plus, sa variance vaut

$$\mathrm{Var}(\mathbb{1}_{\{X_1 \leq x\}}) = \mathbb{E}\mathbb{1}_{\{X_1 \leq x\}}^2 - \mathbb{P}(X_1 \leq x)^2 = F(x) - F(x)^2.$$

Ainsi, on voit que $\mathbb{1}_{\{X_1 \leq x\}}$ suit une loi de Bernoulli de paramètre $F(x)$. D'après la loi forte des grands nombres, $F_n(x)$ converge presque sûrement vers la moyenne de $\mathbb{1}_{\{X_1 \leq x\}}$, qui n'est autre que $F(x)$.

2. Le résultat découle directement de la loi forte des grands nombres.

3. Remarquons que $S_n^2 = \left(\frac{1}{n}\sum\limits_{i=1}^{n} X_i^2\right) - \bar{X}_n^2$. Le premier terme converge presque sûrement vers $\mathrm{Var}(X_1) + (\mathbb{E}X_1)^2$ tandis que le second tend presque sûrement vers $(\mathbb{E}X_1)^2$. On en déduit que S_n^2 converge presque sûrement vers $\mathrm{Var}(X_1) + (\mathbb{E}X_1)^2 - (\mathbb{E}X_1)^2 = \mathrm{Var}(X_1)$.

□

Définition. *L'estimateur T_n, construit à partir d'un $n-$échantillon, de $f(\theta)$ est consistant (ou convergeant) si T_n converge en probabilité vers $f(\theta)$ pour tout $\theta \in \Theta$. Il est dit fortement consistant si T_n converge $\mathbb{P}_\theta$-presque sûrement vers $f(\theta)$ pour tout $\theta \in \Theta$.*

Par le théorème 13.2, on voit que les estimateurs $F_n(x), \bar{X}_n$ et S_n^2 sont fortement consistants.

Exemple: Supposons qu'on observe un n-échantillon de la loi gaussienne $\mathcal{N}(\theta, 1)$. Un estimateur raisonnable de θ est la moyenne empirique

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n).$$

13.1.2 Théorème de Glivenko–Cantelli

Théorème 13.3 (Glivenko–Cantelli)**.** *Soit $(X_1, \ldots, X_n)$ un n–échantillon de la loi μ, de fonction de répartition F, et F_n la fonction de répartition empirique associée :*

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{]-\infty,x]}(X_k).$$

Alors,

$$\lim_{n \to +\infty} \|F_n - F\|_\infty = 0 \ \mathbb{P} - p.s.$$

Démonstration. Commençons d'abord par expliquer comment on peut, en pratique, calculer la quantité $\|F_n - F\|_\infty$. Cela assurera en particulier la mesurabilité de $\|F_n - F\|_\infty$.

Réordonnons les nombres $X_1, \ldots, X_n$ en $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$ de telle sorte que $X^{(1)} \leq X^{(2)} \leq \cdots \leq X^{(n)}$ (c'est ce que l'on appelle une statistique d'ordre). Les fonctions ainsi définies sont bien des variables aléatoires puisque l'on a l'identité

$$\{X^{(k)} \leq t\} = \left\{ \sum_{i=1}^{n} \mathbb{1}_{[0,t]}(X_i) \geq k \right\}.$$

Pour $0 \leq k \leq n$, la fonction $F_n - F$ vaut $\frac{k}{n} - F(t)$ sur $[X^{(k)}, X^{(k+1)}[$, avec la convention $X^{(0)} = -\infty$ et $X^{(n+1)} = +\infty$. Comme $F_n - F$ est décroissante et continue à droite sur $[X^{(k)}, X^{(k+1)}[$, on a

$$\sup_{t \in [X^{(k)}, X^{(k+1)}]} \left| \frac{k}{n} - F(t) \right| = \max \left(\left| \frac{k}{n} - F(X^{(k)}) \right|, \left| \frac{k}{n} - F(X^{(k+1)} - 0) \right| \right).$$

Ici, $F(x - 0)$ désigne la limite de F en x à gauche. Comme F_n et F ont les mêmes limites en $-\infty$ et en $+\infty$, on a simplement

$$\|F_n - F\|_\infty = \max \left(\max_{1 \leq k \leq n} \left| \frac{k}{n} - F(X^{(k)}) \right|, \max_{1 \leq k \leq n} \left| \frac{k-1}{n} - F(X^{(k)} - 0) \right| \right),$$

ce qui montre que $\|F_n - F\|_\infty$ est bien mesurable. On peut aller plus loin : le même raisonnement que ci-dessus montre que l'application

$$\psi_n : (x_1, \ldots, x_n) \mapsto \sup_{x \in \mathbb{R}} \left| F(x) - \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{]-\infty,x]}(x_k) \right|.$$

est $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) - (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable et on a $\|F_n - F\|_\infty = \psi_n(X_1, \dots, X_n)$. Ainsi $\mathbb{P}(\|F_n - F\|_\infty \to 0) = \mathbb{P}_X(\psi_n(\Pi_1, \dots, \Pi_n) \to 0)$: ainsi le résultat recherché ($\mathbb{P}(\|F_n - F\|_\infty \to 0) = 1$) est une propriété de la loi du processus $(X_n)_{n\geq 1}$: pour montrer que le théorème est vrai, on peut donc choisir les X_n sur l'espace de notre choix, pourvu qu'elles forment une suite de variables aléatoires identiquement distribuées de loi μ.

Soit donc $(U_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. D'après le théorème 5.36, les variables aléatoires $X_k = Q^*(U_k)$, avec $Q^*(u) = \inf\{x \in \mathbb{R}; 1 - F(x) \leq u\}$ forment un échantillon de la loi μ. On a

$$\begin{aligned} \|F_n - F\|_\infty &= \sup_{x\in\mathbb{R}} \left|\frac{1}{n}\sum_{k=1}^n \mathbb{1}_{]-\infty,x]}(X_k) - F(x)\right| \\ &= \sup_{x\in\mathbb{R}} \left|\frac{1}{n}\sum_{k=1}^n \mathbb{1}_{]-\infty,x]}(Q^*(U_k)) - F(x)\right| \\ &= \sup_{x\in\mathbb{R}} \left|\frac{1}{n}\sum_{k=1}^n \mathbb{1}_{]-\infty,F(x)]}(U_k) - F(x)\right| \\ &\leq \sup_{y\in[0,1]} \left|\frac{1}{n}\sum_{k=1}^n \mathbb{1}_{]-\infty,y]}(U_k) - y\right|. \end{aligned}$$

Notons que la dernière inégalité est en réalité une égalité lorsque F est continue. Ce résultat sera réutilisé plus tard.

Ainsi on est ramené à étudier le cas où μ est la loi uniforme sur $[0,1]$, puisque $\tilde{F}_n(y) = \frac{1}{n}\sum_{k=1}^n \mathbb{1}_{]-\infty,y]}(U_k)$ est la fonction de répartition empirique associée à l'échantillon $U_1, \dots, U_n$. Notons $D = [0,1] \cap \mathbb{Q}$. Par la loi forte des grands nombres, $\mathbb{P}(\tilde{F}_n(q) \to q) = 1$ pour $q \in D$. Par intersection dénombrable, l'événement $\tilde{\Omega} = \bigcap_{q\in D} \{\tilde{F}_n(q) \to q\}$ est encore de probabilité un. On reconnait ici les conditions d'applications du théorème 1.27 de Dini–Polyà dans sa version étendue : la convergence aux points rationnels d'une suite de fonctions croissantes de $[0,1]$ dans $\mathbb{R}$ vers une fonction continue croissante sur $[0,1]$ entraîne la convergence uniforme, ce qui achève la preuve.

□

Remarque 13.4. *Si on veut utiliser la convergence pour un seul x ou pour un ensemble dénombrable de valeurs de x, il n'est pas nécessaire d'invoquer Glivenko-Cantelli : la loi forte des grands nombres suffit.*

Remarque 13.5. *L'identité*

$$\{X^{(k)} \leq t\} = \left\{\sum_{i=1}^n \mathbb{1}_{[0,t]}(X_i) \geq k\right\},$$

établie au cours de la preuve, permet également de calculer la loi de la statistique d'ordre : comme les événements $\{X_i \leq t\}$ *sont indépendants et de même probabilité* $F(t)$*, on a*

$$\begin{aligned}\mathbb{P}(X^{(k)} \leq t) &= \mathcal{B}(n, F(t))([k, +\infty[) \\ &= \sum_{i=k}^{n} \binom{n}{i} F(t)^i (1 - F(t))^{n-i}.\end{aligned}$$

Dans le cas où les X_k *suivent la loi uniforme sur* $[0, 1]$*, il n'est pas très difficile, en dérivant la fonction de répartition, de vérifier que* $X^{(k)}$ *suit la loi Bêta de paramètres* k *et* $n + 1 - k$ *(exercice laissé au lecteur).*

13.1.3 Choix d'un estimateur

Le problème de l'estimation consiste à optimiser le choix de l'estimateur (qui est supposé être réel ici). Il faut ainsi introduire un critère de qualité. *A priori*, on a envie de dire que l'estimateur S est meilleur que T si l'erreur commise par S est plus petite en valeur absolue que celle commise par T. Remarquons que l'erreur est $T(\omega) - f(\theta)$: elle dépend à la fois du paramètre inconnu θ et du résultat ω de l'expérience, connu mais aléatoire. Ainsi, deux estimateurs S et T ne seront pratiquement jamais comparables, au sens où par exemple $|S(\omega) - f(\theta)| \leq |T(\omega) - f(\theta)|$ pour tous les ω et θ.

L'idée sous-jacente aux divers critères de qualité possibles consiste à choisir ce qu'on appelle parfois une fonction de perte. Il s'agit d'une fonction $h : \mathbb{R}_+ \to \mathbb{R}_+$ croissante et nulle en 0. La perte de l'estimateur T est alors $h(|T(\omega) - f(\theta)|)$ et le risque associé est l'espérance de cette fonction par rapport à $\mathbb{P}_\theta$: $R_T(\theta) = \mathbb{E}_\theta h(|T(\omega) - f(\theta)|)$. Le risque est une fonction de θ, mais qui ne dépend plus de l'aléa ω. Un estimateur S est dit meilleur que T si leurs fonctions de risque respectives satisfont $R_S(\theta) \leq R_T(\theta)$ pour tout θ.

Remarque 13.6. *Pour* h *fixé, dire que* S *est meilleur que* T *revient à dire qu'en moyenne, si on répète souvent l'expérience statistique, la fonction de perte de* S *sera plus petite que celle de* T*. Ce n'est bien entendu pas le cas pour une expérience donnée. D'autre part, le choix de* h *est arbiraire. Si on prend par exemple une fonction puissance* $h(x) = x^\alpha$*, plus* α *est grand et plus on privilégie les "grandes" erreurs par rapport aux "petites". On utilise en général la fonction* $h(x) = x^2$*.*

Définition. *1. Le risque quadratique de l'estimateur* T *de* $f(\theta)$ *est*

$$R_T(\theta) = \mathbb{E}_\theta[(T - f(\theta))^2]$$

où l'espérance $\mathbb{E}_\theta$ *dépend de la probabilité* $\mathbb{P}_\theta$*.*

2. Si S *et* T *sont deux estimateurs de* $f(\theta)$*, on dit que* S *est meilleur que* T *(au sens du risque quadratique) si* $R_S(\theta) \leq R_T(\theta)$ *pour tout* $\theta \in \Theta$*. Il est* *strictement meilleur s'il est meilleur et si de plus $R_S(\theta) < R_T(\theta)$ pour au moins une valeur de θ.*

Exemple:

1. Revenons à la fabrication de pièces. Dans la première modélisation où $\Omega = \{0, \ldots, n\}$ et $\mathbb{P}_\theta = \mathcal{B}(n, \theta)$, la variable aléatoire $T(\omega) = \omega/n$ est le meilleur estimateur de θ. Le carré T^2 est un estimateur raisonnable de θ^2, mais ce n'est pas le meilleur.
2. On observe un n-échantillon de la loi gaussienne $\mathcal{N}(\theta, 1)$. On a vu qu'un estimateur raisonnable de θ est la moyenne empirique
$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n).$$
Comme $f(\theta) = \theta$ est la moyenne $\mathbb{E}_\theta(\bar{X})$, le risque $R_{\bar{X}}(\theta)$ est la variance de $\bar{X}$ sous $\mathbb{P}_\theta$. On a donc $R_{\bar{X}}(\theta) = 1/n$. On peut aussi considérer X_1 comme un estimateur de θ, de moyenne θ et de risque quadratique $R_{X_1}(\theta) = 1$. Donc $\bar{X}$ est strictement meilleur que X_1 dès que $n \geq 2$.

Remarque 13.7. *La relation "S est meilleur que T" est une relation d'ordre partiel sur la famille de tous les estimateurs, qui est la famille de toutes les variables aléatoires. Deux estimateurs donnés ne sont en général pas comparables et il n'existe pas d'estimateur meilleur que tous les autres. En effet, dans l'exemple précédent de la gaussienne, si $T(\omega) = c$ constante arbitraire, alors le risque quadratique pour estimer $f(\theta)$ est $R_T(\theta) = (c-\theta)^2$. On a donc $R_T(\theta) < R_{\bar{X}}(\theta)$ pour certaines valeurs de θ et $R_T(\theta) > R_{\bar{X}}(\theta)$ pour d'autres valeurs de θ. Dans ce cas, T n'est pas un estimateur raisonnable de θ, car il ne dépend pas de l'observation, mais son risque est nul quand le paramètre inconnu θ vaut c.*

La détermination d'un meilleur estimateur (c'est-à-dire tel qu'il n'en existe pas de strictement meilleur) est un problème mathématique extrêmement difficile, car la classe de tous les estimateurs est trop vaste. Dans la plupart des cas, on se retreint à une classe particulière d'estimateurs.

Définition. *1. L'estimateur T de $f(\theta)$ est dit sans biais (ou non-biaisé) si on a $\mathbb{E}_\theta(T) = f(\theta)$ pour tout $\theta \in \Theta$. Dans ce cas, le risque quadratique $R_T(\theta)$ est la variance de T sous $\mathbb{P}_\theta$.*

2. T est un meilleur estimateur sans biais de $f(\theta)$ s'il est sans biais et s'il est meilleur que tout autre estimateur sans biais.

3. Soient S et T deux estimateurs, construits à partir d'un $n-$échantillon, sans biais du paramètre θ. On dit que S est un estimateur préférable à T si $\mathrm{Var}(S) < \mathrm{Var}(T)$.

Exemple: Soient $X_1, X_2, \ldots, X_n$ un n-échantillon d'une loi inconnue admettant un moment d'ordre deux.

Notre but est d'estimer m et σ^2 à partir de $X_1, X_2, \dots, X_n$.

Une idée naturelle est d'approcher m par la moyenne empirique

$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

C'est une bonne idée ! En effet

$$\mathbb{E}\bar{X}_n = \mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{n} X_k\right] = \frac{1}{n}\mathbb{E}\left[\sum_{k=1}^{n} X_k\right] = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}X_k = \frac{1}{n}\sum_{k=1}^{n} m = \frac{1}{n}nm = m.$$

$\bar{X}_n$ est donc un estimateur non biaisé de m.

Maintenant, comment estimer σ^2 ? Une autre idée naturelle est d'approcher σ^2 par la variance empirique

$$\sigma_n^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X}_n)^2 = \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)^2.$$

On a d'une part,

$$\mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{n} X_k^2\right] = \mathbb{E}[X_1^2] = \operatorname{Var} X_1 + (\mathbb{E}X_1)^2 = \sigma^2 + m^2,$$

et d'autre part,

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)^2\right] = \frac{1}{n^2}\mathbb{E}\left[\left(\sum_{k=1}^{n} X_k\right)^2\right] = \frac{1}{n^2}\mathbb{E}\left(\sum_{k=1}^{n} X_k^2 + 2\sum_{1\le i<j\le n} X_i X_j\right).$$

Pour tout k, on a $\mathbb{E}X_k^2 = \sigma^2 + m^2$ et, comme X_i et X_j sont indépendantes pour $i \neq j$, on a $\mathbb{E}X_iX_j = \mathbb{E}X_i\mathbb{E}X_j = m^2$. On en déduit

$$\mathbb{E}\left[\left(\sum_{k=1}^{n} X_k\right)^2\right] = \sum_{k=1}^{n}\mathbb{E}\left[X_k^2\right] + 2\sum_{1\le i<j\le n}\mathbb{E}X_iX_j = n(\sigma^2 + m^2) + n(n-1)m^2.$$

Ainsi $\mathbb{E}(\sigma_n^2) = \sigma^2 + m^2 - \frac{1}{n^2}\left(n(\sigma^2 + m^2) + n(n-1)m^2\right) = \frac{n-1}{n}\sigma^2$. Ce n'est pas tout à fait ce que l'on voulait, car on avait espéré trouver σ^2. Ce n'est pas grave : il suffit de considérer

$$s_n^2 = \frac{n}{n-1}\sigma_n^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X}_n)^2\right) = \left(\frac{1}{n-1}\sum_{k=1}^{n} X_k^2\right) - \frac{n}{n-1}\bar{X}_n^2.$$

s_n^2 est un estimateur non biaisé de σ car $\mathbb{E}s_n^2 = \frac{n}{n-1}\mathbb{E}\sigma_n^2 = \sigma^2$.

Remarque 13.8. *Il n'existe pas de justification satisfaisante à l'usage d'estimateurs sans biais, si ce n'est la méthode des moments qui est assez efficace (et repose sur cette notion). Néanmoins, quand on observe un n-échantillon $(X_1, \dots, X_n)$ de loi μ_θ pour estimer $f(\theta)$, on suppose qu'il existe une fonction g*

telle que $\mathbb{E}_\theta g(X_i) = f(\theta)$ *pour tout* $\theta \in \Theta$. *Alors,* $T_n = \frac{1}{n}(g(X_1)+\cdots+g(X_n))$ *est un estimateur sans biais de* $f(\theta)$. *D'après la loi des grands nombres,* T_n *converge* $\mathbb{P}_\theta$*-p.s. vers* $f(\theta)$. *Plus généralement, on peut montrer que sous des conditions assez faibles, les estimateurs sans biais construits sur un* n*-échantillon convergent vers la valeur à estimer (mais dans la pratique,* n *est bien entendu fini).*

Exemple:

1. Dans le cas des pièces défectueuses, $T = \frac{X}{n}$ est un estimateur sans biais de $\theta = p$.
2. On observe une variable aléatoire de loi exponentielle de paramètre θ. On a alors $\Omega = \mathbb{R}_+$, $\Theta =]0, \infty[$ et $\mathbb{P}_\theta$ est la loi exponentielle de paramètre θ. On observe la variable aléatoire $X(\omega) = \omega$. Si T est un estimateur sans biais de θ, on a alors (par définition du biais) $\mathbb{E}_\theta T(X) = \theta$. Par le théorème de transfert, cela revient à satisfaire
$$\int_0^\infty e^{-\theta x} T(x)\, dx = 1 \quad \forall \theta > 0,$$
ce qui n'est possible pour aucune fonction T ne dépendant pas de θ : en effet, d'après le théorème de convergence dominée, on a
$$\lim_{\theta \to +\infty} \int_0^\infty e^{-\theta x} T(x)\, dx = 0.$$
Ainsi, dans ce cas, il n'existe aucun estimateur sans biais.

Proposition 13.9. *Soit* $(X_1, \ldots, X_n)$ *un* n*-échantillon d'une loi admettant une variance. Alors la moyenne empirique* $\bar{X}_n$ *est le meilleur (au sens du risque quadratique) estimateur de la moyenne parmi tous les estimateurs linéaires sans biais.*

Démonstration. On a un n-échantillon $(X_1, \ldots, X_n)$ de loi μ_θ sur $\mathbb{R}$. Notons $\mathbb{P}_\theta$ l'unique probabilité sous laquelle les X_i sont indépendantes de loi μ_θ. Notons m_θ la moyenne et σ_θ^2 la variance de μ_θ. Un estimateur affine T est de la forme $T = b + \sum a_i X_i$, avec b et a_i constantes. On veut estimer $f(\theta) = m_\theta$. Comme $\mathbb{E}_\theta T = b + \sum a_i f(\theta)$, T est sans biais si et seulement si $b = 0$ et $\sum a_i = 1$. Le risque quadratique est donc $R_T(\theta) = \sigma_\theta^2 \sum a_i^2$. Considérons la mesure de probabilité $\mu = \frac{1}{n} \sum_{i=1}^n \delta_{a_i}$. Appliquons maintenant l'inégalité de Cauchy-Schwarz à la fonction identité. On obtient alors, car $\sum_{i=1}^n a_i = 1$
$$\frac{1}{n^2} = \left(\frac{1}{n}\sum_{i=1}^n a_i\right)^2 = \left(\int_{\mathbb{R}} x\, d\mu(x)\right)^2 \leq \int_{\mathbb{R}} x^2\, d\mu(x) = \frac{1}{n}\sum_{i=1}^n a_i^2.$$

On a donc $\sum_{i=1}^{n} a_i^2 \geq \frac{1}{n}$ avec égalité si et seulement si $a_i = 1/n$ et dans ce cas, $T = \bar{X}_n$. □

13.2 Intervalle de confiance

On se place ici en dimension 1. Comme nous l'avons déjà vu, l'erreur $T(\omega) - f(\theta)$ commise en remplaçant $f(\theta)$ par $T(\omega)$ est à la fois aléatoire et dépendante du paramètre inconnu θ. Le risque quadratique est une mesure "déterministe" du carré de cette erreur, mais il dépend encore de la valeur inconnue θ. On utilise donc souvent une "fourchette d'estimation".

Définition. *Soient T un estimateur de $f(\theta)$ et $0 < \alpha < 1$ (ce nombre est fixé a priori, proche de 1). On appelle intervalle de confiance de niveau α un intervalle aléatoire $I(\omega)$ dont les extrémités ne dépendent pas de θ et dans lequel "$f(\theta)$ se trouve avec une probabilité au moins égale à α" :*

$$\mathbb{P}_\theta(\theta \in I) \geq \alpha, \quad \forall \theta \in \Theta.$$

On parle alors d'intervalle de confiance bilatéral.

Dans le cas où l'intervalle de confiance est semi-infini, on parle alors d'intervalle de confiance unilatéral.

Exemple:

1. Pièces défectueuses. Sous $\mathbb{P}_\theta$, la variable $nT = nT_n$ suit la loi $\mathcal{B}(n, \theta)$, donc la variance de T est $\theta(1-\theta)/n$. Par l'inégalité de Chebychev , on a
$$\mathbb{P}_\theta(|T - \theta| > a) \leq \frac{\theta(1-\theta)}{na^2} \leq \frac{1}{4na^2} \quad \forall \theta \in \Theta.$$
Donc un intervalle de confiance de niveau $0,95$ pour θ est donné par $\left[T - \frac{1}{\sqrt{0,2n}}; T + \frac{1}{\sqrt{0,2n}}\right]$. Mais cette inégalité est en réalité une approximation grossière. En fait, si n est grand, alors le TCL nous dit que la variable $(T - \theta)\sqrt{n/\theta(1-\theta)}$ suit approximativement la loi $\mathcal{N}(0, 1)$ sous $\mathbb{P}_\theta$. Ainsi, pour tout $a > 0$ et tout $\theta \in [0, 1]$, on a
$$\mathbb{P}_\theta\left(\left|\frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}}\right| \geq a\right) = \mathbb{P}_\theta(\theta \in [T_n - \frac{a\sqrt{\theta(1-\theta)}}{\sqrt{n}}, T_n - \frac{a\sqrt{\theta(1-\theta)}}{\sqrt{n}}])$$
$$\geq \mathbb{P}_\theta(\theta \in [T_n - \frac{a}{2\sqrt{n}}\sqrt{n}, T_n - \frac{a}{2\sqrt{n}}])$$
Pour $n \geq 30$, le membre de droite se laisse approcher par $\mathbb{P}(|X| \geq a)$ avec $X \sim \mathcal{N}(0, 1)$. Les tables nous donnent $\mathbb{P}(X \leq 1,96) = 0,975$, soit $\mathbb{P}(|X| \geq 1,96) = 0,05$.

Un intervalle de confiance de niveau $0,95$ est donné par

$$[T - 0,98/\sqrt{n}, T + 0,98/\sqrt{n}], \tag{13.1}$$

qui est plus petit donc meilleur que le précédent lorsque n est grand. Pour simplifier les calculs et la mémorisation, on lui préfère parfois l'intervalle de confiance plus grand

$$[T - \frac{1}{\sqrt{n}}, T + \frac{1}{\sqrt{n}}]. \tag{13.2}$$

2. Soit un n-échantillon de loi $\mathcal{N}(\theta, \sigma^2)$ avec σ^2 connu. La loi de $\bar{X}_n - \theta$ sous $\mathbb{P}_\theta$ est la loi $\mathcal{N}(0, \sigma^2/n)$, donc $\sqrt{n}(\bar{X}_n - \theta)/\sigma$ suit la loi $\mathcal{N}(0,1)$. Fixons un niveau $\alpha = 0,95$. On lit sur la table de la loi gaussienne que

$$\mathbb{P}_\theta(|\sqrt{n}(\bar{X}_n - \theta)/\sigma| > 1,96) = 0,05$$

et donc un intervalle de confiance de niveau $0,95$ pour la moyenne est $\left[\bar{X}_n - 1,96\frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1,96\frac{\sigma}{\sqrt{n}}\right]$.

Rappelons que l'estimateur sans biais de la variance d'un $n-$échantillon $(X_1, \ldots, X_n)$ est définie par

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2. \tag{13.3}$$

Proposition 13.10. *Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi $\mathcal{N}(\mu, \sigma^2)$. Alors*

— *Les variables$\bar{X}_n$ et s_n^2 sont indépendantes.*
— *La variable $\bar{X}_n$ suit la loi $\mathcal{N}(\mu, \sigma^2/n)$.*
— *La variable $(n-1)s_n^2/\sigma^2$ suit la loi χ^2_{n-1}.*

Démonstration. On sait déjà que $\bar{X}_n$ suit la loi $\mathcal{N}(\mu, \sigma^2/n)$. Il suffit de montrer le résultat lorsque $\mu = 0$ et $\sigma^2 = 1$. Notons $f = \frac{1}{\sqrt{n}}(1, \ldots, 1)$ et F le sous-espace vectoriel de $\mathbb{R}^n$ engendré par f. D'après le théorème de Cochran, $p_F X$ et $p_{F^\perp}$ sont indépendantes, avec $\|p_F X\|_2^2$ suit la loi χ^2_1 et $\|p_{F^\perp}\|_2^2$ suit la loi χ^2_{n-1}. On a $p_F X = \langle X, f\rangle f = \bar{X}_n(1, \ldots, 1)$ et

$$\|p_F X\|_2^2 = \|X - P_F X\|_2^2 = \sum_{k=1}^{n} (X_k - \bar{X}_n)^2 = (n-1)s_n^2,$$

ce qui donne les résultats voulus.

□

Définition. *Soient X et Y deux variables indépendantes. Si X suit la loi $\mathcal{N}(0,1)$ et Y suit la loi χ^2_n, alors on dit 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.*

Proposition 13.11. *Soit $(X_1, \dots, X_n)$ un n-échantillon de loi $\mathcal{N}(\mu, \sigma^2)$, où μ et σ^2 sont inconnues. Alors $\sqrt{n}(\bar{X}_n - \mu)/S$ suit la loi de Student t_{n-1} à $n-1$ degrés de liberté.*

Démonstration. Il suffit d'utiliser le résultat précédent. On sait que $\bar{X}_n$ et S^2 sont indépendantes. De plus, $\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}$ suit la loi gaussienne standard, tandis que $\frac{n-1}{\sigma^2}S^2$ suit la loi χ^2_{n-1}. Par définition de la loi de Student, on en déduit que $\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}\sqrt{\frac{n-1}{\frac{n-1}{\sigma^2}S^2}} = \sqrt{n}\frac{\bar{X}_n-\mu}{S}$ suit la loi t_{n-1}. □

Exemple: Ce résultat peut être utilisé pour trouver un intervalle de confiance pour la moyenne d'une loi de moyenne et de variance finis, mais inconnus, à partir d'un échantillon.
Cherchons par exemple un intervalle de confiance de niveau $\alpha = 0,95$ pour μ lorsque $n = 25$ et σ^2 est inconnue. Si T est une variable aléatoire de Student t_{24}, on lit dans la table que $\mathbb{P}(|T| > 2,06) \simeq 0,05$. L'intervalle cherché est donc

$$\left[\bar{X}_n - 2,06\frac{S}{\sqrt{n}}, \bar{X}_n + 2,06\frac{S}{\sqrt{n}}\right] = [\bar{X}_n - 0,41S, \bar{X}_n + 0,41S].$$

Revenons maintenant à la méthode de Monte-Carlo (vue dans le paragraphe 10.4.3). Rappelons qu'afin d'approcher la valeur d'une intégrale $I = \int_{[0,1]^d} g(x)\,dx$ où g est une fonction de carré intégrable, on simule des variables aléatoires $(X_i)_{i\geq 1}$ de loi uniforme sur $[0,1]^d$ puis on utilise la loi forte des grands nombres qui dit que

$$\frac{1}{n}\sum_{i=1}^n g(X_i) \underset{n\to+\infty}{\overset{p.s.}{\longrightarrow}} \mathbb{E}[g(X_1)] = I.$$

On remarque que $\hat{I}_n = \frac{1}{n}\sum\limits_{i=1}^n g(X_i)$ est un estimateur de I : c'est l'estimateur de la moyenne empirique. Comme $\mathbb{E}[\hat{I}_n] = I$, il est sans biais. Nous avons vu précédemment, dans la section 10.4.3, une grossière estimation de l'erreur d'approximation entre $\hat{I}_n$ et I en utilisant la loi faible des grands nombres. Nous allons maintenant en donner une meilleure estimation en utilisant le théorème central de la limite.
Notons $\sigma^2 = \mathrm{Var}(g(X_1))$. La variance de $\hat{I}_n$ est donnée par :

$$\mathrm{Var}(\hat{I}_n) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(g(X_1)) = \frac{1}{n}\mathrm{Var}(g(X_1)) = \frac{\sigma^2}{n}.$$

Le théorème central de la limite assure alors que

$$\sqrt{\frac{n}{\sigma^2}}(\hat{I}_n - I) \quad \text{converge en loi vers une variable aléatoire de loi } \mathcal{N}(0,1).$$

Comme la variance σ^2 est généralement inconnue (elle dépend de I), on ne peut pas donner un intervalle de confiance en fonction de σ^2 et on travaille avec l'estimateur de la variance

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(g(X_i) - \hat{I}_n)^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} g^2(X_i) - \hat{I}_n^2\right).$$

s_n^2 est un estimateur sans biais de la variance de $g(X_1)$. On remarque que comme $\hat{I}_n$ converge p.s. vers I, on en déduit que s_n^2 converge p.s. vers σ^2. Pour estimer l'erreur standard (ou encore écart-type) et l'erreur relative, on remplace σ^2 par s_n^2. Ainsi, l'erreur standard est proche de $\sqrt{\frac{s_n^2}{n}}$ et l'erreur relative est proche de $\frac{1}{\hat{I}_n}\sqrt{\frac{s_n^2}{n}}$. Nous pouvons maintenant donner un intervalle de confiance pour notre estimateur $\hat{I}_n$. En effet, le lemme de Slutsky combiné avec le théorème central limite implique que la quantité

$$\sqrt{\frac{n}{s_n^2}}(\hat{I}_n - I) = \sqrt{\frac{\sigma^2}{s_n^2}}\sqrt{\frac{n}{\sigma^2}}(\hat{I}_n - I)$$

converge en loi vers une variable aléatoire de loi gaussienne standard. On obtient ainsi pour n suffisamment grand et pour tout $\varepsilon > 0$:

$$\mathbb{P}(|\hat{I}_n - I| \leq \varepsilon) = \mathbb{P}\left(\sqrt{\frac{n}{s_n^2}}(\hat{I}_n - I) \leq \sqrt{\frac{n}{s_n^2}}\varepsilon\right) \simeq 2\int_{-\infty}^{\sqrt{\frac{n}{s_n^2}}\varepsilon} \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx - 1.$$

Pour un intervalle de confiance de niveau 95%, ce qui correspond à

$$2\int_{-\infty}^{\sqrt{\frac{n}{s_n^2}}\varepsilon} \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx - 1 = 0.95,$$

on trouve que $\sqrt{\frac{n}{s_n^2}}\varepsilon = 1.96$ soit encore $\varepsilon = 1.96\sqrt{\frac{s_n^2}{n}}$. En conclusion, l'intervalle de confiance de I au niveau de confiance 95% est

$$\left[\hat{I}_n - 1.96\sqrt{\frac{s_n^2}{n}}, \hat{I}_n + 1.96\sqrt{\frac{s_n^2}{n}}\right].$$

La vitesse de convergence est donc de l'ordre de $\frac{1}{\sqrt{n}}$. Cette vitesse peut paraître faible en petite dimension mais elle présente un double avantage :

1. elle ne dépend pas de la dimension (ce qui est utile quand la dimension est grande),
2. elle ne dépend pas de la régularité de g, à condition qu'elle soit de carré intégrable.

On remarque ici l'importance de la variance σ^2 dans la qualité de l'approximation. Pour améliorer l'erreur, on peut chercher des méthodes pour réduire cet écart-type afin d'accélérer la convergence. Nous ne les étudierons pas ici.

13.3 Modèles paramétriques (non-bayésiens)

13.3.1 Maximum de vraisemblance

Définition. *1. Soit $(X_1, \ldots, X_n)$ un $n-$échantillon de loi discrète $\mathbb{P}_\theta$, où le paramètre θ est inconnu. La fonction $L(X_1, \ldots, X_n, \theta) = \prod_{i=1}^{n} \mathbb{P}_\theta(X_i)$ est appelée fonction de vraisemblance de la loi $\mathbb{P}_\theta$.*

2. Soit $(X_1, \ldots, X_n)$ un $n-$échantillon de loi de densité $f(x, \theta)$ dépendant du paramètre θ inconnu. La fonction $L(X_1, \ldots, X_n, \theta) = \prod_{i=1}^{n} f(X_i, \theta)$ est appelée fonction de vraisemblance de la loi $\mathbb{P}_\theta$.

Exemple:

1. Soit $(X_1, \ldots, X_n)$ un $n-$échantillon d'une loi de Bernoulli de paramètre p inconnu. On sait que $\mathbb{P}(X = 1) = p = p^1(1-p)^{1-1}$ et $\mathbb{P}(X = 0) = 1 - p = p^0(1-p)^1$. La fonction de vraisemblance pour une observation est donc $L(X, p) = p^X(1-p)^{1-X}$ et la fonction de vraisemblance de l'échantillon est
$$L(X_1, \ldots, X_n, p) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i}.$$
2. Soit $(X_1, \ldots, X_n)$ un $n-$échantillon d'une loi de Poisson de paramètre λ inconnu. La fonction de vraisemblance pour une observation est donc $L(X, p) = e^{-\lambda}\lambda^X/X!$ (ce qui est bien défini car X prend des valeurs entières positives). La fonction de vraisemblance de l'échantillon est
$$L(X_1, \ldots, X_n, \lambda) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{X_i}}{X_i!} = e^{-n\lambda}\frac{\lambda^{X_1+\cdots+X_n}}{\prod_{i=1}^{n} X_i!}.$$
3. Soit $(X_1, \ldots, X_n)$ un $n-$échantillon d'une loi gaussienne de paramètres (μ, σ^2). On vérifie que la fonction de vraisemblance de l'échantillon est
$$L(X_1, \ldots, X_n, \mu, \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2\right\}.$$

Passons maintenant à l'estimation par la méthode du maximum de vraisemblance.

Définition. *Soit $(X_1, \ldots, X_n)$ un $n-$échantillon de fonction de vraisemblance $L(X_1, \ldots, X_n, \theta)$, où θ est inconnue. On appelle estimateur du maximum de vraisemblance tout estimateur $\hat{\theta}$ de θ tel que*
$$\sup_{\theta} L(X_1, \ldots, X_n, \theta) = L(X_1, \ldots, X_n, \hat{\theta}). \tag{13.4}$$

Cette équation est équivalente, lorsque $\log L$ *est bien défini, à*

$$\sup_{\theta} \log L(X_1, \ldots, X_n, \theta) = \log L(X_1, \ldots, X_n, \hat{\theta}).$$

Remarque 13.12. *Pour calculer le maximum de vraisemblance, on étudie le sens de variation de la fonction* $\theta \mapsto L(X_1, \ldots, X_n, \theta)$. *Si cette fonction est régulière, cela revient à étudier le signe* $\frac{d}{d\theta} L(X_1, \ldots, X_n, \theta)$ *(rappelons que la connaissance des points d'annulation de la dérivée ne suffit pas à connaître les variations de la fonction).*

Notons qu'il n'y a *a priori* ni existence, ni unicité d'un estimateur du maximum de vraisemblance (car l'équation (13.4) peut avoir une ou plusieurs solutions, voire aucune solution). Néanmoins, dans la plupart des cas concrets, un tel estimateur existe, est unique et assez facile à calculer. Il n'y a pas vraiment de justification théorique à cette méthode. Notons tout de même que, sous des hypothèses générales, l'estimateur du maximum de vraisemblance d'un $n-$échantillon converge vers la vraie valeur du paramètre lorsque n tend vers l'infini.

Remarque 13.13. *Cet estimateur est invariant par changement de paramètre. Par exemple, si on se place sur* $\mathbb{R}_+$, *il est équivalent de paramètrer le modèle par* θ^2 *au lieu de* θ. *L'estimateur du maximum de vraisemblance pour* θ^2 *sera alors le carré de celui pour* θ. *En revanche, si* T *est le meilleur estimateur sans biais de* θ, *alors* T^2 *ne sera pas le meilleur estimateur sans biais de* θ^2.

Exemple:

1. Dans le cas d'une loi de Bernoulli de paramètre p, on a vu que

   $$L(X_1, \ldots, X_n, p) = \prod_{i=1}^{n} p^{X_i} (1-p)^{1-X_i}$$

   et donc

   $$\log L(X_1, \ldots, X_n, p) = (\log p) \ \sum_{i=1}^{n} X_i + \left(n - \sum_{i=1}^{n} X_i\right) \log(1-p).$$

   On trouve alors que

   $$\frac{d}{dp} \log L(X_1, \ldots, X_n, p) = \frac{1}{p} \sum_{i=1}^{n} X_i - \left(n - \sum_{i=1}^{n} X_i\right) \frac{1}{1-p}.$$

   Cette dérivée est de la forme $\frac{a}{p} - \frac{b}{1-p} = \frac{a-(a+b)p}{p(1-p)}$, avec $a \geq 0$ et $b \geq 0$: elle a le signe de $a-(a+b)p$. La fonction de vraisemblance atteint donc son maximum en $p = a/(a+b)$.
   On obtient donc que $\hat{p} = \frac{1}{n}(X_1 + \cdots + X_n)$ est l'estimateur du maximum de vraisemblance. Ainsi, l'estimateur du maximum de vraisemblance

coïncide avec la moyenne empirique ; c'est un estimateur consistant (par la loi des grands nombres) et sans biais.

2. Dans le cas d'une loi gaussienne de paramètres (μ, σ^2), où σ^2 est connue alors que μ est inconnue, on a vu que
$$L(X_1, \ldots, X_n, \mu) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2\right\}.$$
Pour maximiser $L(X_1, \ldots, X_n, \mu)$, il faut minimiser $\sum_{i=1}^{n} (X_i - \mu)^2$. L'application $\mu \mapsto \sum_{i=1}^{n} (X_i - \mu)^2$ est un polynôme de degré deux, de coefficient dominant positif : il est minimal là où sa dérivée s'annule, c'est-à-dire lorsque $\sum_{i=1}^{n} 2(X_i - \mu) = 0$, soit $\mu = \bar{X}_n$. Ainsi, l'estimateur du maximum de vraisemblance coïncide encore avec la moyenne empirique.

3. Dans le cas d'une loi gaussienne de paramètres (μ, σ^2), où cette fois σ^2 est inconnue alors que μ est connue, on a vu que
$$L(X_1, \ldots, X_n, \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2\right\}.$$
On passe à nouveau au logarithme comme précédemment et on obtient
$$\frac{d}{d\sigma} \log L(X_1, \ldots, X_n, \sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (X_i - \mu)^2 = \frac{n}{\sigma^3}(\sigma_n^2 - \sigma^2),$$
où σ_n^2 est la variance empirique
$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2.$$
Il est alors clair que la fonction de vraisemblance atteint son maximum pour $\sigma^2 = \sigma_n^2$. σ_n^2 est donc ici l'estimateur du maximum de vraisemblance. C'est un estimateur consistant mais qui, comme on l'a déjà vu, est biaisé.

4. Dans le cas d'une loi gaussienne de paramètres (μ, σ^2), où σ^2 est inconnue et μ est inconnue, on refait comme précédemment. On a vu que
$$L(X_1, \ldots, X_n, \mu, \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2\right\}.$$
Le calcul fait au 3. montre qu'on a toujours
$$L(X_1, \ldots, X_n, \mu, \sigma^2) \leq \frac{1}{(\sqrt{2\pi}\sigma_n)^n} \exp\left\{-\frac{1}{2\sigma_n^2} \sum_{i=1}^{n} (X_i - \mu)^2\right\},$$

avec égalité pour $\sigma = \sigma_n$. D'après le calcul fait en 2., ce majorant est maximal pour $\mu = \bar{X}_n$.

Finalement, le couple $\theta = (\mu, \sigma^2)$ réalisant le maximum de vraisemblance est formé par la moyenne empirique $\bar{X}_n$ et la variance empirique σ_n^2. Comme on l'a déjà signalé, $\bar{X}_n$ est consistant sans biais, tandis que σ_n^2 est consistant, mais biaisé. Pour avoir un estimateur sans biais de σ^2, il faut considérer l'estimateur $s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$.

13.3.2 Méthode des moments

Soit un modèle statistique indexé par un ensemble Θ. Supposons donné un $n-$échantillon $X_1, \ldots, X_n$ d'une loi intégrable sous $\mathbb{P}_\theta$ pour chaque $\theta \in \Theta$. On cherche un estimateur $\hat{\theta}$ de θ.

Supposons que Θ est une partie de $\mathbb{R}$. On suppose également qu'on a un intervalle I de $\mathbb{R}$ tel que $\mathbb{P}_\theta(X_1 \in I) = 1$ pour tout $\theta \in \Theta$. L'idée de la méthode des moments est de prendre comme estimateur $\hat{\theta}$ une valeur de θ telle que $\mathbb{E}_\theta X = m(\theta)$ coïncide avec la moyenne empirique observée.

Définition. *Un estimateur par la méthode des moments est une solution de l'équation*

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i = m(\hat{\theta}).$$

Si m réalise une bijection de Θ dans I , alors $\hat{\theta} = m^{-1}(\bar{X}_n)$.

Si Θ est une partie de $\mathbb{R}^d$, on calcule $\mathbb{E}_\theta X = m_1(\theta), \ldots, \mathbb{E}_\theta X^d = m_d(\theta)$. Un estimateur par la méthode des moments est une solution du système d'équations

$$\frac{1}{n} \sum_{i=1}^{n} X_i = m_1(\hat{\theta}), \quad \ldots \ , \quad \frac{1}{n} \sum_{i=1}^{n} X_i^d = m_d(\hat{\theta}).$$

Si m est inversible, alors $(\hat{\theta}_1, \ldots, \hat{\theta}_d)' = m^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} X_i, \ldots, \frac{1}{n} \sum_{i=1}^{n} X_i^d \right)'$.

Exemple: On veut estimer la moyenne et la variance par la méthode des moments. Posons $\theta = (m, \sigma^2) \in \mathbb{R} \times]0, +\infty[$. On a $m = \mathbb{E}_\theta X$ et $\sigma^2 = \mathrm{Var}_\theta(X)$. On sait de plus que $\mathbb{E}_\theta(X^2) = \sigma^2 + m^2$, ce qui donne alors

$$\frac{1}{n} \sum_{i=1}^{n} X_i = \hat{m} \quad \text{et} \quad \frac{1}{n} \sum_{i=1}^{n} X_i^2 = \hat{\sigma}^2 + \hat{m}^2.$$

On en déduit donc que $\hat{m}$ est la moyenne empirique et

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 .$$

Remarque 13.14. — *Dans certains cas, l'estimation par la méthode des moments est moins bonne que l'estimation par maximum de vraisemblance. Néanmoins, dans le cas de la loi Gamma par exemple, le calcul de la fonction de vraisemblance peut poser des problèmes (l'utilisation de l'ordinateur et d'algorithmes numériques est indispensable) tandis que l'estimation des moments est très facilement accessible.*

— *La méthode des moments peut s'utiliser comme point de départ pour maximiser la (log-)vraisemblance. En effet, on doit alors utiliser des algorithmes numériques, comme la Méthode de Newton, qui nécessitent des points de départ.*

— *Lorsque la taille de l'échantillon n'est pas suffisamment grande, la loi des grands nombres ne s'applique pas et par conséquent, les moments empiriques n'approchent pas suffisamment les moments théoriques. Ainsi, la méthode des moments n'est pas une bonne méthode dans ce cas, car les estimateurs obtenus peuvent sortir du support des paramètres. Par exemple, pour la loi $\Gamma(a)$, un petit échantillon peut conduire à $a < 0$.*

— *Enfin, nous avons vu que la méthode des moments consiste à résoudre une équation du type $\bar{X}_n = m(\hat{\theta})$, ce qui n'est pas toujours possible.*

13.3.3 Méthode des moindres carrés (régression linéaire)

En statistiques, étant donné un échantillon $(x_i)_{1\le i\le n}$, un modèle de régression simple suppose que les résultats observés y_i sont liés aux x_i. On représente dans un graphe l'ensemble des observations (x_i, y_i). On peut alors proposer un modèle linéaire, c'est-à-dire chercher la droite dont l'équation est $y_i = a + bx_i$ et qui passe au plus près des points du graphe. Passer au plus près signifie ici rendre minimale la somme des carrés des écarts des points à la droite

$$\sum_{i=1}^{n} (y_i - a - bx_i)^2.$$

Cela revient donc à trouver les valeurs des paramètres a et b qui minimisent cette somme.

Voyons maintenant la formulation statistique du problème. Supposons données des variables Y_i associées aux points $x_1, \dots, x_n$. Les résultats observés sont liés aux x_i comme suit

$$Y_i = a + bx_i + \varepsilon_i,$$

où a et b sont les paramètres inconnus, ε_i sont des erreurs indépendantes, de même loi centrée et de variance σ^2 (connue ou non).

On dit alors que les Y_i suivent un modèle linéaire.

Définition. *On dit que a et b sont les coefficients de régression linéaire.*

La méthode des moindres carrés consiste à minimiser la distance euclidienne, dans $\mathbb{R}^n$, entre le vecteur Y et $a + bx$, où a est le vecteur constant $(a, \dots, a)$ et $b \in \mathbb{R}$. On cherche donc à déterminer

$$\min_{a,b \in \mathbb{R}} \sum_{i=1}^{n} (Y_i - (a + bx_i))^2.$$

Les estimateurs $\hat{a}$ et $\hat{b}$ réalisant ce minimum sont appelés estimateur des moindres carrés.

Ainsi, les estimateurs des moindres carrés $\hat{a}$ et $\hat{b}$ sont les coefficients de la droite des moindres carrés associée aux points $(x_1, Y_1), \dots, (x_n, Y_n)$.

On est donc ramené à un problème purement déterministe. $(x_i, y_i)_{1 \le i \le n}$ étant donné, on veut calculer

$$\operatorname*{argmin}_{a,b \in \mathbb{R}} \sum_{i=1}^{n} (y_i - (a + bx_i))^2.$$

Pour calculer les coefficients $\hat{a}$ et $\hat{b}$, nous avons deux méthodes à notre disposition, que nous allons présenter ici. Bien entendu, ces deux méthodes reviennent au même.

Méthode 1 : par l'espérance

Bien que ce problème soit déterministe, des techniques probabilistes peuvent être utiles : soit (X, Y) un vecteur aléatoire suivant la loi $\frac{1}{n} \sum_{i=1}^{n} \delta_{(x_i, y_i)}$. Comme

$$\sum_{i=1}^{n} (y_i - (a + bx_i))^2 = n\mathbb{E}[(Y - (a + bX))^2],$$

on a

$$(\hat{a}, \hat{b}) = \operatorname*{argmin}_{a,b \in \mathbb{R}} \mathbb{E}[(Y - (a + bX))^2].$$

Tout d'abord, montrons que pour toute variable aléatoire Z de carré intégrable, on a $\mathbb{E}(Z - \alpha)^2 \ge \operatorname{Var} Z$, le cas d'égalité ayant lieu pour $\alpha = \mathbb{E}[Z]$. En effet, en développant le carré et en utilisant la linéarité de l'espérance, on voit que

$$\mathbb{E}(Z - \alpha)^2 = \mathbb{E}(X - \mathbb{E}[X])^2 + (\alpha - \mathbb{E}[X])^2 = \operatorname{Var} X + (\alpha - \mathbb{E}[X])^2.$$

Ainsi, si on pose $Z_b = Y - bX$, on a

$$\mathbb{E}(Y - (a + bX))^2 = (a - \mathbb{E}[Z_b])^2 + \operatorname{Var} Z_b.$$

Si on trouve $\hat{b}$ réalisant le minimum de $\operatorname{Var} Z_b$, en posant

$$\hat{a} = \mathbb{E}[Z_{\hat{b}}] = \mathbb{E}[Y] - \hat{b}\mathbb{E}[X] = \overline{y} - \hat{b}\overline{x},$$

on aura alors trouvé le couple $(\hat{a}, \hat{b})$ réalisant le minimum Remarquons de plus que, pour $b \in \mathbb{R}$, on a toujours $\operatorname{Var}(Y - bX) = \operatorname{Var}(Y - \mathbb{E}Y - b(X - \mathbb{E}[X]))$, ce qui implique donc qu'on peut toujours supposer ici les variables X et Y centrées. On cherche le minimum de

$$\operatorname{Var}(Y - bX) = \operatorname{Var}(Y) - 2b \operatorname{Covar}(X, Y) + b^2 \operatorname{Var}(X),$$

qui est un polynôme du second degré en b, atteignant son minimum en

$$\hat{b} = \frac{\operatorname{Covar}(X, Y)}{\operatorname{Var}(X)} = \frac{\sum y_i(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}.$$

Pour conclure, en posant $\mathbf{1} = (1, \ldots, 1)^*$, il suffit de voir que

$$\min_{a,b \in \mathbb{R}} \mathbb{E}\|Y - (a\mathbf{1} + bX)\|^2 = \min_{a \in \mathbb{R}} \mathbb{E}\|Y - (a\mathbf{1} + \hat{b}X)\|^2$$

et donc ce dernier minimum est atteint en

$$\hat{a} = \frac{1}{n} \sum_{1 \leq i \leq n} (y_i - \hat{b}x_i) = \bar{y} - \hat{b}\bar{x} = \bar{y} - \frac{\sum y_i(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \bar{x}.$$

Méthode 2 : par la projection

Si l'on pose $\mathbf{1} = (1, \ldots, 1)^*$, on a

$$\sum_{i=1}^{n} (y_i - (a + bx_i))^2 = \|y - (a\mathbf{1} + bx)\|_2^2,$$

et la méthode des moindres carrés consiste à minimiser la distance, dans $\mathbb{R}^n$, entre le vecteur y et le sous-espace $H = \{a\mathbf{1} + bx : a, b \in \mathbb{R}\}$ engendré par $\mathbf{1}$ et x.

Remarque 13.15. *Le vecteur $\hat{a} + \hat{b}x$ réalisant le minimum de cette distance est la projection de y sur H. Le vecteur $y - (\hat{a}x + \hat{b})$ est orthogonal à H, ce qui implique*

$$\langle y - (\hat{a}x + \hat{b}), \mathbf{1} \rangle = 0 \quad \textit{et} \quad \langle y - (\hat{a}x + \hat{b}), x \rangle = 0. \tag{13.5}$$

En résolvant le système, on identifie $\hat{a}$ et $\hat{b}$.

Une méthode plus agréable pour trouver $\hat{a}$ et $\hat{b}$ est la suivante. On introduit une base orthogonale dans H : $(\mathbf{1}, x - (\bar{x} \cdot \mathbf{1}))$. Remarquons que le vecteur orthogonal à $\mathbf{1}$ obtenu par l'orthogonalisation de Gram-Schmidt est

$$x - \langle x - \frac{1}{\sqrt{n}} \cdot \mathbf{1} \rangle \cdot \frac{1}{\sqrt{n}} \mathbf{1} = x - (\bar{x} \cdot \mathbf{1}).$$

On écrit le système (13.5) dans cette base orthogonale avec

$$\hat{y} = \hat{a}\mathbf{1} + \hat{b}x = (\hat{a} + \hat{b}\bar{x})\mathbf{1} + \hat{b}(x - \bar{x}\mathbf{1}).$$

Posons maintenant $\hat{\alpha} = \hat{a} + \hat{b}\bar{x}$. L'équation (13.5) devient alors

$$\langle y - (\hat{\alpha}\mathbf{1} + \hat{b}(x - \bar{x} \cdot \mathbf{1})), \mathbf{1} \rangle = 0 \quad \text{et} \quad \langle y - (\hat{\alpha}\mathbf{1} + \hat{b}(x - \bar{x} \cdot \mathbf{1})), x - \bar{x} \cdot \mathbf{1} \rangle = 0.$$

Or $x - \bar{x} \cdot \mathbf{1}$ et $\mathbf{1}$ sont orthogonaux, donc le système précédent se réduit à

$$\langle y - \hat{\alpha}\mathbf{1}, \mathbf{1} \rangle = 0 \quad \text{et} \quad \langle y - \hat{b}(x - \bar{x} \cdot \mathbf{1}), x - \bar{x} \cdot \mathbf{1} \rangle = 0.$$

On obtient donc finalement

$$\begin{aligned} \hat{\alpha} &= \frac{\langle y, \mathbf{1} \rangle}{\langle \mathbf{1}, \mathbf{1} \rangle} = \frac{\sum y_i}{n} = \bar{y} \\ \text{et} \quad \hat{b} &= \frac{\langle y, x - \bar{x} \cdot \mathbf{1} \rangle}{\|x - \bar{x}\mathbf{1}\|^2} = \frac{\sum y_i(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}. \end{aligned}$$

À partir de ces expressions, on trouve la valeur des estimées $\hat{a} = \hat{\alpha} - \hat{b}\bar{x}$, soit

$$\hat{a} = \bar{y} - \frac{\sum y_i(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \bar{x}.$$

Enfin, en substituant Y à y dans les expressions trouvées, on a les estimateurs

$$\hat{a} = \bar{Y} - \frac{\sum Y_i(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \bar{x}.$$

13.3.4 Exercice : recherche d'estimateurs

Soit $X_1, \ldots, X_n$ un $n-$échantillon de loi $\Gamma(\theta, \mu)$ où $\theta > 0$ est inconnu alors que μ est connu. Soit b un paramètre donné, $0 < b < n\mu$, et considérons les variables aléatoires

$$S = X_1 + \cdots + X_n \quad \text{et} \quad \tau_b = \frac{\Gamma(n\mu)}{\Gamma(n\mu - b)} S^{-b}.$$

1. Quelle est la loi de la variable S ?
2. Montrer que τ_b est un estimateur sans biais du paramètre θ^b pour tout $0 < b < n\mu$.
3. L'estimateur τ_b est-il consistant pour le paramètre θ^k, où k est un entier positif tel que $0 < k < n\mu$? Sinon, trouver un estimateur consistant de θ^k.

Solution

1. Considérons X et Y deux variables indépendantes suivant respectivement les lois $\Gamma(\alpha,\mu_1)$ et $\Gamma(\alpha,\mu_2)$. En calculant la fonction caractéristique de la variable $X+Y$, on voit que $X+Y$ suit la loi $\Gamma(\alpha,\mu_1+\mu_2)$. Comme les variables X_i sont toutes indépendantes et de même loi, on en déduit que S suit la loi $\Gamma(\theta,n\mu)$.
2. S est un estimateur de θ^b. Montrons que S est sans biais. Pour ce faire, calculons son espérance. Rappelons que la variable S admet la densité
$$f_S(s)=\theta^{n\mu}s^{n\mu-1}e^{-\theta s}\frac{1}{\Gamma(n\mu)}\,\mathbb{1}_{s\geq 0}.$$
On a alors par le théorème de transfert
$$\mathbb{E}\tau_b=\frac{\Gamma(n\mu)}{\Gamma(n\mu-b)}\mathbb{E}S^{-b}=\frac{\Gamma(n\mu)}{\Gamma(n\mu-b)}\int_0^{+\infty}s^{-b}\theta^{n\mu}s^{n\mu-1}e^{-\theta s}\frac{1}{\Gamma(n\mu)}\,ds.$$
Procédons au changement de variable $u=\theta s$.
Comme $\Gamma(\mu)=\int_0^{+\infty}t^{\mu-1}e^{-t}\,dt$, on obtient $\mathbb{E}\tau_b=\theta^b$. Cela montre que τ_b est un estimateur sans biais de θ^b.
3. On se demande maintenant si l'estimateur τ_k converge en probabilité vers θ^k (lorsque $n\to+\infty$). Remarquons tout d'abord que par la loi forte des grands nombres, la moyenne empirique $\frac{X_1+\cdots+X_n}{n}$ converge p.s. vers $\mathbb{E}X_1=1/\theta$. De plus, il est très facile de voir que pour tout $x>0$, on a $\Gamma(x+1)=x\Gamma(x)$. En itérant cette identité, on trouve que
$$\Gamma(n\mu)=(n\mu-1)\ldots(n\mu-k)\Gamma(n\mu-k).$$
On en déduit donc que $\tau_k=\frac{(n\mu-1)\ldots(n\mu-k)}{(X_1+\cdots+X_n)^k}$ converge presque sûrement vers $\frac{\mu^k}{(\mathbb{E}X_1)^k}=(\mu\theta)^k$. Ainsi, τ_k n'est pas un estimateur consistant. Un estimateur (fortement) consistant de θ^k est $\frac{\tau_k}{\mu^k}$.

13.4 Modèles non-paramétriques

13.4.1 Les tests d'hypothèse du chi-deux

Soit $(\Omega,\mathcal{F},(\mathbb{P}_\theta)_{\theta\in\Theta})$ un modèle statistique. L'ensemble Θ est ici divisé en une partie Θ_0 et son complémentaire Θ_1. L'objectif d'un test est (au vu de l'observation ω) de "décider" si la vraie valeur de θ se trouve dans Θ_0 ou dans Θ_1.

Si on reprend l'exemple des pièces défectueuses, notons θ_0 la valeur limite de la proportion de pièces défectueuses qui est acceptable. On veut décider, au vu du nombre X de pièces défectueuses observées dans un échantillon de n pièces, si $\theta>\theta_0$ ou non. Comme dans la plupart des problèmes pratiques de ce type, il est dissymétrique : on veut être “raisonnablement sûr” que $\theta \leq \theta_0$. On veut rejeter l'hypothèse $\theta > \theta_0$ avec une “erreur” faible si elle est vraie (car si la machine est mal réglée, les clients vont refuser les lots de pièces qui contiendront en moyenne trop de pièces défectueuses). En revanche, si la vraie valeur est $\theta \leq \theta_0$ et si on décide à tort qu'elle est plus grande, ce n'est pas très grave : on fera juste un réglage supplémentaire (et inutile) de la machine.

Définition. *1. Dans un problème de test, on veut tester l'hypothèse H_0 selon laquelle $\theta \in \Theta_0$ contre l'alternative H_1 selon laquelle $\theta \in \Theta_1$.*

2. *La région critique est la partie (ou l'événement) D de Ω sur laquelle on rejette l'hypothèse H_0. Si $\omega \in D$, on décide que H_1 n'est pas rejetée, alors que si $\omega \notin D$, alors H_0 n'est pas rejetée.*
3. *Si $\theta \in \Theta_0$, la probabilité de rejeter H_0 alors qu'elle est vraie $\mathbb{P}_\theta(D)$ est l'erreur de première espèce.*
4. *Si $\theta \in \Theta_1$, la probabilité de ne pas rejeter H_0 alors qu'elle est fausse $1 - \mathbb{P}_\theta(D) = \mathbb{P}_\theta(D^c)$ est l'erreur de seconde espèce.*
5. *Le nombre $\alpha = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(D)$ est le niveau du test ou de la région critique et la fonction $\mathbb{P}_\cdot(D) : \Theta \to [0,1]$, $\theta \mapsto \mathbb{P}_\theta(D)$ est la fonction puissance du test.*

Il faut donc construire un test. Cela consiste à trouver une région critique qui minimise autant que possible les erreurs.Toutefois, on ne pourra jamais rendre petits à la fois le risque de première espèce $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(D)$ et le risque de deuxième espèce $\sup_{\theta \in \Theta_1} \mathbb{P}_\theta(D^c)$. En effet, il est fréquent que Θ soit une partie de $\mathbb{R}^d$ et pour n'importe quelle région critique D, la fonction puissance est continue. Or on cherche à la rendre aussi proche que possible de 1 sur Θ_1 et de 0 sur Θ_0. Ceci est manifestement contradictoire, en général, au voisinage de la frontière entre Θ_0 et Θ_1.

En fait, les tests sont construits de la manière suivante. On commence par fixer une borne supérieure au niveau α, en général 0,1 ou 0,05 ou 0,01. Ensuite, parmi toutes les régions critiques de niveau α (ou $\leq \alpha$), on cherche à maximiser la fonction puissance sur Θ_1 (donc à minimiser les erreurs de seconde espèce). Implicitement, cela signifie qu'on considère comme plus grave une erreur de première espèce qu'une erreur de seconde espèce : les premières sont majorées uniformément par α, tandis que les secondes sont souvent proches de $1 - \alpha$ aux points de Θ_1 qui sont “proches” de la frontière avec Θ_0.

En pratique, on cherche une statistique (une fonction des observations) dont on connaît la loi si H_0 est vraie et qui ne se comporte pas de la même manière selon que H_0 ou H_1 est vraie. Les tests du χ^2 sont un exemple relativement simple de tests statistiques qui vont permettre de tester :

1. l'adéquation à une loi de probabilité sur un ensemble fini ou dénombrable : est-il raisonnable de penser que les résultats observés sont des réalisations indépendantes identiquement distribuées d'une loi donnée $(p_1, \ldots, p_k)$ sur $\{1, \ldots, k\}$?
2. l'indépendance de deux caractères mesurés sur un même individu,
3. l'homogénéité de plusieurs échantillons : deux médicaments ont-ils le même effet (guérison, amélioration, état stationnaire) sur la population atteinte ?

Test d'adéquation à une loi

Le test d'adéquation est aussi appelé test d'ajustement par certains auteurs.
On est essentiellement dans un cadre non-paramétrique. L'espace Θ est très grand. Il s'agit de l'ensemble de toutes les probabilités sur un espace donné E. Plus précisément, on considère un n-échantillon $(X_1, \ldots, X_n)$ de loi p sur un ensemble E. On se donne une loi p_0. On veut tester
H_0 : "on a $p = p_0$", contre H_1 : "on a $p \neq p_0$".

Nous allons décrire le test du χ^2. Tout d'abord, on choisit une partition finie $E_1, \ldots, E_q$ de l'espace E et on pose pour $1 \leq j \leq q$

$$\mu_j = p_0(E_j).$$

On suppose que $\mu_j > 0$ pour tout j (dans la pratique, on choisit la partition de sorte que les μ_j soient le plus proche possible de $1/q$). Ensuite, on note N_j^n le nombre de variables X_i qui tombent dans E_j (on a donc $N_1^n + \cdots + N_q^n = n$) et on pose

$$T_n = \sum_{j=1}^{q} \frac{1}{n\mu_j}(N_j^n - n\mu_j)^2 = n \sum_{j=1}^{q} \frac{(\frac{N_j^n}{n} - \mu_j)^2}{\mu_j}.$$

N_j^n est appelée variable de comptage. L'idée du test du χ^2 d'adéquation est la suivante. Si $p = p_0$, on a convergence presque sûre de N_j^n/n vers μ_j par la loi des grands nombres. On a vu au chapitre 12 que T_n converge en loi vers une variable aléatoire finie. En revanche, si $p \neq p_0$, on a $\lim_{n \to +\infty} N_j^n/n = p(E_j)$ p.s. Par suite, si $p(E_j) \neq \mu_j$ pour au moins une valeur de j, alors les variables T_n tendent vers l'infini.

Remarque 13.16. *La quantité T_n quantifie l'écart entre la distribution empirique $\sum_{j=1}^{q} \frac{N_j^n}{n}\delta_j$ et la loi μ_j qui correspond à l'hypothèse nulle.*

Proposition 13.17. *Soit ν une loi de probabilité sur E. Alors*

$$\frac{T_n}{n} = \sum_{j=1}^{q} \frac{1}{n^2\nu(E_j)}(N_j^n - n\nu(E_j))^2 \underset{n\to+\infty}{\to} \sum_{j=1}^{q} \frac{(\mu_j - \nu(E_j))^2}{\nu(E_j)} \quad p.s.$$

En particulier, on retrouve que T_n tend vers l'infini si $\nu \neq p_0$.

Démonstration. En effet, $\dfrac{T_n}{n} = \displaystyle\sum_{j=1}^{q} \frac{1}{\nu(E_j)}\left(\frac{N_j^n}{n} - \nu(E_j)\right)^2$. D'après la loi forte des grands nombres, $\frac{N_j^n}{n}$ converge presque sûrement vers $\nu(E_j)$, d'où le résultat voulu. □

Cela conduit à choisir une région critique de la forme $D = \{T_n \geq a\}$.

Il nous reste à déterminer le niveau du test. Cela est essentiel. Ce niveau est $\mathbb{P}_{p_0}(D)$ où $\mathbb{P}_{p_0}$ est la probabilité sur notre espace d'états lorsque la loi des X_i est p_0. Comme p_0 est connu, on peut calculer ce niveau : sous $\mathbb{P}_{p_0}$, la variable $S_n = (N_1^n, \dots, N_q^n)$ suit une loi multinomiale donnée par

$$\mathbb{P}_{p_0}(N_1^n = i_1, \dots, N_q^n = i_q) = n! \prod_{j=1}^{q} \frac{(\mu_j)^{i_j}}{i_j!} \quad \text{si} \quad i_1 + \dots + i_q = n.$$

On peut alors en déduire la loi de T_n et calculer $\mathbb{P}_{p_0}(T_n \geq a)$. Mais les calculs sont compliqués et numériquement impossibles dès que n est grand. On rappelle donc le théorème 12.17.

Théorème 13.18. *Si $p = p_0$, alors la suite $(T_n)_{n\geq 1}$ converge en loi vers une variable aléatoire de loi χ^2_{q-1} à $q-1$ degrés de liberté.*

Pratiquement, on choisit les E_j de sorte que pour chaque j, le produit $n\mu_j$ soit assez grand (≥ 10) et les μ_j aussi voisins que possible les uns des autres. Dans ce cas, l'approximation de la loi de T_n par χ^2_{q-1} est bonne. Le niveau étant fixé, on lit alors dans les tables de loi du χ^2 la valeur de a dans $D = \{T_n \geq a\}$. On rejette alors H_0 si $T_n \geq a$ et on ne rejette pas H_0 sinon (remarquons qu'en général, on ne dit pas en statistique qu'on accepte H_0).

Remarque 13.19. *Le test du χ^2 s'appuie sur deux résultats asymptotiques (une convergence en loi si H_0 est vraie et une convergence presque sûre si H_1 est vraie). Or on ne dispose jamais que d'un nombre fini d'observations. Toute la question est de savoir si l'on a le droit de faire comme si la limite en loi était une égalité. En pratique, pour que le test soit **valide**, il faut que, pour tout $i = 1, \dots, k$, np_i soit supérieur ou égal à 5. Si ce n'est pas le cas, il faut regrouper des classes à trop faible effectif pour atteindre le seuil exigé. Le test du χ^2 peut aussi être utilisé pour tester l'adéquation à une loi sur un ensemble infini, même non-dénombrable : il suffit d'en faire une partition finie et de faire un test du χ^2 sur le numéro de la classe à laquelle appartient la variable.*

Remarque 13.20. *Il faut penser que les hypothèses sont posées dans l'espoir d'être rejetées. C'est dans ce cas qu'elles donnent les informations les plus probantes. En effet, si H_0 n'est pas rejetée, il se peut que cela soit dû au fait que la loi p de l'échantillon ne vérifie pas H_0 mais est toutefois très proche (par exemple au sens de la distance du chi-deux) de p_0. Cela arrive fréquemment si le nombre de classes est petit, par exemple suite à un regroupement.*

Remarque 13.21. *La loi du χ^2 n'étant pas centrée, on pourrait s'étonner que les grandes valeurs de la statistique conduisent au rejet, mais pas les valeurs petites, pourtant peu probables sous H_0. C'est simplement dû au fait que les petites valeurs de H_1 ne sont pas non plus probables sous H_1, empêchant le test sur les petites valeurs d'être discriminant.*

Les hypothèses des tests traitent de propriétés que possèdent (ou pas) la mesure de probabilité ayant servi à générer l'échantillon. En ce sens, on peut les voir comme des parties de l'ensemble des mesures sur $\mathbb{R}$ ou $\mathbb{R}^d$. Dans le cas le plus simple, H_0 est un singleton et H_1 son complémentaire. Cependant, dans certains cas, il n'y a pas de loi précise qui apparaisse simplement comme hypothèse naturelle, mais plutôt une famille de lois (par exemples des gaussiennes ou des lois de Poisson).

Test d'adéquation à une famille de lois

On souhaite maintenant mettre en place un test permettant de décider si la loi de l'échantillon appartient ou non à une famille de lois $(p_\theta)_{\theta\in\Theta}$ indexée par un paramètre θ à valeurs dans $\Theta \subset \mathbb{R}^d$. On suppose que l'on dispose de $\hat{\theta}_n$, l'estimateur du maximum de vraisemblance de θ. On a besoin d'un théorème un peu plus puissant (et difficile) que le précédent. Sa démonstration est délicate car elle fait intervenir les propriétés des estimateurs du maximum de vraisemblance.

Théorème 13.22. *Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes de même loi p_θ (avec $\theta \in \Theta$ inconnue). Notons $\hat{\theta}_n$ l'estimateur du maximum de vraisemblance de θ et $\mu_j(\hat{\theta}_n) = p_{\hat{\theta}_n}(E_j)$. Alors*

$$T'_n = \sum_{j=1}^{q} \frac{(N_j^n - n\mu_j(\hat{\theta}_n))^2}{n\mu_j(\hat{\theta}_n)} = n \sum_{j=1}^{q} \frac{(\frac{N_j^n}{n} - \mu_j(\hat{\theta}_n))^2}{\mu_j(\hat{\theta}_n)}.$$

converge en loi vers une variable de loi du χ^2_{q-d-1} à $q-d-1$ degrés de liberté.

Remarque 13.23. *La quantité T'_n quantifie l'écart entre la distribution empirique $\sum_{j=1}^q \frac{N_j^n}{n}\delta_j$ et la loi qui, sous l'hypothèse nulle, maximise la vraisemblance.*

Remarque 13.24. *Pour les familles de lois classiques comme* $(\mathcal{B}(n,p))_{0<p<1}$*,* $(\mathcal{N}(\theta,1))_{\theta\in\mathbb{R}}$*, ou* $(\mathcal{P}(\theta))_{\theta>0}$*, l'estimateur du maximum de vraisemblance est l'estimateur de la méthode des moments (voir §13.3.2).*

Remarque 13.25. *La loi limite est toujours une loi du* χ^2 *car l'estimateur du maximum de vraisemblance possède dans une très grande majorité des cas le comportement suivant :* $\hat{\theta}_n$ *converge p.s. vers* θ *(on dit qu'il est fortement consistant) et* $\sqrt{n}(\hat{\theta}_n-\theta)$ *converge en loi vers une mesure gaussienne. On peut toutefois faire un commentaire qualitatif sur le nombre de degrés de liberté de la loi limite. En effet, il est naturel que celui-ci soit plus petit que* $q-1$ *puisque l'on compare les fréquences empiriques, non plus à une loi fixée, mais à la loi la plus vraisemblable dans une famille paramétrée au vu des observations. Il paraîtrait logique que* T_n' *soit d'une certaine façon plus petit que* T_n*. C'est bien ce qui se passe puisque la fonction de répartition de la loi* χ^2_{q-1} *est inférieure à celle de la loi* χ^2_{q-d-1} *(on peut penser à l'interprétation en terme de somme de carrés de variables aléatoires gaussiennes indépendantes). On dit que* T_n' *est stochastiquement inférieur à* T_n*.*

Corollaire 13.26. *Pour tester*

$$H_0 : \quad p \in \{p(\theta), \theta\in\Theta\} \text{ contre } H_1 : \quad p \notin \{p(\theta),\theta\in\Theta\},$$

on utilise la statistique T_n' *qui a les comportements suivants :*

1. T_n' *converge en loi vers une variable de loi du* χ^2_{q-d-1} *si* H_0 *est vraie,*
2. T_n' *converge presque sûrement vers* $+\infty$ *sinon.*

Test d'homogénéité

Les tests du χ^2 permettent aussi de tester l'homogénéité de plusieurs échantillons. On étudie un caractère pouvant prendre q valeurs $A_1,\dots,A_q$ (ou q modalités, ou à valeurs dans q classes). On dispose de ℓ échantillons $E_1,\dots,E_\ell$ différents. Pour tout $i\in\{1,\dots,q\}$, pour tout $j\in\{1,\dots,\ell\}$, on connaît l'effectif observé n_{ij} de la valeur A_i dans l'échantillon E_j (dit autrement, n_{ij} est le nombre d'individus possédant la modalité i de la variable dans l'échantillon E_j). On souhaite tester l'hypothèse

H_0 : les échantillons sont issus de la même population, contre H_1 : les échantillons n'ont pas même loi.

La mise en place pratique du test est la suivante. On définit

$$\begin{aligned} n_{i\cdot} &= n_{i1}+\cdots+n_{iq} \quad \text{qui est l'effectif de l'échantillon } i,\\ n_{\cdot j} &= n_{1j}+\cdots+n_{\ell j} \quad \text{qui est le nombre total des individus possédant } A_j,\\ n &= \sum_{i=1}^{\ell}\sum_{j=1}^{q} n_{ij} = \sum_{i=1}^{\ell} n_{i\cdot} = \sum_{j=1}^{q} n_{\cdot j} \quad \text{qui est le nombre total d'individus.} \end{aligned}$$

On suppose de plus que les proportions des effectifs restent constants : il existe des réels positifs $\theta_1, \dots, \theta_\ell$ de somme 1 tels que $n_{i.}/n = \theta_i$ pour tout n.

Théorème 13.27. *Définissons*

$$T_n = \sum_{i=1}^{\ell} \sum_{j=1}^{q} \frac{\left(n_{ij} - \frac{n_{i.} n_{.j}}{n}\right)^2}{\frac{n_{i.} n_{.j}}{n}}.$$

Si H_0 est vraie, alors T_n converge en loi vers une variable suivant la loi du $\chi^2_{(q-1)(\ell-1)}$ à $(q-1)(\ell-1)$ degrés de liberté. Et T_n tend p.s. vers $+\infty$ sinon.

Démonstration. On prouvera juste le premier point, ce qui permettra de réviser les vecteurs gaussiens. Sous l'hypothèse H_0, les échantillons sont issus de la même loi et il s'agit donc de variables aléatoires indépendantes, de même loi p.

D'après la preuve du théorème du χ^2, pour tout échantillon i, on a

$$\left(\frac{n_{i,j} - p_j n_{i.}}{\sqrt{n_{i.} p_j}}\right)_{1 \le j \le q} \Longrightarrow \mathcal{N}(0, M),$$

où M est un projecteur orthogonal de rang $q-1$.

Comme $\frac{n_{.j}}{n}$ tend en probabilité vers p_j, on en déduit, grâce au théorème de Slutsky, que

$$\left(\frac{n_{i,j} - p_j n_{i.}}{\sqrt{\frac{n_{i.} n_{.j}}{n}}}\right)_{1 \le j \le q} \Longrightarrow \mathcal{N}(0, M) \text{ et } \left(\frac{n_{i,j} - p_j n_{i.}}{\sqrt{\frac{n_{i.} n_{.j}}{n}}}\right)_{1 \le i \le \ell, 1 \le j \le q} \Longrightarrow \mathcal{N}(0, M)^{\otimes \ell}.$$

Cependant, on remarque que

$$n_{ij} - \frac{n_{i.} n_{.j}}{n} = (n_{ij} - p_j n_{i.}) - \frac{n_{i.}}{n} \sum_{k=1}^{\ell} (n_{kj} - p_j n_{k.}),$$

donc T_n est l'image de $\left(\frac{n_{i,j} - p_j n_{i.}}{\sqrt{\frac{n_{i.} n_{.j}}{n}}}\right)_{i,j}$ par l'application de $(\mathbb{R}^q)^\ell$ dans $\mathbb{R}$ qui à $(x_1, \dots, x_\ell)$ associe

$$\sum_{i=1}^{\ell} \sum_{j=1}^{q} \left(x_i^j - \frac{n_{i.}}{n} \sum_{k=1}^{\ell} x_k^j\right)^2 = \sum_{i=1}^{\ell} \left\| x_i - \frac{n_{i.}}{n} \sum_{k=1}^{\ell} x_k \right\|_2^2.$$

Soient donc $V_1, \dots, V_\ell$ des variables aléatoires indépendantes de loi $\mathcal{N}(0, M)$. D'après ce qui précède, T_n converge en loi vers $\sum_{i=1}^{\ell} \left\| V_i - \theta_i \sum_{k=1}^{\ell} V_k \right\|_2^2$.

Comme M est un projecteur orthogonal de rang $q-1$, il existe une base orthonormale $(f_1, \dots, f_q)$ de $\mathbb{R}^q$ telle que les variables $(\langle V_i, f_j \rangle)_{1 \le i \le \ell, 1 \le j \le q-1}$ soient des gaussiennes $\mathcal{N}(0,1)$ indépendantes et que $(\langle V_i, f_q \rangle) = 0$ presque sûrement pour tout i. On a alors

$$\sum_{i=1}^{\ell}\left\|V_i-\theta_i\sum_{k=1}^{\ell}V_k\right\|_2^2=\sum_{j=1}^{q-1}\sum_{i=1}^{\ell}\left(\langle V_i,f_j\rangle-\theta_i\sum_{k=1}^{\ell}\langle V_k,f_j\rangle\right)^2.$$

En utilisant le théorème de Cochran , on voit, comme dans la preuve du théorème du chi-deux, que $\sum_{i=1}^{\ell}\left(\langle V_i,f_j\rangle-\theta_i\sum_{k=1}^{\ell}\langle V_k,f_j\rangle\right)^2$ suit la loi du χ^2 à $\ell-1$ degrés de libertés. Par indépendance, $\sum_{i=1}^{\ell}\left\|V_i-\theta_i\sum_{k=1}^{\ell}V_k\right\|_2^2$ suit donc la loi du χ^2 à $(\ell-1)(q-1)$ degrés de liberté, ce qui achève la preuve. □

13.4.2 Les tests de Kolmogorov-Smirnov

Le test du χ^2 n'est pas bien adapté à la comparaison d'une statistique à une loi continue. Dans ce cas, on peut se tourner vers le test de Kolmogorov-Smirnov.

Test d'adéquation

On va commencer par énoncer un lemme un peu étonnant :

Lemme 13.28. *Soient $(X_1,\dots,X_n)$ un $n-$échantillon de la loi μ, de fonction de répartion F et F_n la fonction de répartition empirique associée :*

$$F_n(x)=\frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{]-\infty,x]}(X_k).$$

On suppose que F est continue. Alors, la loi de $D_n=\|F_n-F\|_\infty$ ne dépend pas de μ. On dit alors que D_n est une statistique libre.

Démonstration. Le résultat a déjà été démontré au cours de la preuve du théorème de Glivenko-Cantelli. □

Remarque 13.29. *Le calcul numérique de D_n est très simple, car il ne fait intervenir qu'un nombre fini de points.*

Prenant pour hypothèse H_0 : “l'échantillon suit la loi μ”, si $\mathbb{P}(D_n\le x)=\alpha$, on rejettera avec un risque α l'hypothèse H_0 si le D_n observé dépasse x.

À la statistique D_n, qui est sans doute la plus naturelle, on préfère parfois l'emploi d'autres statistiques assez proches, à savoir

$$D_n^+=\sup_{x\in\mathbb{R}}\left(F(x)-F_n(x)\right)\text{ et }D_n^-=\sup_{x\in\mathbb{R}}\left(F_n(x)-F(x)\right).$$

Il n'est pas difficile de voir que D_n^+ et D_n^- ont même distribution, et, bien sûr, que $D_n = \max(D_n^+, D_n^-)$.

Un résultat dû à Birnbaum et Tingey [2] nous dit que

$$\mathbb{P}(D_n^+ > \varepsilon) = \sum_{i=0}^{\lfloor n(1-\varepsilon)\rfloor} \binom{n}{i} \varepsilon \left(\varepsilon + \frac{i}{n}\right)^{i-1} \left(1 - \varepsilon - \frac{i}{n}\right)^{n-i}.$$

Il est également théoriquement possible de calculer explicitement la fonction de répartition de D_n (c'est une fonction polynomiale par morceau), mais on n'obtient pas des formules fermées comme pour D_n^+. À partir de cela, on peut construire des tables de la loi de D_n pour les petites valeurs de n. Pour les grandes valeurs de n, le calcul de la fonction de répartition de D_n est numériquement plus compliqué. Heureusement, on a le théorème suivant, dit de Kolmogorov-Smirnov, qui permet de bonnes approximations.

Théorème 13.30 (Kolmogorov–Smirnov)**.** *Soit* $(X_1, \dots, X_n)$ *un échantillon de la loi* μ *et* F_n *la fonction de répartition empirique associée :*

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{]-\infty,x]}(X_k).$$

On suppose de plus que F *est continue. Alors*

$$\sqrt{n}\|F_n - F\|_\infty \Longrightarrow \mu_{KS},$$

où μ_{KS} *est la loi de Kolmogorov-Smirnov, qui est caractérisée par*

$$\mu_{KS}(]-\infty, t]) = 1 + 2 \sum_{k=1}^{+\infty} (-1)^k \exp(-2k^2t^2).$$

Ainsi, comme la série $\sum_{k=1}^{+\infty} (-1)^k \exp(-2k^2t^2)$ converge très rapidement, il suffit de calculer les trois premiers termes pour obtenir quatre chiffres significatifs dès que $t > 0,56$.

En revanche, pour les petites valeurs de t, la convergence n'est, en pratique, pas suffisamment rapide. Heureusement, on dispose alors d'une formule bien plus efficace (voir Durbin [10]) :

$$\mu_{KS}(]-\infty, t]) = \frac{\sqrt{2\pi}}{t} \sum_{k=1}^{+\infty} \exp(-\frac{(2k-1)^2\pi^2}{8t^2}).$$

Remarque 13.31. *Notons que Smirnov a également démontré que*

$$\forall \lambda > 0 \quad \lim_{n\to+\infty} \mathbb{P}(\sqrt{n}D_n^+ > \lambda) = e^{-2\lambda^2}.$$

On peut se demander ce qui se passe lorsque l'échantillon suit une loi μ' différente de celle que l'on teste. La réponse découle facilement du théorème de Glivenko-Cantelli : si l'échantillon a la fonction de répartition F_1, et non F, alors $\|F_n - F\|_\infty$ converge vers $\|F_1 - F\|_\infty$, et donc $\sqrt{n}\|F_n - F\|_\infty$ tend vers l'infini.

Test d'homogénéité

Le test de Kolmogorov-Smirnov pour tester l'homogénéité de deux échantillons est très semblable au précédent. Supposons qu'on ait un premier échantillon $X_1, \dots, X_n$ de loi μ ayant la fonction de répartition continue F, et un deuxième échantillon $X_1', \dots, X_n'$ de loi ν ayant la fonction de répartition continue G. On veut tester l'hypothèse $H_0 : \mu = \nu$ contre l'hypothèse H_1 : $\mu \neq \nu$.

La statistique

$$D_{m,n} = \sup_{x \in \mathbb{R}} |F_n(x) - G_m(x)|$$

est encore une statistique libre et, lorsque m et n tendent vers l'infini, la quantité $\sqrt{\frac{mn}{m+n}} D_{m,n}$ converge en loi vers la distribution de Kolmogorov-Smirnov (résultat admis).

Un résultat dû à Gnedenko et Korolyuk [19] et Drion [8](voir également de Oliveira Carvalho [7]) nous dit que pour tout entier k, on a

$$\mathbb{P}\left(D_{n,n} > \frac{k}{n}\right) = 2 \sum_{i=1}^{\lfloor n/k \rfloor} (-1)^{i-1} \frac{\binom{2n}{n+i}}{\binom{2n}{n}}.$$

En admettant ce résultat, il n'est pas très difficile de démontrer que $\sqrt{n/2} D_{n,n}$ converge en loi vers la distribution de Kolmogorov-Smirnov (laissé en exercice au lecteur).

13.4.3 Exercice : test du χ^2

On étudie l'évolution de la taille d'une tumeur bénigne de forme circulaire (sans prise de médicament). Pour ce faire, on mesure le diamètre de la tumeur, en millimètres. On dispose d'un 500-échantillon, composé d'individus atteints par cette tumeur. Notons T la taille de la tumeur et E_T l'effectif correspondant à cette taille. Le résultat des mesures est donné ci-dessous :

T	4,5-4,6	4,6-4,7	4,7-4,8	4,8-4,9	4,9-5,0	5,0-5,1	5,1-5,2	5,2-5,3
E_T	20	33	69	137	121	67	42	11

1. On veut tester l'hypothèse H_0 : X suit une loi gaussienne, contre H_1 : X n'est pas gaussienne. Les paramètres de la loi gaussienne seront estimés par ceux de l'échantillon.

(a) Déterminer la moyenne $\bar{x}$ et la variance s^2 de cet échantillon.

(b) Peut-on rejeter l'appartenance de la loi de X à la famille des lois gaussiennes au risque de 1% ?

2. Un autre échantillon, de même taille, est relevé avec les résultats suivants :

T	4,5-4,6	4,6-4,7	4,7-4,8	4,8-4,9	4,9-5,0	5,0-5,1	5,1-5,2	5,2-5,3
E_T	15	53	82	105	117	76	39	13

Peut-on estimer qu'au risque de 5%, ces deux échantillons appartiennent à la même population ? On utilisera pour cela un test du χ^2.

Solution 1. Testons l'hypothèse H_0 : "X suit une loi gaussienne", contre H_1 : "X n'est pas gaussienne".

(a) On obtient par le calcul la moyenne de l'échantillon :

$$\bar{x} = \frac{4,55 \times 20 + 4,65 \times 33 + \cdots + 5,25 \times 11}{500} = 4,896$$

et sa variance est :

$$s^2 = \frac{4,55^2 \times 20 + 4,65^2 \times 33 + \cdots + 5,25^2 \times 11}{500} - \bar{x}^2 = 2,4364.10^{-2}.$$

(b) Il s'agit de tester l'appartenance de la loi observée à la famille des lois gaussiennes. On a vu plus haut que $\bar{x}$ et s^2 sont les estimateurs du maximum de vraisemblance, ce qui permet d'utiliser le corollaire 13.26. On a, en notant E_o l'effectif observé, p_{th} la probabilité théorique de l'intervalle et $E_{\text{th}} = 500p_{\text{th}}$ l'effectif théorique :

T	$<$ 4,6	4,6-4,7	4,7-4,8	4,8-4,9	4,9-5,0	5,0-5,1	5,1-5,2	$>$ 5,2
E_o	20	33	69	137	121	67	42	11
F	0,029	0,1046	0,2693	0,5102	0,7474	0,9044	0,9743	1
p_{th}	0,029	0,0757	0,1647	0,241	0,2372	0,157	0,0699	0,0257
E_{th}	14,48	37,83	82,33	120,48	118,58	78,5	34,94	12,87

Précisons aussi que pour trouver les valeurs de F, il ne faut surtout pas oublier de centrer et réduire notre variable. Ainsi, $0,029$ correspond au fractile de la loi normale standard $\mathcal{N}(0,1)$ évalué en $(4,6-\bar{x})/s$. En formant les différences, on construit les probabilités théoriques. L'effectif théorique correspondant s'obtient en multipliant par 500.

On calcule alors la valeur du χ^2 observé, qui s'écrit tout simplement comme

$$\chi^2 = \frac{(20-14,48)^2}{14,48} + \frac{(33-37,83)^2}{37,83} + \cdots + \frac{(11-12,87)^2}{12,87} = 10,57.$$

De plus, on sait qu'asymptotiquement, on a une loi du χ^2 dont le degré de liberté est "nombre de classes - nombre de paramètres inconnus-1", soit $8-2-1=5$ degrés de liberté (car ici, on a 8 classes et deux paramètres inconnus, qui sont la moyenne et la variance de la gaussienne). On obtient alors dans la table du χ^2 que la valeur théorique à 5 degrés de liberté au risque de 1% est $\chi_0^2 = 15,09$.

Afin de conclure si on rejette l'hypothèse H_0 ou pas, il faut comparer la valeur observée χ^2 et la valeur théorique χ_0^2. On a ici

$$\chi^2 = 10,57 < 15,09 = \chi_0^2.$$

On conclut donc qu'au risque 1%, on ne rejette pas l'hypothèse de normalité de la loi. Cependant, il faudrait encore déterminer la valeur maximale α_m de non-rejet de H_0. En général, il convient d'être prudent (et éventuellement de procéder à une nouvelle série de mesures). Remarquons également que la bonne approximation de la variance n'est pas s^2, mais $\frac{n}{n-1}s^2$ qui est un estimateur sans biais de la variance.

Bien sûr, de nos jours, les logiciels de calcul remplacent avantageusement l'usage des tables et libèrent de calculs fastidieux qui sont sources d'erreurs. Ci-dessous un exemple en Julia.

```
using Distributions
function Calcul_chi_deux(obs,ref)
    th=ref./sum(ref) ; effectif=sum(obs)
    dobs=obs./effectif
    return(effectif*sum(((th-dobs).^2)./th))
end

a=collect(4.55:0.1:5.25); b=[20,33,69,137,121,67,42,11]
xbarre=sum(a.*b)/sum(b)
sdeux=sum(a.*a.*b)/sum(b)-xbarre^2; s=sqrt(sdeux)
F=(cdf.(Normal(xbarre,s),a.+0.05))
dth=[F[1:7];1]-[0;F[1:7]]
chi_deux=Calcul_chi_deux(b,dth); risque=0.01
chi_deux_zero=cquantile(Chisq(length(b)-3),risque)
```

2. Faisons un test du χ^2 d'homogénéité. L'hypothèse nulle à tester est

H_0 : les deux échantillons proviennent de la même population.

On obtient le tableau suivant en notant $E1$ l'effectif 1 et $E2$ l'effectif 2 :

T	4,5-4,6	4,6-4,7	4,7-4,8	4,8-4,9	4,9-5,0	5,0-5,1	5,1-5,2	5,2-5,3
$E1$	20	33	69	137	121	67	42	11
$E2$	15	53	82	105	117	76	39	13
Total	35	86	151	242	238	143	81	24

Remarquant que le total de chaque effectif vaut 500 (et la somme 1000),

$$\chi^2 = 1000\left(\frac{20^2}{500\times 35}+\frac{33^2}{500\times 86}+\cdots+\frac{11^2}{500\times 24}\right)+$$
$$+1000\left(\frac{15^2}{500\times 35}+\frac{53^2}{500\times 86}+\cdots+\frac{13^2}{500\times 24}-1\right)=11,63.$$

Or on sait que si H_0 est vraie, alors la loi suit une loi du χ^2 à "(nombre de colonnes-1)$\times$ (nombre de lignes -1)" degrés de liberté, soit $(2-1)\times(8-1)=7$ degrés de liberté. On trouve dans la table du χ^2 à 7 degrés de liberté, au niveau de risque 5%, $\chi_0^2 = 14,07$.
Comme $\chi^2 = 11,63 < 14,07 = \chi_0^2$, on conclut qu'on ne rejette pas l'hypothèse H_0 : au risque de 5%, les deux échantillons proviennent de la même population.

13.5 Exercices de statistiques

13.5.1 Exercices corrigés

Exercice 280. On effectue une enquête, durant une épidémie de grippe, dans le but de connaître la proportion p de personnes présentant des complications graves. On observe un échantillon représentatif de 400 personnes et pour un tel échantillon, 40 personnes ont présenté des complications.

1. Donner un intervalle de confiance pour p au risque 5%.
2. On désire que la valeur estimée $\hat{p}$ diffère de la proportion inconnue exacte p de moins de $0,005$ avec une probabilité égale à 95%. Quel sera l'effectif d'un tel échantillon ?
3. Quel devrait être le risque pour obtenir le même intervalle qu'à la question précédente en conservant l'effectif $n = 400$. Quelles conclusions peut-on en tirer ?

$\rightarrow$ indication $\rightarrow$ solution

Exercice 281. *Concentration pour la méthode de Monte-Carlo.*
Soient $g : [0,1]^d \to \mathbb{R}$ une fonction dans $L^\infty([0,1]^d)$ et $(X_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes et identiquement distribuées de loi uniforme sur $[0,1]^d$. On se propose ici de montrer une inégalité de concentration pour la méthode de Monte-Carlo.
Montrer que pour tout $\varepsilon \in]0, \frac{\|g\|_2^2}{\|g\|_\infty}[$, on a

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n g(X_i) - \int_{[0,1]^d} g(x)\,dx\right| > \varepsilon\right) \leq 2\exp\left(-\frac{n\varepsilon^2}{4\|g\|_2^2}\right).$$

$\rightarrow$ indication $\rightarrow$ solution

Exercice 282. Déterminer, pour une loi de Poisson de paramètre λ inconnu, l'estimateur du maximum de vraisemblance. $\rightarrow$ indication $\rightarrow$ solution

Exercice 283. On a enregistré, pendant $n = 200$ jours, le nombre d'appels téléphoniques qui ont eu lieu entre 18h00 et 18h15 dans un standard téléphonique. On a obtenu les résultats suivants :

Nb d'appels	0	1	2	3	4	5	6	7	8	9	10	11	12
Nb de jours	1	16	31	37	41	30	23	13	6	1	0	1	0

1. Calculer la moyenne et la variance de cette distribution.
2. Par quelle loi de probabilité peut-on ajuster la distribution observée ?
3. Quel test statistique utiliseriez-vous pour vérifier le bien-fondé de votre choix ? Quelle est votre conclusion, au risque de 5 % ?

$\rightarrow$ indication $\rightarrow$ solution

Exercice 284. Une enquête épidémiologique portant sur 1000 diabétiques hospitalisés a été réalisée sur une période de trois ans afin de déterminer la fréquence de l'hypertension artérielle (HTA) et les facteurs de risque cardiovasculaires associés. Les patients âgés de 30 à 80 ans ont été répartis en deux groupes : 100 patients entrent dans le groupe I, les diabétiques hypertendus, et 900 patients entrent dans le groupe II, les diabétiques non hypertendus. Certains résultats ont été réunis dans le tableau ci-dessous.

	Groupe I	Groupe II
Hommes / Femmes	45 / 55	465 / 435
- de 50 ans / + 50 ans	30 / 70	420 / 480

1. Tester l'hypothèse selon laquelle la population étudiée de diabétiques est composée d'autant d'hommes que de femmes.
2. Le pourcentage de patients de plus de 50 ans est-il significativement différent chez les hypertendus que chez les non hypertendus ?

$\rightarrow$ indication $\rightarrow$ solution

13.5.2 Exercices non corrigés

Exercice 285. Soit $X = (X_1, \ldots, X_n)'$ un vecteur gaussien de composantes indépendantes, où X_i suit la loi $\mathcal{N}(m, \sigma^2)$. Posons

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2, \quad S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

1. Quelle est la loi de $\bar{X}$?
2. Quelle est la loi de $n\frac{\tilde{S}^2}{\sigma^2}$?
3. Montrer que $n\frac{S^2}{\sigma^2}$ suit une loi χ^2_{n-1}.
4. Montrer que $\bar{X}$ et S^2 sont indépendantes.

$\rightarrow$ indication

Exercice 286. *Convergence de la loi multinomiale.*
Soient $p_1, \ldots, p_k$ des réels tels que $0 < p_i < 1$ et $p_1 + \cdots + p_k = 1$. Soient $n, n_1, \ldots, n_k$ des entiers positifs tels que $0 \leq n_i \leq n$ et $\sum\limits_{i=1}^{k} n_i = n$. Soit $X = (X_1, \ldots, X_k)'$ un vecteur aléatoire à valeurs dans $\mathbb{N}^k$ tel que

$$\mathbb{P}(X_1 = n_1, \ldots, X_k = n_k) = \frac{n!}{n_1! \ldots n_k!} p_1^{n_1} \ldots p_k^{n_k}.$$

On dit que X suit la loi multinomiale de paramètres $p = (p_1, \ldots, p_k)$ et n. Montrer que $\sqrt{n}\left(\frac{X}{n} - p\right)$ converge en loi vers une variable de loi $\mathcal{N}(0, P - pp')$,

où $P = \begin{pmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & p_k \end{pmatrix}$ et $p = \begin{pmatrix} p_1 \\ \vdots \\ p_k \end{pmatrix}$. $\rightarrow$ indication

Exercice 287. Soient deux populations de chevaux : les bons sauteurs et les mauvais sauteurs. On étudie la hauteur du garrot que l'on suppose distribuée normalement dans les deux populations. Pour cela, on prélève un échantillon dans chacune des deux populations, ce qui donne les résultats suivants :

	Effectif	Moyenne	Écart-type estimé
Bons sauteurs	$n_1 = 55$	$\bar{X}_1 = 164$	$\hat{\sigma}_1 = 4,7$
Mauvais sauteurs	$n_2 = 50$	$\bar{X}_2 = 161,5$	$\hat{\sigma}_2 = 5,2$

1. La différence des moyennes observées est-elle significative (au risque de 5%) ?
2. Même question mais en supposant que les effectifs sont : $n_1 = 12$ et $n_2 = 10$.

$\rightarrow$ indication

Exercice 288. Il arrive que le coucou vienne pondre subrepticement dans le nid d'un autre oiseau, comme la fauvette ou le roitelet, qui élève alors ces petits intrus plutôt que ses propres petits.

Dans un article de la revue Biometrica, le biologiste Latter donne la longueur L (en mm) des œufs de coucou trouvés dans les nids de ces deux espèces d'oiseaux :

— Dans les nids de petite taille (roitelet) :

19,8	22,1	21,5	20,9	22,0	21,0	22,03	21,0
20,3	20,9	22,0	22,0	20,8	21,2	21,0	

— Dans les nids de taille plus grande (fauvette) :

22,0	23,9	20,9	23,8	25,0	24,0	23,8
21,7	22,8	23,1	23,5	23,0	23	23,1

1. Donner une estimation ponctuelle de la moyenne et de la variance de L pour chacune des deux populations : œufs de coucou dans des nids de roitelet, œufs de coucou dans des nids de fauvette.
2. En supposant que L suive une loi normale dans chacune des deux populations et que les variances ne soient pas significativement différentes, tester l'hypothèse “le coucou adapte la taille de ses œufs à la taille du nid dans lequel il pond”.
3. Tester l'hypothèse d'homogénéité des variances de la question précédente.

$\rightarrow$ indication

Exercice 289. Lors d'une mission en mer, deux méthodes sont employées pour estimer la concentration de chlorophylle a (en mg/l) dans la masse d'eau. La première (S) plus grossière et plus ancienne, mais servant toujours de référence, est un dosage par spectrométrie. La seconde (H), plus précise mais encore expérimentale, est un dosage par HPLC. On obtient

Essai	1	2	3	4	5	6	7	8
S	5,7	8,2	6,9	6,0	3,8	3,9	4,3	2,7
H	4,0	5,8	4,9	4,8	3,6	3,5	2,9	1,2

Les résultats des deux méthodes diffèrent-ils significativement (à 1% de risque) ?

$\rightarrow$ indication

Exercice 290. Sur un échantillon de 40 mollusques, 9 indiquent la présence d'un parasite associé à la bilharziose.

1. Estimer la proportion de mollusques porteurs du parasite sur la population entière.
2. Donner l'intervalle de confiance de cette proportion au risque $\alpha = 5\%$.
3. Si le nombre de mollusques infectés avait été de 3, aurait-il encore été possible de calculer un intervalle de confiance ? (Justifier votre réponse.)

$\rightarrow$ indication

Chapitre 14

Sommes de variables aléatoires indépendantes

14.1 Théorèmes de Lindeberg et de Lyapounov

14.1.1 Théorème de Lindeberg

Théorème 14.1. *Soient $(X_{n,k})_{n,k\geq 1}$ des variables aléatoires centrées admettant un moment d'ordre 2 et $(N_n)_{n\geq 1}$ une suite d'entiers tendant vers l'infini. On suppose que pour tout n, la suite $(X_{n,k})_{1\leq k\leq N_n}$ est formée de variables indépendantes. On pose*

$$S_n = \sum_{k=1}^{N_n} X_{n,k} \quad \textit{et} \quad \sigma_n^2 = \mathrm{Var}\, S_n.$$

On suppose que la condition de Lindeberg est satisfaite, c'est-à-dire :

$$\forall \varepsilon > 0, \qquad \lim_{n\to+\infty} \sum_{k=1}^{N_n} \frac{1}{\sigma_n^2} \mathbb{E}\left(X_{n,k}^2 \mathbb{1}_{\{|X_{n,k}|\geq \varepsilon\sigma_n\}}\right) = 0.$$

Alors S_n/σ_n converge en loi vers une variable aléatoire de loi $\mathcal{N}(0,1)$.

Démonstration. Notons $\phi_{n,k}$ la fonction caractéristique de $\frac{X_{n,k}}{\sigma_n}$. La fonction caractéristique de S_n/σ_n est alors $\prod_{k=1}^{N_n} \phi_{n,k}$. L'idée est bien sûr de montrer que, pour tout t réel, $\prod_{k=1}^{N_n} \phi_{n,k}(t)$ converge vers $\exp(-\frac{t^2}{2})$ et d'appliquer le théorème de Lévy 11.15.
Notons $\psi_{n,k}$ la fonction caractéristique d'une variable gaussienne qui a même

espérance et même variance que $\frac{X_{n,k}}{\sigma_n}$, soit $\psi_{n,k}(t) = \exp(-\frac{t^2}{2}\frac{\operatorname{Var} X_{n,k}}{\sigma_n^2})$. On a

$$\prod_{k=1}^{N_n} \phi_{n,k}(t) - \exp\left(-\frac{t^2}{2}\right) = \prod_{k=1}^{N_n} \phi_{n,k}(t) - \prod_{k=1}^{N_n} \psi_{n,k}(t)$$

car $\frac{1}{\sigma_n^2}\sum_{k=1}^{n} \operatorname{Var} X_{n,k} = 1$. On va maintenant procéder comme dans la preuve du théorème central limite : la différence de deux produits de nombres complexes, dont les modules ne dépassent pas 1, n'excède pas la somme des différences [1], d'où

$$\left|\prod_{k=1}^{N_n} \phi_{n,k}(t) - \exp(-\frac{t^2}{2})\right| \leq \sum_{k=1}^{N_n} |\phi_{n,k}(t) - \psi_{n,k}(t)|.$$

Soit α un réel positif [2] tel que

$$\forall x \in \mathbb{R}, \quad |e^{ix} - (1 + ix - x^2/2)| \leq \alpha \min(x^2, |x|^3).$$

Appliquons cette inégalité à $x = \frac{tX_{n,k}}{\sigma_n}$, puis prenons l'espérance. On obtient alors que

$$\left|\phi_{n,k}(t) - \left(1 - \frac{1}{2}t^2\frac{\operatorname{Var} X_{n,k}}{\sigma_n^2}\right)\right| \leq \alpha\mathbb{E}Y,$$

où

$$Y = \min\left(t^2\left(\frac{X_{n,k}}{\sigma_n}\right)^2, t^3\left(\frac{|X_{n,k}|}{\sigma_n}\right)^3\right).$$

On peut écrire :

$$\mathbb{E}Y = \mathbb{E}\left[Y\ \mathbb{1}_{\{|X_{n,k}|/\sigma_n < \varepsilon\}}\right] + \mathbb{E}\left[Y\ \mathbb{1}_{\{|X_{n,k}|/\sigma_n \geq \varepsilon\}}\right].$$

Lorsque $\frac{|X_{n,k}|}{\sigma_n}$ est petit, on a intérêt à l'élever à une plus grande puissance, donc on écrit :

$$\begin{aligned}
\mathbb{E}Y &\leq |t|^3\mathbb{E}\left[\frac{|X_{n,k}|^3}{\sigma_n^3}\ \mathbb{1}_{\left\{\frac{|X_{n,k}|}{\sigma_n} < \varepsilon\right\}}\right] + t^2\mathbb{E}\left[X_{n,k}^2\frac{1}{\sigma_n^2}\ \mathbb{1}_{\left\{\frac{|X_{n,k}|}{\sigma_n} \geq \varepsilon\right\}}\right]\\
&\leq \varepsilon|t|^3\frac{\operatorname{Var} X_{n,k}}{\sigma_n^2} + t^2\frac{1}{\sigma_n^2}\mathbb{E}\left[X_{n,k}^2\ \mathbb{1}_{\{|X_{n,k}| \geq \varepsilon\sigma_n\}}\right].
\end{aligned}$$

Ainsi

$$\sum_{k=1}^{N_n}\left|\phi_{n,k}(t) - \left(1 - \frac{1}{2}t^2\frac{\operatorname{Var} X_{n,k}}{\sigma_n^2}\right)\right| \leq \alpha\varepsilon|t|^3 + \alpha t^2\sum_{k=1}^{N_n}\frac{1}{\sigma_n^2}\mathbb{E}\Big[X_{n,k}^2\ \mathbb{1}_{\{|X_{n,k}| \geq \varepsilon\sigma_n\}}\Big].$$

1. Lorsqu'il n'y a que deux termes, la preuve résulte de $zu - z'u' = z(u - u') + u'(z - z')$; dans le cas général, on procède par récurrence.
2. Vérifier l'existence…

Comme ε est arbitraire, on voit en utilisant la condition de Lindeberg que

$$\forall t \in \mathbb{R}, \quad \lim_{n\to+\infty} \sum_{k=1}^{N_n} \left| \phi_{n,k}(t) - \left(1 - \frac{t^2}{2}\frac{\operatorname{Var} X_{n,k}}{\sigma_n^2}\right)\right| = 0.$$

Il ne nous reste donc plus qu'à montrer que

$$\forall t \in \mathbb{R}, \quad \lim_{n\to+\infty} \sum_{k=1}^{N_n} \left| \psi_{n,k}(t) - \left(1 - \frac{t^2}{2}\frac{\operatorname{Var} X_{n,k}}{\sigma_n^2}\right)\right| = 0,$$

c'est-à-dire que

$$\forall t \in \mathbb{R}, \quad \lim_{n\to+\infty} \sum_{k=1}^{N_n} \left| \exp\left(-\frac{t^2}{2}\frac{\operatorname{Var} X_{n,k}}{\sigma_n^2}\right) - \left(1 - \frac{t^2}{2}\frac{\operatorname{Var} X_{n,k}}{\sigma_n^2}\right)\right| = 0.$$

Mais pour tout $z \in \mathbb{C}$, on a en considérant le développement en série entière de l'exponentielle : $|e^z - 1 - z| \leq \frac{e^{|z|}}{2}|z|^2$, d'où

$$\sum_{k=1}^{N_n} \left| \exp\left(-\frac{t^2}{2}\frac{\operatorname{Var} X_{n,k}}{\sigma_n^2}\right) - \left(1 - \frac{t^2}{2}\frac{\operatorname{Var} X_{n,k}}{\sigma_n^2}\right)\right| \leq \frac{t^4}{8} e^{t^2/2} \frac{1}{\sigma_n^4} \sum_{k=1}^{N_n} (\operatorname{Var} X_{n,k})^2.$$

Par ailleurs, si l'on pose $M_n = \max(\operatorname{Var} X_{n,k}; 1 \leq k \leq N_n)$, on a alors

$$\frac{1}{\sigma_n^4} \sum_{k=1}^{N_n} (\operatorname{Var} X_{n,k})^2 \leq \frac{1}{\sigma_n^4} \sum_{k=1}^{N_n} M_n \operatorname{Var} X_{n,k} = \frac{M_n}{\sigma_n^2}.$$

Ainsi, pour conclure la preuve, il suffit de montrer que $\frac{M_n}{\sigma_n^2}$ tend vers 0.

Soient $\varepsilon \in]0, 1/2]$ et n un entier. Pour tout $1 \leq i \leq N_n$, on a

$$\begin{aligned} \mathbb{E}\left[\frac{X_{n,i}^2}{\sigma_n^2}\right] &= \frac{1}{\sigma_n^2}\mathbb{E}\left[X_{n,i}^2 \, \mathbb{1}_{\{|X_{n,i}|<\varepsilon\sigma_n\}}\right] + \frac{1}{\sigma_n^2}\mathbb{E}\left[X_{n,i}^2 \, \mathbb{1}_{\{|X_{n,i}|\geq\varepsilon\sigma_n\}}\right] \\ &\leq \varepsilon^2 + \frac{1}{\sigma_n^2}\mathbb{E}\left[X_{n,i}^2 \, \mathbb{1}_{\{|X_{n,i}|\geq\varepsilon\sigma_n\}}\right], \end{aligned}$$

d'où

$$\frac{M_n}{\sigma_n^2} \leq \varepsilon^2 + \sum_{k=1}^{N_n} \frac{1}{\sigma_n^2}\mathbb{E}\left[X_{n,k}^2 \, \mathbb{1}_{\{|X_{n,k}|\geq\varepsilon\sigma_n\}}\right].$$

D'après la condition de Lindeberg, pour n assez grand, le deuxième terme ne dépasse pas ε^2 et donc $\frac{M_n}{\sigma_n^2} \leq 2\varepsilon^2 \leq \varepsilon$, ce qui montre que $\frac{M_n}{\sigma_n^2}$ tend vers 0. $\square$

Comme corollaire, on retrouve évidemment le Théorème Central Limite.

Théorème 14.2. *Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires réelles indépendantes identiquement distribuées admettant un moment d'ordre 2. On note m*

leur espérance commune, σ^2 leur variance commune. Alors

$$\frac{(X_1 + \cdots + X_n) - nm}{\sqrt{n}} \underset{n \to +\infty}{\Longrightarrow} \mathcal{N}(0, \sigma^2).$$

Démonstration. La preuve est laissée en exercice. □

14.1.2 Condition de Lyapounov

Soient $(X_{n,k})_{n,k\geq 1}$ des variables aléatoires centrées admettant un moment d'ordre 2 et $(N_n)_{n\geq 1}$ une suite d'entiers tendant vers l'infini. On suppose que pour tout n, la suite $(X_{n,k})_{1\leq k\leq N_n}$ est formée de variables indépendantes. Soit $\delta > 0$. Si les variables $X_{n,k}$ admettent un moment d'ordre $2+\delta$ et satisfont la *condition de Lyapounov*, c'est-à-dire

$$\lim_{n\to+\infty} \sum_{k=1}^{N_n} \sigma_n^{-(2+\delta)} \mathbb{E}\left[|X_{n,k}|^{2+\delta}\right] = 0,$$

alors la condition de Lindeberg

$$\lim_{n\to+\infty} \sum_{k=1}^{N_n} \frac{1}{\sigma_n^2} \mathbb{E}\left[X_{n,k}^2 \, \mathbb{1}_{\{|X_{n,k}|\geq \varepsilon\sigma_n\}}\right] = 0$$

est vérifiée et le théorème de Lindeberg s'applique. En effet, il suffit de remarquer que

$$\mathbb{1}_{\{|X_n|\geq \varepsilon\sigma_n\}} \leq \left(\frac{|X_n|}{\varepsilon\sigma_n}\right)^{\delta}.$$

Proposition 14.3. *Soient $(X_{n,k})_{n,k\geq 0}$ une suite de variables aléatoires centrées indépendantes bornées par une constante M, et $(N_n)_{n\geq 1}$ une suite d'entiers tendant vers l'infini. On pose*

$$S_n = \sum_{k=1}^{N_n} X_{n,k} \quad \textit{et} \quad \sigma_n^2 = \operatorname{Var} S_n.$$

Si σ_n tend vers l'infini lorsque n tend vers l'infini, alors S_n/σ_n converge en loi vers la loi $\mathcal{N}(0,1)$.

Démonstration. Soit $\delta > 0$ quelconque. On a :

$$\sum_{k=1}^{N_n} \frac{1}{\sigma_n^{2+\delta}} \mathbb{E}\left[X_{n,k}^{2+\delta}\right] \leq \sum_{k=1}^{N_n} \frac{M^\delta}{\sigma_n^{2+\delta}} \mathbb{E}\left[X_{n,k}^2\right] = \frac{M^\delta}{\sigma_n^\delta},$$

qui tend bien vers 0 lorsque n tend vers l'infini. Ainsi le critère de Lyapounov est vérifié, donc celui de Lindeberg aussi, et le théorème 14.1 s'applique. □

14.2 Sommes et séries de variables indépendantes

14.2.1 Loi du zéro-un

Théorème 14.4. *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes. L'événement "La série de terme général X_n converge" est de probabilité 0 ou 1.*

Démonstration. Pour $k \geq 0$, notons A_k l'événement "La série de terme général $(X_{n+k}, n \geq 0)$ converge". Par construction, A_k est $\sigma(X_i, i \geq k)$-mesurable. Comme la nature d'une série ne dépend pas des premiers termes, on a pour tout k, $A_k = A_0$. Ainsi, pour tout $k \geq 0$, $A_0 \in \sigma(X_i, i \geq k)$, ce qui signifie que $A_0 \in \mathcal{Q} = \bigcap_{k\geq 1} \sigma(X_i, i \geq k)$. Cette tribu est la tribu de queue associée à une famille de variables aléatoires indépendantes sous $\mathbb{P}$. D'après la loi du $0-1$ de Kolmogorov, elle est triviale sous $\mathbb{P}$.

Ainsi, la probabilité $\mathbb{P}(A_0)$ ne peut valoir que 0 ou 1. □

14.2.2 Une inégalité maximale

Théorème 14.5. *Soient $X_1, \dots, X_n$ n variables aléatoires indépendantes. On note $(S_k)_{0\leq k\leq n}$ la suite des sommes partielles : $S_0 = 0$, $S_k = X_1 + \dots + X_k$. Pour tout $\alpha > 0$, on a alors*

$$\mathbb{P}\left(\max_{1\leq k\leq n} |S_k| \geq 4\alpha \right) \leq 4 \max_{1\leq k\leq n} \mathbb{P}(|S_k| \geq \alpha).$$

Démonstration. Notons $T = \inf\{k \in \{1, \dots, n\}; |S_k| \geq 4\alpha\}$, où nous convenons que $T = +\infty$ si $\{k \in \{1, \dots, n\}; |S_k| \geq 4\alpha\} = \varnothing$. T est un temps d'arrêt, ce qui signifie que pour tout $k \geq 1$, $\{T = k\}$ est $\sigma(X_1, \dots, X_k)-$mesurable.

1) On peut noter que $\mathbb{P}\left(\max_{1\leq k\leq n} |S_k| \geq 4\alpha \right) = \mathbb{P}(T < +\infty)$. Montrons tout d'abord que

$$\mathbb{P}\left(\max_{1\leq k\leq n} |S_k| \geq 4\alpha \right) = \mathbb{P}(T < +\infty) \leq 2\mathbb{P}\left(\max_{1\leq k\leq n} |S_n - S_k| \geq 2\alpha \right).$$

On a

$$\begin{aligned}
\{T < +\infty\} &= (\{T < +\infty\} \cap \{|S_n| \geq 2\alpha\}) \cup (\{T < +\infty\} \cap \{|S_n| < 2\alpha\}) \\
&\subset \{|S_n| \geq 2\alpha\} \cup (\{T < +\infty\} \cap \{|S_n| < 2\alpha\}) \\
&= \{|S_n| \geq 2\alpha\} \bigcup \bigcup_{k=1}^{n-1} (\{T = k\} \cap \{|S_n| < 2\alpha\}) \\
&\subset \{|S_n| \geq 2\alpha\} \bigcup \bigcup_{k=1}^{n-1} (\{T = k\} \cap \{|S_n - S_k| \geq 2\alpha\}).
\end{aligned}$$

Cela implique alors

$$\mathbb{P}(T < +\infty) \leq \mathbb{P}(|S_n| \geq 2\alpha) + \sum_{k=1}^{n-1} \mathbb{P}(\{T = k\} \cap \{|S_n - S_k| \geq 2\alpha\}).$$

Mais $\{T = k\}$ appartient à la tribu $\sigma(X_1, \ldots, X_k)$ tandis que $S_n - S_k$ est $\sigma(X_{k+1}, \ldots, X_n)$-mesurable. L'indépendance de ces deux tribus donne alors

$$\begin{aligned}\mathbb{P}(T < +\infty) &\leq \mathbb{P}(|S_n| \geq 2\alpha) + \sum_{k=1}^{n-1} \mathbb{P}(T = k)\mathbb{P}(|S_n - S_k| \geq 2\alpha) \\ &\leq \beta + \sum_{k=1}^{n-1} \mathbb{P}(T = k)\beta \leq 2\beta,\end{aligned}$$

où l'on a posé $\beta = \max\limits_{0 \leq k \leq n} \mathbb{P}(|S_n - S_k| \geq 2\alpha)$. On en déduit que

$$\mathbb{P}(T < +\infty) = \mathbb{P}\left(\max_{1 \leq k \leq n} |S_k| \geq 4\alpha\right) \leq 2 \max_{0 \leq k \leq n} \mathbb{P}(|S_n - S_k| \geq 2\alpha).$$

Mais par ailleurs, on a

$$\max_{0 \leq k \leq n} \mathbb{P}(|S_n - S_k| \geq 2\alpha) \leq \mathbb{P}\left(\max_{0 \leq k \leq n} |S_n - S_k| \geq 2\alpha\right).$$

On a donc bien $\mathbb{P}\left(\max\limits_{1 \leq k \leq n} |S_k| \geq 4\alpha\right) \leq 2\mathbb{P}\left(\max\limits_{1 \leq k \leq n} |S_n - S_k| \geq 2\alpha\right)$.

2) Afin d'améliorer l'inégalité précédente, nous posons maintenant, pour tout $k \in \{0, \ldots, n\}$, $S'_k = S_n - S_{n-k}$.
En d'autres termes, on a $S'_k = X'_1 + X'_2 + \cdots + X'_k$, avec $X'_k = X_{n-k}$. On a bien sûr $S'_0 = 0$ et $S'_n = S_n$. On répète le raisonnement précédent, cette fois avec $T' = \inf\{k \in \{1, \ldots, n\}; |S'_k| \geq 2\alpha\}$, où on définit à nouveau $T' = +\infty$ si $\{k \in \{1, \ldots, n\}; |S'_k| \geq 2\alpha\} = \varnothing$. On obtient ainsi

$$\begin{aligned}&\{\max_{0 \leq k \leq n} |S_n - S_k| \geq 2\alpha\} \\ = \ &\{T' < +\infty\} \\ = \ &(\{T' < +\infty\} \cap \{|S'_n| \geq \alpha\}) \cup (\{T' < +\infty\} \cap \{|S'_n| < \alpha\}) \\ \subset \ &\{|S'_n| \geq \alpha\} \cup (\{T' < +\infty\} \cap \{|S'_n| < \alpha\}) \\ = \ &\{|S'_n| \geq \alpha\} \cup \bigcup_{k=1}^{n-1} (\{T' = k\} \cap \{|S'_n| < \alpha\}) \\ \subset \ &\{|S'_n| \geq \alpha\} \cup \bigcup_{k=1}^{n-1} (\{T' = k\} \cap \{|S'_n - S'_k| < \alpha\}).\end{aligned}$$

D'où

$$\mathbb{P}(T' < +\infty) \leq \mathbb{P}(|S'_n| \geq \alpha) + \sum_{k=1}^{n-1} \mathbb{P}(\{T' = k\} \cap \{|S'_n - S'_k| < \alpha\}).$$

Comme $\{T'=k\}$ et $S'_n - S'_k$ sont indépendants, cela donne

$$\begin{aligned}
\mathbb{P}(T'<+\infty) &\leq \mathbb{P}(|S'_n|\geq\alpha)+\sum_{k=1}^{n-1}\mathbb{P}(\{T'=k\})\mathbb{P}(|S'_n-S'_k|\geq\alpha)\\
&= \mathbb{P}(|S_n|\geq\alpha)+\sum_{k=1}^{n-1}\mathbb{P}(\{T'=k\})\mathbb{P}(|S_{n-k}|\geq\alpha)\\
&\leq \beta'+\sum_{k=1}^{n-1}\mathbb{P}(\{T'=k\})\beta'\leq 2\beta',
\end{aligned}$$

où l'on a posé $\beta'=\max_{0\leq k\leq n}\mathbb{P}(|S_k|\geq\alpha)$. Finalement, on conclut que

$$\mathbb{P}\left(\max_{1\leq k\leq n}|S_k|\geq 4\alpha\right)=\mathbb{P}(T<+\infty)\leq 4\beta'=4\max_{1\leq k\leq n}\mathbb{P}(|S_k|\geq\alpha).$$

□

14.2.3 Lien entre les modes de convergence

Théorème 14.6. *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes. On note $(S_n)_{n\geq 0}$ la suite des sommes partielles : $S_0=0$, et pour $n\geq 1$: $S_n=X_1+\cdots+X_n$.*
Alors, S_n converge presque sûrement si et seulement si S_n converge en probabilité.

Démonstration. Comme la convergence presque sûre implique la convergence en probabilité, il suffit de démontrer l'autre implication. Supposons donc que S_n converge en probabilité vers une variable notée S. Soit $\varepsilon>0$. On pose $Y_n=\sup\{|S_i-S_j|; i,j\geq n\}$. Remarquons que la suite $(Y_n)_{n\geq 0}$ est à valeurs dans $\overline{\mathbb{R}}_+$, décroissante et minorée par 0. Posons donc $Y=\lim_{n\to+\infty}Y_n$. Soient $\varepsilon>0$ et n un entier. On a $\mathbb{P}(Y>\varepsilon)\leq\mathbb{P}(Y_n>\varepsilon)$ et comme pour tous les i,j, on a $\{|S_i-S_j|>\varepsilon\}\subset\{|S_n-S_i|>\frac{\varepsilon}{2}\}\cup\{|S_n-S_j|>\frac{\varepsilon}{2}\}$, on obtient :

$$\mathbb{P}(Y_n>\varepsilon)=\mathbb{P}(\exists i,j\geq n;\quad |S_i-S_j|>\varepsilon)\leq\mathbb{P}(\exists i\geq n;\quad |S_n-S_i|>\varepsilon/2).$$

D'après le théorème de continuité séquentielle croissante, on a

$$\mathbb{P}(\exists i\geq n;\ |S_n-S_i|>\varepsilon/2)=\lim_{p\to+\infty}\mathbb{P}\left(\max_{1\leq k\leq p}|S_{n+k}-S_n|>\varepsilon/2\right).$$

Mais d'après le théorème 14.5, on sait que

$$\mathbb{P}\left(\max_{1\leq k\leq p}|S_{n+k}-S_n|>\varepsilon/2\right)\leq 4\max_{1\leq k\leq p}\mathbb{P}(|S_{n+k}-S_n|>\varepsilon/8).$$

On voit par ailleurs que

$$\begin{aligned}\mathbb{P}(|S_{n+k}-S_n|>\varepsilon/8) &\leq \mathbb{P}(|S_n-S|>\varepsilon/16)+\mathbb{P}(|S_{n+k}-S|>\varepsilon/16)\\ &\leq 2\max_{r\geq n}\mathbb{P}(|S_r-S|>\varepsilon/16).\end{aligned}$$

Ainsi, on obtient $\mathbb{P}(Y_n>\varepsilon)\leq 8\,\max_{r\geq n}\,\mathbb{P}(|S_r-S|>\varepsilon/16)$, d'où

$$\forall n\geq 1,\quad \mathbb{P}(Y>\varepsilon)\leq 8\max_{r\geq n}\mathbb{P}(|S_r-S|>\varepsilon/16).$$

En faisant tendre n vers $+\infty$, on obtient que $\mathbb{P}(Y>\varepsilon)=0$. On conclut donc que $\mathbb{P}(Y>0)=\mathbb{P}\left(\bigcup_{n=1}^{+\infty}\{Y>1/n\}\right)=0$. Ainsi $Y=0$ presque sûrement, et la suite $(S_n)_{n\geq 0}$ est presque sûrement de Cauchy, donc presque sûrement convergente. □

14.2.4 Critères de convergence

Convergence L^2

Théorème 14.7. *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes, centrées, admettant un moment d'ordre 2. On note $(S_k)_{k\geq 0}$ la suite des sommes partielles : $S_0=0$, $S_k=X_1+\cdots+X_k$. Si*

$$\sum_{k=1}^{+\infty}\operatorname{Var}X_k<+\infty,$$

alors S_n converge presque sûrement.

Démonstration. D'après le théorème précédent, il suffit de montrer que S_n converge en probabilité. Il suffit donc de montrer que S_n converge dans L^2. Comme L^2 est complet, il suffit de montrer que $(S_n)_{n\geq 0}$ est une suite de Cauchy dans L^2. Soient n,p des entiers. Comme les variables sont centrées et indépendantes, on a

$$\|S_{n+p}-S_n\|_2^2=\mathbb{E}\left(\sum_{k=n+1}^{n+p}X_k\right)^2=\operatorname{Var}\left(\sum_{k=n+1}^{n+p}X_k\right)=\sum_{k=n+1}^{n+p}\operatorname{Var}X_k.$$

Soit $\varepsilon>0$. On peut choisir N tel que $\sum_{k=N+1}^{+\infty}\operatorname{Var}X_k\leq\varepsilon^2$ car la série de terme général $\operatorname{Var}X_k$ converge. Ainsi, pour tout $n\geq N$ et tout p entier, on a $\|S_{n+p}-S_n\|_2\leq\varepsilon$. Comme $\varepsilon>0$ est quelconque, cela montre que $(S_n)_{n\geq 0}$ est une suite de Cauchy dans L^2, ce qui achève la preuve. □

Théorème des trois séries

Théorème 14.8. *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes. Soit c un réel strictement positif quelconque. Notons $X_k^{(c)} = X_k \mathbb{1}_{\{|X_k|\leq c\}}$.*

Alors, la série de terme général X_n converge presque sûrement si et seulement si les séries de terme général $\mathrm{Var}(X_k^{(c)})$, $\mathbb{P}(|X_k| \geq c)$ et $\mathbb{E}[X_k^{(c)}]$ convergent.

Démonstration. Montrons d'abord que la condition est suffisante. La série de terme général $\mathbb{P}(|X_k| \geq c)$ converge, donc d'après le premier lemme de Borel-Cantelli 10.12

$$\mathbb{P}\left(\overline{\lim_{n\to+\infty}} \{|X_n| \geq c\}\right) = 0, \text{ soit } \mathbb{P}\left(\underline{\lim_{n\to+\infty}} \{|X_n| < c\}\right) = 1,$$

ou encore $\mathbb{P}\left(\exists N : k \geq N \Longrightarrow |X_k| < c\right) = 1$. Cela implique que presque sûrement, à partir d'un certain rang, $X_k = X_k^{(c)}$, ce qui signifie que les deux séries sont de même nature.

Bien sûr, comme la série de terme général $\mathbb{E}[X_k^{(c)}]$ converge, la série de terme général $X_k^{(c)}$ et celle de terme général $X_k^{(c)} - \mathbb{E}[X_k^{(c)}]$ sont de même nature. Or cette dernière série converge presque sûrement d'après le théorème 14.7, ce qui achève la preuve de la suffisance.

Réciproquement, supposons que la série de terme général S_n converge presque sûrement. Cela implique que pour tout $c > 0$,

$$\mathbb{P}(|X_n| > c \text{ infiniment souvent}) = 0,$$

et donc la série de terme général $\mathbb{P}(|X_n| > c)$ converge d'après le deuxième lemme de Borel-Cantelli 10.16. À partir d'un certain rang, X_n et X_n^c sont égales, donc la série de terme général X_n^c converge également presque sûrement. Maintenant, raisonnons par l'absurde et supposons que la série de terme général $\mathrm{Var}(X_k^c)$ diverge. On obtient alors :

$$\frac{\sum\limits_{k=1}^{n} X_k^c}{\sqrt{\sum\limits_{k=1}^{n} \mathrm{Var}(X_k^c)}} \to 0 \quad \text{p.s.}$$

Posons

$$Z_n = \frac{\sum\limits_{k=1}^{n} (X_k^c - \mathbb{E}[X_k^c])}{\sqrt{\sum\limits_{k=1}^{n} \mathrm{Var}(X_k^c)}} \quad \text{et} \quad \alpha_n = \frac{\sum\limits_{k=1}^{n} \mathbb{E}[X_k^c]}{\sqrt{\sum\limits_{k=1}^{n} \mathrm{Var}(X_k^c)}}.$$

D'après la proposition 14.3, Z_n converge en loi vers la loi $\mathcal{N}(0,1)$, ce qui contredit que $Z_n + \alpha_n$ converge presque sûrement vers 0.

En effet, $Z_n \Longrightarrow \mathcal{N}(0,1)$ entraîne que $\phi_{Z_n}(1)$ tend vers $\exp(-1/2)$ alors que $Z_n + \alpha_n \xrightarrow{\text{p.s.}} 0$ entraîne que $\phi_{Z_n}(1) \sim e^{-i\alpha_n}$, ce qui est clairement incompatible en comparant les modules. Il y a une contradiction. On en déduit que la série de terme général $\mathrm{Var}(X_k^c)$ converge. On peut maintenant appliquer le théorème 14.7 : la série de terme général $X_k^c - \mathbb{E}[X_k^c]$ converge presque sûrement. Comme la série de terme général X_k^c converge presque sûrement, on en déduit que la série de terme général $\mathbb{E}[X_k^c]$ converge. □

14.2.5 Inégalité de Hoeffding

L'inégalité de Hoeffding est une inégalité de concentration pour des sommes de variables aléatoires indépendantes et bornées. Dans ce cas, elle est plus précise que l'inégalité de Chebychev, comme nous le verrons sur un exemple à la fin de ce paragraphe.

Théorème 14.9. *Soit $(X_n)_n$ une suite de variables aléatoires réelles indépendantes. Supposons qu'il existe deux suites de réels $(a_n)_{n\geq 0}$ et $(b_n)_{n\geq 0}$ telles que, pour tout $n \geq 0$, $a_n < b_n$ et*

$$\mathbb{P}(a_n \leq X_n \leq b_n) = 1.$$

Posons $S_n = X_1 + \cdots + X_n$. On a alors pour tout $x > 0$ et tout $n \geq 0$:

$$\mathbb{P}(|S_n - \mathbb{E}S_n| > x) \leq 2\exp\left\{-\frac{2x^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right\}.$$

Afin de démontrer ce résultat, nous avons besoin d'un lemme intermédiaire.

Lemme 14.10. *Soit Y une variable aléatoire réelle bornée, centrée et telle qu'il existe deux réels a et b tels que $a < b$ et $\mathbb{P}(a \leq Y \leq b) = 1$. Alors pour tout réel $t > 0$, on a*

$$\mathbb{E}[e^{tY}] \leq \exp\left\{(a-b)^2\frac{t^2}{8}\right\}.$$

Démonstration. Comme la fonction $x \mapsto e^{tx}$ est convexe, on voit que lorsqu'on a $a \leq Y \leq b$, alors

$$e^{tY} \leq \frac{b-Y}{b-a}e^{ta} + \frac{Y-a}{b-a}e^{tb}.$$

Comme on sait que $\mathbb{P}(a \leq Y \leq b) = 1$ et que $\mathbb{E}(Y) = 0$, l'inégalité précédente nous donne, en prenant l'espérance :

$$\mathbb{E}[e^{tY}] \leq \frac{b}{b-a}e^{ta} - \frac{a}{b-a}e^{tb}.$$

On pose alors $u = t(b-a)$ et on applique la formule de Taylor-Lagrange à la fonction

$$\psi(u) = \log\left(\frac{b}{b-a}e^{ta} + \frac{-a}{b-a}e^{tb}\right) = \frac{a}{b-a}u + \log\left(1 + \frac{a}{b-a}(1-e^u)\right),$$

en remarquant que $\psi(0) = \psi'(0) = 0$ et

$$\psi''(u) = -\frac{a}{b-a}e^u \frac{1 + \frac{a}{b-a}}{(1 + \frac{a}{b-a}(1-e^u))^2}.$$

Comme $\psi''(u)$ est de la forme $\alpha\beta/(\alpha+\beta)^2$, on en déduit que $|\psi''(u)| \leq 1/4$. Ainsi, par Taylor-Lagrange, il existe un réel $0 \leq s \leq u$ tel que

$$\psi(u) = \psi(0) + \psi'(0)u + \psi''(s)\frac{u^2}{2} \leq \frac{(a-b)^2}{8}t^2.$$

□

Nous pouvons maintenant prouver l'inégalité de Hoeffding.

Démonstration du théorème 14.9. Posons, pour tout $1 \leq i \leq n$,
$Y_i = X_i - \mathbb{E}X_i$, $\alpha_i = a_i - \mathbb{E}X_i$ et $\beta_i = b_i - \mathbb{E}X_i$.
On remarque que $\mathbb{P}(\alpha_i \leq Y_i \leq \beta_i) = 1$ et $S_n - \mathbb{E}S_n = Y_1 + \cdots + Y_n$. De plus, comme les variables X_i sont indépendantes, il en est de même pour Y_i. Par l'inégalité de Markov, on a donc pour tout $x \geq 0$ et tout $u > 0$

$$\mathbb{P}(S_n - \mathbb{E}S_n \geq x) \leq \frac{\mathbb{E}(e^{u(S_n - \mathbb{E}S_n)})}{e^{ux}} = \mathbb{E}(e^{u(Y_1+\cdots+Y_n)})e^{-ux} = e^{-ux}\prod_{i=1}^n \mathbb{E}(e^{uY_i}).$$

On obtient alors grâce au lemme précédent appliqué à Y_i en remarquant que $\beta_i - \alpha_i = b_i - a_i$:

$$\mathbb{P}(S_n - \mathbb{E}S_n \geq x) \leq \exp\left\{-ux + \frac{u^2}{8}\sum_{i=1}^n (b_i - a_i)^2\right\}.$$

Cette inégalité étant vraie pour tout u, nous choisissons alors u tel que cette inégalité soit la meilleurs. Le minimum de la borne de droite est atteint en

$$u = \frac{4x}{\sum_{i=1}^n (b_i - a_i)^2},$$

ce qui montre que

$$\mathbb{P}(S_n - \mathbb{E}S_n \geq x) \leq \exp\left\{-\frac{2x^2}{\sum_{i=1}^n (b_i - a_i)^2}\right\}.$$

En remplaçant X_i par $-X_i$ dans ce qui précède, on obtient l'inégalité similaire

$$\mathbb{P}(S_n - \mathbb{E}S_n \leq -x) \leq \exp\left\{-\frac{2x^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right\}.$$

L'inégalité cherchée est donc une conséquence directe de ces deux derniers résultats. □

Exemple: Supposons que $(X_1, \ldots, X_n)$ soit un vecteur aléatoire composé de variables indépendantes de même loi de Bernoulli $\mathcal{B}(p)$, avec $0 < p < 1$. La variable S_n suit alors la loi binomiale $\mathcal{B}(n, p)$. Par l'inégalité de Chebychev , on obtient

$$\mathbb{P}(|S_n - \mathbb{E}S_n| \geq x) \leq \frac{np(1-p)}{x^2},$$

alors que par l'inégalité de Hoeffding

$$\mathbb{P}(|S_n - \mathbb{E}S_n| \geq x) \leq 2\exp\left\{-\frac{2}{n}x^2\right\}.$$

On voit donc que l'inégalité de Hoeffding est nettement plus précise dès que x est suffisamment grand.

14.3 Grandes déviations

Soit $(X_n, n \geq 1)$ une suite de variables aléatoires indépendantes identiquement distribuées de loi μ, supposée intégrable. On note la somme partielle $S_n = X_1 + \cdots + X_n$. Par la loi (faible) des grands nombres, on sait que :

— si $x > \int y \, d\mu(y)$, alors $\lim_{n\to+\infty} \mathbb{P}(S_n \geq xn) = 0$,

— si $x < \int y \, d\mu(y)$, alors $\lim_{n\to+\infty} \mathbb{P}(S_n \geq xn) = 1$.

On se demande ici à quelle vitesse a lieu la convergence. Il s'agit d'un problème de grandes déviations (de la moyenne empirique par rapport à la moyenne théorique).

Nous allons présenter ci-dessous un résultat permettant de minorer la vitesse de convergence dans la loi faible des grands nombres pour les sommes de variables aléatoires indépendantes de même loi, admettant des moments exponentiels.

Commençons par donner un résultat simple concernant des suites déterministes.

Définition. *La suite de réels $(a_n, n \geq 1)$ est appelée suite sous-additive si pour tous $m, n \in \mathbb{N}$, on a $a_{m+n} \leq a_m + a_n$.*

Lemme 14.11 (lemme de Fekete [3]). *Pour toute suite sous-additive de réels* $(a_n, n \geq 1)$*, la suite* $(\frac{a_n}{n})_{n\geq 1}$ *converge dans* $\overline{\mathbb{R}}$ *et on a*

$$\lim_{n\to+\infty} \frac{a_n}{n} = \inf_{n\geq 1} \frac{a_n}{n} \in [-\infty, +\infty).$$

Démonstration. On itère la relation de sous-additivité, ce qui nous donne pour tous $k, n \in \mathbb{N}$:

$$a_n \leq \lfloor n/k \rfloor a_k + a_{n-k\lfloor n/k \rfloor} \leq \lfloor n/k \rfloor a_k + \max(0, a_1, \ldots, a_{k-1}).$$

Comme $\lfloor n/k \rfloor$ est équivalent à n/k lorsque n tend vers l'infini, on trouve pour tout $k \geq 0$

$$\overline{\lim_{n\to+\infty}} \frac{a_n}{n} \leq \frac{a_k}{k}$$

et donc

$$\inf_{n\geq 1} \frac{a_n}{n} \leq \underline{\lim_{n\to+\infty}} \frac{a_n}{n} \leq \overline{\lim_{n\to+\infty}} \frac{a_n}{n} \leq \inf_{n\geq 1} \frac{a_n}{n}.$$

□

Proposition 14.12. *Soit* $(X_n)_{n\geq 1}$ *une suite de variables aléatoires indépendantes identiquement distribuées. Notons* $S_n = X_1 + \cdots + X_n$*. Alors si on pose, pour* $x \in \mathbb{R}$*,* $h(x) = -\sup_n \frac{1}{n} \log \mathbb{P}(S_n \geq nx)$*, on a*

1. *Pour tout* x *réel,* $\lim_{n\to+\infty} -\frac{1}{n} \log \mathbb{P}(S_n \geq nx) = h(x)$*,*
2. h *est une fonction à valeurs dans* $[0, +\infty]$*, croissante et convexe,*
3. $h(x) < \infty$ *si et seulement si* $\mathbb{P}(X_1 \geq x) > 0$*.*

Démonstration. 1. Remarquons l'indépendance et l'équi-distribution des variables X_i nous donne, pour tous $m, n \in \mathbb{N}$

$$\begin{aligned} \mathbb{P}(S_{m+n} \geq (m+n)x) &\geq \mathbb{P}(\{S_m \geq mx\} \cap \{S_{m+n} - S_m \geq nx\}) \\ &= \mathbb{P}(S_m \geq mx)\mathbb{P}(S_n \geq nx). \end{aligned}$$

On passe au logarithme dans cette expression et on conclut que la suite $(-\log \mathbb{P}(S_n \geq nx))$ est sous-additive. On obtient donc le résultat voulu en appliquant le lemme 14.11.

3. En réalité, le lemme de Fekete devrait s'appeler lemme de Fekete–Pólya–Szegö. Ce lemme est démontré dans un cas particulier dans un article de Fekete [15] (1923), et sera rendu célèbre, sous la forme qu'on lui connaît maintenant, grâce au fameux livre d'exercices de Pólya et Szegö : *Aufgaben und Lehrsätze aus der Analysis I* (1925), que l'on connaît plus aujourd'hui dans sa traduction anglaise citée en référence [28]. Pour une généralisation, voir Hammersley [20].

2. Il est évident que h est à la fois à valeurs positives et croissante. Montrons la convexité de h. Soit $\lambda \in [0,1]$. On suppose que $x \geq y$, que les suites d'entiers $(n_k), (p_k)$ tendent vers l'infini et que $\frac{n_k}{n_k+p_k}$ tend vers λ par valeurs supérieures. On a alors comme précédemment
$$\begin{aligned}
&-\frac{\log \mathbb{P}(S_{n_k+p_k} \geq (n_k+p_k)(\lambda x + (1-\lambda)y))}{n_k+p_k} \\
\leq\ & -\frac{\log \mathbb{P}(S_{n_k+p_k} \geq (n_k+p_k)(\frac{n_k}{n_k+p_k}x + \frac{p_k}{n_k+p_k}y))}{n_k+p_k} \\
\leq\ & \frac{n_k}{n_k+p_k}\left(-\frac{1}{n_k}\log \mathbb{P}(S_{n_k} \geq n_k x)\right) \\
& + \frac{p_k}{n_k+p_k}\left(-\frac{1}{p_k}\log \mathbb{P}(S_{p_k} \geq p_k y)\right)
\end{aligned}$$
En faisant tendre k vers l'infini, on obtient
$$h(\lambda x + (1-\lambda)y) \leq \lambda h(x) + (1-\lambda)h(y).$$

3. Si $\mathbb{P}(X_1 \geq x) = 0$, alors $\mathbb{P}(S_n \geq nx) = 0$ pour tout n et on a donc $h(x) = +\infty$.
Réciproquement, le premier point implique que $\log \mathbb{P}(X_1 \geq x) \leq -h(x)$ et donc si $h(x) = +\infty$, alors $\mathbb{P}(X_1 \geq x) = 0$.

$\square$

Afin d'expliciter la limite h dans la proposition précédente, nous avons besoin d'une notation supplémentaire.

Définition. *Soit X une variable aléatoire de loi de probabilité μ sur $\mathbb{R}$. On appelle cumulant de μ la fonction*
$$\Lambda_\mu : \mathbb{R} \to (-\infty, +\infty],\ \lambda \mapsto \log \mathbb{E}[e^{\lambda X}],$$
avec la convention $\log(+\infty) = +\infty$.

Remarque 14.13. *Λ_μ est une fonction convexe. Soient $x, y \in \mathbb{R}$ et $0 < \theta < 1$. En appliquant l'inégalité de Hölder avec $p = \frac{1}{\theta}$ et $q = \frac{1}{1-\theta}$, on a*
$$\begin{aligned}
\Lambda_\mu(\theta x + (1-\theta)y) &= \log \mathbb{E}[\exp\left(\theta x + (1-\theta)y\right)X] \\
&= \log \mathbb{E}[\exp(\theta x X)\exp((1-\theta)yX)] \\
&\leq \log\left((\mathbb{E}\exp(xX))^\theta(\mathbb{E}\exp(yX))^{1-\theta}\right) \\
&\quad = \theta\Lambda_\mu(x) + (1-\theta)\Lambda_\mu(y).
\end{aligned}$$

À toute fonction convexe f, on peut associer sa transformée de Fenchel–Legendre f^*. Celle-ci est définie par
$$f^*(x) = \sup_{\lambda \in \mathbb{R}}(\lambda x - f(\lambda)).$$

La fonction f^* est aussi convexe, car

$$\begin{aligned} f^*(px+(1-p)y) &= \sup_\lambda \{p(\lambda x - f(\lambda)) + (1-p)(\lambda y - f(\lambda))\} \\ &\leq p \sup_\lambda \{\lambda x - f(\lambda)\} + (1-p) \sup_\lambda \{\lambda y - f(\lambda)\} \\ &= pf^*(x) + (1-p)f^*(y). \end{aligned}$$

On va voir que h coïncide avec la transformée de Fenchel–Legendre Λ_μ^* de la fonction cumulant.

On peut parfois déterminer explicitement la transformée de Legendre, comme nous allons le voir dans quelques exemples. **Exemple:** Si X suit la loi de Bernoulli de paramètre p, alors $\Lambda^*(x) = +\infty$ pour $x \notin [0,1]$ et

$$\Lambda_X^*(x) = x \log\left(\frac{x}{p}\right) + (1-x)\log\left(\frac{1-x}{1-p}\right) \text{ pour } x \in [0,1].$$

(Les calculs sont laissés au lecteur.)

Remarque 14.14. *La définition du cumulant des variables aléatoires réelles peut être étendue à des vecteurs aléatoires à valeurs dans* $\mathbb{R}^d$.
On a alors $\Lambda_\mu : \mathbb{R}^d \to (-\infty, +\infty],\ \lambda \mapsto \log \mathbb{E}[e^{\langle \lambda, X\rangle}]$ *et*

$$\Lambda_\mu^*(x) = \sup_{\lambda \in \mathbb{R}^d} (\langle \lambda, x\rangle - \Lambda_\mu(\lambda)).$$

Nous laissons au lecteur le soin de vérifier que les résultats énoncés précédemment s'étendent en dimension d.
Toujours à titre d'exercice, on pourra remarquer que si $N = (N_1, \dots, N_d)$ *est un vecteur gaussien standard (centré et réduit), alors* $\Lambda_N^*(x) = \|x\|_2^2/2$.

Le principal résultat est le suivant, qui identifie la fonction h de la proposition 14.12. Supposons à partir de maintenant que μ est intégrable, donc $\mathbb{E}X_1$ existe, et que la dimension est $d = 1$.

Théorème 14.15. *Soit* $(X_n)_{n\geq 1}$ *une suite de variables aléatoires indépendantes identiquement distribuées, intégrables, de loi commune* μ.
Alors pour tout $x \geq \mathbb{E}X_1$, *on a*

$$\lim_{n\to+\infty} -\frac{1}{n}\log \mathbb{P}(S_n \geq nx) = \Lambda_\mu^*(x).$$

La preuve que nous présentons suit dans ses grandes lignes celle de Cerf et Petit [5], mais certains éléments de la preuve apparaissaient déjà chez Hammersley [21] et Kingman [23].

Démonstration. Afin de simplifier les écritures, on notera dans ce qui suit Λ et Λ^* pour Λ_μ et Λ_μ^*.

1) Montrons que $h(x) \geq \Lambda^*(x)$. Par l'inégalité de Chebychev , on a pour tout $\lambda \geq 0$,

$$\mathbb{P}(S_n \geq xn) \leq \mathbb{P}(e^{\lambda S_n} \geq e^{\lambda nx}) \leq e^{-\lambda nx}\mathbb{E}[e^{\lambda S_n}] = e^{n\Lambda(\lambda)-\lambda nx}$$

et donc

$$\frac{1}{n}\log \mathbb{P}(S_n \geq nx) \leq \Lambda(\lambda) - \lambda x. \tag{14.1}$$

Ainsi, $\forall\lambda \geq 0 \quad \forall x \in \mathbb{R} \quad \Lambda(\lambda) + h(x) \geq \lambda x$, ce qui nous donne

$$\forall\lambda \geq 0 \quad \Lambda(\lambda) \geq h^*(\lambda) = \sup_{x\in\mathbb{R}} (\lambda x - h(x)).$$

Si $x \geq \mathbb{E}X_1$, l'inégalité (14.1) demeure valide pour $\lambda \leq 0$, car on a alors $\Lambda(\lambda) - \lambda x \geq 0$ pour $x \geq \mathbb{E}X_1$ (grâce à l'inégalité de Jensen). Ainsi

$$\forall\lambda \in \mathbb{R} \quad \forall x \geq \mathbb{E}X_1 \quad \Lambda(\lambda) + h(x) \geq \lambda x,$$

d'où l'on tire

$$h(x) \geq \Lambda^*(x) \text{ pour } x \geq \mathbb{E}[X_1].$$

2) On va commencer par montrer que $\Lambda(\lambda) \leq h^*(\lambda)$ pour tout $\lambda \geq 0$. Traitons à part le cas $\lambda = 0$: on a

$$h^*(0) = \sup_{x\in\mathbb{R}} -h(x) \geq \sup_{x\in\mathbb{R}} -\log \mathbb{P}(S_1 \geq x) = 0 = \Lambda(0).$$

Supposons maintenant $\lambda > 0$. On a

$$\begin{aligned} \Lambda(\lambda) &= \log \mathbb{E}[e^{\lambda X}] = \overline{\lim_{M\to+\infty}} \log \mathbb{E}[e^{\lambda X}\mathbb{1}_{\{|X|\leq M\}}] \\ &\leq \overline{\lim_{M\to+\infty}} \inf_{n\geq 1} \frac{1}{n}\log \mathbb{E}[e^{\lambda S_n}\mathbb{1}_{\{|S_n|\leq nM\}}]. \end{aligned}$$

On peut écrire

$$\begin{aligned} &\mathbb{E}[e^{\lambda S_n}\mathbb{1}_{\{|S_n|\leq nM\}}] = \int_{[-nM,nM]} e^{\lambda u} d\mathbb{P}_{S_n}(u) \\ = &\int_{[-nM,nM]} \left(e^{-\lambda nM} + \int_{[-nM,nM]} \mathbb{1}_{\{x\leq u\}}\lambda e^{\lambda x}\, d\lambda(x)\right) d\mathbb{P}_{S_n}(u) \\ \leq &\ e^{-\lambda nM} + \int_{[-nM,nM]} \lambda e^{\lambda x}\mathbb{P}(S_n \geq x)\, d\lambda(x) \\ &= e^{-\lambda nM} + \int_{[-M,M]} n\lambda e^{\lambda nx}\mathbb{P}(S_n \geq nx)\, d\lambda(x) \\ \leq &\ e^{-\lambda nM} + \int_{[-M,M]} n\lambda e^{\lambda nx} e^{-nh(x)}\, d\lambda(x) \\ &\leq e^{-\lambda nM} + 2Mn\lambda e^{nh^*(\lambda)} \end{aligned}$$

Comme $h^*(0) = \sup_{x\in\mathbb{R}} - h(x) \geq 0$, il existe $x \in \mathbb{R}$ avec $h(x) < +\infty$, ce qui entraîne que $h^*(\lambda) > -\infty$. On peut donc choisir M tel que $-M\lambda \le h^*(\lambda)$. On a alors pour tout $n \ge 1$

$$\log \mathbb{E}[e^{\lambda S_n} \mathbb{1}_{\{|S_n| \le nM\}}] \le \log(2Mn\lambda + 1) + nh^*(\lambda),$$

d'où

$$\inf_{n \ge 1} \frac{1}{n} \log \mathbb{E}[e^{\lambda S_n} \mathbb{1}_{\{|S_n| \le nM\}}] \le h^*(\lambda),$$

et finalement $\Lambda(\lambda) \le h^*(\lambda)$.

3) Pour conclure, il nous reste à montrer que $h(x) \le \Lambda^*(x)$.
Notons $c = \inf\{x \in \mathbb{R} : \mathbb{P}(X_1 \ge x) = 0\}$.

— si $x < c$. Il suffit de montrer qu'il existe λ réel avec $h(x) \le \lambda x - \Lambda(\lambda)$. Or, comme h est convexe croissante, finie sur $]-\infty, c[$, il existe λ réel positif tel que pour tout y, $h(y) \ge h(x) + \lambda(y - x)$, soit $\lambda x - h(x) \ge \lambda y - h(y)$. En prenant le supremum en y, on obtient $\lambda x - h(x) \ge h^*(\lambda)$, mais comme on a montré que $h^*(\lambda) = \Lambda(\lambda)$ pour $\lambda \ge 0$, cela achève la preuve dans ce cas.

— si $x > c$, comme $\Lambda(\lambda) \le c\lambda$ pour tout $\lambda \ge 0$, on a

$\Lambda^*(x) \ge \sup_{\lambda \in \mathbb{R}_+} (x - c)\lambda = +\infty$, et l'inégalité est évidente.

— si $x = c$, il est facile de voir que $\mathbb{P}(S_n \ge nc) = \mathbb{P}(X_1 = c)^n$, d'où $h(c) = -\log \mathbb{P}(X_1 = c)$. Par ailleurs, pour $\varepsilon, \lambda > 0$, on a

$$\mathbb{E}(e^{\lambda X_1}) \le e^{\lambda(c-\varepsilon)} + e^{\lambda c}\mathbb{P}(c - \varepsilon < X_1 \le c),$$

d'où $\lambda c - \Lambda(\lambda) \ge -\log(e^{-\lambda\varepsilon} + \mathbb{P}(c - \varepsilon < X_1 \le c))$ et

$$\begin{aligned}\Lambda^*(c) \ge \overline{\lim_{\lambda \to +\infty}}\ \lambda c - \Lambda(\lambda) &\ge \overline{\lim_{\lambda \to +\infty}}\ -\log(e^{-\lambda\varepsilon} + \mathbb{P}(c - \varepsilon < X_1 \le c)) \\ &= -\log(\mathbb{P}(c - \varepsilon < X_1 \le c)).\end{aligned}$$

En faisant tendre ε vers 0, on obtient donc $\Lambda^*(c) \ge h(c)$.

□

Le résultat utilisé dans la pratique est le suivant.

Corollaire 14.16 (inégalité de Chernov). *Soit $(X_n, n \ge 1)$ une suite de variables aléatoires indépendantes de même loi, intégrable .*
Alors, pour tout $x > \mathbb{E}(X_1)$,

$$\mathbb{P}(X_1 + \cdots + X_n \ge xn) \le e^{-n\Lambda^*(x)}. \tag{14.2}$$

On a $\Lambda^*(x) > 0$ dès qu'il existe $\varepsilon > 0$ tel que $\Lambda(\varepsilon) < +\infty$.

Démonstration. (14.2) vient de l'identité $h(x) = \Lambda^*(x)$ pour $x \ge \mathbb{E}(X_1)$ et de la définition de h. En appliquant le théorème de convergence monotone à la partie négative de X_1, on a $\lim_{N\to+\infty} \mathbb{E}(\max(-N, X_1)) = \mathbb{E}(X_1)$, donc on peut trouver N tel que $x > \mathbb{E}(\max(-N, X_1))$. Notons ν la loi de $\max(-N, X_1)$ et prenons a avec $x > a > \mathbb{E}\max(-N, X_1)$. Pour tout $\lambda > 0$, on a

$$\Lambda^*(x) \geq -\Lambda(\lambda) + \lambda x \geq -\Lambda_\nu(\lambda) + \lambda x,$$

donc pour avoir $\Lambda^*(x) > 0$, il suffit de trouver $\lambda > 0$ tel que $\Lambda_\nu(\lambda) < \lambda a$. Posons $\phi(\lambda) = \int e^{\lambda t}\, d\nu(t)$. Pour $\lambda \in]0, \varepsilon/2[$, on a

$$|\frac{e^{\lambda t} - 1}{\lambda}| \leq |t| \max(1, e^{\lambda t}) \leq N + \frac{2}{\varepsilon} e^{\varepsilon t} \text{ pour } \nu \text{ presque tout } t,$$

d'où, avec le théorème de convergence dominée : $\phi(\lambda) = 1 + \lambda \int t\, d\nu(t) + o(\lambda)$, lorsque λ tend vers 0 par valeurs supérieures. Finalement

$$\Lambda_\nu(\lambda) = \lambda \mathbb{E}(\max(-N, X_1)) + o(\lambda),$$

ce qui donne le résultat voulu en prenant $\lambda > 0$ suffisamment petit.

□

Remarque 14.17. *Les estimées de grandes déviations s'étendent aux intervalles de la forme $[x, +\infty[$, aux intervalles ouverts, fermés ou semi-ouverts. Il s'agit du principe de grandes déviations, prouvé par Cramér pour une suite de variables indépendantes de même loi sur $\mathbb{R}$. Nous ne l'énonçons pas ici, car il s'agit d'un résultat compliqué, qui a déjà fait l'objet de divers livres (pas toujours faciles d'accès), comme celui de Varadhan.*

Exemple: Voyons maintenant sur un exemple comment utiliser ces résultats. Étudions le cas simple où les variables $(X_n)_{n\geq 1}$ sont indépendantes de loi de Bernoulli de paramètre p. Dans ce cas, S_n correspond au nombre de "pile" obtenu lors de n lancers indépendants d'une pièce de monnaie.

Le théorème précédent implique le résultat suivant : pour tout $m \in [p, 1[$, on a

$$\begin{aligned}
\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}(S_n \geq nm) &= \sup_{s\geq 0} (m \log s - \log G_X(s)) \\
&= m \log \frac{m(1-p)}{p(1-m)} - \log\left(1 - \left(1 - \frac{m(1-p)}{p(1-m)}\right) p\right),
\end{aligned}$$

où $G_X(s) = 1 - (1-s)p$ est la fonction génératrice de la loi de Bernoulli de paramètre $0 < p < 1$ (car pour une loi de Bernoulli, $\Lambda(\lambda) = \log G_X(e^\lambda)$).

14.4 Exercices sur les sommes de variables indépendantes

14.4.1 Exercices corrigés

Exercice 291. *Séries à termes positifs.*
Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes à termes positifs. Soit $c > 0$ quelconque. Montrer que la série de terme général X_n converge si et seulement si les séries de termes généraux respectifs $\mathbb{P}(X_n > c)$ et $\mathbb{E}[X_n \mathbb{1}_{\{X_n \leq c\}}]$ convergent. → indication → solution

Exercice 292. *Permutations aléatoires – nombres de Stirling de première espèce.*
Pour a et b deux entiers naturels non nuls, on note $(a\ b)$ la permutation de $\mathbb{N}^*$ envoyant a sur b, b sur a, et laissant les autres points invariants. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes, où X_n suit la loi uniforme sur $\{1,\dots,n\}$. On pose $\sigma_n = (1\ X_1) \circ \dots \circ (n\ X_n)$. On peut observer que σ_n laisse stables les éléments de $\{n+1, n+2, \dots\}$ et peut ainsi s'identifier à un élément de $\mathfrak{S}_n$.

1. Montrer que la restriction de σ_n à $\{1,\dots,n\}$ suit la loi uniforme sur $\mathfrak{S}_n$.
2. Pour $n \geq 1$ et $k \in \{1,\dots,n\}$, on pose $Y_{n,k} = 1$ si k est le plus petit entier de son orbite dans σ_n, 0 sinon.
   On rappelle que l'orbite d'un point par une permutation σ est l'ensemble de ses itérées : $\{\sigma^k(x); k \in \mathbb{N}\}$.
   Montrer qu'à n fixé, $Y_{n,1},\dots,Y_{n,n}$ sont des variables aléatoires indépendantes.
3. Les nombres de Stirling de première espèce non signés, notés $c(n,k)$, sont, par définition, les coefficients dans la base canonique du développement de la factorielle croissante :
   $$(x)^n = x(x+1)(x+2)\dots(x+n-1) = \sum_{k=0}^{n} c(n,k)x^k.$$
   On pose $C_n = \sum_{k=1}^{n} Y_{n,k}$. Notons que C_n est le nombre de cycles de σ_n.
   Montrer que pour tout k entre 0 et n, $c(n,k)$ est le nombre d'éléments de $\mathfrak{S}_n$ ayant exactement k orbites et que $\mathbb{P}(C_n = k) = \frac{c(n,k)}{n!}$.
4. Montrer que
   $$\frac{C_n - \log n}{\sqrt{\log n}} \Longrightarrow \mathcal{N}(0,1).$$

→ indication → solution

Exercice 293. *Théorème de Steinhaus.*
Soit $\sum\limits_{n=0}^{+\infty} a_n z^n$ une série entière de rayon de convergence $R < +\infty$. On dit que le point z de module R est un point régulier s'il existe une fonction holomorphe g définie sur un voisinage V de z qui coïncide avec f sur $V \cap B(0,R)$. Les points de module R qui ne sont pas réguliers sont dits singuliers. On note A_r l'ensemble des points réguliers et A_s l'ensemble des points singuliers. Un théorème classique de la théorie des fonctions holomorphes dit que A_s est toujours non-vide (voir par exemple le théorème 16.2 dans Rudin [32]). Soient $\sum\limits_{n=0}^{+\infty} a_n z^n$ une série entière de rayon de convergence 1 et $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. Pour $|z|<1$, on pose

$$F(z) = \sum_{n=0}^{+\infty} a_n e^{2i\pi X_n} z^n.$$

Le but de l'exercice est de montrer qu'avec probabilité 1, l'ensemble A_s des points singuliers de F est le cercle unité tout entier : on dit alors que le cercle unité est une coupure. Ce résultat constitue le théorème de Steinhaus [4].

1. Pour z dans le disque unité ouvert, on pose
$$E(z) = \left\{ \overline{\lim_{n\to+\infty}} \left(\frac{|F^{(n)}(z)|}{n!} \right)^{1/n} < \frac{1}{1-|z|} \right\}.$$
Montrer que $\{A_r \neq \varnothing\} = \bigcup\limits_{\substack{r\in[0,1[\cap\mathbb{Q} \\ \theta\in[0,1]\cap\mathbb{Q}}} E(re^{i2\pi\theta})$.
2. Montrer que $\mathbb{P}(E(z))$ ne peut valoir que 0 ou 1, puis que $\mathbb{P}(E(z))$ ne dépend que de $|z|$.
3. Conclure.

→ indication → solution

Exercice 294. *Loi des événements rares.*
Soit $(N_n)_{n\geq 1}$ une suite d'entiers tendant vers l'infini.
On suppose que pour tout n, $A_{n,1}, \dots, A_{n,N_n}$ sont des événements indépendants avec $\mathbb{P}(A_{n,k}) = p_{n,k}$. On suppose également que

$$\lim_{n\to+\infty} \sum_{k=1}^{N_n} p_{n,k} = \lambda > 0 \text{ et que } \lim_{n\to+\infty} \sup_{1\leq k\leq N_n} p_{n,k} = 0.$$

4. Paul Lévy attribue la preuve présentée ici à Laurent Schwartz [26] (qui était son gendre !)

Montrer que $\sum_{k=1}^{N_n} \mathbb{1}_{A_{n,k}}$ converge en loi vers la loi de Poisson de paramètre λ.
Indication : on pourra s'inspirer de la preuve du théorème de Lindeberg.

$\rightarrow$ indication $\rightarrow$ solution

Exercice 295. *Produit eulérien aléatoire.*
Rappels : soit $(z_n)_{n\geq 1}$ une suite de nombres complexes. On dit que le produit $\prod_{n=1}^{+\infty} (1+z_n)$ est convergent si la suite $(\prod_{n=1}^{N} (1+z_n))_{N\geq 1}$ admet une limite non nulle lorsque N tend vers l'infini. Il est bien connu (et on ne demande pas de le redémontrer) que si la série de terme général (z_n) est absolument convergente et que $1+z_n \neq 0$ pour tout $n \geq 1$, alors le produit $\prod_{n=1}^{+\infty}(1+z_n)$ est convergent.

1. Soit (Z_n) une suite de variables aléatoires complexes indépendantes centrées. On suppose qu'il existe une suite $(c_n)_{n\geq 1}$ telle que $\forall n \geq 1$, $c_n < 1$, $\mathbb{P}(|Z_n| \leq c_n) = 1$ et $\sum_{n=1}^{+\infty} c_n^2 < +\infty$. Montrer que le produit infini
$$\prod_{n=1}^{+\infty}(1+Z_n)$$
est presque sûrement convergent.
2. On appelle loi uniforme sur le disque unité la loi image de la loi uniforme sur $[0, 2\pi]$ par l'application $x \mapsto e^{ix}$. On note $\mathcal{P}$ l'ensemble des nombres premiers et $(X_k)_{k\in\mathcal{P}}$ une suite de variables indépendantes suivant la loi uniforme sur le disque unité. Montrer que pour $s > 1/2$, le produit aléatoire infini
$$\zeta(s, X) = \prod_{p\in\mathcal{P}}(1 - \frac{X_p}{p^s})^{-1}$$
converge presque sûrement.

$\rightarrow$ indication $\rightarrow$ solution

14.4.2 Exercices non corrigés

Exercice 296. Faire la preuve du théorème central limite 14.2 qui a été laissée en exercice.
$\rightarrow$ indication

Exercice 297. Soit $(a_n)_{n\geq 1}$ une suite de réels positifs. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes. Déterminer des conditions nécessaires et suffisantes pour avoir presque sûrement :
— convergence de la série de terme général $a_n X_n$
— absolue convergence de la série de terme général $a_n X_n$
dans les situations suivantes :

1. *séries de Rademacher* : les X_n suivent la loi uniforme sur $[-1, 1]$.

2. les X_n suivent la loi normale $\mathcal{N}(0,1)$.
3. les X_n suivent la loi de Cauchy $\mathcal{C}(0,1)$.

$\rightarrow$ indication

Exercice 298. *Séries de variables symétriques.*
Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes symétriques. Soit $c > 0$ quelconque. Montrer que la série de terme général X_n converge si et seulement si les séries de termes généraux respectifs $\mathbb{P}(X_n > c)$ et $\mathrm{Var}[X_n \, \mathbb{1}_{\{X_n \leq c\}}]$ convergent. $\rightarrow$ indication

Exercice 299. Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires indépendantes suivant la loi uniforme sur $[0, 2\pi]$.
Montrer que le rayon de convergence de la série entière

$$z \mapsto \sum_{n=0}^{+\infty} \cos(X_n) z^n$$

est presque sûrement 1. $\rightarrow$ indication

Exercice 300. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes identiquement distribuées. Montrer qu'il existe une suite $(a_n)_{n\geq 1}$ de réels non nuls telle que la série de terme général $a_n X_n$ converge presque sûrement. $\rightarrow$ indication

Exercice 301. *Nombre d'inversions d'une permutation aléatoire.*
Le nombre d'inversions d'une permutation $\sigma \in \mathfrak{S}_n$ est le nombre de couples (i, j) avec $1 \leq i < j \leq n$ tels que $\sigma(i) > \sigma(j)$. On suppose que σ_n suit la loi uniforme sur $\mathfrak{S}_n$. En s'inspirant de l'exercice 292, établir un théorème central limite pour le nombre $N(\sigma_n)$ d'inversions d'une permutation aléatoire. $\rightarrow$ indication

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 on veut montrer que $x \in \underset{i \in I}{\cup} A_i$, 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 $\{1, \ldots, n\}$.

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 $\{1, \ldots, n\}$.

Cet entier n est unique : on l'appelle le *cardinal* de l'ensemble Ω. On le note $|\Omega|$. De manière intuitive, c'est le nombre d'éléments de Ω.
Le cardinal de l'ensemble vide est zéro.
Pour Ω fini de cardinal n, et $p \in \{0, \ldots, n\}$, on note $\mathcal{B}_p(\Omega)$l'ensemble des parties de Ω de cardinal p. Par exemple $\mathcal{B}_2(\{a, b, c\}) = \{\{a, b\}, \{b, c\}, \{a, c\}\}$.
On note de plus $\mathcal{P}(\Omega)$l'ensemble des parties de Ω, quel que soit leur cardinal.
Par exemple $\mathcal{P}(\{a, b, c\}) = \{\varnothing, \{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{a, c\}, \{a, b, c\}\}$.

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 $A = \bigcup_{i\in I} A_i$ et $i \neq j \Longrightarrow A_i \cap A_j = \varnothing$.[1]
On a alors

$$|A| = \sum_{i\in I} |A_i|.$$

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 $A_i \cap A_j = \varnothing$. $\square$

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.

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.

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 $D = \{1, \ldots, p\}$. Or, un p-uplet $(x_1, x_2, \ldots, x_p)$ dont les composantes sont des éléments de A peut être considéré comme la donnée d'une application de $\{1, \ldots, p\}$ dans A. Le nombre de p-uplets $(x_1, x_2, \ldots, x_p)$ 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 $A = \{0, \ldots, 20\}$ 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 $D = \{1, \ldots, 300\}$ 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*

$$\binom{n}{p} = \frac{n(n-1)\ldots(n-p+1)}{p(p-1)\ldots 1} = \frac{n!}{p!(n-p)!}.$$

Démonstration. Il suffit d'appliquer le lemme des bergers à

— D : ensemble des injections de $\{1, \ldots, p\}$ dans Ω,
— $A = \mathcal{B}_p(\Omega)$,
— ϕ définie par $\phi(f) = \text{Image}(f) = \{f(k); k \in \{1, \ldots, p\}\}$.

On a vu précédemment que $|A| = n(n-1)\ldots(n-p+1)$. Il n'est pas difficile de voir que ϕ est surjective. Une partie $\{e_1, \ldots, e_p\}$ de Ω étant donnée, combien existe-t-il d'injections (en fait de bijections) de $\{1, \ldots, p\}$ dans Ω telles que $\{f(1), \ldots, f(p)\} = \{e_1, \ldots, e_p\}$? C'est évidemment le nombre d'injections de $\{1, \ldots, n\}$ dans $\{e_1, \ldots, e_p\}$, 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. $\square$

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 :*

$$x_1 + x_2 + \dots + x_p \leq n. \tag{A.1}$$

Démonstration. Il suffit de remarquer que l'application

$$(x_1, \dots, x_p) \quad \mapsto \quad (x_1, x_1 + x_2, x_1 + x_2 + x_3, \dots, x_1 + x_2 + \dots + x_p)$$

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 $\{1, \dots, n\}$. $\square$

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*

$$x_1 + x_2 + \dots + x_p = n. \tag{A.2}$$

Démonstration. Il suffit de remarquer que les solutions $(x_1, \ldots, x_p) \in (\mathbb{N}\backslash\{0\})^p$ de l'équation (A.2) sont exactement les solutions de l'inéquation (A.2) qui ne sont pas solutions de l'inéquation

$$x_1 + x_2 + \cdots + x_p \leq n - 1. \tag{A.3}$$

Il y en a donc $\binom{n}{p} - \binom{n-1}{p} = \binom{n-1}{p-1}$. □

Théorème A.13. *Pour n, p des entiers positifs tels que $p \geq 1$, il existe exactement $\binom{n+p-1}{p-1}$ p-uplets $(x_1, \ldots, x_p) \in \mathbb{N}^p$ solutions de l'équation*

$$x_1 + x_2 + \cdots + x_p = n. \tag{A.4}$$

Démonstration. Il suffit de remarquer que l'application

$$(x_1, \ldots, x_p) \quad \mapsto (x_1 + 1, \ldots, x_p + 1)$$

réalise une bijection entre les solutions $(x_1, \ldots, x_p) \in \mathbb{N}^p$ de l'équation (A.4) et les solutions $(x_1, \ldots, x_p) \in (\mathbb{N}\backslash\{0\})^p$ de l'équation

$$x_1 + x_2 + \cdots + x_p = n + p. \tag{A.5}$$

□

Exemple: 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 $(x_1, \ldots, x_p) \in \mathbb{N}^p$ 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

$$\begin{aligned}\left|\bigcup_{i=1}^{n} A_i\right| &= \sum_{B\in\mathcal{P}(\{1,\ldots,n\})\setminus\varnothing} (-1)^{1+\mathrm{Card}(B)} \left|\bigcap_{j\in B} A_j\right| \qquad (A.7)\\ &= \sum_{i=1}^{n}|A_i| - \sum_{1\le i_1<i_2\le n}|A_{i_1}\cap A_{i_2}| + \cdots + \\ &\quad + \cdots + (-1)^{k+1}\sum_{1\le i_1<i_2<\ldots<i_k\le n}|A_{i_1}\cap\ldots\cap A_{i_k}| + \cdots + \\ &\quad + \cdots + (-1)^{n+1}|A_1\cap\ldots\cap A_n|.\end{aligned}$$

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 le chapitre 6).

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 $\binom{n}{a_1,a_2,\dots,a_m}$, considérons l'application de $\Psi^n(a_1,\dots,a_m)$ dans $\Psi^n(a_1,\dots,a_{m-2},a_{m-1}+a_m)$ qui à ϕ associe la fonction $\phi' = \phi\wedge(m-1)$: on remplace chaque occurence de m par $m-1$.
L'image réciproque de $\phi' \in \Psi^n(a_1,\dots,a_{m-2},a_{m-1}+a_m)$ 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

$$\binom{n}{a_1,a_2,\dots,a_m} = \binom{a_m+a_{m-1}}{a_m}\binom{n}{a_1,a_2,\dots,a_{m-2},a_{m-1}+a_m}.$$

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

$$\binom{n}{a_1,a_2,\dots,a_m} = \frac{n!}{a_1!a_2!\dots a_m!}.$$

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 donne un ensemble Ω de 15 entiers compris entre 1000 et 2000. Montrer que l'on peut en extraire deux sous-ensembles disjoints non-vides A et B tels que la somme des éléments de A soit égale à la somme des éléments de B. Indication : montrer que l'application de $\mathcal{P}(\Omega)$ dans $\mathbb{N}$ qui à A associe $\sum\limits_{x\in A} x$ n'est pas injective.
   Raffiner le raisonnement pour montrer que si l'on remplace 15 par 14, le résultat est encore vrai. Indication : remplacer $\mathcal{P}(\Omega)$ par $\bigcup\limits_{i=5}^{8} \mathcal{B}_i(\Omega)$.
3. On considère l'ensemble Ω des suites de n chiffres (les chiffres sont pris dans $\{0,1,\dots,9\}$). Combien vaut $|\Omega|$? Combien y-a-il de chiffres comportant un nombre pair de zéros ?

Annexe B

Compléments

B.1 Équi-intégrabilité

Définition. *On dit qu'une famille $\mathcal{A}$ de variables aléatoires définies sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ est équi-intégrable (ou uniformément intégrable) si*

$$\lim_{M\to+\infty} \sup_{X\in\mathcal{A}} \mathbb{E}[|X|\, \mathbb{1}_{\{|X|\geq M]\}}] = 0.$$

Remarque B.1.
– Une famille constituée d'une seule variable intégrable est équi-intégrable.
– La réunion de deux familles équi-intégrables est équi-intégrable.
Par suite une famille finie de variables intégrables est équi-intégrable.
– Une famille équi-intégrable est toujours bornée dans L^1.

En effet, si M est choisi tel que $\sup_{X\in\mathcal{A}} \mathbb{E}[|X|\, \mathbb{1}_{\{|X|\geq M]\}}] \leq 1$, alors comme $|X| \leq M + |X|\, \mathbb{1}_{\{|X|\geq M\}}$, on a $\mathbb{E}[|X|] \leq M+1$ pour tout $X \in \mathcal{A}$.
– Si la famille $\mathcal{A}$ est équi-intégrable et que pour tout $Y \in \mathcal{B}$, il existe $X \in \mathcal{A}$ avec $|Y| \leq |X|$, alors la famille $\mathcal{B}$ est équi-intégrable.
– Si la famille $\mathcal{A}$ est équi-intégrable, la famille $(\max(|X|,|Y|))_{(X,Y)\in\mathcal{A}^2}$ est équi-intégrable.
En effet, $\max(|X|,|Y|)\, \mathbb{1}_{\{\max(|X|,|Y|\})\geq M\}} \leq |X|\, \mathbb{1}_{\{|X|\geq M\}} + |Y|\, \mathbb{1}_{\{|Y|\geq M\}}$ entraîne

$$\mathbb{E}[\max(|X|,|Y|)\, \mathbb{1}_{\{\max(|X|,|Y|\})\geq M\}}] \leq \mathbb{E}[|X|\, \mathbb{1}_{\{|X|\geq M\}}] + \mathbb{E}[|Y|\, \mathbb{1}_{\{|Y|\geq M\}}].$$

– Par suite, si la famille $\mathcal{A}$ est équi-intégrable, la famille $(X+Y)_{(X,Y)\in\mathcal{A}^2}$ est équi-intégrable.
En effet, il suffit de remarquer que $|X+Y| \leq 2\max(|X|,|Y|)$ et d'appliquer les remarques précédentes.

Le résultat principal est le suivant.

Théorème B.2. *Soit $(X_n)_{n\geq 1}$ une suite équi-intégrable de variables aléatoires. On suppose que X_n converge en loi vers X lorsque n tend vers l'infini. Alors X est intégrable et la suite $(\mathbb{E}X_n)_{n\geq 1}$ converge vers $\mathbb{E}X$.*

Pour ce résultat, on va avoir besoin d'un lemme intermédiaire.

Lemme B.3. *Si $(X_n)_{n\geq 1}$ converge en loi vers X, alors $\mathbb{E}|X| \leq \varliminf_{n\to+\infty} \mathbb{E}|X_n|$.*

Démonstration. Comme $|X_n|$ converge en loi vers $|X|$, on peut se ramener au cas où les variables X_n sont positives. On a pour tout n,

$$\mathbb{E}X_n = \int_{[0,+\infty[} \mathbb{P}(X_n > t)\, d\lambda(t).$$

On sait que $\mathbb{P}(X_n > t)$ converge vers $\mathbb{P}(X > t)$ en tous les points de continuité de F_X. Or les points de discontinuité de F_X sont au plus dénombrables, donc $\mathbb{P}(X_n > t)$ converge λ-presque partout vers $\mathbb{P}(X > t)$. On peut donc appliquer le lemme de Fatou

$$\begin{aligned}
\mathbb{E}X = \int_{[0,+\infty[} \mathbb{P}(X > t)\, d\lambda(t) &= \int_{[0,+\infty[} \varliminf_{n\to+\infty} \mathbb{P}(X_n > t)\, d\lambda(t) \\
&\leq \varliminf_{n\to+\infty} \int_{[0,+\infty[} \mathbb{P}(X_n > t)\, d\lambda(t) \\
&= \varliminf_{n\to+\infty} \mathbb{E}X_n.
\end{aligned}$$

□

Remarque B.4. *Certains auteurs invoquent ce théorème sous le nom de "lemme de Fatou".*

On peut maintenant passer à la démonstration du théorème.

Démonstration du théorème B.2. D'après le lemme X est intégrable, donc la famille $\{X_n; n \geq 1\} \cup \{X\}$ est équi-intégrable. ε étant fixé, on peut trouver M tel que

$$\sup_{n\geq 1} \mathbb{E}[X_n\, \mathbb{1}_{\{X_n \geq M\}}] \leq \varepsilon \text{ et } \mathbb{E}[X \mathbb{1}_{\{X \geq M\}}] \leq \varepsilon. \tag{B.1}$$

Notons que

$$\begin{aligned}
&\mathbb{E}[X_n] = \mathbb{E}[X_n \wedge M] + \mathbb{E}[(X_n - M)\mathbb{1}_{\{X_n \geq M\}}] \\
\text{et}\quad &\mathbb{E}[X] = \mathbb{E}[X \wedge M] + \mathbb{E}[(X - M)\mathbb{1}_{\{X \geq M\}}].
\end{aligned}$$

Comme $0 \le (X_n - M)\mathbb{1}_{\{X_n \ge M\}} \le X_n \mathbb{1}_{\{X_n \ge M\}}$ et
$0 \le (X - M)\mathbb{1}_{\{X \ge M\}} \le X\mathbb{1}_{\{X \ge M\}}$, on en déduit que

$$|\mathbb{E}[X_n] - \mathbb{E}[X]| \le |\mathbb{E}[X_n \wedge M] - \mathbb{E}[X \wedge M]| + \varepsilon.$$

Mais la fonction $x \mapsto x \wedge M$ est continue bornée sur $\mathbb{R}_+$, donc, par définition de la convergence en loi $\lim\limits_{n\to+\infty} \mathbb{E}[X_n \wedge M] = \mathbb{E}[X \wedge M]$, d'où

$$\overline{\lim_{n\to+\infty}} |\mathbb{E}[X_n] - \mathbb{E}[X]| \le \varepsilon,$$

et comme ε est quelconque, $\overline{\lim\limits_{n\to+\infty}} |\mathbb{E}[X_n] - \mathbb{E}[X]| = 0$, ce qui donne le résultat voulu. □

Corollaire B.5. *Soit $(X_n)_{n\ge1}$ une suite de variables aléatoires. On suppose que X_n converge en probabilité vers X lorsque n tend vers l'infini. Si la famille $(|X_n|^p)_{n\ge1}$ est équi-intégrable, alors $X \in L^p$ et X_n converge dans L^p vers X.*

Démonstration. Si X_n converge en probabilité vers X, alors la suite (Y_n) définie par $Y_n = |X_n - X|^p$ converge en probabilité, donc en loi, vers 0. La suite $|X_n|^p$ est équi-intégrable, donc $|X|^p$ est intégrable, soit $X \in L^p$.
Il reste à voir que (Y_n) est équi-intégrable, ce qui découle de l'inégalité
$Y_n \le 2^p \max(|X_n|^p, |X|^p)$ et des remarques faites plus haut. Il suffit alors d'appliquer le théorème à la suite $(Y_n)_{n\ge1}$. □

Remarque B.6. *La manière la plus simple de montrer l'équi-intégrabilité d'une famille est de montrer sa bornitude dans L^p pour un certain $p > 1$.*
En effet si $\mathbb{E}[|X|^p] \le C$ pour tout $X \in \mathcal{A}$, on a pour tout $X \in \mathcal{A}$,

$$\mathbb{E}[|X|\mathbb{1}_{\{|X|\ge M\}}] \le \mathbb{E}[\frac{|X|^p}{M^{p-1}}\mathbb{1}_{\{X\ge M\}}] \le \frac{\mathbb{E}[|X|^p]}{M^{p-1}} \le \frac{C}{M^{p-1}}.$$

Ainsi, si $(X_n)_{n\ge1}$ converge en probabilité vers X et que la suite $(X_n)_{n\ge1}$ est bornée dans L^q, on sait que X_n converge vers X dans L^p pour $p < q$.

On dit parfois que l'équi-intégrabilité est ce qui manque à la convergence presque sûre pour avoir la convergence L^1. Le théorème suivant précise (et renforce) cet énoncé.

Théorème B.7. *Soit $(X_n)_{n\ge1}$ une suite de variables aléatoires positives, intégrables, convergeant en loi vers une variable aléatoire X intégrable. On suppose que $\mathbb{E}X_n$ tend vers $\mathbb{E}X$. Alors les variables aléatoires $(X_n)_{n\ge1}$ sont uniformément intégrables.*

Démonstration. Comme

$$\mathbb{E}[X \, \mathbb{1}_{\{X \geq M\}}] = \mathbb{E}[X] - \int_{[0,M]} \mathbb{P}(t < X < M) \, d\lambda(t)$$

est vraie pout toute variable aléatoire positive X, on a la convergence de $\mathbb{E}[X_n \, \mathbb{1}_{\{X_n \geq M\}}]$ vers $\mathbb{E}[X \, \mathbb{1}_{\{X \geq M\}}]$ si M est un point de continuité de X. On peut donc trouver M tel que $\mathbb{E}[X \, \mathbb{1}_{\{X \geq M\}}] < \varepsilon$. Si M n'est pas un point de continuité de F_X, on le remplace par un point de continuité M' de F_X tel que $M' \geq M$. Pour n_0 assez grand, on a $\mathbb{E}[X_n \, \mathbb{1}_{\{X_n \geq M'\}}] < \varepsilon$ pour $n \geq n_0$. Comme la famille finie $X_1, \dots, X_{n_0 - 1}$ est équi-intégrable, il existe M'' tel que $\mathbb{E}[X_n \, \mathbb{1}_{\{X_n \geq M''\}}] < \varepsilon$ pour tout $n < n_0$.
Ainsi si on prend $M_1 = \max(M', M'')$, on a $\mathbb{E}[X_n \, \mathbb{1}_{\{X_n \geq M_1\}}] < \varepsilon$ pour tout $n \geq 1$. □

Remarque B.8. *On pourrait développer la notion d'équi-intégrabilité sur un espace mesuré (pas nécessairement un espace probabilisé) en disant que (f_n) est équi-intégrable si*

$$\lim_{M \to +\infty} \sup_{n \geq 1} \int |f_n| \, \mathbb{1}_{\{|f_n| \geq M\}} \, d\mu = 0.$$

Dans ce cas, le theorème B.2 s'appelle théorème de Vitali. Mais ce cadre est beaucoup moins intéressant car la convergence presque partout jointe à l'équi-intégrabilité n'implique pas la convergence des intégrales.

Exemple: On prend pour μ la mesure de Lebesgue et $f_n = \frac{1}{n} \mathbb{1}_{[n,2n]}$. Pour $M > 1$, on a $\sup_{n \geq 1} \int |f_n| \, \mathbb{1}_{\{|f_n| \geq M\}} \, d\lambda = 0$ et f_n converge partout vers 0. Cependant, l'intégrale de f_n est constante égale à 1.

B.2 Lemme de recouvrement de Vitali et applications

B.2.1 Lemme de recouvrement de Vitali

Lemme B.9. *Soit X une partie de $\mathbb{R}^d$ et $(r_x)_{x \in X}$ une famille de réels strictement positifs. On suppose que M est un ensemble borné avec*

$$X \subset M \subset \bigcup_{x \in X} B(x, r_x).$$

Alors, il existe $Y \subset X$ avec Y dénombrable, tel que les boules $(B(y, r_y))_{y \in Y}$ soient deux à deux disjointes et

$$M \subset \bigcup_{y \in Y} B(y, 5r_y).$$

Démonstration. Soit K tel que $M \subset B(0,K)$. Si la famille des r_x n'est pas bornée, elle contient alors un élément dépassant $2K$: cette seule boule suffit à recouvrir M (et forme donc une famille d'éléments deux à deux disjoints).

On suppose donc maintenant que la famille des rayons est bornée. On pose $R_0 = \sup\{r_x; x \in X\}$, avec la convention $\sup\{\varnothing\} = -\infty$. Soit $x_0 \in X$ tel que $r_{x_0} \geq R_0/2$. On pose $C_0 = \{x_0\}$. On procède alors par récurrence :

— Si $R_n = -\infty$, alors $R_{n+1} = -\infty$ et $x_{n+1} = 0$.

— Sinon, on pose

$$R_{n+1} = \sup\left\{r_x; x \in X; B(x,r_x) \bigcap \bigcup_{i=0}^{n} B(x_i, r_{x_i}) = \varnothing\right\}.$$

Si $R_{n+1} = -\infty$, on pose $x_{n+1} = 0$, sinon on choisit $x_{n+1} \in X$ tel que

$$B(x_{n+1}, r_{x_{n+1}}) \bigcap \bigcup_{i=0}^{n} B(x_i, r_{x_i}) = \varnothing$$

et

$$r_{x_{n+1}} \geq \frac{1}{2}\sup\left\{r_x; x \in X; B(x,r_x) \bigcap \bigcup_{i=0}^{n} B(x_i, r_{x_i}) = \varnothing\right\}.$$

On remarque donc que soit $R_n \geq 0$, soit $R_n = -\infty$.
Soit $N = \inf\{n : R_n = -\infty\}$. On pose $Y = \{x_i; 0 \leq i < N\}$. Montrons que Y convient. Soit $x \in X$ quelconque. Montrons que $B(x,r_x)$ rencontre $\bigcup_{y \in Y} B(y, r_y)$. Si $N < +\infty$, c'est évident d'après la définition de N.

Si $N = +\infty$, posons $R = \sup_{x \in M} \|x\|$. Comme $\bigcup_{y \in Y} B(y,r_y) \subset B(0, R+R_0)$, on a $\sum_{i=0}^{+\infty} \lambda(B(x_i, r_{x_i})) \leq \lambda(B(0, R+R_0))$, soit $\sum_{i=0}^{+\infty} r_{x_i}^d \leq (R+R_0)^d$, ce qui entraîne que $(r_{x_n})_{n \geq 1}$ tend vers 0. Ainsi, si n est choisi tel que $2r_{x_{n+1}} < r_x$, la définition de x_{n+1} montre que $B(x,r_x)$ doit rencontrer $\bigcup_{i=0}^{n} B(x_i, r_{x_i})$ Soit donc $B(x_n, r_{x_n})$ la boule de plus petit indice qui rencontre $B(x,r_x)$. On a nécessairement $2r_{x_n} \geq r_x$ sinon cela contredirait la définition de x_n. Cela entraîne que $B(x,r_x) \subset B(x_n, 5r_{x_n})$, ce qui achève la preuve. □

Remarque B.10. — *Dans la preuve, on a utilisé la mesure de Lebesgue pour montrer qu'un ensemble borné ne peut contenir qu'un nombre fini uniformément borné de boules de rayon $\ell > 0$ disjointes. Cela pourrait paraître comme un raisonnement circulaire dans la mesure où on peut utiliser le lemme de Vitali dans la construction de la mesure de Lebesgue. Donnons donc une preuve ne faisant pas intervenir la mesure de Lebesgue.*
Soit K tel que $B(0,\ell) \supset B_\infty(0, K\ell)$ (équivalence des normes). Toute boule $B(x,\ell)$ contient un élément $c_x \in K\ell\mathbb{Z}^d$. Soit $(B(x,\ell))_{x \in A}$ une famille de boules disjointes dont la réunion est incluse dans $[-M,M]^d$.

L'application $x \mapsto c_x$ *est injective, à valeurs dans* $[-M, M]^d \cap K\ell\mathbb{Z}^d$, *dont le cardinal est majoré par* $(2M/(K\ell)+1)^d$, *donc*

$$|A| \leq (2M/(K\ell)+1)^d.$$

— *Si* M *n'est pas borné, le résultat est faux. En effet, il suffit de considérer* $\mathbb{R}^d$, *qui est la réunion des boules* $B(0, n)$ *où* n *décrit* $\mathbb{N}$.

B.2.2 Inégalité maximale de Hardy–Littlewood

À toute fonction intégrable au sens de Lebesgue, $f \in L^1(\mathbb{R}^d)$, on peut associer la fonction maximale de Hardy–Littlewood $Mf : \mathbb{R}^d \to \overline{\mathbb{R}}_+$ définie par

$$Mf(x) = \sup_{r>0} \frac{1}{\lambda\left(B(x,r)\right)} \int_{B(x,r)} |f(t)|\, d\lambda(t). \tag{B.2}$$

Théorème B.11. *Pour toute fonction* $f \in L^1(\mathbb{R}^d)$ *et tout* $\eta > 0$, *on a*

$$\lambda\left(\{Mf > \eta\}\right) \leq \frac{5^d ||f||_1}{\eta}.$$

Démonstration. Par définition de la borne supérieure, si $Mf(x) > \eta$, il existe $r_x > 0$ tel que $\int_{B(0,r_x)} |f|\, d\lambda > \eta\lambda(B(0, r_x))$.

Soit $K > 0$. On a évidemment

$$B(0, K) \cap \{Mf > \eta\} \subset \bigcup_{\substack{x : x \in B(0,K), \\ Mf(x) > \eta}} B(x, r_x).$$

Comme $B(0, K) \cap \{Mf(x) > \eta\}$ est borné, on peut considérer la famille associée à cette famille de boules par le lemme de Vitali. Comme la famille des boules $(B(y, r_y))_{y\in Y}$ est disjointe et dénombrable, on a

$$\int |f|\, d\lambda \geq \sum_{y\in Y} \int_{B(y,r_y)} |f|\, d\lambda \geq \sum_{y\in Y} \eta\lambda(B(y, r_y)),$$

d'après la propriété définissant les r_x. D'autre part

$$\sum_{y\in Y} \eta\lambda(B(y, r_y)) = \sum_{y\in Y} \frac{1}{5^d}\eta\lambda(B(y, 5r_y)) \geq \frac{\eta}{5^d}\lambda\left(\bigcup_{y\in Y} B(y, 5r_y)\right).$$

Mais

$$\bigcup_{y\in Y} B(y, 5r_y) \supset B(0, K) \cap \{Mf > \eta\},$$

d'où finalement

$$\|f\|_1 \geq \frac{1}{5^d}\eta\lambda(B(0, K) \cap \{Mf > \eta\}).$$

Avec le théorème de continuité séquentielle croissante, on obtient

$$\|f\|_1 \geq \frac{1}{5^d}\eta\lambda(Mf > \eta),$$

ce qui donne le résultat voulu. □

B.2.3 Théorème de différentiation de Lebesgue

Théorème B.12. *Soit $f \in L^1$. On pose, pour $r > 0$*

$$Q_r f = \frac{1}{r^d V_d} \int_{B(x,r)} f(t)\, d\lambda(t),$$

où V_d est la volume de la boule unité de $\mathbb{R}^d$. Alors

$$\lim_{r\to 0} Q_r f = f \quad \lambda\text{-presque partout.}$$

Démonstration. On peut déjà noter que

$$\begin{aligned} |Q_r f(x) - f(x)| &= \left| \frac{1}{r^d V_d} \int_{B(x,r)} (f(y) - f(x))\ d\lambda(y) \right| \\ &\leq \frac{1}{r^d V_d} \int_{B(x,r)} |f(y) - f(x)|\, d\lambda(y). \end{aligned}$$

Pour $f \in L^1$, on pose $\Delta f(x) = \varlimsup_{r\to 0^+} \frac{1}{r^d V_d} \int_{B(x,r)} |f(y) - f(x)|\, d\lambda(y)$. Notre but est de montrer que Δf est nulle presque partout.
Notons qu'il n'est pas difficile de voir que quelles que soient les fonctions g et h dans L^1, on a $\Delta(g+h) \leq \Delta g + \Delta h$.

Considérons d'abord le cas où f est continue à support compact. Alors, f est uniformément continue sur $\mathbb{R}^d$. Notons

$$\omega_f(\eta) := \sup\{|f(x) - f(y)|; |x - y| \leq \eta\}.$$

Alors ω_f a une limite nulle en 0. On a aisément

$$\frac{1}{r^d V_d} \int_{B(x,r)} |f(y) - f(x)|\, d\lambda(y) \leq \omega_f(r), \text{ donc } \Delta f(x) = 0.$$

Soient maintenant f quelconque dans L^1 et $\eta > 0$. On veut montrer que $\lambda(\Delta f > 2\eta) = 0$. Soit $\varepsilon > 0$. Comme l'ensemble des fonctions continues à support compact est dense dans L^1, il existe g continue à support compact et h dans L^1 avec $f = g + h$ et $\|h\|_1 \leq \varepsilon$. Ainsi $\Delta f = \Delta(g+h) \leq \Delta g + \Delta h = \Delta h$. Cependant, si l'on pose $h_x(y) = h(x) - h(y)$, il vient aisément

$$\Delta h(x) \leq Mh_x(x) \leq Mh(x) + |h(x)|,$$

où M est l'opérateur défini par (B.2). Ainsi

$$\begin{aligned}\lambda(\Delta f > 2\eta) &\le \lambda(\Delta h > 2\eta) \le \lambda(Mh > \eta) + \lambda(h > \eta)\\ &\le \frac{5^d\|h\|_1}{\eta} + \frac{\|h\|_1}{\eta} \le \frac{5^d+1}{\eta}\varepsilon,\end{aligned}$$

grâce aux inégalités de Hardy–Littlewood et de Markov. Comme ε peut être pris arbitrairement petit, on a $\lambda(\Delta f > 2\eta) = 0$, et par continuité séquentielle décroissante $\lambda(\Delta f > 0) = 0$, ce qui achève la preuve. □

B.3 Régularité des mesures

Soit X un espace métrique, muni de sa tribu borélienne et d'une mesure m. On dit que la mesure m est régulière si, pour toute partie mesurable A, on a :

$$m(A) = \inf\{m(O); O \text{ ouvert }, O \supset A\} = \sup\{m(F); F \text{ compact }, F \subset A\}.$$

Théorème B.13. *Une mesure sur $\mathbb{R}^d$ qui assigne une masse finie aux ensembles bornés est régulière.*

Démonstration. On commence par traiter le cas où m est une mesure finie. On pose

$$\mathcal{C} = \left\{A \in \mathcal{B}(\mathbb{R}^d)|_{\substack{\forall\eta>0,\exists O \text{ ouvert et } F \text{ fermé}\\ \text{tels que } F\subset A\subset O \text{ et } m(O\backslash F)<\eta}}\right\}.$$

1) On montre que $\mathcal{C}$ est stable par passage au complémentaire.
Soit $A \in \mathcal{C}$. Montrons que $A^c \in \mathcal{C}$. Soit $\eta > 0$. Par définition de $\mathcal{C}$, il existe O ouvert et F fermé tels que $F \subset A \subset O$ et $m(O\backslash F) < \eta$. On a $O^c \subset A^c \subset F^c$, où O^c est fermé, F^c est ouvert, et $m(F^c\backslash O^c) = m(O\backslash F) < \eta$. Comme on trouve un tel couple quel que soit $\eta > 0$, on a bien $A^c \in \mathcal{C}$.

2) On montre que $\mathcal{C}$ contient les fermés.
Soient F un fermé et $\eta > 0$. Notons $O_\varepsilon = \{x \in \mathbb{R}; d(x,F) < \varepsilon\}$. Comme l'application $x \mapsto d(x,F)$ est continue (elle est même lipschitzienne de constante 1), O_ε est un ouvert. Comme $\bigcap_{n=1}^{+\infty} O_{1/n} = F$, on a donc $\bigcap_{n\ge1} \left(O_{1/n}\backslash F\right) = \varnothing$. D'après le théorème de continuité séquentielle décroissante [1], on a

$$\lim_{n\to+\infty} m(O_{1/n}\backslash F) = m(\varnothing) = 0,$$

donc il existe n tel que $m(O_{1/n}\backslash F) < \eta$. Comme η est quelconque, on a bien $F \in O$.

1. C'est ici que la finitude de la mesure sert.

3) On montre que $\mathcal{C}$ est une tribu.
Soit $(A_n)_{n\geq 1}$ une famille d'éléments de $\mathcal{C}$. Posons $A = \bigcup_{n=1}^{+\infty} A_n$. On va montrer que $A \in \mathcal{C}$. Soit $\eta > 0$. Pour tout $n \geq 1$, il existe un fermé F_n et un ouvert O_n tels que $F_n \subset A_n \subset O_n$ et $m(O_n \backslash F_n) < \eta/2^{n+1}$. Posons $O = \bigcup_{n\geq 1} O_n$. Par construction, O est ouvert et $A \subset O$. Posons

$$R = \bigcup_{n\geq 1} (O_n \backslash F_n).$$

Notons que $m(R) \leq \sum_{n=1}^{+\infty} m(O_n \backslash F_n) \leq \eta/2$.

Posons $A'_n = \bigcup_{i=1}^{n} A_i$. (A'_n) est une suite croissante d'ensembles dont la réunion est A. D'après le théorème de continuité séquentielle croissante, on a donc $m(A) = \lim_{n\to+\infty} m(A'_n)$. Il existe donc n_0 tel que $m(A'_{n_0}) > m(A) - \eta/2$. Posons $F = \bigcup_{i=1}^{n_0} F_i$. Par construction, F est fermé. On a

$$O \backslash F = (O \backslash A) \cup (A \backslash F) \quad = \quad (O \backslash A) \cup (A \backslash A'_{n_0}) \cup (A'_{n_0} \backslash F)$$

Or on a d'une part

$$O \backslash A = \bigcup_{n=1}^{+\infty} (O_n \backslash A) \subset \bigcup_{n=1}^{+\infty} (O_n \backslash A_n) \subset \bigcup_{n=1}^{+\infty} (O_n \backslash F_n) = R$$

et d'autre part

$$A'_{n_0} \backslash F = \bigcup_{i=1}^{n_0} A_i \cap F^c \subset \bigcup_{i=1}^{n_0} A_i \cap F_i^c \subset \bigcup_{i=1}^{n_0} O_i \cap F_i^c \subset R.$$

Ainsi $O \backslash F \subset (A \backslash A'_{n_0}) \cup R$ et $m(O \backslash F) \leq m(A \backslash A'_{n_0}) + m(R) < \eta$.

Par les trois points précédents, $\mathcal{C}$ est donc une sous-tribu de la tribu borélienne de $\mathbb{R}^d$. Mais elle contient tous les fermés, qui engendrent la tribu borélienne de $\mathbb{R}^d$, donc $\mathcal{C}$ est la tribu borélienne de $\mathbb{R}^d$.

Pour tout fermé F inclus dans A, on a $m(F) \leq m(A)$. Mais d'après ce qui précède, pour tout $\eta > 0$ on peut trouver un fermé F et un ouvert O tels que $F \subset A \subset O$ et $m(O \backslash F) < \eta$:

$$m(F) = m(A) - m(A \backslash F) \geq m(A) - m(O \backslash F) > m(A) - \eta.$$

Cela montre que $m(A) = \sup\{m(F); F \text{ fermé } \subset A\}$.

L'identité avec les ouverts se traite de la même manière. D'autre part, pour tout F fermé, on a d'après le théorème de continuité séquentielle croissante $m(F) = \lim_{n\to+\infty} m(F \cap B(0,n))$, ce qui entraîne qu'on a aussi pour tout borélien A

$$m(A) = \sup\{m(K);\ K \text{ compact }, K \subset A\}.$$

La preuve est achevée dans le cas d'une mesure finie. Passons au cas général. Soit $\alpha < m(A)$. D'après le théorème de continuité séquentielle croissante, il existe N tel que $m(A \cap B(0,N)) > \alpha$.
La mesure m' définie par $m'(B) = m(B \cap B_F(0,N))$ est une mesure finie, car $B_F(0,N)$ est borné. Comme $m'(A) > \alpha$, on déduit de la première partie de la preuve qu'il existe K compact avec $K \subset A$, et $m'(K) > \alpha$. La mesure sous m de $K \cap B_F(0,N)$ dépasse donc α. Comme $K \cap B_F(0,N)$ est compact, cela donne la propriété voulue d'approximation par des compacts.

Passons à l'approximation par des ouverts. Si $m(A) = +\infty$, il n'y a rien à démontrer. Sinon, fixons $\varepsilon > 0$ et, pour n dans $\mathbb{Z}^d$, posons $U_n = n + [0,1[^d$. L'adhérence de U_n est compacte, donc comme précédemment, la mesure m'_n définie par $m'_n(B) = m(B \cap U_n)$ est une mesure finie, et il existe un ouvert O_n tel que $O_n \supset A \cap U_n$ et $m'_n(O_n) \le m'_n(A \cap U_n) + \frac{\varepsilon}{3^d 2^{\|n\|_1}}$, soit encore $m(O_n) \le m(A \cap U_n) + \frac{\varepsilon}{3^d 2^{\|n\|_1}}$. L'ensemble $O = \bigcup_{n \in \mathbb{Z}^d} O_n$ est ouvert, contient A et on a

$$m(O) \le \sum_{n \in \mathbb{Z}^d} m(O_n) \le \sum_{n \in \mathbb{Z}^d} m(A \cap U_n) + \frac{\varepsilon}{3^d 2^{\|n\|_1}} = m(A) + \varepsilon.$$

□

B.4 Exercices

1. Soit (X_n) une suite de variables aléatoires positives avec

$$\sup_{n \ge 1} \mathbb{E}[X_n \log(1 + X_n)] < +\infty.$$

Montrer que cette famille est équi-intégrable.

2. *Une caractérisation utile de l'équi-intégrabilité.*
Montrer que la famille $\mathcal{A}$ de variables aléatoires définies sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ est équi-intégrable si et seulement si elle est bornée dans L^1 et vérifie

$$\lim_{\eta \to 0} \sup_{\substack{X \in \mathcal{A} \\ A \in \mathcal{F} : \mathbb{P}(A) \le \eta}} \mathbb{E}[|X|\, \mathbb{1}_A] = 0.$$

3. *Loi des grands nombres L^1.*
Soit $(X_n)_{n \ge 1}$ une suite de variables aléatoires indépendantes identiquement distribuées avec un moment d'ordre 1. La loi faible des grands nombres nous dit que la suite $\frac{X_1 + \cdots + X_n}{n}$ converge en probabilité vers

$\mathbb{E}[X_1]$. À l'aide de l'exercice précédent, montrer qu'elle converge aussi dans L^1 vers $\mathbb{E}[X_1]$.

Annexe C

Indications des exercices

C.1 Exercices 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 6

1. On pourra d'abord traiter le cas où $\{s_N\} \leq x$, puis s'y ramener.
2. Commencer par montrer que les points de $]0, 1[$ sont valeurs d'adhérence.
3. Noter que la suite (z^n) a au moins une valeur d'adhérence et en déduire qu'on peut trouver une puissance de z arbitrairement proche de 1.
4. Noter que $z = \exp(2i\pi\theta)$ n'est pas une racine de l'unité.

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 ...et b_n ...donc $a_n + b_n$ ...

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 $R_n = \sum\limits_{k=n+1}^{+\infty} \frac{u_k}{k^x}$.

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, $(1 - \varepsilon)a \leq (a_n)^n \leq (1 + \varepsilon)a$.

C.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 $A_K = \underset{k\in K}{\cup} \Omega_k$, 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. Vérifier que vous connaissez les définitions des mots intervenant dans l'énoncé, en particulier la notion de tribu engendrée par une application et celle de mesurabilité par rapport à un couple d'espaces mesurés.

2. Si Y prend les valeurs $y_1, \dots, y_n$, on cherchera à écrire $Y = f \circ X$, avec $f(x) = \sum_{i=1}^{n} y_k \mathbb{1}_{A_k}$.
3. On peut poser $\phi_n(x) = \frac{\lfloor nx \rfloor}{n} \mathbb{1}_{[0,n]}(|x|)$ et construire soigneusement une fonction f_n telle que $f_n \circ X = \phi_n \circ Y$. Si vous séchez, on peut trouver la preuve dans [18] dans un cadre plus général.

Indication 38 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.

C.3 Exercices sur le formalisme probabiliste

Indication 39 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 40 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 41 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 42 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 43 1. $A_d = d\mathbb{N} \cap \Omega_n$.

2. On déterminera un entier d tel que $\bigcap\limits_{i=1}^{r} A_{d_i} = A_d$.
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 44 1. On pose $n_0 = 0$, $s_0 = 0$, puis pour $k \geq 0$:
$n_{k+1} = \inf\{n > n_k; s_k + u_n < \ell\}$ et $s_{k+1} = s_k + u_{n_{k+1}}$.

2. Utiliser la formule donnant $\phi(n)/n$ en fonction de ses facteurs premiers et utiliser la divergence de la série des inverses des nombres premiers.

Indication 45 On pourra poser $p_n = \mu(\{n\})$ et $r_n = \mu([n, +\infty[)$.

Indication 46 Remarquer qu'une union dénombrable d'événements est de probabilité nulle si et seulement si chacun est de probabilité nulle.

Indication 47 1. Appliquer la formule du crible. 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.

2. Bien sûr $\sum\limits_{k \in A} \frac{1}{k^2} \leq \sum\limits_{k=1}^{+\infty} \frac{1}{k^2} < +\infty$. On conclut à l'aide d'un raisonnement classique sur les séries à paramètre : on se fixe M tel que $\sum\limits_{k \in A; k > M} \frac{1}{k^2} \leq \varepsilon$, 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 $\varlimsup\limits_{n \to +\infty} |\mathbb{P}_n(A) - \sum_{k \in A} \frac{\mu(k)}{k^2}| \leq \varepsilon$. Pour un énoncé général, on pourra voir également le théorème 4.36.

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 $N_k(E) = |E \cap \{1, \dots, k\}|$, 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
$$\begin{aligned} & \sum_{k=M}^{+\infty} \frac{|N_k(E) - k\ell|}{\zeta(s)} \left(\frac{1}{k^s} - \frac{1}{(k+1)^s}\right) \\ \leq \quad & \varepsilon \sum_{k=M}^{+\infty} \frac{|N_k(\mathbb{N}^*)|}{\zeta(s)} \left(\frac{1}{k^s} - \frac{1}{(k+1)^s}\right) \\ \leq \quad & \varepsilon\mu_s(\mathbb{N}^*) = \varepsilon. \end{aligned}$$
Pour conclure, il faut utiliser que $\lim_{s \to 1^+} \zeta(s) = +\infty$.
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].

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

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 48 Utiliser la formule de Bayes.

Indication 49 Commencer par choisir clairement l'espace Ω.

Indication 50 On pourra conditionner par la valeur prise par l'*ensemble* des trois nombres tirés au sort.

Indication 51

1. L'ensemble
$$\Omega = \{(x_1, \dots, x_{a+b}) \in \{\overrightarrow{i}\,;\overrightarrow{j}\}^{a+b} :\ x_1 + \cdots + x_{a+b} = a\overrightarrow{i} + b\overrightarrow{j}\}$$
est un choix possible.
2. Si l'on note $I = \{$ le graphe coupe la diagonale.$\}$, on pourra montrer par une bijection que les événements $I \cap \{$A gagne le premier échange$\}$ et $I \cap \{$B gagne le premier échange$\}$ ont la même probabilité.

Indication 52 Faire en sorte que le résultat soit nul.

C.4 Exercices sur les intégrales

Indication 53 On pourra noter que $\{\frac{1}{x}\} = \lim\limits_{n\to+\infty} \{\frac{1}{x}\}\mathbb{1}_{]1/n,1]}(x)$.

Indication 54 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 55

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 56

1. On peut montrer que pour $n \geq 1$ et $x \geq 0$, $\log(n+x) = \log(n) + \log(1 + x/n)$.
2. Minorer $(1+x)^n$.

l'ouvrage de Bordellès [3].

3. Pour $n \geq 3$, $\frac{1+nx}{(1+x)^n} \leq \frac{3}{(1+x)^2}$.

Indication 57

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 58

1. $\cos^2 = 1 - \sin^2$ et intégrer par parties.
2. Calculer $W_0 W_1$.

Indication 59

1. Appliquer le théorème de convergence dominée.
2. Faire (au moins) un changement de variable.

Indication 60 Faire un changement de variable.

Indication 61

1. Bien contrôler le reste.
2. Revoir les techniques usuelles du calcul des primitives.

Indication 62 Fubini est ton ami.

Indication 63

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 64 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 65 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 66 Penser à la formule d'intégration des fonctions radiales ; utiliser les symétries.

Indication 67 1. Commencer par exhiber une partition de $\mathbb{R}^d$.

2. Commencer par montrer que $\lambda(B) = \sum_{s\in\mathbb{Z}^d} \lambda((s+B)\cap T)$.
3. Revoir le lien entre mesure de Lebesgue et transformations affines.
4. (a) Pour $C, D \subset E$, $(C \cap D = \varnothing) \Longrightarrow |C| + |D| = |C \cup D| \leq |E|$, donc par contraposée…
   (b) Raisonner modulo p, puis appliquer le théorème de Minkowski.
   (c) Commencer par montrer que tout nombre premier s'écrit comme somme de 4 carrés.

Indication 68 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 69 On pourra remarquer que $\frac{1}{e^x-1} = \frac{e^{-x}}{1-e^{-x}}$, puis penser à son développement en série.

Indication 70 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 71 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 72 1. Il suffit de vérifier que T est C^1, injective, et que sa différentielle est toujours inversible.

2. Appliquer le théorème de changement de variable C^1.
3. Déterminer les coordonnées du centre de gravité g_0 de $V \times \{0\}$, et reconnaître la courbe décrite par $M_\theta g_0$ quand θ varie entre α et β.
4. Appliquer à nouveau le théorème de changement de variable C^1.

Indication 73 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 74 On peut appliquer la technique de la fonction test et le théorème de Tonelli.

Indication 75 1. Écrire $F_n(t)$ sous forme d'une intégrale qui ressemble à l'intégrale de Dirichlet.

2. Faire une transformation d'Abel.
3. Penser à une série de Bertrand.

Indication 76 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 77 Utiliser la transformation d'Abel.

Indication 78 On peut trouver une constante A telle que $g \leq A|f|$.

Indication 79 1. Vérifier les axiomes.

2. Commencer par le cas où f est étagée.

Indication 80 Utiliser le théorème de convergence monotone.

Indication 81 Utiliser le théorème de convergence dominée.

Indication 82 On pourra prendre $A = \{f > 0\}$.

Indication 83
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 84 Remarquer que $\log x + \frac{1}{x} \geq 0$ et développer $e^x - 1$ en série.

Indication 85 Développer en série et utiliser le théorème de convergence dominée.

Indication 86 On pourra remarquer que $\frac{1}{e^x-1} = \frac{e^{-x}}{1-e^{-x}}$.

Indication 87 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 88 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 89 On pourra intégrer des fonctions par rapport à la mesure de comptage et utiliser le théorème de convergence dominée.

Indication 90 On pourra procéder par récurrence sur n, ou commencer par prolonger la fonction Γ.

Indication 91 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 92 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 93 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 $\alpha = 2$.

Indication 94

1. Utiliser une partition de l'espace $\mathbb{R}^d$.
2. On pourra montrer qu'il existe des constantes positives C et D telles que pour tout x avec $\|x\|_2 \geq 1$, on ait
$$\frac{C}{\|x\|_2^\alpha} \leq \frac{1}{\text{Ent}(\|x\|_\infty)^\alpha} \leq \frac{D}{\|x\|_2^\alpha}.$$
3. Le plus simple ici est de montrer que f est C^1 par des théorèmes généraux, puis de calculer la différentielle dans une direction donnée.
4. On pourra utiliser le théorème de C^1-difféomorphisme. Pour le calcul de la différentielle, il pourra être intéressant d'écrire la matrice de Df_x dans une base orthogonale bien choisie.

Indication 95

1. Faire une intégration par parties.
2. Oui !
3. (a) Faire une intégration par parties.
   (b) Pour le premier point, noter que Ψ est continue, de limite 1 en l'infini. Enfin, appliquer le théorème de convergence dominée.
   (c) On peut noter que $\Psi(x)/x^2$ est continue sur $\mathbb{R}_+^*$, avec des limites finies en $+\infty$ et en 0. Pour la dernière inégalité, on pourra écrire
$$\int_x^{+\infty} \frac{\Psi(\lambda u)}{u^2}\,du = \int_x^1 \frac{\Psi(\lambda u)}{u^2}\,du + \int_1^{+\infty} \frac{\Psi(\lambda u)}{u^2}\,du.$$
   (d) Dans un premier temps, dominer $|R_\lambda(x) - R(x)|$ par une fonction intégrable. Dans un second temps, pour $\lambda \leq 1$, dominer la fonction $|R_\lambda(x) - R(x)|$ par une fonction intégrable ne dépendant pas de λ.

4. Utiliser le théorème de Fubini.
5. On pourra écrire
$$\int_0^{+\infty} \frac{R_\lambda(x)}{\sqrt{x}}\, dx = \int_0^1 \frac{R_\lambda(x)}{\sqrt{x}}\, dx + \int_1^{+\infty} \frac{R_\lambda(x)}{\sqrt{x}}\, dx.$$
La première intégrale se traite aisément par convergence dominée, tandis que la deuxième nécessite une intégration par parties comme précédemment.

Indication 96 1. Séparer les problèmes en 0 et en l'infini.

2. On pourra calculer la somme $e^{-t} + e^{-2t} + \cdots + e^{-nt}$.
3. Se placer sur un intervalle $]a, +\infty[$, avec $a > 0$.
4. Utiliser le théorème de convergence dominée.
5. On pourra remarquer que $\psi(1) = \log n + J_n - H_{n-1}$ pour tout $n \geq 1$.
6. Utiliser $\Gamma(x+1) = x\Gamma(x)$ et la continuité de Γ.
7. Exprimer $\beta(1-\varepsilon, \varepsilon) - \beta(t-\varepsilon, \varepsilon)$ sous forme intégrale, puis appliquer le théorème de convergence dominée.
8. On pourra d'abord montrer par un changement de variable que
$$\frac{\Gamma'(t)}{\Gamma(t)} - \frac{\Gamma'(1)}{\Gamma(1)} = \psi(t) - \psi(1).$$

C.5 Exercices sur les lois

Indication 97 On pourra commencer par déterminer la loi de X_3.

Indication 98 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 99 Pour montrer que les F_i sont des fonctions de répartition, il faut vérifier qu'elles sont continues à droites, croissantes et satisfont

$\lim\limits_{n\to-\infty} F_i(n) = 0$ et $\lim\limits_{n\to+\infty} F_i(n) = 1$.

Indication 100 Transformer la condition sur y en condition sur Φ.

Indication 101 Il s'agit de déterminer la constante de normalisation.

Indication 102 Penser aux valeurs effectivement prises par X.

Indication 103 Introduire le discriminant.

Indication 104 On pourra éventuellement introduire l'angle entre les droites (AB) et (BC).

Indication 105 Reconnaître une loi binomiale. Former le quotient des probabilités respectives de deux entiers consécutifs.

Indication 106 Pour la dernière question, il pourra être utile de se souvenir qu'une fonction de répartition est continue à droite.

Indication 107 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 108

1. On pourra par exemple calculer la densité du couple (X, Y).
2. Se ramener à un calcul d'aire.

Indication 109 Penser au théorème de Tonelli.

Indication 110 On rappelle que la fonction de répartition caractérise la loi des variables aléatoires réelles.

Indication 111

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 112 Si X n'est pas constante, il existe a tel que $\mathbb{P}(X \le a) > 0$ 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 113

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 114

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 115 Calculer $\mathbb{P}(1 + \nu_{p_1}(X) > k_1, \dots, 1 + \nu_{p_n}(X) > k_n)$.

Indication 116 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 117 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 118 Traduire les événements considérés en fonction de Y.

Indication 119 Poser $x = AM$ et résoudre l'inéquation.

Indication 120
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 121 À k fixé, il faut déterminer les valeurs de X qui sont telles que $Y = k$.

Indication 122 Représenter graphiquement les domaines considérés.

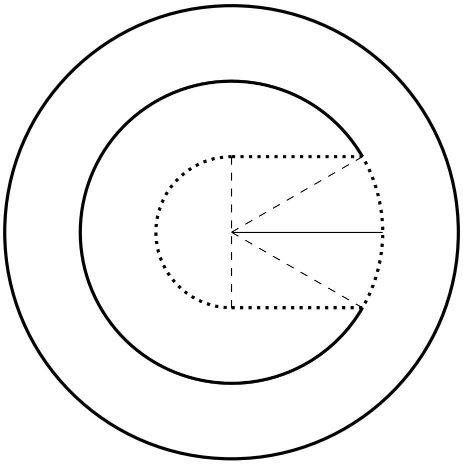
Indication 123
1. On pourra montrer que X est $\mathcal{T}$-mesurable, puis que les A_p sont $\sigma(X)$-mesurables.
2. On montrera que l'ensemble cherché est $\{\omega \in \mathbb{N}; X(\omega) = 2\times3\times5\times11\}$.
3. Le sens direct est le plus simple. Pour la réciproque, remarquer que si A_n est $\mathcal{T}$-mesurable, il doit contenir $\{\omega \in \mathbb{N}^*; X(\omega) = X(n)\}$.
4. On montrera que $A_m \cap A_n = A_{m\wedge n}$.
5. Remarquer que pour p assez grand $1 - \frac{1}{p^{2s}} \geq e^{-\frac{2}{p^{2s}}}$.

Indication 124 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 125
1. On pourra montrer que $\mathcal{Q}$ est la tribu de queue associée à la famille $(\mathcal{A}_n)_{n\geq1}$.
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 126 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 127 1. On pourra remarquer que pour $i \neq j$, $X_i - X_j$ est une variable à densité.

2. Pour la première égalité, utiliser la question précédente ; ensuite utiliser le théorème de transfert.
3. Appliquer le théorème de Fubini.
4. On pourra par exemple calculer $\mathbb{P}_{(T,N)}$ sur des ensembles de type $]a, +\infty[\times]b, +\infty[$.

Indication 128 Commencer par calculer la fonction de répartition.

C.6 Exercices sur les espérances

Indication 129 On peut utiliser la méthode de la fonction test, ou calculer la fonction de répartition.

Indication 130 1. Faire apparaître un polynôme du second degré.

2. Considérer une borne des réels vérifiant une des deux conditions.

Indication 131 On pourra montrer que pour toute fonction H mesurable bornée, on a $\mathbb{E}[H(\{\frac{1}{X}\})] = \mathbb{E}[H(X)]$.

Indication 132 Appliquer le théorème de transfert.

Indication 133 Penser à utiliser la fonction de queue.

Indication 134 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 135 On pourra appliquer la formule du crible et revoir les résultats en annexe sur le nombre de solutions d'équations en entiers.

Indication 136 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 137 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 138 Pour la première question, on peut se ramener au cas où les variables suivent des lois de Bernoulli.

Indication 139 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 140 On pourra établir que $\mathbb{E}|X - Y| = \mathbb{E}(f(Y))$.

Indication 141 Pour $\min(X, Y)$, penser à utiliser la fonction de queue.

Indication 142 1. Penser à utiliser la fonction de queue.

2. Faire apparaître une somme de Bernoulli.

Indication 143 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 144 1. Utiliser un C^1-difféomorphisme.

2. On peut noter que X/Y n'est pas de carré intégrable.

Indication 145 On peut remarquer que $G_{k+1} - G_k = -b + (a+b+c)\mathbb{1}_{\{N>k\}}$.

Indication 146 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 147 1. Utiliser un C^1-difféomorphisme.

2. La question est difficile sans quelque familiarité avec les coordonnées barycentriques : si les points a, b, c du plan forment un triangle non plat, alors ils constituent une base barycentrique du plan, c'est-à-dire que tout point P du plan s'écrit de manière unique sous la forme
$$P = U_1 a + U_2 b + U_3 c,$$
avec $U_1 + U_2 + U_3 = 1$. Les points dont les trois coordonnées barycentriques U_1, U_2, U_3 sont positives forment l'enveloppe convexe de a, b, c : c'est le triangle plein. On sait que l'aire du triangle abP est la moitié de la valeur absolue du déterminant $[\overrightarrow{Pa}, \overrightarrow{Pb}] = [\overrightarrow{aP}, \overrightarrow{ba}]$. Un calcul simple donne
$$\begin{aligned}[\overrightarrow{aP}, \overrightarrow{ba}] &= [(U_1 - 1)a + U_2 b + U_3 c, \overrightarrow{ba}] \\ &= [(-U_2 - U_3)a + U_2 b + U_3 c, \overrightarrow{ba}] \\ &= U_2[-a + b, \overrightarrow{ba}] + U_3[c - a, \overrightarrow{ba}] \\ &= U_2[\overrightarrow{ab}, \overrightarrow{ba}] + U_3[\overrightarrow{ac}, \overrightarrow{ba}] \\ &= U_3[\overrightarrow{ac}, \overrightarrow{ba}]\end{aligned}$$

Indication 148 Pour disqualifier l'indépendance, observer le support de la loi. Faire ensuite un changement de variables polaire.

Indication 149 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 150 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 151 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 152 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 153 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 154 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 155 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 $\{N_t = n\} = \{S_n < t \leq S_{n+1}\}$.

Indication 156 Si X désigne le nombre de fois où l'on a obtenu le nombre choisi, le gain est $X - \mathbb{1}_{\{X=0\}}$.

Indication 157 Bien observer que X et Y sont indépendantes.

Indication 158 Une probabilité est l'espérance d'une indicatrice.

Indication 159 Interpréter le membre de gauche comme la valeur absolue d'une covariance.

Indication 160 Appliquer le théorème de transfert.

Indication 161 Si on note C_x l'ensemble des cases contrôlées par la dame i, et $C = \bigcup_{i=1}^{d} C_i$, on peut minorer $|C|$ grâce aux inégalités de Bonferroni. Dans un second temps, on remarquera que si ϕ est une bijection de l'échiquier dans lui-même, les ensembles C et $\phi(C)$ se coupent dès que $|C| > n^2/2$.

Indication 162

1. On montrera que $\mathbb{E}(X-a)^2 - \sigma^2 = (m-a)^2$.
2. Prendre $a = \frac{a_1+a_n}{2}$.

Indication 163 Effectuer une transformation d'Abel.

Indication 164 Appliquer le théorème de transfert, et penser à la forme canonique des polynômes du second degré.

Indication 165 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 166 Fonction de répartition, ou transformation : on a le choix !

Indication 167 On peut raisonner en terme de loi.

Indication 168 S'inspirer de l'exercice précédent et utiliser une transformation.

Indication 169 Appliquer l'exercice précédent.

Indication 170

1. Se rappeler la définition d'une densité. Il s'agit d'une fonction mesurable, positive et intégrable d'intégrale valant 1.
2. On aura intérêt à prendre k la plus petite possible.
   On trouvera $k = \sqrt{\frac{2e}{\pi}} = 1,3155$.
3. Relire le cours et écrire un programme informatique.
4. La méthode de Box-Muller est une méthode directe.

Indication 171 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 172

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 173 1. On prendra $\Omega = \mathcal{B}_r(\{1,\ldots,n\})$.

2. Utiliser les relations entre l'espérance et la queue de distribution, puis utiliser la relation de récurrence du triangle de Pascal.

Indication 174 1. On pourra remarquer qu'une rotation est une application linéaire isométrique.

2. Remarquer que $|X-Y| = \sqrt{2}U$.

Indication 175 1. On trouvera que les X_i sont des variables de Bernoulli indépendantes, avec $X_i \sim \mathrm{Ber}(p_i)$.

2. $\mathbb{P}(Y_i \neq X_i) = \mathbb{P}(Y_i \geq 2) + \mathbb{P}(Y_i = 0, Z_i = 1)$.
3. (a) Grâce aux propriétés de convolution des lois de Poisson, on trouvera que $Y \sim \mathcal{P}(p_1 + \cdots + p_n)$.

   (b) Remarquer que $|\mathbb{1}_{\{X\in A\}} - \mathbb{1}_{\{Y\in A\}}| \leq \mathbb{1}_{\{X\neq Y\}}$ et aussi que $\mathbb{P}(X \in A) - \mathbb{P}(Y \in A) = \mathbb{E}[\mathbb{1}_{\{X\in A\}} - \mathbb{1}_{\{Y\in A\}}]$.

C.7 Exercices sur les espaces L^p

Indication 176 1. On prend des exposants α et β conjugués pour appliquer l'inégalité de Hölder, puis on remarque que $p\alpha > p$.

2. $N(p)$ est minorée par 0. On utilise le théorème de transfert et un équivalent du $\log$ en 0.
3. On applique le théorème de Tonelli car tout est positif.
4. (a) X est comprise en 0 et M p.s.

   (b) On utilise la question précédente.

   (c) Inégalité de Markov

   (d) On remet tout ensemble.
5. Cas où $\log X$ est intégrable : on remarque que pour tout x, on a

$$|e^x - 1| \leq |x| \max(1, e^x)$$

et on utilise la convergence dominée. Sinon, on remarque que $(\log X)^+$ est intégrable (car X^{p_0} l'est) et on utilise la convergence monotone.

Indication 177 On veut montrer que pour tout $N \in \mathbb{N}$, il existe $n \geq N$ avec $x \in A_n$. Cela signifie que pour $N \in \mathbb{N}$, on veut montrer qu'il existe $n \geq N$ tel que $\{s_n\} \geq x \geq \{s_{n+1}\}$. C'est ce qu'on a fait dans la première question de l'exercice 6. On trouve ensuite que la limite supérieure de $f_n(x)$ est 1. On notera aussi que l'ensemble des valeurs d'adhérence d'une suite donnée est un fermé.

Indication 178 1. Suivre l'indication...

2. Idem.
3. On applique la question précédente à $|f|\mathbb{1}_{\{|f|\leq n\}}$ (pourquoi est-elle intégrable ?)

Indication 179 1. On pourra montrer la convexité de $\log \Gamma$ ou regarder la dérivée de $\frac{\Gamma'}{\Gamma}$.

2. Comparer $\frac{\Gamma'(x+1)}{\Gamma(x+1)}$ et $\frac{\Gamma'(x)}{\Gamma(x)}$.
3. On pourra commencer par observer le signe de $\Gamma(x)$ et procéder par récurrence.

Indication 180 On pourra raisonner par l'absurde.

Indication 181 1. Cherchez un peu plus...

2. Découper $\mathbb{R}$ en intervalles de longeur 2π.
3. Utiliser des équivalents.

Indication 182 Étudier d'abord la convergence ponctuelle.

Indication 183 On pourra utiliser des sous-suites.

Indication 184 Retrousser ses manches (ou équivalents).

Indication 185 Qu'est-ce que la mesure de comptage ? Qu'intégrer par rapport à la mesure de comptage ?

Indication 186 Majorer $\sqrt{|f^2+g^2|}$ par une fonction intégrable.

Indication 187 Passer à l'intégrale de Riemann.

Indication 188 1. Prendre $X = [0,1]$ et pour μ la mesure de Lebesgue sur $[0,1]$, choisir ensuite (f_n) telle que $f_n(x) \to 0$ pour tout $x \in [0,1[$ et que l'on ait $\int f_n = 1$.

2. Pour p entier s'écrivant $p = 2^n + k$, avec $0 \leq k < 2^n$, poser $u_p = \frac{k}{2^n}$. Ensuite, poser $\phi_n(x) = \max(1 - \frac{n}{4}|x|, 0)$ et $f_n(x) = \phi_n(x - u_n)$.
3. Symétriser l'exemple trouvé à la première question.

Indication 189 Utiliser l'inégalité de Cauchy-Schwartz.

Indication 190 Utiliser l'inégalité de Hölder.

Indication 191 1. Utiliser l'inégalité de Holder.

2. (a) Considérer l'intégrale $\int_{]0,x[} f d\lambda$ comme une intégrale de Riemann.
   (b) Remarquer que $T(f)(x)$ est bornée et décroit suffisamment vite à l'infini.
   (c) Remarquer que $f(x) = T(f)(x) + xT(f)'(x)$ et faire une intégration par parties.
   (d) Cherchez un peu plus...
   (e) $f = f^+ - f^-$.
3. (a) Pour le premier point, on pourra utiliser l'inégalité de Hölder.
   (b) Utiliser la densité des fonctions continues à support compact dans l'ensemble $\mathcal{L}^p$.

Indication 192 Considérer la suite (g_n) définie par $g_n = \dfrac{|f|^q}{f} \mathbb{1}_{\{|f| \leq n\}}$ sur $\{|f| > 0\}$ et $g_n = 0$ sur $\{f = 0\}$.

C.8 Exercices sur la convolution et Fourier

Indication 193 1. On peut remarquer que f et g sont presque des densités de variables gaussiennes (il suffit de normaliser ces fonctions) et donc le produit de convolution correspond à la loi de la somme de ces deux gaussiennes indépendantes.

2. On se retrousse les manches et on calcule...

Indication 194 1. On remarque que pour $x \in [-1, 1]$, on a $0 \leq 1 - x^2 \leq 1$ et pour $0 \leq x \leq 1$, on a $1 - x^2 \geq 1 - x$.

2. On majore $|f * k_n - f|$ en remarquant que $f(x) = \left(\int_{\mathbb{R}} k_n(t) d\lambda(t)\right) f(x)$ et on utilise les hypothèses sur f.
3. On développe $(1 - (x - t)^2)^n$ par la formule de Leibnitz, puis idem pour $(x - t)^{2k}$.
4. Commencer par montrer le résultat pour $a = -1/2$ et $b = 1/2$.

Indication 195 1. Pour $g \in L^1$ l'application $x \mapsto T_x g$ est continue de $\mathbb{R}$ dans L^1.

2. Utiliser l'indication de l'énoncé.

Indication 196 1. Pour éviter d'oublier des cas, se souvenir que le support de la convolée est inclus dans la "somme" des supports ; la parité peut également permettre de simplifier des choses

2. Remarquer que $f^{(*)n}$ est positive.

3. Procéder par récurrence.

Indication 197 Si $a = 0$ ou $b = 0$, alors $ab = 0$.

Indication 198 1. Ah bah non ! Vous l'avez déjà eue, l'indication.

2. Réduire A dans une base orthonormale.

Indication 199 1. On pourra remarquer que la transformée de Fourier est injective dans $L^1(\mathbb{R}^n)$.

2. Utiliser la transformation de Fourier et une fonction bien choisie.

Indication 200 On pourra utiliser la formule d'inversion.

Indication 201 On pourra à nouveau utiliser la formule d'inversion.

Indication 202 On pourra utiliser le théorème de Fubini.

C.9 Exercices sur les fonctions caractéristiques

Indication 203 1. Penser à faire apparaître des densités.

2. Revoir les propriétés de base de la transformée de Laplace (caractérisation, lien avec l'indépendance).

3. Par récurrence.

4. Calculer la transformée de Laplace.

Indication 204 1. On peut noter que $\mathbb{E}[t^S \mathbb{1}_{\{T=n\}}] = G_{X_1}(t)^n \mathbb{E}[\mathbb{1}_{\{T=n\}}]$.

2. Penser au lien fonction génératrice/moments.

Indication 205 1. Appliquer la technique de la fonction test.

2. On peut remarquer que $\phi_{\varepsilon X}(t) = \frac{1}{2}(\phi_X(t) + \phi_X(-t))$.

Indication 206 1. Utiliser le théorème de Fubini et la valeur de la fonction caractéristique d'une gaussienne.

2. On rappelle que la fonction caractéristique. . .caractérise !

Indication 207 1. C'est un problème d'interversion d'intégrales, ou de limite et d'intégrale, suivant que l'on choisit d'écrire l'espérance sous forme d'une série ou comme une intégrale par rapport à $\mathbb{P}$.

2. (a) Penser au théorème de convergence monotone, ou au lemme de Fatou.

   (b) On pourra faire apparaître une intégrale de Wallis ou appliquer la méthode de Laplace.

Indication 208 1. Appliquer le théorème de Fubini.

2. Majorer le sinus.
3. Couper l'intégrale en deux.

Indication 209 1. Remarquer que $t^Z = t^X \,\mathbb{1}_A + t^Y \,\mathbb{1}_{A^c}$.

2. Remarquer que le score est une variable aléatoire fabriquée suivant le principe de la première question.

Indication 210 Utiliser la première question de l'exercice précédent.

Indication 211 1. S'inspirer de l'exercice 204.

2. Commencer par déterminer la fonction génératrice de K_1L_1. On pourra s'inspirer de l'exercice 209.
3. Relire le cours.

Indication 212 1. On trouvera $f^n \circ f$, où $f(z) = (1+z)/2$.

2. Remarquer que $f^n \circ f = (f \circ f)^n$.

Indication 213 Penser à la fonction caractéristique.

Indication 214 Regarder la liste des fonctions caractéristiques connues, ou/et chercher l'équation fonctionnelle que doit vérifier ϕ_X.

Indication 215 Utiliser encore les liens entre fonction caractéristique et moments.

Indication 216 1. Utiliser les hypothèses.

2. Utiliser les liens entre fonction caractéristique et moments.

3. Idem.

Indication 217 1. Remarquer que Re $(1 - \exp(itX - i\theta)) \geq 0$.

2. Prendre t et t' rationnels différents.

Indication 218 On pourra donner un équivalent de la fonction caractéristique en l'origine.

Indication 219 1. On remarquera que tout demi-plan est réunion croissante de disques.

2. Si $(a, b) \neq (0, 0)$, considérer le demi-plan $\{ax + by > t\}$.
3. On pourra éventuellement regarder la fonction caractéristique.

Indication 220 1. Remarquer que $e^{itZ}\mathbb{1}_{\{Y=n\}} = \sum_{k=0}^{n-1} e^{ikt}\mathbb{1}_{\{Z=k,Y=n\}}$.

2. Il faut appliquer plusieurs fois le théorème de convergence dominée.
3. On pourra remarquer et prouver que

$$\phi_Z(t) - 1 = \frac{i}{e^{it} - 1} \int_0^t (\phi_Y(x) - e^{ix})\, dx.$$

4. Revoir les liens entre fonction caractéristique et indépendance.
5. La fonction caractéristique...caractérise !

C.10 Exercices sur la convergence presque sûre

Indication 221 1. Penser à l'inégalité de Chebychev.

2. Elle est dans l'énoncé.
3. Idem.

Indication 222 1. La convergence presque sûre implique la convergence en probabilité (vers la même limite).

2. Ne pas oublier que l'espérance est linéaire.
3. Que vaut $|Y_n - X|$?

Indication 223 La série de terme général $\mathbb{P}(X_n \neq 0)$ converge-t-elle ?

Indication 224 1. On revient à la définition : Y_k suit une loi de Bernoulli et est indépendante de Y_{k+i} pour tout $i \geq 2$.

2. Utiliser la question précédente.

3. On travaille d'un côté avec tous les indices impairs et de l'autre avec tous les indices pairs. Ainsi, on a de l'indépendance au sein de chaque groupe et on utilise donc la loi forte des grands nombres.

Indication 225 Utiliser le second lemme de Borel–Cantelli pour une sous-suite d'événements indépendants.

Indication 226 1. Montrer que $\mathbb{P}\left(\frac{|X_n|}{n} \geq a \quad \text{i.s.}\right) = 1$.

2. Remarquer que $\left\{ \sup\limits_{n\geq 1} \frac{|S_n|}{n} < +\infty \right\} \subset \left\{ \sup\limits_{n\geq 1} \frac{|X_n|}{n} < +\infty \right\}$.

Indication 227 1. Elles sont données dans l'énoncé.

2. Utiliser la question précédente.
3. Idem.

Indication 228 1. On remarque que $gX_i = A_i + X_{i+1}$.

2. Elle est donnée dans l'énoncé.
3. On utilise la question précédente et on fait une projection sur la coordonnée voulue.
4. Il suffit de montrer que pour tout n, les variables $A_0, \dots, A_n$ sont indépendantes.
5. Revenir à la définition de la $g-$normalité.
6. Tout le travail est déjà fait dans la question précédente.

Indication 229 1. Si $i \in \{0, \dots, 2^n - 1\}$ s'écrit $i = \sum\limits_{k=0}^{n-1} a_k 2^{n-k}$, on a alors $\mathbb{P}(2^n S_n = i) = \mathbb{P}\left(\bigcap\limits_{k=1}^{n} \{X_k = a_k\}\right)$

2. Utiliser la question précédente.
3. On sait tout faire à i fixé.
4. Tout est dit dans l'énoncé.

Indication 230 Penser à utiliser les deux lemmes de Borel-Cantelli pour calculer $\mathbb{P}\left(\overline{\lim\limits_{n\to+\infty}} \{\frac{X_n}{\log n} \geq a\}\right)$ selon que $a > 1$ ou $a \leq 1$.

Indication 231 On remarque d'abord que pour tout entier a, on a :
$\{a$ est valeur d'adhérence de $X_n\} = \overline{\lim\limits_{n\to+\infty}} \{X_n = a\}$. Puis on pense à la loi du $0 - 1$ de Borel.

Indication 232 1. Elle est déjà donnée !

2. Poser $f_n = 1 - \mathbb{1}_{B_n}$ et penser à Fatou (en justifiant tout, évidemment !)
3. Montrer et utiliser que pour $\lambda \in]0,1[$, on a

$$\overline{\lim_{n\to+\infty}}\ B_n \supset \{N_n \geq \lambda\mathbb{E}[N_n] \text{ infiniment souvent}\}.$$

4. Utiliser la question précédente pour conclure.

Indication 233 1. Penser que $\mathbb{1}_{B_n}^2 = \mathbb{1}_{B_n}$. On utilisera aussi l'équivalent à l'infini rappelé dans l'énoncé.

2. Elle est déjà donnée !
3. Utiliser les questions précédentes et le fait que la fonction $x \mapsto \frac{\log x}{x}$ décroît sur $[e, +\infty[$.
4. On utilise Kochen-Stone ici !

Indication 234 On raisonne par l'absurde.

Indication 235 1. On pourra se ramener à une étude de fonction.

2. Que dire si aucun X_n n'est nul et qu'une infinité de termes sont différents de 1 ?
3. (a) On peut se donner N_1 tel que $\lambda_n \leq \log n$ pour $n \geq N_1$, et majorer alors les produits partiels.
   (b) Ici, les produits partiels peuvent se calculer explicitement.

Indication 236 1. Il faut connaître (ou retrouver) les moments de la loi de Poisson.

2. On peut noter que $\mathbb{1}_{\{X_n \geq k\}} \leq \frac{X_n(X_n-1)\dots(X_n-k+1)}{k!}$.

Indication 237 1. Si L_r est l'ensemble des familles libres de cardinal r, on peut appliquer le lemme des bergers à la projection canonique de L_{r+1} dans L_r, de manière à avoir une formule de récurrence. On pourra encore appliquer ce lemme à l'application qui à une famille libre associe l'espace vectoriel qu'elle engendre.

2. (a) On pourra remarquer que le produit infini $\prod_{i=1}^{+\infty}(1-q^{-i})$ converge, et en déduire un équivalent de $Q(n)$.
   (b) Il faut réutiliser les estimées trouvées et penser à Borel–Cantelli.

Indication 238 Considérer $S_n = X_1 + \cdots + X_n$ où (X_n) est une suite de variables de Bernoulli de paramètre s, et observer la suite (x_{S_n}).

Indication 239 Appliquer la loi forte des grands nombres.

Indication 240 On pourra montrer que N_n/n tend presque sûrement vers $1/2$.

Indication 241 Pour le sens direct, appliquer le premier lemme de Borel-Cantelli. Pour la réciproque, appliquer le deuxième lemme de Borel-Cantelli à une suite de variables aléatoires bien choisie.

Indication 242 On peut remplacer $(\log n)^{\frac{3}{2}}$ par $(\log n)u_n$, où (u_n) est une suite quelconque de limite infinie. Pour tout $\varepsilon > 0$, appliquer le lemme de Borel-Cantelli aux événements $\left\{\frac{X_n}{(\log n)u_n} > \varepsilon\right\}$.

Indication 243

1. S'inspirer de l'exercice sur le calcul de $\overline{\lim\limits_{n\to+\infty}} \frac{X_n}{\log n}$ pour des variables exponentielles, et utiliser un équivalent pour la queue de la gaussienne.
2. On pourra remarquer que la suite $(\frac{(X_n+Y_n)}{\sqrt{2}})_{n\geq 1}$ est une suite de variables aléatoires indépendantes suivant la loi $\mathcal{N}(0,1)$.

Indication 244 On a $T_n = \sum\limits_{k=2}^{n} \mathbb{1}_{\{U_{k-1}<p\}\Delta\{U_k<p\}}$. On pourra découper T_n en deux sommes de variables aléatoires indépendantes et calculer $\mathbb{P}((U_{k-1}-p)(U_k-p)<0)$.

Indication 245

1. Passer au logarithme.
2. On peut écrire M_n sous la forme $M_n = PD_nP^{-1}$ et introduire la norme $\|x\|_* = \|P^{-1}x\|_\infty$.

Indication 246

1. Commencer par calculer $f_1([0,1])$ et $f_2([0,1])$.
2. On notera que
$$\mathbb{1}_{K_{n+1}}(x) = \mathbb{1}_{[0,1/3]}(x)\mathbb{1}_{K_n}(3x) + \mathbb{1}_{[2/3,1]}(x)\mathbb{1}_{K_n}(3x-2),$$
puis on appliquera le théorème du point fixe avec $C([0,1],\|.\|_\infty)$.
3. Refaire les grandes lignes de l'exercice 228.
4. Remarquer que
$$\omega = X_0(\omega) = \sum_{i=0}^{+\infty} \frac{A_i^g(\omega)}{3^{i+1}} \quad \text{avec } A_i^g \in \{0,1,\ldots,g-1\}.$$

5. Pour le premier point, on montrera que T est $\max(\frac{3^\alpha}{2}, \beta)$-contractante, avec
$$\beta = \sup_{\substack{x\in[0,1]\\ y\in[0,1]}} \frac{\min(x,1-x)^\alpha + \min(y,1-y)^\alpha}{2(\frac{x+y+1}{3})^\alpha}.$$
Pour le deuxième point, on peut noter que
$$F\left(\frac{1}{3^{n+1}}(1+2\sum_{i=0}^{n} a_i 3^{n-1})\right) = \frac{1}{2^{n+1}}(1+\sum_{i=0}^{n} a_i 3^{n-1}).$$
Ainsi, si $\omega \in \mathbb{K}$ s'écrit $\omega = \sum_{i=0}^{+\infty} 2a_i 3^{-(i+1)}$ avec $a_n = 0$, on aura
$$\begin{aligned} F\left(\omega + \frac{2}{3^{n+1}}\right) - F(\omega) \quad &\geq \quad F\left(\sum_{i=0}^{n-1} 2a_i 3^{-(i+1)} + \frac{2}{3^{n+1}}\right) \\ &\quad -F\left(\sum_{i=0}^{n-1} 2a_i 3^{-(i+1)} + \frac{1}{3^{n+1}}\right) = \frac{1}{2^{n+1}}. \end{aligned}$$

C.11 Exercices sur la convergence en loi

Indication 247 1. On peut noter que $X_n = \sum_{k=0}^{n} \theta^k U_{n-k}$, où l'on a posé $U_0 = X_0$.

2. Appliquer les théorèmes de Lévy.

Indication 248 Appliquer le théorème de Lévy.

Indication 249 1. On pourra commencer par étudier la convergence en probabilité.

2. Évaluer $\mathbb{E}(\phi(X_n))$ pour ϕ bien choisie.

Indication 250 1. Calculer la fonction de répartition.

2. Penser au lemme de Borel–Cantelli.
3. Calculer la fonction de répartition.

Indication 251 1. Penser au théorème de Slutsky.

2. Utiliser le TCL et la question précédente.
3. Revoir les propriétés de convolutions des lois Gamma.
4. Utiliser le TCL.
5. Penser au théorème de convergence dominée.
6. Recoller les morceaux.

Indication 252 1. Un calcul un peu délicat qui mélange changements d'indices et regroupement de paquets.

2. Utiliser divers développements du cosinus pour obtenir des majorations appropriées à respectivement $]0,1]$ et $[1,+\infty[$.
3. Appliquer un théorème de Lévy judicieusement choisi.
4. Comparer les fonctions caractéristiques.

Indication 253 1. On pourra penser au théorème de Portmanteau.

2. Noter que $|e^{ixs}-e^{ixt}| \leq \min(|s-t||x|,2)$.

Indication 254 Réécrire la condition $\frac{N_t-t}{\sqrt{t}} \leq a$ à l'aide de la suite (S_n) et penser au lemme de Slutsky.

Indication 255 1. On pourra considérer une suite $(X_n)_{n\in\mathbb{N}^*}$ de variables aléatoires indépendantes de même loi de Poisson de paramètre 1, poser $S_n = \sum_{k=1}^{n} X_k$ et étudier la quantité $\mathbb{P}\left(\frac{S_n-\mathbb{E}S_n}{\sqrt{n}} \leq 0\right)$.

2. La suite $(Y_k)_{k\geq 1}$ des couleurs tirées se modélise par une suite de variables aléatoires indépendantes suivant la loi uniforme sur $\{1,\dots,n\}$. L'espérance étudiée peut se calculer à l'aide de la fonction de queue.

Indication 256 On peut regarder de deux manières la suite $(Z_{2n}-Z_n)_{n\geq 1}$

Indication 257 Écrire $\mathbb{E}(\frac{|S_n|}{\sqrt{n}}$ à l'aide de la fonction de queue de $\frac{|S_n|}{\sqrt{n}}$.

Indication 258 1. (a) On peut noter que

$$\begin{aligned}\mathbb{1}_{\{Y^1=a_1,\dots,Y^N=a_N\}} &= \prod_{k=1}^{N}(\mathbb{1}_{\{Y^k\geq a_k\}}-\mathbb{1}_{\{Y^k\geq a_k+1\}}) \\ &= \prod_{k=1}^{N}\sum_{j=0}^{1}(-1)^j\mathbb{1}_{\{Y^k\geq a_k+j\}} \\ &= \sum_{(b_1,\dots,b_N)\in\{0,1\}^N}\prod_{k=1}^{N}(-1)^{b_1+\dots+b_N}\mathbb{1}_{\{\forall k\in\{1,\dots,n\},Y^k\geq a_k+b_k\}}\end{aligned}$$

(b) On rappelle que $|\mathbb{P}(C)-\mathbb{P}(D)| \leq \mathbb{P}(C\Delta D)$.

(c) Utiliser l'inégalité triangulaire sur les probabilités.

2. Les quantités $\mathbb{P}(N|Z_n)$ et $\mathbb{P}(N|W_n)$ s'exprimant simplement, il suffit d'appliquer le critère de la question précédente.

Indication 259 Utiliser le premier théorème de Lévy.

Indication 260 1. Commencer par déterminer la loi de X_n^x.

2. Pour montrer la croissance, noter que $f(x) = \lim_{n\to+\infty} \mathbb{E}[u \circ X_n^x]$. Pour déterminer la limite, il est commode de se ramener au cas où $(u_n)_{n\geq 0}$ est à valeurs positives et de remarquer qu'alors, pour tout n, il existe un polynôme P_n tel que

$$\forall x > 0, \quad f(x) \geq \left(1 - \frac{P_n(x)}{e^x}\right) u_n.$$

Indication 261 Notons $M_{x,y} = 1$ si la pièce en (x, y) est face, -1 sinon. Si on note $C'_n = \sum_{k\neq X} M_{k,X}$ et $L'_n = \sum_{k\neq X} M_{X,k}$, on a

$$D_n = \sum_{(k,l)\in\{1,\ldots,n\}\backslash\{X\}} M_{k,l} + M_{X,X} + |C'_n - L'_n|.$$

Indication 262 Utiliser le lemme de Fatou.

Indication 263 On pourra déterminer ν_n telle que $\mu_n = \nu_n * \nu_n$.

C.12 Exercices sur les vecteurs gaussiens

Indication 264 Utiliser l'invariance de la gaussienne $\mathcal{N}(0, \mathrm{Id}_n)$ par le groupe orthogonal.

Indication 265 Appliquer la loi forte des grands nombres.

Indication 266 Utiliser l'invariance de la gaussienne $\mathcal{N}(0, \mathrm{Id}_2)$ par une rotation bien choisie.

Indication 267 1. Réaliser μ et ν à l'aide de leurs fonctions de répartition et d'une variable aléatoire uniforme sur $[0, 1]$.

2. Introduire des sommes de variables indépendantes.
3. Écrire les lois considérées comme sommes de variables indépendantes.
4. (a) On pourra noter que cette loi est la loi de $\|Y_a\|_2^2$, où $Y_a \sim \mathcal{N}(a, \mathrm{Id}_n)$.
   (b) On pourra noter que $\chi^2(n, \lambda) = \chi^2(1, \lambda/n)^{*n}$, ce qui permet de se ramener au cas où $n = 1$.

Indication 268 Utiliser le théorème de Cochran et l'exercice sur la loi de Dirichlet de la section 6.12.1.

Indication 269 Remarquer que $\sqrt{\frac{\mathrm{Var}(U)}{\mathrm{Var}(V)}}V$ a même loi que U.

Indication 270 On rappelle que si X suit la loi $\mathcal{N}(0,1)$, alors $\mathbb{E}(X^4) = 3$.

Indication 271 Trouver une matrice orthogonale O telle que la matrice de covariance de OX soit diagonale.

Indication 272
1. Déterminer le spectre de C.
2. On pourra commencer par trouver une constante K telle que
$$\forall a \in \mathbb{R}^3, \quad \mathbb{E}|\langle X, a\rangle| = \sqrt{\mathbb{E}(|\langle X, a\rangle|^2)}.$$

Indication 273 On pourra montrer qu'il existe un vecteur gaussien dont la matrice de covariance est
$$\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}.$$

Indication 274
1. Relire le cours !
2. Se souvenir qu'une matrice de covariance est symétrique positive.
3. Si (Z, T) est gaussien, alors $(\frac{Z}{\sqrt{\mathrm{Var}(Z)}}, T)$ aussi.
4. Utiliser la question précédente.

Indication 275
$$\mathbb{E}\|X\|_2^4 = \mathbb{E}\left(\sum_{k=1}^{n} X_k^2\right)^2.$$

Indication 276 Trouver une matrice orthogonale O telle que la matrice de covariance de OX soit diagonale.

Indication 277 Revoir l'image d'un vecteur gaussien par une transformation affine.

Indication 278 On pourra diagonaliser A dans une base orthogonale.

Indication 279 Introduire une "racine carrée" A de K^{-1} et déterminer la loi de AX.

C.13 Exercices sur les statistiques

Indication 280
1. On utilise la formule (13.2).
2. Utiliser la formule de l'intervalle de confiance une fois de plus.
3. Ne pas oublier que $\sigma\sqrt{\frac{\theta(1-\theta)}{N}} \leq \frac{\sigma}{2}\sqrt{1/N}$.

Indication 281 D'abord majorer $\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i) - \int_{[0,1]^d} g(x)\,dx > \varepsilon\right)$ en utilisant l'inégalité de Markov et en remarquant que cette probabilité est inchangée si on remplace chaque membre de l'inégalité par la fonction $x \mapsto e^{\alpha x}$ avec $\alpha > 0$. On optimise ensuite en α pour obtenir le résultat recherché.

Indication 282 Commencer par retrouver la fonction de vraisemblance.

Indication 283
1. C'est juste un peu de calcul...
2. La moyenne et la variance sont proches, donc...
3. On fait un test du χ^2 d'adéquation à l'appartenance à une famille.

Indication 284
1. On compare un pourcentage observé avec un pourcentage théorique.
2. On compare deux pourcentages observés sur échantillons indépendants.

Indication 285
1. Ne pas oublier que les variables X_i sont indépendantes et gaussiennes.
2. Relire le cours.
3. Commencer par supposer $m = 0$, puis faire le cas m quelconque.
4. Relire le cours.

Indication 286 Considérer chaque composante X_i comme une somme de n variables aléatoires de Bernoulli.

Indication 287 Faire un test de comparaison sur échantillons indépendants.

Indication 288
1. Utiliser les tableaux.
2. C'est encore un test du χ^2.
3. Bien relire la question : tout y est.

Indication 289 Idem que pour l'exercice 287.

Indication 290 Chercher encore un peu...

C.14 Exercices sur les sommes de variables aléatoires indépendantes

Indication 291 Remarquer que $(X_n \mathbb{1}_{\{X_n \le c\}})^2 \le cX_n \, \mathbb{1}_{\{X_n \le c\}}$.

Indication 292
1. Procéder par récurrence sur n.
2. Observer que $Y_{n,k} = Y_{k,k} = \mathbb{1}_{\{X_k = k\}}$.
3. On pourra calculer une fonction caractéristique de C_n (sa fonction génératrice $z \mapsto \mathbb{E}(z^{C_n})$ ou sa fonction caractéristique $t \mapsto \mathbb{E}(e^{itC_n})$).
4. Appliquer le TCL de Lindeberg.

Indication 293
1. Noter que l'événement $E(z)$ signifie que le disque du développement en série de F au point z déborde le disque unité.
2. Penser à la loi du 0–1. Pour montrer l'invariance, le résultat de l'exercice 129 sera utile.
3. On peut commencer par montrer l'existence d'une constante $\ell(z)$ telle que $\overline{\lim}_{n \to +\infty} |\frac{F^n(z)}{n!}|^{1/n} = \ell(z)$ presque sûrement.

Indication 294 On a

$$\begin{aligned} & \left| \prod_{k=1}^{N_n} \Big((1 - p_{n,k}) + p_{n,k} e^{it} \Big) - \prod_{k=1}^{N_n} \Big(\exp(p_{n,k}(e^{it} - 1)) \Big) \right| \\ \le & \sum_{k=1}^{N_n} |g(p_{n,k}(e^{it} - 1))|, \end{aligned}$$

avec $g(z) = e^z - 1 - z$. On pourra noter que $|g(z)| \le \frac{e^2}{2}|z|^2$, pour $|z| \le 2$.

Indication 295 Penser à écrire $1 + z = e^z((1+z)e^{-z}) = e^z(1 + \psi(z))$.

Indication 296 Pas d'indication.

Indication 297
1. On peut commencer par montrer que la série diverge si (a_n) ne tend pas vers zéro.
2. On peut remarquer que si $Y_n = a_n X_n$ avec (a_n) de limite nulle, alors $\mathbb{E}|Y_n^{(c)}| \sim a_n \mathbb{E}|X_0|$.
3. Remarquer que $\mathbb{P}(|a_n X_n| \ge 1) = \arctan a_n$.

Indication 298 Appliquer le théorème des trois séries.

Indication 299 La série diverge pour $z = 1$.

Indication 300 Il suffit d'avoir $\mathbb{P}\left(|a_n X_n| > \frac{1}{n^2}\right) \leq \frac{1}{n^2}$.

Indication 301 En utilisant la représentation $\sigma_n = (1\ X_1) \circ \ldots \circ (n\ X_n)$ de l'exercice 292, on trouve

$$N(\sigma_n) = \sum_{i=1}^{n} (i - X_i)$$

et on peut alors utiliser le TCL de Lindeberg.

Annexe D

Solutions des exercices corrigés

D.1 Solutions sur les compléments

Solution 1 Remarquons tout d'abord qu'il s'agit de séries à termes positifs.

a) On a : $\frac{u_{n+1}}{u_n} = \frac{2((n+1)^2+1)}{(2n+3)(2n+2)(n^2+1)} \leq 1/4$. D'après la règle de d'Alembert, la série de terme général $u_n = \frac{(n^2+1)2^n}{(2n+1)!}$ converge.

b) On fait un développement limité lorsque n tend vers l'infini :

$$\frac{n^2-5n+1}{n^2-4n+2} = 1 - \frac{1}{n} + o\left(\frac{1}{n}\right).$$

On passe alors au logarithme :

$$\log\left(\frac{n^2-5n+1}{n^2-4n+2}\right) = -\frac{1}{n} + o\left(\frac{1}{n}\right) = -\frac{1}{n} + o\left(\frac{1}{n}\right).$$

On peut ainsi calculer la racine n-ième de u_n, qui vaut :

$$\sqrt[n]{u_n} = e^{1/n \log u_n} = e^{-1+o(1)} \to e^{-1} < 1.$$

On conclut donc grâce au critère de Cauchy que la série de terme général $u_n = \left(\frac{n^2-5n+1}{n^2-4n+2}\right)^{n^2}$ converge.

c) On passe également au logarithme :

$$u_n = e^{\log u_n} = e^{n^2 \log(1+1/n) - n} = e^{n-1/2-n+o(1)} \to e^{-1/2} \neq 0.$$

Comme la limite n'est pas nulle, on conclut immédiatement que la série de terme général $u_n = \left(1+\frac{1}{n}\right)^{n^2} e^{-n}$ diverge.

d) La technique est identique à celle de la question b) et on trouve :

$$\sqrt[n]{u_n} = e^{1/n \log u_n} = e^{-1+1/(n+1)+o(1/(n+1))} \to e^{-1} < 1$$

et la série de terme général $u_n = \left(1+\frac{1}{n}\right)^{-n^2}$ converge.

e) Le terme général de cette série est positif et équivalent en l'infini à $2/n^3$. Par la règle de Riemann, on obtient que cette dernière série converge.

Solution 2 Remarquons qu'il y a toujours différentes manières de démontrer la convergence des séries ci-dessous et nous donnons ici celle qui permet de calculer la somme (lorsqu'il y a convergence).

a) On décompose la fraction en éléments simples et on trouve :

$$\sum_{n=2}^{N} \frac{1}{n(n-1)} = \sum_{n=2}^{N} \left(\frac{1}{n-1} - \frac{1}{n}\right) = 1 - \frac{1}{N}.$$

En faisant tendre N vers l'infini, on obtient que la série $\sum_{n=2}^{+\infty} \frac{1}{n(n-1)}$ est donc convergente, de somme 1.

Pour la convergence de la série $\sum_{n=2}^{+\infty} \frac{1}{n^2}$, il suffit de remarquer l'encadrement $0 \leq \frac{1}{n^2} \leq \frac{1}{n(n-1)}$.

b) On fait à nouveau une décomposition en éléments simples :

$$\sum_{n=1}^{+\infty} \frac{1}{(n+\alpha)(n+\alpha+1)} = \sum_{n=1}^{+\infty} \left(\frac{1}{n+\alpha} - \frac{1}{n+\alpha+1}\right).$$

Similairement à la question précédente, on calcule la somme partielle

$$\sum_{n=1}^{N} \frac{1}{(n+\alpha)(n+\alpha+1)} = \frac{1}{1+\alpha} - \frac{1}{N+1+\alpha}.$$

En faisant tendre N vers l'infini, on obtient que la série converge et sa somme vaut $\frac{1}{1+\alpha}$.

c) On utilise toujours la même technique de décomposition pour trouver que la série converge et sa somme vaut $117/20$.

d) Comme $8/9 < 1$, la série $\sum_{n=10}^{+\infty} 100 \left(\frac{8}{9}\right)^n$ converge. Il s'agit en fait d'une série géométrique de rapport 8/9. On peut écrire :

$$\sum_{n=10}^{+\infty} 100 \left(\frac{8}{9}\right)^n = 100 \left(\frac{8}{9}\right)^{10} \sum_{n=0}^{+\infty} \left(\frac{8}{9}\right)^n = 900 \left(\frac{8}{9}\right)^{10}.$$

e) Grâce à l'indication, on remarque que pour tout $N \geq 1$, on a

$$\begin{aligned} \sum_{n=1}^{N} \arctan \frac{1}{n^2+n+1} &= \sum_{n=1}^{N} (\arctan(n+1) - \arctan(n)) \\ &= \arctan(N+1) - \arctan(1) \\ &= \arctan(N+1) - \frac{\pi}{4}. \end{aligned}$$

Or on sait que $\arctan(N+1)$ tend vers $\pi/2$ quand N tend vers l'infini, d'où la convergence de la série et $\sum_{n=1}^{+\infty} \arctan \frac{1}{n^2+n+1} = \pi/4$.

Solution 3 Notons $S_n = \sum_{i=0}^{n} u_i$ et $T_n = \sum_{i=0}^{n} v_i$. On remarque que, par définition de v_n, $T_n = S_{2n+1}$. Ainsi, si S_n converge vers un réel ℓ, alors la suite extraite T_n aussi. Réciproquement, si T_n converge vers un réel ℓ, montrons que S_n tend vers ℓ : la sous-suite des termes impairs (S_{2n+1}) tend vers ℓ, et celle des termes pairs aussi car $S_{2n} = S_{2n+1} - u_{2n+1}$ tend vers $\ell - 0$. Finalement la suite (S_n) elle-même tend vers ℓ.

Application : $u_n := \log\left(1 + \frac{(-1)^n}{n+2}\right)$,

$$v_n := \log\left(1 + \frac{1}{2n+2}\right) + \log\left(1 + \frac{-1}{2n+3}\right) = \log\left(\frac{2n+3}{2n+2}\frac{2n+2}{2n+3}\right) = 0.$$

Donc, la série de terme général $\log\left(1 + \frac{(-1)^n}{n+2}\right)$ converge.

Solution 4 On pose $S = \sup_{N\geq 1} \sum_{n=1}^{N} \sum_{k=1}^{N} a_{n,k}$.

Si on montre que $S = \sum_{n=1}^{+\infty} \sum_{k=1}^{+\infty} a_{n,k}$, comme on peut intervertir les sommes doubles finies, on aura l'égalité voulue. Pour tout $N \geq 1$, on a facilement $\sum_{n=1}^{N} \sum_{k=1}^{N} a_{n,k} \leq \sum_{n=1}^{+\infty} \sum_{k=1}^{+\infty} a_{n,k}$, d'où en passant au supremum $S \leq \sum_{k=1}^{+\infty} a_{n,k}$. Il reste à montrer l'inégalité inverse.

i) S'il existe n_0 tel que $\sum_{k=1}^{+\infty} a_{n_0,k} = +\infty$, alors pour $M < +\infty$, il existe k_0 tel que $\sum_{k=1}^{k_0} a_{n_0,k} \geq M$, et si l'on pose $N = \max(n_0, k_0)$, on a alors $S \geq \sum_{n=1}^{N} \sum_{k=1}^{N} a_{n,k} \geq M$, d'où $S = +\infty$, puisque M est quelconque.

ii) Sinon, posons $s_n = \sum_{k=0}^{+\infty} a_{n,k}$ et soient $m < \sum_{n=1}^{+\infty} s_n$ et $\varepsilon > 0$. Il existe n_0 tel que $m < \sum_{n=1}^{n_0} s_n$. Mais comme $s_n < +\infty$ pour tout n, il existe k_n tel que $\sum_{k=1}^{k_n} a_{n,k} \geq s_n - \frac{\varepsilon}{2^n}$.

Maintenant si l'on pose $N = \max(n_0, k_1, \ldots, k_{n_0})$, on a

$$\sum_{n=1}^{N} \sum_{k=1}^{N} a_{n,k} \geq \sum_{n=1}^{n_0} \sum_{k=1}^{k_n} a_{n,k} \geq \sum_{n=1}^{n_0} \left(s_n - \frac{\varepsilon}{2^n}\right) \geq m - \varepsilon,$$

d'où $S \geq m - \varepsilon$. Comme ε peut être pris arbitrairement petit, on a $S \geq m$: S majore tous les minorants de $\sum_{n=1}^{+\infty} s_n$, donc $M \geq \sum_{n=1}^{+\infty} s_n$.

Solution 5 Commençons par montrer la convergence de la série de terme général $\frac{\log n}{n^a}$. On voit immédiatement que cette série diverge si $a \leq 1$ (par comparaison avec la série harmonique). Supposons $a > 1$, soit $a = 1+\varepsilon$ et donc $\frac{\log n}{n^{1+\varepsilon}} = o\left(\frac{1}{n^{1+\varepsilon/2}}\right)$. Comme la série de terme général $n^{-(1+\varepsilon/2)}$ converge, on conclut que la série recherchée converge.
La fonction est positive, donc l'intégrale est de même nature que la série de terme général $a_n = \int_{n\pi-\frac{\pi}{2}}^{n\pi+\frac{\pi}{2}} \frac{x^\gamma}{1+x^\alpha|\sin x|^\beta}\, dx = \int_{-\pi/2}^{\pi/2} \frac{(n\pi+x)^\gamma\ dx}{1+(n\pi+x)^\alpha|\sin x|^\beta}$.
Comme pour $x \in [-\pi/2, \pi/2]$, on a $\frac{2}{\pi}|x| \leq |\sin x| \leq |x|$, on a

$$\int_{-\pi/2}^{\pi/2} \frac{(n\pi-\frac{\pi}{2})^\gamma dx}{1+(n\pi+\frac{\pi}{2})^\alpha|x|^\beta} \leq a_n \leq \int_{-\pi/2}^{\pi/2} \frac{(n\pi+\frac{\pi}{2})^\gamma dx}{1+(n\pi-\frac{\pi}{2})^\alpha(2/\pi)^\beta|x|^\beta},$$

$$\text{soit}\quad K_1 n^\gamma \int_{-\pi/2}^{\pi/2} \frac{dx}{1+K_2 n^\alpha|x|^\beta} \leq a_n \leq K_1' n^\gamma \int_{-\pi/2}^{\pi/2} \frac{dx}{1+K_2' n^\alpha|x|^\beta}.$$

Posons ainsi

$$\begin{aligned} I_n(K) &= \int_{-\pi/2}^{\pi/2} \frac{dx}{1+Kn^\alpha|x|^\beta} = 2\int_0^{\pi/2} \frac{dx}{1+Kn^\alpha x^\beta} \\ &= 2K^{-\frac{1}{\beta}} n^{-\frac{\alpha}{\beta}} \int_0^{\frac{\pi}{2}K^{\frac{1}{\beta}} n^{\frac{\alpha}{\beta}}} \frac{dy}{1+y^\beta}. \end{aligned}$$

— Si $\beta > 1$, alors $I_n(K) \sim 2K^{-\frac{1}{\beta}} n^{-\frac{\alpha}{\beta}} \int_0^{+\infty} \frac{dy}{1+y^\beta}$.
— Si $\beta = 1$, alors $I_n(K) \sim 2K^{-\frac{1}{\beta}} n^{-\alpha} \log(1+\frac{\pi}{2}Kn^\alpha) \sim 2K^{-\frac{1}{\beta}} n^{-\alpha}\alpha\ \log n$.
— Si $\beta < 1$, alors $I_n(K) \sim 2K^{-\frac{1}{\beta}} n^{-\frac{\alpha}{\beta}} \int_1^{\frac{\pi}{2}K^{\frac{1}{\beta}} n^{\frac{\alpha}{\beta}}} \frac{dy}{y^\beta} \sim \frac{2}{K}(\frac{\pi}{2})^{1-\beta} n^{-\alpha}$.

La nature de la série de terme général $I_n(K)$ ne dépend pas de K. Par comparaison avec des séries de Riemann ou du type $\frac{\log n}{n^a}$, on obtient :
— Si $\beta > 1$, l'intégrale converge si et seulement si $\frac{\alpha}{\beta} - \gamma > 1$.
— Si $\beta = 1$, l'intégrale converge si et seulement si $\alpha - \gamma > 1$.
— Si $\beta < 1$, l'intégrale converge si et seulement si $\alpha - \gamma > 1$.

Solution 6 1. On pose $x_{n+1} = s_{n+1} - s_n$. Premier cas : $\{s_N\} \leq x$. On pose $L = \lfloor s_N \rfloor$ On a donc $s_N \leq L + x$. Soit n le plus grand entier tel que $s_n \leq L + x$. On a

$$L = \lfloor s_N \rfloor \leq s_N \leq s_n \leq s_{n+1} = s_n + (x_{n+1}) < (L+x) + (1-x) = L+1,$$

donc $\lfloor s_n \rfloor = \lfloor s_{n+1} \rfloor = L$, ce qui fait que les inégalités $s_n - L \leq x$ et $s_{n+1} - L > x$ se traduisent respectivement en $\{s_n\} \leq x$ et $x < \{s_{n+1}\}$. Par ailleurs

$$s_{n+1} - s_n = (L + \{s_{n+1}\}) - (L + \{s_n\}) = \{s_{n+1}\} - \{s_n\}.$$

Deuxième cas : $\{s_N\} > x$. Posons $L = \lfloor s_N \rfloor$ Bien sûr, $s_N < L + 1$.

Prenons pour n le dernier terme avec $s_n < L+1$. On a

$$L+1 \leq s_{n+1} = s_n + x_{n+1} \leq L+1+x_{n+1} < L+1+x < L+2,$$

d'où $\lfloor s_{n+1} \rfloor = L+1$, et l'inégalité $s_{n+1} < L+1+x$ devient $\{s_{n+1}\} < x$: on est ramené au cas précédent.

2. Soit $x \in]0,1[$. Soit $N \geq 1$ fixé et $\varepsilon > 0$. On peut trouver $N_0 \geq N$ tel que pour $n \geq N_0$, $s_{n+1} - s_n < \min(x, 1-x, \varepsilon)$. Alors d'après le lemme, il existe $n \geq N_0$, avec $\{s_n\} \leq x \leq \{s_{n+1}\}$ et $s_{n+1} - s_n = \{s_{n+1}\} - \{s_n\}$. On a donc $0 \leq x - \{s_n\} \leq \{s_{n+1}\} - \{s_n\} = s_{n+1} - s_n \leq \varepsilon$. En particulier, on a trouvé $n \geq N$ avec $|x - \{s_n\}| \leq \varepsilon$. On a donc montré que x est une valeur d'adhérence de $(\{s_n\})$. Ainsi, l'ensemble des valeurs d'adhérence de $(\{s_n\})$ contient $]0,1[$. Comme c'est un fermé, il contient $[0,1]$. Bien sûr, il est inclus dans $[0,1]$, donc c'est $[0,1]$.

3. On écrit $z = e^{2i\pi\theta}$, avec θ irrationnel. Soit $Z \in \mathbb{U}$, avec $Z \neq 1$. On peut écrire $Z = e^{2i\pi\alpha}$, avec $0 < \alpha < 1$. Soit $N \geq 1$, $\varepsilon > 0$. On doit montrer qu'il existe $n \geq N$, avec $|z^n - Z| \leq \varepsilon$. Quitte à remplacer ε par un réel plus petit, on peut supposer que $\varepsilon < \min(\alpha, 1-\alpha)$.

   La suite (z^n) est à valeurs dans $\mathbb{U}$ qui est compact : on peut trouver une suite strictement croissante d'entiers (n_k) de limite Z^* avec $Z^* \in \mathbb{U}$. On a $\lim \frac{z^{n_{k+1}}}{z^{n_k}} = 1$, donc il existe $k \geq 1$ tel que $|z^{n_{k+1}-n_k} - 1| \leq \frac{\varepsilon}{2}$. On pose $K_0 = n_{k+1} - n_k$. On écrit $\{\theta K_0\} = L + r$, avec $L \in \mathbb{Z}$ et $|r| \leq \frac{1}{2}$. On a

   $$|e^{2i\pi r} - 1| = |e^{2i\pi\{\theta K_0\}} - 1| = |e^{2i\pi\theta K_0} - 1| \leq \frac{\varepsilon}{2} \leq \frac{1}{2},$$

   donc $|r| \leq \frac{1}{4}$. La concavité du sinus sur $[0, \pi/2]$ nous donne

   $$\frac{2}{\pi} 2\pi|r| \leq |\sin(2\pi r)| \leq |1 - e^{2i\pi r}| \leq \frac{\varepsilon}{2},$$

   soit $|r| \leq \frac{\varepsilon}{8}$.

   Si $r > 0$, on considère la suite $(s_n)_{n \geq N}$ définie par $s_n = nr$. On a $s_{n+1} - s_n = r \leq \frac{\varepsilon}{8} < \varepsilon < \min(1-\alpha, \alpha)$, donc d'après le lemme, il existe $n \geq N$, avec $\{s_n\} \leq \alpha \leq \{s_{n+1}\}$ et $s_{n+1} - s_n = \{s_{n+1}\} - \{s_n\}$. On a donc $0 \leq \alpha - \{s_n\} \leq \{s_{n+1}\} - \{s_n\} = s_{n+1} - s_n = r \leq \frac{\varepsilon}{8}$. On en déduit

   $$\begin{aligned} |e^{2i\pi rn} - e^{2i\pi\alpha}| &= |e^{2i\pi\{rn\}} - e^{2i\pi\alpha}| \\ &\leq 2\pi|\{rn\} - \alpha| = 2\pi(\alpha - \{rn\}) \leq \frac{2\pi\varepsilon}{8} \leq \varepsilon. \end{aligned}$$

   Si $r < 0$, on applique le même raisonnement avec $(-rn)$ à la place de (rn) et $1-\alpha$ à la place de α : on trouve alors $n \geq N$, avec

   $$|e^{2i\pi(-r)n} - e^{2i\pi(1-\alpha)}| \leq \varepsilon,$$

ce qui en conjuguant, donne encore $|e^{2i\pi rn} - e^{2i\pi\alpha}| \leq \varepsilon$. On a traité tous les cas : r ne peut être nul, sinon z serait une racine K_0-ième de l'unité. Finalement, on a trouvé $n \geq N$ avec

$$|z^{nK_0} - Z| = |e^{2i\pi\theta K_0 n} - e^{2i\pi\alpha}| = |e^{2i\pi rn} - e^{2i\pi\alpha}| \leq \varepsilon,$$

et on a évidemment $nK_0 \geq N.1 = N$. On a donc montré que Z est une valeur d'adhérence. Par suite, l'ensemble des valeurs d'adhérence contient $\mathbb{U}\backslash\{1\}$, donc $\mathbb{U}$; c'est donc $\mathbb{U}$ puisque la suite prend ses valeurs dans $\mathbb{U}$.

4. On remarque que $z = \exp(2i\pi\theta)$ n'est pas une racine de l'unité. Pour $x \in [-1, 1]$, on peut trouver α réel, avec $x = \cos\alpha$. Si $(z^{n_k})_{k\geq 1}$ est une suite d'entiers de limite $e^{i\alpha}$, alors en prenant la partie réelle, on obtient $\lim\cos(2\pi\theta n_k) = \cos(\alpha) = x$: x est bien valeur d'adhérence de la suite $(\cos(2\pi\theta n)_{n\geq 1}$. L'ensemble de ces valeurs d'adhérences est donc $[-1, 1]$ tout entier, la $\overline{\lim}$ vaut 1 et la $\underline{\lim}$ vaut -1. Le sinus se traite de manière semblable.

D.2 Exercices sur la théorie de la mesure

Solution 16 1. Soit $\mathcal{A} = \{A \subset \mathbb{N} : A \text{ ou } A^c \text{ est fini}\}$. Montrons que $\mathcal{A}$ est une algèbre, mais pas une tribu. Comme $\varnothing$ est fini, on voit immédiatement que $\varnothing \in \mathcal{A}$ et $\mathbb{N} \in \mathcal{A}$. Il est évident que si $A \in \mathcal{A}$, alors $A^c \in \mathcal{A}$. Vérifions maintenant la stabilité par réunion finie. Soient $A_1, \ldots, A_n$ des éléments de $\mathcal{A}$. S'il existe un indice i_0 tel que $A^c_{i_0}$ est fini, alors $\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A^c_i \subset A^c_{i_0}$ est fini et donc $\bigcup_{i=1}^{n} A_i$ appartient à $\mathcal{A}$. Sinon, A_i est fini pour tout i et il en est de même pour leur réunion. On conclut donc que $\mathcal{A}$ est une algèbre.

Pour voir que $\mathcal{A}$ n'est pas une tribu, il suffit de poser $A_i = \{2i\}$. On remarque alors que $\bigcup_{i\geq 1} A_i$ est l'ensemble des entiers pairs positifs, qui n'est pas fini, et son complémentaire non plus.

2. Soit $\mathcal{A} = \{A \subset \mathbb{R} : A \text{ ou } A^c \text{ est fini}\}$. On fait de même que précédemment et on obtient le même résultat (et le même contre-exemple convient).

Solution 17 1. - La tribu engendrée par les intervalles $[a, b]$ est incluse dans la tribu borélienne car tout fermé est un borélien.

Réciproquement, comme $]a, b[= \bigcup_{n=1}^{+\infty} [a + 1/n, b - 1/n]$, la tribu engendrée par les intervalles $[a', b']$ avec a', b' rationnels contient les ensembles de la forme $]a, b[$, avec a et b dans $\mathbb{Q}$, donc la tribu borélienne.

- Les intervalles semi-ouverts $[a,b[$ sont des boréliens car $[a,b[=]-\infty,b[\setminus]-\infty,a[$ et tous les ouverts sont des boréliens. Réciproquement, on a comme précédemment $]a,b[=\bigcup_{n=1}^{+\infty}]a,b-1/n]$.

- Les demi-droites fermées sont fermées, donc boréliennes. On a également $[a,b[=[a,+\infty[\setminus[b,+\infty[$. Or les ensembles de la forme $[a,b[$ engendrent la tribu borélienne, d'où le résultat.

2. Les ouverts et les fermés étant des boréliens, les tribus respectivement engendrées par les boules ouvertes et les pavés sont des sous-tribus de la tribu borélienne. Voyons qu'elles coïncident avec la tribu borélienne.
   — La tribu engendrée par les pavés fermés contient les pavés ouverts : en effet $]a,b[\times]c,d[=\bigcup_{n=1}^{+\infty}[a+1/n,b-1/n]\times[c+1/n,d-1/n]$, donc elle contient la tribu borélienne qui est engendrée par les pavés ouverts.
   — Chaque rectangle ouvert est intersection de quatre demi-plans ouverts. Donc pour montrer que la tribu engendrée par les boules contient les pavés, il suffit de montrer qu'elle contient les demi-plans ouverts de frontière parallèle aux axes principaux. On a par exemple
   $$]a,+\infty[\times\mathbb{R}=\bigcup_{n=1}^{+\infty}B((a+n,0),n).$$
   Les trois autres types de demi-plans se traitent de la même manière.

Solution 18 1. Comme $\mathcal{A}$ est une tribu, elle est stable par réunion dénombrable et par intersection dénombrable. On en déduit ainsi que
$$\varliminf_{n\to+\infty} A_n=\bigcup_{n=1}^{+\infty}\bigcap_{m=n}^{+\infty}A_m \quad \text{et} \quad \varlimsup_{n\to+\infty} A_n=\bigcap_{n=1}^{+\infty}\bigcup_{m=n}^{+\infty}A_m$$
appartiennent à $\mathcal{A}$. On montre de plus que si $\omega\in\varliminf A_n$, alors ω appartient à tous les A_n sauf un nombre fini. Ainsi, ω appartient à une infinité de A_n et donc $\omega\in\varlimsup A_n$, ce qui prouve l'inclusion demandée.

2. Comme $(\mathbb{1}_{A_n}(\omega))_{n\geq1}$ est une suite à valeurs dans $\{0,1\}$ qui est fermé, la plus grande et la plus petite valeur d'adhérence, soit $\varlimsup \mathbb{1}_{A_n}(\omega)$ et $\varliminf \mathbb{1}_{A_n}(\omega)$, ne peuvent valoir que 0 et 1. Une suite d'entiers ne peut converger qu'en étant constante à partir d'un certain rang : ainsi $\varlimsup \mathbb{1}_{A_n}(\omega)$ vaut 1 si et seulement si il existe une suite strictement croissante d'entiers n_k avec $\mathbb{1}_{A_{n_k}}(\omega)=1$, soit encore $\omega\in A_{n_k}$: c'est exactement dire que $\omega\in\varlimsup A_n$, soit $\mathbb{1}_{\{\varlimsup A_n\}}(\omega)=1$. Ainsi, la limite supérieure des indicatrices est l'indicatrice de la limite supérieure. De même, dire que la limite inférieure de $\mathbb{1}_{A_n}(\omega)$ est 1, c'est dire que 0 n'est pas valeur d'adhérence, c'est-à-dire qu'il n'y a qu'un nombre fini

de n avec $\mathbb{1}_{A_n}(\omega) = 0$. C'est exactement dire qu'à partir d'un certain rang on a $\omega \in A_n$, ce qui signifie que ω est dans la limite inférieure des A_n. Ainsi, la limite inférieure des indicatrices est l'indicatrice de la limite inférieure.

Comme $\mathbb{1}_{\{A\setminus B\}} = \mathbb{1}_A - \mathbb{1}_B$ lorsque $B \subset A$, on en déduit que $\overline{\lim}\ \mathbb{1}_{A_n} - \underline{\lim}\ \mathbb{1}_{A_n} = \mathbb{1}_{\{\overline{\lim}\ A_n \setminus \underline{\lim}\ A_n\}}$.

Solution 19 1. On remarque $\mu(\varnothing) = \sum_{n=0}^{+\infty} \mu_n(\varnothing) = 0$. De plus, si $(A_k)_{k\geq 1}$ est une suite d'éléments de $\mathcal{T}$ deux-à-deux disjoints, alors comme μ_n est une mesure positive pour tout n, on a :

$$\mu\left(\bigcup_{k=1}^{+\infty} A_k\right) = \sum_{n=0}^{+\infty} \mu_n\left(\bigcup_{k=1}^{+\infty} A_k\right) = \sum_{n=0}^{+\infty}\ \sum_{k=1}^{+\infty} \mu_n(A_k).$$

Or, comme on l'a vu à l'exercice corrigé 4, on peut toujours intervertir les sommes doubles de termes positifs. Ainsi, on a

$$\begin{aligned}\mu\left(\bigcup_{k=1}^{+\infty} A_k\right) &= \sum_{n=0}^{+\infty}\ \sum_{k=1}^{+\infty} \mu_n(A_k)\\ &= \sum_{k=1}^{+\infty}\ \sum_{n=0}^{+\infty} \mu_n(A_k) = \sum_{k=1}^{+\infty} \mu(A_k).\end{aligned}$$

Ceci prouve bien que μ est une mesure.

2. Les $(p_n\mu_n)$ sont des mesures, donc d'après la question précédente, leur somme infinie est encore une mesure. De plus, on a par définition de p_n, $\mu(E) = \sum_{n=0}^{+\infty} p_n\mu_n(E) = \sum_{n=0}^{+\infty} p_n = 1$, donc μ est une probabilité.

3. (a) Si $n = 1$, on obtient alors $\mu_1(A_1) = \sum_{p=1}^{+\infty} \delta_p\left([1, 1+1+1]\right) = 3$. Si $n \geq 2$, on trouve $\mu_1(A_n) = \sum_{p=1}^{+\infty} \delta_p\left([n, n+1+1/n^2]\right) = 2$. On calcule identiquement $\mu_2(A_1) = 1+2+3 = 6$ et, pour $n \geq 2$, $\mu_2(A_n) = n+n+1 = 2n+1$.

   (b) On remarque que $B_n = \left[1, n+1+\frac{1}{n^2}\right]$.
   Ainsi, cela permet de voir que $\mu_1(B_1) = 3$, $\mu(B_n) = n+1$ si $n \geq 2$, $\mu_2(B_1) = 6$ et $\mu_2(B_n) = 1+2+\cdots+n+(n+1) = \frac{(n+1)(n+2)}{2}$.
   Comme $B = [1, +\infty[$, on trouve $\mu_1(B) = \mu_2(B) = +\infty$.

   (c) On a $C_1 = [1,3]$ et donc $\mu_1(C_1) = 3$, $\mu_2(C_1) = 6$. Or $C_2 = [2,3]$, on obtient donc $\mu_1(C_2) = 2$ et $\mu_2(C_2) = 5$. À partir de $C_3 = \{3\}$, on voit que $\mu_1(C_3) = 1$ et $\mu_2(C_3) = 3$. Enfin, comme $C_n = \varnothing$ pour tout $n \geq 4$, on obtient $C = \varnothing$ et $\mu_1(C_n) = \mu_1(C) = \mu_2(C_n) = \mu_2(C) = 0$ pour tout $n \geq 4$.

Solution 20 1. Soient $A_1, \dots, A_n$ des parties de Ω et $\mathcal{A}$ la tribu engendrée par ces parties. On note F l'application de Ω dans $\{0,1\}^n$, qui à ω associe $(\mathbb{1}_{A_1}(\omega), \dots, \mathbb{1}_{A_n}(\omega))$. F est $\sigma(A_1, \dots, A_n)$-mesurable car ses composantes le sont, donc les ensembles $F^{-1}(x)$, où x décrit $\{0,1\}^n$, forment une partition $\sigma(A_1, \dots, A_n)$-mesurable de Ω. D'autre part, tout A_i s'écrit comme réunion de 2^{n-1} tels ensembles, puisque

$$A_i = \bigcup_{\substack{x \in \{0;1\}^n \\ x_i = 1}} F^{-1}(x),$$

donc les deux tribus engendrées coïncident.

2. (a) — $\varnothing \cap A_j = \varnothing$ pour tout j, donc $\varnothing \in \mathcal{T}$.
   — Soit $(A_n)_{n\geq 1}$ une suite d'éléments de $\mathcal{T}$. On a $\left(\bigcup_{n=1}^{+\infty} A_n\right) \cap \Omega_j = \bigcup_{n=1}^{+\infty} (A_n \cap \Omega_j)$. Comme les termes de la réunion sont tous égaux à $\varnothing$ ou Ω_j, la réunion est donc égale à Ω_j si l'un des termes vaut Ω_j, $\varnothing$ sinon.
   — Soient $A \in \mathcal{T}$ et $j \in J$. Si $A \cap \Omega_j = \varnothing$, alors $A^c \cup \Omega_j^c = \Omega$, donc $A \cap (A^c \cup \Omega_j^c) = A$, soit $A \cap \Omega_j^c = A$. Inversement, si $A \cap \Omega_j = \Omega_j$, alors $A^c \cup \Omega_j^c = \Omega_j^c$, d'où $\varnothing = \Omega_j \cap (A^c \cup \Omega_j^c) = \Omega_j \cap A^c$.

   Finalement, $\mathcal{T}$ est une tribu.

   (b) Il est facile de voir que tous les Ω_j sont dans $\mathcal{T}$, donc la tribu $\mathcal{F}$ engendrée par les Ω_j est une sous-tribu de $\mathcal{T}$. Soit $A \in \mathcal{F}$: on a

$$A = \bigcup_{j \in J} (\Omega_j \cap J) = \bigcup_{\substack{j \in J \\ \Omega_j \cap A \neq \varnothing}} \Omega_j.$$

   Ainsi tout élément de $\mathcal{F}$ s'écrit sous la forme $A_K = \bigcup_{k \in K} \Omega_k$, où K est une partie de J. Réciproquement, il est facile de voir que tout élément de cette forme est dans la tribu engendrée par les Ω_j, puisque la réunion est finie ou dénombrable.

   (c) Bien sûr, $\Omega_{i_0} \in \mathcal{G}$. Soit $A \in \mathcal{G}$ tel que $x \in A$. Comme $A \in \mathcal{G} \subset \mathcal{T}$, on a $A \cap \Omega_{i_0} = \varnothing$ ou $A \cap \Omega_{i_0} = \Omega_{i_0}$. Mais $x \in A \cap \Omega_{i_0}$, donc $A \cap \Omega_{i_0} \neq \varnothing$. On conclut donc que $A \cap \Omega_{i_0} = \Omega_{i_0}$, soit $\Omega_{i_0} \subset A$.

3. $\mathcal{B}(\mathbb{R})$ est la tribu engendrée par les intervalles de $\mathbb{R}$ de la forme $]a, b[$ avec a, b rationnels, qui forment donc une famille dénombrable. Supposons que $\mathcal{B}(\mathbb{R})$ soit engendrée par une partition dénombrable $(\Omega_j)_{j \in J}$. Soit x réel. D'après la question précédente, l'élément de la partition contenant x est le plus petit élément de $\mathcal{G} = \mathcal{B}(\mathbb{R})$ contenant x : c'est donc $\{x\}$. Ainsi, l'application qui à x associe la partition le contenant est injective. On a ainsi créé une injection de $\mathbb{R}$ dans J qui est dénombrable : contradiction.

Solution 21 Si la suite $(f_n(x))_n$ n'est pas de Cauchy, cela signifie qu'il existe $\varepsilon > 0$ tel que pour tout $N \geq 0$, il existe $n, m \geq N$ tels que $|f_n(x) - f_m(x)| \geq \varepsilon$. On a ainsi :

$$\begin{aligned} A &= \{x \in X : f_n(x) \text{ n'est pas de Cauchy}\} \\ &= \{x \in X : \exists \varepsilon > 0; \forall N \in \mathbb{N}, \exists n, m \geq N; |f_n(x) - f_m(x)| \geq \varepsilon\} \\ &= \bigcup_{p \in \mathbb{N}^*} \bigcap_{N \in \mathbb{N}} \bigcup_{n \geq N} \bigcup_{m \geq N} \left\{x \in X : |f_n(x) - f_m(x)| \geq \frac{1}{p}\right\}. \end{aligned}$$

Pour f et g mesurables, $f-g$ est mesurable et $|f-g|$ aussi car pour $h(x) = |x|$, on a $|f-g| = h \circ (f-g)$ et la fonction valeur absolue est continue, donc mesurable. On en déduit que $\{x \in X : |f_n(x) - f_m(x)| \geq \frac{1}{p}\}$ est mesurable et donc A est mesurable.

Solution 22

1. Notons S' l'image de S par $s \mapsto \{s\}$.
   Si x s'écrit $x = q + s$ avec $(q, s) \in \mathbb{Q} \times S$, on a $x = q + s = q + \lfloor s \rfloor + \{s\}$: on a bien $q + \lfloor s \rfloor \in \mathbb{Q}$ et $\{s\} \in S'$, ce qui donne l'existence de la décomposition voulue. Si $x = q'_1 + s'_1 = q'_2 + s'_2$ avec q'_1, q'_2 dans $\mathbb{Q}$ et s'_1, s'_2 dans S', il existe s_1, s_2 avec $\{s_1\} = s'_1$ et $\{s_2\} = s'_2$. On a
   $$\{s_2\} - \{s_1\} = s'_2 - s'_1 = q'_1 - q'_2 \in \mathbb{Q},$$
   donc $s_2 - s_1 \in \mathbb{Q} \cap S = \{0\}$, d'où $s'_1 = s'_2$ et $q'_1 = q'_2$.

2. On raisonne par l'absurde. Considérons la réunion
   $$U = \bigcup_{q' \in \mathbb{Q} \cap [0,1]} (q' + S') \subset [0,1] + [0,1[= [0,2[.$$
   D'après la question précédente, la réunion est dénombrable et disjointe, donc on a $\mu(U) = \sum_{q' \in \mathbb{Q} \cap [0,1]} \mu(q' + S') = \sum_{q' \in \mathbb{Q} \cap [0,1]} \mu(S')$.
   Comme $U \subset [0,2[$, on a $\mu(U) \leq \mu([0,2[) = 2\mu([0,1[) < +\infty$. Ainsi, on a nécessairement $\mu(S') = 0$.

   Mais alors, comme $\mathbb{R} = \bigcup_{q' \in \mathbb{Q}} (q' + S')$, $\mathbb{R}$ est réunion dénombrable d'ensembles de mesure nulle, donc il est de mesure nulle. Contradiction.

3. Si $\mathcal{B}(\mathbb{R})$ était égale à $\mathcal{P}(\mathbb{R})$, la mesure de Lebesgue fournirait un contre-exemple au résultat de la question précédente.On peut également noter que l'ensemble S' est un exemple (pas très explicite) d'ensemble non borélien.

Solution 23

1. Posons $A_i = \{i\}$ pour $i \in \mathbb{N}^*$. Les A_i sont deux-à-deux

disjoints. On sait que $\mu\left(\bigcup_{i=1}^{+\infty} A_i\right) = \mu(\mathbb{N}^*) = +\infty$, alors que

$$\sum_{i=1}^{+\infty} \mu(A_i) = \sum_{i=1}^{+\infty} \frac{1}{i^2} = \frac{\pi^2}{6}.$$

Cela met en défaut la σ-additivité et implique que μ n'est pas une mesure.

2. On remarque que $\mu(A) = \sum_{n=1}^{+\infty} \frac{1}{n^2}\delta_n(A)$, et donc, d'après l'exercice 19, μ est la somme (infinie) des mesures $\frac{1}{n^2}\delta_n$.

Solution 24 $(A_n)_{n\geq 0}$ est une suite décroissante et chaque A_n est mesurable. On a de plus $\mu(A_n) = +\infty$. Comme $\bigcap_{n=0}^{+\infty} A_n = \bigcap_{n=0}^{+\infty} \{p : \ p \geq p\} = \varnothing$, on a $\mu\left(\bigcup_{n=0}^{+\infty} A_n\right) = 0$. On sait de plus que pour tout $n \geq 0$, $\mu(A_n) = +\infty$, ce qui prouve bien que $\mu\left(\bigcap_{n=0}^{+\infty} A_n\right) \neq \lim_{n\to+\infty} \mu(A_n)$.

Le théorème de continuité séquentielle décroissante n'est pas pris en défaut car celui-ci requiert que les élements considérés soient de mesure finie, ce qui n'est pas le cas ici. Remarquons toutefois que comme $\bigcap_{i=0}^{+\infty} A_i \subset A_n$ pour tout $n \geq 0$, l'inégalité $\mu\left(\bigcap_{n=0}^{+\infty} A_n\right) \leq \inf_{n\geq 0} \mu(A_n)$ demeure.

Solution 25 Comme $A \subset B$, on a les réunions disjointes $B = A \cup (B\backslash A)$ et $B \cap C = (A \cap C) \cup ((B\backslash A) \cap C)$, d'où l'on déduit que

$$\mu(B \cap C) = \mu(A \cap C) + \mu((B\backslash A) \cap C).$$

En particulier, pour $C = \Omega$, on a $\mu(B) = \mu(A) + \mu(B\backslash A)$, d'où on déduit que $\mu(B\backslash A) = 0$.
Dans le cas général, par inclusion, on a $0 \leq \mu((B\backslash A) \cap C) \leq \mu(B\backslash A) = 0$, ce qui entraîne $\mu((B\backslash A) \cap C) = 0$, puis $\mu(B \cap C) = \mu(A \cap C)$.

Solution 26 1. On a $\mathcal{T} = \mathcal{B}(\mathbb{R}) \cap \mathcal{T}'$, où $\mathcal{T}' := \{A \in \mathcal{P}(\mathbb{R}); A = -A\}$ Montrons que $\mathcal{T}'$ est une tribu. Posons $F(x) = -x$. On peut réécrire $\mathcal{T}' := \{A \in \mathcal{P}(\mathbb{R}); A = F^{-1}(A)\}$. Comme $F^{-1}(\varnothing) = \varnothing$, donc $\varnothing \in \mathcal{T}'$. Soit $A \in \mathcal{T}'$. On a $F^{-1}(A^c) = F^{-1}(A)^c = A^c$, donc $A^c \in \mathcal{T}$. Soit maintenant $(A_n)_n$ une suite d'éléments de $\mathcal{T}$. On a

$$F^{-1}\left(\bigcup_{n=1}^{+\infty} A_n\right) = \bigcup_{n=1}^{+\infty} F^{-1}(A_n) = \bigcup_{n=1}^{+\infty} A_n,$$

donc $\bigcup_{n=1}^{+\infty} A_n \in \mathcal{T}'$: $\mathcal{T}'$ est stable par réunion dénombrable. Finalement, $\mathcal{T}'$ est une tribu. Donc $\mathcal{T} = \mathcal{B}(\mathbb{R}) \cap \mathcal{T}'$ est l'intersection de deux tribus : c'est une tribu.

2. Étudions la mesurabilité des fonctions f, g et h.
   — Comme f, g et h sont continues, elles sont boréliennes. Ainsi, pour tout $A \in \mathcal{T}$, on a $f^{-1}(A) \in \mathcal{B}(\mathbb{R})$, $g^{-1}(A) \in \mathcal{B}(\mathbb{R})$ et $h^{-1}(A) \in \mathcal{B}(\mathbb{R})$. Donc f, g, h sont bien $(\mathcal{B}(\mathbb{R}), \mathcal{T})-$mesurables.
   — On sait que tout singleton appartient à la tribu borélienne $\mathcal{B}(\mathbb{R})$. Or $f^{-1}(\{e\}) = \{1\} \notin \mathcal{T}$, donc f n'est pas $(\mathcal{T}, \mathcal{B}(\mathbb{R}))-$mesurable. De même, $g^{-1}(\{1\}) = \{1\} \notin \mathcal{T}$, ce qui conduit à la même conclusion. En ce qui concerne h, remarquons que h est paire, soit $h = h \circ F$. Ainsi pour tout borélien A , $h^{-1}(A) = (h \circ F)^{-1}(A) = F^{-1}(h^{-1}(A))$, donc $h^{-1}(A) \in \mathcal{T}'$. Comme h est borélienne, $h^{-1}(A) \in \mathcal{B}(\mathbb{R})$. On trouve finalement que $h^{-1}(A) \in \mathcal{B}(\mathbb{R}) \cap \mathcal{T}' = \mathcal{T}$, donc h est bien $(\mathcal{T}, \mathcal{B}(\mathbb{R}))-$mesurable.
   — Comme h est $(\mathcal{T}, \mathcal{B}(\mathbb{R}))$-mesurable, elle est aussi $(\mathcal{T}, \mathcal{T})$-mesurable. Soit $A \in \mathcal{T}$. On a $g = F \circ g \circ F$, donc $g^{-1}(A) = F^{-1}(g^{-1}(F^{-1}(A)))$. Mais $F^{-1}(A) = A$, donc $g^{-1}(A) = F^{-1}(g^{-1}(A))$, et $g^{-1}(A) \in \mathcal{T}'$. Comme g est mesurable et A borélien, $g^{-1}(A)$ est un borélien et finalement $g^{-1}(A) \in \mathcal{T}$: g est bien $(\mathcal{T}, \mathcal{T})-$mesurable.
   En revanche, bien que $\{e\}$ soit un élément de $\mathcal{T}$, on remarque que $f^{-1}(\{-e\} \cup \{e\}) = \{1\} \notin \mathcal{T}$ et donc f n'est pas $(\mathcal{T}, \mathcal{T})-$mesurable.
3. En reprenant la preuve montrant que h est $(\mathcal{T}, \mathcal{B}(\mathbb{R}))-$mesurable, on prouve que toute fonction mesurable paire est $(\mathcal{T}, \mathcal{B}(\mathcal{T}))-$mesurable. Réciproquement, soit φ une fonction $(\mathcal{T}, \mathcal{B}(\mathcal{T}))-$mesurable. Elle est alors borélienne car pour tout $B \in \mathcal{B}(\mathbb{R})$, on a $\varphi^{-1}(B) \in \mathcal{T} \subset \mathcal{B}(\mathbb{R})$ et pour tout $x \in \mathbb{R}$, on a $\varphi^{-1}(\{\varphi(x)\}) \in \mathcal{T}$ et donc $-x \in \varphi^{-1}(\{\varphi(x)\})$, ce qui implique que $\varphi(x) = \varphi(-x)$. Cela prouve qu'une fonction est $(\mathcal{T}, \mathcal{B}(\mathcal{T}))-$mesurable si et seulement si elle est borélienne paire.

D.3 Exercices sur le formalisme probabiliste

Solution 39 $\mathbb{1}_\Omega(\omega) = 1$ pour tout $\omega \in \Omega$ donc $\mathbb{1}_\Omega \equiv 1$. Au contraire, il n'existe pas de ω dans l'ensemble $\varnothing$, donc $\mathbb{1}_\varnothing \equiv 0$.

On obtient facilement $\mathbb{1}_{A^c} = 1 - \mathbb{1}_A$.

Il est facile de voir que $\mathbb{1}_A\mathbb{1}_B$ est à valeurs dans $\{0, 1\}$. On a

$$\begin{aligned}(\mathbb{1}_A(\omega)\mathbb{1}_B(\omega) \neq 0) \quad &\Longleftrightarrow \quad (\mathbb{1}_A(\omega) \neq 0) \text{ et } (\mathbb{1}_B(\omega) \neq 0) \\ &\Longleftrightarrow \quad (\omega \in A) \text{ et } (\omega \in B) \iff (\omega \in A \cap B),\end{aligned}$$

donc $\mathbb{1}_{A\cap B} = \mathbb{1}_A\mathbb{1}_B$.

$\mathbb{1}_{A\setminus B} = \mathbb{1}_{A\cap B^c} = \mathbb{1}_A(1 - \mathbb{1}_B)$ en combinant les formules précédentes.

Si A et B sont disjoints $\mathbb{1}_A + \mathbb{1}_B$ est à valeurs dans $\{0, 1\}$ puisque les deux termes de la somme ne peuvent simultanément être égaux à 1.

$$\begin{aligned}(\mathbb{1}_A(\omega)+\mathbb{1}_B(\omega)>0) &\iff (\mathbb{1}_A(\omega)>0) \text{ ou } (\mathbb{1}_B(\omega)>0)\\ &\iff (\omega\in A) \text{ ou } (\omega\in B) \iff (\omega\in A\cup B),\end{aligned}$$

donc $\mathbb{1}_{A\cup B}=\mathbb{1}_A+\mathbb{1}_B$ si A et B sont disjoints.

Comme $A\Delta B$ est la réunion disjointe de $A\setminus B$ et $B\setminus A$, on a

$$\begin{aligned}\mathbb{1}_{A\Delta B} &= \mathbb{1}_{A\setminus B}+\mathbb{1}_{B\setminus A}=\mathbb{1}_A-\mathbb{1}_A\mathbb{1}_B+\mathbb{1}_B-\mathbb{1}_B\mathbb{1}_A\\ &= \mathbb{1}_A^2-2\mathbb{1}_A\mathbb{1}_B+\mathbb{1}_B^2=(\mathbb{1}_A-\mathbb{1}_B)^2\end{aligned}$$

Comme $A\cup B$ est la réunion disjointe de $A\cap B$ et $A\Delta B$, on a

$$\mathbb{1}_{A\cup B}=\mathbb{1}_{A\cap B}+\mathbb{1}_{A\cup B}=\mathbb{1}_A\mathbb{1}_B+\mathbb{1}_A+\mathbb{1}_B-2\mathbb{1}_A\mathbb{1}_B=\mathbb{1}_A+\mathbb{1}_B-\mathbb{1}_A\mathbb{1}_B.$$

Comme $(A\cup B)^c=A^c\cap B^c$, on a

$$\mathbb{1}_{A\cup B}=1-\mathbb{1}_{A^c\cap B^c}=1-\mathbb{1}_{A^c}\mathbb{1}_{B^c}=1-(1-\mathbb{1}_A)(1-\mathbb{1}_B).$$

Solution 40 On va montrer que si on remplace un seul élément de la famille par son complémentaire, on a encore une famille d'événements indépendants. Il suffira ensuite d'itérer. On fixe donc $i_0\in I$, on pose $B_{i_0}=A_{i_0}^c$ et $B_i=A_i$ pour $i\in I\backslash\{i_0\}$. Soit $J\subset I$. On doit montrer que

$$\mathbb{P}\left(\underset{j\in J}{\cap} B_j\right)=\prod_{j\in J}\mathbb{P}(B_j).$$

Si $i_0\notin J$, il n'y a rien à démontrer car l'identité est une simple conséquence de l'indépendance des A_j. Sinon

$$\underset{j\in J}{\cap} B_j=\left(\underset{j\in J\backslash\{i_0\}}{\cap} A_j\right)\setminus\left(\underset{j\in J}{\cap} A_j\right),$$

donc grâce à l'indépendance des A_j, on a

$$\begin{aligned}\mathbb{P}\left(\underset{j\in J}{\cap} B_j\right) &= \left(\prod_{j\in J\backslash\{i_0\}}\mathbb{P}(A_j)\right)-\left(\prod_{j\in J}\mathbb{P}(A_j)\right)\\ &= \left(\prod_{j\in J\backslash\{i_0\}}\mathbb{P}(A_j)\right)(1-\mathbb{P}(A_{j_0}))\\ &= \prod_{j\in J}\mathbb{P}(B_j),\end{aligned}$$

ce qui est le résultat voulu.

Solution 41 1. Soit A un réel quelconque. Comme la série harmonique diverge, il existe un entier n tel que $\sum_{k=1}^{n}\frac{1}{k}\geq A$.

Pour tout $s > 1$, on a $\zeta(s) \geq \sum_{k=1}^{n} \frac{1}{k^s}$, donc

$$\varliminf_{s \to 1^+} \zeta(s) \geq \varliminf_{s \to 1^+} \sum_{k=1}^{n} \frac{1}{k^s} = \lim_{s \to 1^+} \sum_{k=1}^{n} \frac{1}{k^s} = \sum_{k=1}^{n} \frac{1}{k} \geq A.$$

Comme A peut être pris arbitrairement grand, $\varliminf_{s \to 1^+} \zeta(s) \geq +\infty$, d'où le résultat voulu.

2. Notons que $p\mathbb{N}^* = \bigcup_{k=1}^{+\infty} \{pk\}$. Bien sûr, la réunion est disjointe, donc

$$\begin{aligned} \mu_s(p\mathbb{N}^*) &= \sum_{k=1}^{+\infty} \mu_s(\{kp\}) = \sum_{k=1}^{+\infty} \frac{1}{\zeta(s)} \frac{1}{(kp)^s} \\ &= \frac{1}{\zeta(s)} \frac{1}{p^s} \sum_{k=1}^{+\infty} \frac{1}{k^s} = \frac{1}{\zeta(s)} \frac{1}{p^s} \zeta(s) = \frac{1}{p^s}. \end{aligned}$$

En prenant $p = 1$, on a $\mu_s(\mathbb{N}^*) = 1$ et μ_s est bien une mesure de probabilité.

3. $\bigcap_{k=1}^{+\infty} A_k^c$ est l'ensemble des entiers naturels non nuls qui ne sont multiples d'aucun nombre premier, autrement dit qui ne sont divisibles par aucun nombre premier. Or 1 est le seul entier naturel qui n'ait pas de facteur premier, d'où l'identité $\{1\} = \bigcap_{k=1}^{+\infty} A_k^c$. Posons $B_n = \bigcap_{k=1}^{n} A_k^c$. La suite $(B_n)_{n \geq 1}$ est décroissante et son intersection vaut $\bigcap_{n=1}^{+\infty} B_n = \bigcap_{k=1}^{+\infty} A_k^c = \{1\}$. Ainsi, d'après le théorème de continuité séquentielle décroissante, $\lim_{n \to +\infty} \mu_s(B_n) = \mu_s(\{1\}) = \frac{1}{\zeta(s)}$, ce qui est le résultat voulu.

4. Soit I fini inclus dans $\mathbb{N}^*$

$$\prod_{i \in I} \mu_s(A_i) = \prod_{i \in I} \mu_s(p_i\mathbb{N}^*) = \prod_{i \in I} \frac{1}{p_i^s}.$$

D'autre part, comme les p_i sont premiers, ils sont premiers entre eux, d'où

$$\bigcap_{i \in I} A_i = \bigcap_{i \in I} p_i\mathbb{N}^* = (\prod_{i \in I} p_i)\mathbb{N}^*,$$

d'où

$$\mu_s\left(\bigcap_{i \in I} A_i\right) = \mu_s\left((\prod_{i \in I} p_i)\mathbb{N}^*\right) = \frac{1}{\left(\prod_{i \in I} p_i\right)^s},$$

d'où l'identité $\prod_{i\in I} \mu_s(A_i) = \mu_s\left(\bigcap_{i\in I} A_i\right)$, ce qui donne la preuve de l'indépendance sous μ_s.

5. On sait que

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

Par continuité de la fonction $x \mapsto -\log x$, cela donne

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

Or les (A_k) sont indépendants, donc les (A_k^c) sont indépendants, ce qui donne

$$\mu_s\left(\bigcap_{k=1}^{n} A_k^c\right) = \prod_{k=1}^{n} \mu_s(A_k^c),$$

d'où

$$\begin{aligned} -\log\mu_s\left(\bigcap_{k=1}^{n} A_k^c\right) &= \sum_{k=1}^{n} -\log\mu_s(A_k^c) = \sum_{k=1}^{n} -\log(1-\mu_s(A_k)) \\ &= \sum_{k=1}^{n} -\log\left(1-\frac{1}{p_k^s}\right) \end{aligned}$$

ce qui donne $\log\zeta(s) = \lim_{n\to+\infty} \sum_{k=1}^{n} -\log\left(1-\frac{1}{p_k^s}\right)$.

6. Soit A un réel quelconque. Comme $\lim_{s\to1} \zeta(s) = +\infty$, il existe s tel que $\zeta(s) > e^A$. Ainsi $\sum_{k=1}^{+\infty} -\log\left(1-\frac{1}{p_k}\right) \geq \sum_{k=1}^{+\infty} -\log\left(1-\frac{1}{p_k^s}\right) > A$. Comme A peut être pris arbitrairement grand, la série de terme général $\left(-\log(1-\frac{1}{p_k})\right)_{k\geq1}$ est donc divergente.
   Lorsque k tend vers l'infini $-\log\left(1-\frac{1}{p_k}\right) \sim \frac{1}{p_k}$, donc d'après le théorème sur les séries à termes positifs équivalentes, la série de terme général $\left(\frac{1}{p_k}\right)_{k\geq1}$ diverge également.

7. (a) En procédant comme au 4., on voit que les ensembles $A_i = p_i\mathbb{N}$ sont indépendants sous μ, par suite leurs complémentaires aussi. Comme on a l'inclusion $\{n\} \subset \bigcap_{i=n}^{\ell} A_i^c$, on obtient l'inégalité voulue.
   (b) On a $\lim_{\ell\to+\infty} \prod_{i=n}^{\ell} (1-\frac{1}{p_i}) = 0$, car d'après le 6., le logarithme tend vers $-\infty$. De plus, $\mu(\{n\}) = 0$ pour $n \geq 1$, d'où $\mu = \delta_0$, ce qui contredit $\mu(2\mathbb{N}) = \frac{1}{2}$.

Solution 42 1. On prend comme espace de probabilité $\Omega = \mathfrak{S}_n$, l'ensemble des permutations de $\{1, \dots, n\}$, que l'on munit de l'équiprobabilité. On sait bien que $\text{Card}(\mathfrak{S}_n) = n!$, d'où le résultat.

2. Se donner un élément de A_k^n revient à se donner un sous-ensemble F de $\{1, \dots n\}$ possédant k éléments (les points fixes), puis à se donner une permutation sans point fixe de $\{1, \dots, n\} \backslash F$, qui est un ensemble à $n-k$ éléments : d'où l'expression de A_k^n. Ensuite, il est aisé de constater que les A_k^n forment une partition de $\mathfrak{S}_n$, d'où

$$n! = \text{Card}(\mathfrak{S}_n) = \sum_{k=0}^{n} \text{Card}(A_k^n).$$

En utilisant $\binom{n}{k} = \binom{n}{n-k}$ puis le changement de variable $k' = n - k$, on obtient l'identité voulue.

3. On applique la formule du binôme : $\Phi(X^j) = (X+1)^j = \sum\limits_{i=0}^{j} \binom{i}{j} X^i$.

Ainsi, si $M = [m_{i,j}]_{0 \le i,j \le n}$, on a $m_{i,j} = \binom{i}{j}$ si $i \le j$ et 0 si $i > j$. Comme M est la matrice de Φ , M^{-1} est la matrice de Φ^{-1}. Mais en regardant la définition de Φ, on voit que $\Phi^{-1}(P) = P(X-1)$. Il suffit encore d'appliquer la formule du binôme, à savoir

$$\Phi^{-1}(X^j) = (X-1)^j = \sum_{i=0}^{j} \binom{i}{j} (-1)^{j-i} X^i,$$

donc si $M^{-1} = [u_{i,j}]_{0 \le i,j \le n}$, on a $u_{i,j} = (-1)^{j-i}\binom{i}{j}$ si $i \le j$ et 0 si $i > j$.

4. En écrivant le produit matriciel, $(d_0, d_1, \dots, d_n).M = (0!, 1!, \dots, n!)$ apparaît comme conséquence de la formule démontrée en 2. On en déduit que $(d_0, d_1, \dots, d_n) = (0!, 1!, \dots, n!).M^{-1}$. On a été capable de calculer M^{-1}, donc en faisant le produit matriciel, on en déduit l'expression de d_n, puis de p_n.

5. On reconnaît ici la somme partielle de rang n du développement en série entière de $\exp(-1)$. D'où $\lim p_n = \frac{1}{e}$. Qui plus est, d'après le critère spécial des séries alternées, on a $|p_n - \frac{1}{e}| \le \frac{1}{(n+1)!}$ (la valeur absolue du reste est inférieure au premier terme négligé). En multipliant l'inégalité par $n!$, on a donc

$$\left| d_n - \frac{n!}{e} \right| \le \frac{1}{n+1} \le \frac{1}{3},$$

d'où le résultat.

Remarquons qu'au lieu de 3., on peut utiliser la technique des séries génératrices pour calculer p_n. Comme $|p_n| \le 1$, la série entière $\sum\limits_{n=0}^{+\infty} p_n x^n$ a un rayon de convergence supérieur ou égal à 1. Notons $f(x)$ sa somme. En développant $\exp$ en série et en faisant un produit de Cauchy, on voit grâce à 2. que

$$\exp(x)f(x) = \sum_{n\geq 0} x^n, \text{ soit } f(x) = \exp(-x) \sum_{n=0}^{+\infty} x^n.$$

En développant $\exp(-x)$ en série et en faisant un produit de Cauchy, on retrouve l'expression de p_n.

Solution 43 1. On a $A_d = \{md; 1 \leq m \leq n/d\}$. Comme $A_d \subset \Omega_n$ et comme $\mathbb{P}$ est la loi uniforme sur Ω_n, on a

$$\mathbb{P}(A_d) = \frac{|A_d|}{\Omega_n} = \frac{n/d}{n} = \frac{1}{d}.$$

2. On a

$$\begin{aligned}\bigcup_{i=1}^{r} A_{d_i} &= \bigcap_{i=1}^{r} (\Omega_n \cap d_i\mathbb{Z}) = \Omega_n \cap \bigcap_{i=1}^{r} d_i\mathbb{Z}\\ &= \Omega_n \cap \prod_{i=1}^{r} d_i\mathbb{Z},\end{aligned}$$

car les d_i sont premiers entre eux ; et le produit est encore un diviseur de n. On a donc

$$\mathbb{P}\left(\bigcap_{i=1}^{r} A_{d_i}\right) = \mathbb{P}\left(A_{\prod_{i=1}^{r} d_i}\right) = \frac{1}{\prod_{i=1}^{r} d_i}.$$

Pour $J \subset I$, on a en particulier

$$\mathbb{P}\left(\bigcap_{i\in J} A_{p_i}\right) = \frac{1}{\prod_{i\in J} p_i} = \prod_{i\in J} \frac{1}{p_i} = \prod_{i\in J} \mathbb{P}(d_i).$$

Comme l'identité a lieu pour $J \subset I$ quelconque, les événements A_{p_i} sont indépendants.

3. Deux entiers non nuls sont premiers entre eux si et seulement si ils n'ont pas de diviseur premier communs. Ainsi le complémentaire de G_n dans Ω_n est formé des éléments de Ω_n qui sont divisibles par au moins un des p_i : on a donc

$$\mathbb{P}(G_n) = 1 - \mathbb{P}(G_n^c) = 1 - \mathbb{P}\left(\bigcup_{i\in I} A_{p_i}\right).$$

4. On a aussi

$$G_n = \left(\bigcup_{i\in I} A_{p_i}\right)^c = \bigcap_{i\in I} A_{p_i}^c.$$

Les événements A_{p_i} sont indépendants, donc leurs complémentaires le sont aussi (voir par exemple l'exercice 40) et on a

$$\begin{aligned}\mathbb{P}(G_n) &= \mathbb{P}(\left(\bigcap_{i\in I} A_{p_i}^c\right) \\ &= \prod_{i\in I} \mathbb{P}(A_{p_i}^c) = \prod_{i\in I} \left(1-\frac{1}{p_i}\right).\end{aligned}$$

Comme $\mathbb{P}(G_n) = \frac{\phi(n)}{n}$, on obtient le résultat voulu.

Solution 44 1. On pose $s_0 = 0$, $n_0 = 0$, puis

$$n_{k+1} = \inf\{p \geq n_k + 1; s_k + u_p < \ell\} \text{ et } s_{k+1} = s_k + u_{n_{k+1}}.$$

Comme (u_n) tend vers 0, on montre par récurrence que pour tout k, $s_k < \ell$ et $n_k < +\infty$. La suite (s_n) est croissante, majorée par ℓ, donc converge vers une limite $r \leq \ell$. Montrons que $r = \ell$. On raisonne par l'absurde et on suppose que $r < \ell$.

Alors $\lim\limits_{k\to+\infty} (s_k + u_{n_k+1}) = r + 0 = r < \ell$, donc il existe k tel que $k \geq k_0$ entraîne $s_k + u_{n_k+1} < \ell$, ce qui entraîne $n_{k+1} = n_k + 1$. Alors $r = \sum\limits_{k=1}^{+\infty} u_{n_k} \geq \sum\limits_{k=k_0}^{+\infty} u_{n_k} = \sum\limits_{p=n_{k_0}}^{+\infty} u_p = +\infty$: contradiction.

Remarquons qu'il n'y a pas unicité. Par exemple, de la série divergente des $1/n$, on peut extraire les sous-séries $\sum_{k\geq 1} \frac{1}{k(k+1)}$ et $\sum_{k\geq 1} \frac{1}{2^k}$ qui ont toutes les deux comme somme 1.

2. Comme $\phi(n)/n$ prend ses valeurs dans le fermé $[0,1]$, les valeurs d'adhérence restent dans $[0,1]$.
Soit $x \in]0,1[$. Il existe $\ell > 0$ tel que $x = e^{-\ell}$. Si $(p_i)_{i\geq 1}$ est la suite des nombres premiers, on sait que la série des $-\log(1-p_i^{-1})$ diverge (voir par exemple l'exercice 41). Donc d'après la question précédente, il existe une suite (n_k) telle que

$$\sum_{k=1}^{+\infty} -\log(1-p_{n_k}^{-1}) = \ell.$$

Ainsi, si on pose $s_n = \sum_{k=1}^{n} -\log(1-p_{n_k}^{-1})$, $s_n \to \ell$, donc

$$e^{-s_n} = \prod_{k=1}^{n} (1-p_{n_k}^{-1}) \to e^{-\ell} = x.$$

Mais si on a posé $E_n = \prod\limits_{k=1}^{n} p_{n_k}$, alors

$$\prod_{k=1}^{n} (1-p_{n_k}^{-1}) = \frac{\phi(E_n)}{E_n},$$

donc x est une valeur d'adhérence de $\phi(n)/n$. Ainsi l'ensemble des valeurs d'adhérence de $\phi(n)/n$ contient $]0,1[$. Mais l'ensemble des valeurs d'adhérence d'une suite est un fermé, donc il contient l'adhérence de $]0,1[$, soit $[0,1]$, ce qui achève la preuve.

D.4 Exercices sur les intégrales

Solution 53 Posons

$$\begin{aligned} f_n(x) &= \{\frac{1}{x}\}\mathbb{1}_{]1/n,1]}(x) = \frac{1}{x}\mathbb{1}_{]1/n,1]}(x) - \sum_{k=1}^{n-1} \lfloor \frac{1}{x} \rfloor \mathbb{1}_{]\frac{1}{k+1},\frac{1}{k}]}(x) \\ &= \frac{1}{x}\mathbb{1}_{]1/n,1]}(x) - \sum_{k=1}^{n-1} k\mathbb{1}_{]\frac{1}{k+1},\frac{1}{k}]}(x) \end{aligned}$$

f_n est une fonction continue par morceaux, donc mesurable. Pour tout x de $]0,1]$, $f_n(x)$ tend vers $f(x)=\{\frac{1}{x}\}$, qui est donc aussi une fonction mesurable. Comme f est bornée par 1, l'intégrale existe. Comme les f_n sont elles-mêmes bornées par 1, la valeur de l'intégrale peut se calculer à l'aide du théorème de convergence dominée. En effet

$$\begin{aligned} \int_{]0,1]} f_n \, d\lambda &= \int_{1/n}^{1} \frac{dx}{x} - \sum_{k=1}^{n-1} \int_{\frac{1}{k+1}}^{\frac{1}{k}} k \, dx \\ &= \log n - \sum_{k=1}^{n-1} \frac{1}{k+1} \\ &= \log n - H_n + 1 \end{aligned}$$

On sait que $H_n - \log n$ admet une limite γ lorsque n tend vers l'infini, appelée constante d'Euler. On en déduit que $\int_{]0,1]}\{\frac{1}{x}\}\, d\lambda(x) = 1-\gamma$.

Solution 54 Posons $I=\int_0^{+\infty} f(t)\,dt$, $R(x)=I-\int_0^x f(t)\,dt=\int_x^{+\infty} f(t)\,dt$ et $S(u)=\sup_{x\geq u}|R(x)|$. S est décroissante, de limite nulle en l'infini. Soient x,y avec $0\leq x\leq y$. Comme $-R$ est une primitive de f, on a par intégration par parties que

$$\int_x^y f(t)e^{-\lambda t}\,dt = [-R(t)e^{-\lambda t}]_x^y - \int_x^y R(t)\lambda e^{-\lambda t}\,dt.$$

On en déduit

$$\left|\int_x^y f(t)e^{-\lambda t}\,dt\right| \leq |R(x)|+|R(y)|+\int_x^y S(x)\lambda e^{-\lambda t}\,dt \leq 3S(x) \tag{D.1}$$

Ainsi, d'après le critère de Cauchy, $\int_0^T e^{-\lambda t} f(t)\,dt$ admet bien une limite lorsque T tend vers l'infini, ce qui donne la convergence de l'intégrale impropre. Posons $F_n(x)=\int_0^n e^{-\lambda t} f(t)\,dt=\int_{[0,n]} e^{-\lambda t} f(t)\,d\lambda(t)$. Comme on a

— Pour tout $t \geq 0$, $\lambda \mapsto e^{-\lambda t} f(t)$ est continue,
— Pour tout $t \geq 0$, pour tout $\lambda \geq 0$, $|e^{-\lambda t} f(t)| \leq |f(t)|$,
— $|f|$ est intégrable sur $[0, n]$,

alors pour tout n, F_n est une fonction continue sur $\mathbb{R}_+$. D'après l'inégalité (D.1), pour $n \leq p$, on a $\sup\limits_{\lambda \in \mathbb{R}_+} |F_n(\lambda) - F_p(\lambda)| \leq 3S(n)$. Ainsi F_n vérifie un critère de Cauchy uniforme : comme limite uniforme de fonctions continues, la fonction $\lambda \mapsto \int_0^{+\infty} e^{-\lambda t} f(t)\, dt$ est continue sur $\mathbb{R}_+$.

Solution 55 1. Posons $N_n = \sum\limits_{k=0}^{n} \mathbb{1}_A(T^k(x))$.

L'application T^k est $(\Omega, \mathcal{F}) - (\Omega, \mathcal{F})$ mesurable, comme composée d'applications $(\Omega, \mathcal{F}) - (\Omega, \mathcal{F})$ mesurables. Comme $\mathbb{1}_A$ est $(\Omega, \mathcal{F}) - (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable, finalement $x \mapsto \mathbb{1}_A(T^k(x)$ est $(\Omega, \mathcal{F}) - (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable et, par addition, N_n l'est aussi. Pour $t \leq 0$, on a $\{Y < t\} = \varnothing \in \mathcal{F}$. Pour $t \geq 0$, on a $\{Y < t\} = \bigcup\limits_{n \geq 1} \{N_n > -\log t\}$: c'est une union dénombrable d'éléments de $\mathcal{F}$, donc un élément de $\mathcal{F}$. Ainsi, pour tout t réel, $Y^{-1}(]-\infty, t[) \in \mathcal{F}$: cela suffit à établir la mesurabilité de Y. Y est intégrable car $0 \leq Y \leq 1$ et on intègre par rapport à une mesure finie.

2. On a
$$N(x) = \sum_{k=0}^{+\infty} \mathbb{1}_A(T^k(x)) = \mathbb{1}_A(x) + \sum_{k=1}^{+\infty} \mathbb{1}_A(T^k(x)) = \mathbb{1}_A(x) + N(Tx),$$
d'où, en composant avec $x \mapsto \exp(-x)$, l'identité
$$Y(x) = e^{-\mathbb{1}_A(x)} Y(Tx) = e^{-\mathbb{1}_A(x)} Z(x).$$
Par composition, Z est mesurable, et aussi intégrable car $0 \leq Z \leq 1$. On a
$$\int Z(x)\, d\mu(x) = \int Y(T(x))\, d\mu(x) = \int Y(x)\, d\mu_T(x),$$
où la dernière égalité vient du théorème de transfert. Mais, par hypothèse, $\mu_T = \mu$, donc
$$\int Z(x)\, d\mu(x) = \int Y(x)\, d\mu_T(x) = \int Y(x)\, d\mu(x).$$

3. Pour tout $x \in \Omega$, on a $Y(x) = e^{-\mathbb{1}_A(x)} Z(x) \leq Z(x)$, donc
$$0 \leq Z(x) - Y(x).$$
Mais d'après la question précédente, $Z - Y$ est d'intégrale nulle par

rapport à μ, donc $Z - Y$ est nulle μ-presque sûrement. On a

$$A \cap \{N < +\infty\} = A \cap \{Y > 0\} \subset \{Z \neq Y\},$$

donc $\mu(A \cap \{N < +\infty\}) \leq \mu(Z \neq Y) = 0$. Ainsi pour μ-presque tout point $x \in A$, on a $N(x) = +\infty$, ce qui signifie que la suite des itérées $(T^n(x))_{n\geq 0}$ passe une infinité de fois dans A.

Solution 56 1. Pour $n \geq 1$ et $x \geq 0$, on a $\log(x+n) = \log(n) + \log(1+\frac{x}{n})$. Comme $|e^{-x}\cos x| \leq e^{-x}$ et que

$$|e^{-x}\log(1+\frac{x}{n})\cos(x)| \leq e^{-x}\log(1+\frac{x}{n}) \leq \frac{1}{n}xe^{-x},$$

l'intégrabilité de $x \mapsto e^{-x}$ et de $x \mapsto xe^{-x}$ permet d'écrire

$$I_n = \frac{C\log n}{n} + \frac{R_n}{n}$$

avec $C = \int_{\mathbb{R}_+} \cos(x)e^{-x}\, d\lambda(x)$ et $R_n = \int_{\mathbb{R}_+} \log(1+x/n)e^{-x}\, d\lambda(x)$. Or

$$|R_n| \leq \int_{\mathbb{R}_+} |\log(1+\frac{x}{n})e^{-x}\cos(x)|\, d\lambda(x) \quad \leq \int_{\mathbb{R}_+} \frac{x}{n}e^{-x}\, dx = C'/n,$$

donc $I_n = \frac{C\log n}{n} + O(1/n^2)$. Cela donne une limite nulle pour I_n en l'infini.

2. Comme $x \mapsto \frac{1+nx}{(1+x)^n}$ est continue sur $[0,1]$, on a

$$\int_0^1 \frac{1+nx}{(1+x)^n}\, dx = \int_{[0,1]} \frac{1+nx}{(1+x)^n}\, d\lambda(x)$$

Pour $x \in]0,1]$, on a $\lim\limits_{n\to+\infty} \frac{1+nx}{(1+x)^n} = 0$ (croissance comparée des puissances et exponentielles), donc $\frac{1+nx}{(1+x)^n}$ tend presque partout vers 0 sur $[0,1]$. On a donc envie d'appliquer le théorème de convergence dominée, et on peut l'appliquer car pour tout $x \geq 0$ et $n \geq 1$, on a $(1+x)^n \geq 1+nx$ (par exemple par la formule du binôme de Newton), d'où une domination de $\frac{1+nx}{(1+x)^n}$ par la fonction 1, qui est intégrable sur $[0,1]$.

3. Comme précédemment, $\frac{1+nx}{(1+x)^n}$ tend partout vers 0 sur $[1,+\infty[$. Ici on a encore la majoration $0 \leq \frac{1+nx}{(1+x)^n} \leq 1$, mais ça ne suffit pas car 1 n'est pas intégrable sur $[1,+\infty[$. En revanche comme $1+nx$ est beaucoup plus petit que $(1+x)^n$, on se dit qu'on a de la marge. Comme $\frac{1}{(1+x)^2}$ est intégrable, on écrit

$$\frac{1+nx}{(1+x)^n} = \frac{1+nx}{(1+x)^{n-2}}\frac{1}{(1+x)^2},$$

et pour $n \geq 3$:

$$\begin{aligned}\frac{1+nx}{(1+x)^n} &= \frac{1+nx}{1+(n-2)x}\frac{1+(n-2)x}{(1+x)^{n-2}}\frac{1}{(1+x)^2} \\ &\leq \frac{1+nx}{1+(n-2)x}\frac{1}{(1+x)^2} \leq \frac{3}{(1+x)^2}.\end{aligned}$$

Ainsi on a bien majoré $\frac{1+nx}{(1+x)^n}$ par une fonction intégrable, ce qui permet d'appliquer le théorème de convergence dominée et d'obtenir la convergence vers 0.

Solution 57 1. On va appliquer le théorème de Tonelli.
On a pour $0 \leq x \leq 1$

$$\begin{aligned}\int_{[0,1]}\left|\frac{x-y}{(x^2+y^2)^{3/2}}\right| d\lambda(y) &\geq \int_{[0,x/2]}\left|\frac{x-y}{(x^2+y^2)^{3/2}}\right| d\lambda(y) \\ &\geq \int_{[0,x/2]}\frac{x-x/2}{(x^2+(x/2)^2)^{3/2}} d\lambda(y) = \frac{C}{x},\end{aligned}$$

qui n'est pas intégrable sur $[0,1]$, donc la réponse est non.

2. La fonction à intégrer est positive. Pour $p \neq -1$ et $x \in]0,1]$, on voit que $\int_{[0,1]}(1-xy)^p\, d\lambda(y) = -\frac{1}{p+1}\frac{1}{x}((1-x)^{p+1}-1)$, qui se prolonge par continuité en $x=0$: la fonction est donc bien intégrable, d'après Tonelli. De même, pour $p=-1$ et $x \in]0,1]$ on a $\int_{[0,1]}(1-xy)^{-1}\, d\lambda(y) = -\frac{\log(1-x)}{x}$, qui est intégrable sur $[0,1]$. On pourrait regarder des équivalents en 0 et en 1, mais le plus simple est de remarquer que

$$-\frac{\log(1-x)}{x} = \sum_{n=1}^{+\infty}\frac{x^{n-1}}{n},$$

ce qui nous donne avec le théorème de Tonelli (ou le théorème de convergence monotone)

$$\int_{[0,1]} -\frac{\log(1-x)}{x}\, d\lambda(x) = \sum_{n=1}^{+\infty}\frac{1}{n^2}.$$

3. $\frac{|x^2-y^2|}{(x^2+y^2)^2} = \frac{|x-y|}{(x^2+y^2)^{3/2}}\frac{x+y}{(x^2+y^2)^{1/2}} = \frac{|x-y|}{(x^2+y^2)^{3/2}}\frac{\|(x,y)\|_1}{\|(x,y)\|_2} \geq \frac{|x-y|}{(x^2+y^2)^{3/2}}$ et l'intégrale diverge grâce à la première question.

4. Idem : $\frac{|x-y|}{(x+y)^3} = \frac{|x-y|}{\|(x,y)\|_1^3} \geq K\frac{|x-y|}{\|(x,y)\|_2^3}$ avec l'équivalence des normes et on conclut grâce à la première question.

Solution 58 1. Pour $n \geq 2$, on a

$$\begin{aligned} W_n &= \int_0^{\pi/2} (\cos\theta)^n \, d\theta = \int_0^{\pi/2} (1-(\sin\theta)^2)(\cos\theta)^{n-2} \, d\theta \\ &= W_{n-2} + \int_0^{\pi/2} (\sin\theta)(-\sin\theta(\cos\theta)^{n-2}) \, d\theta \\ &= W_{n-2} + \left[(\sin\theta)\frac{1}{n-1}(\cos\theta)^{n-1}\right]_0^{\pi/2} - \int_0^{\pi/2} \cos\theta(\cos\theta)^{n-1}\frac{d\theta}{n-1} \\ &= W_{n-2} + 0 - \frac{1}{n-1}W_n, \end{aligned}$$

d'où $(n-1)W_n = (n-1)W_{n-2} - W_n$, puis $nW_n = (n-1)W_{n-2}$ et $W_n = \frac{n-1}{n}W_{n-2}$. Posons $u_n = nW_nW_{n-1}$: on a pour tout $n \geq 1$,

$$u_n = n\frac{n-1}{n}W_{n-2}W_{n-1} = (n-1)W_{n-1}W_{n-2} = u_{n-1},$$

ce qui montre que la suite est constante.

2. Pour tout $\theta \in [0, \pi/2]$, on a

$$0 \leq (\cos\theta)^{n+1} \leq (\cos\theta)^n \leq (\cos\theta)^{n-1},$$

d'où en intégrant $W_{n+1} \leq W_n \leq W_{n-1}$, puis, en multipliant par W_n : $W_nW_{n+1} \leq W_n^2 \leq W_{n-1}W_n$.
En multipliant par n, il vient $nW_nW_{n+1} \leq nW_n^2 \leq nW_{n-1}W_n$, soit $\frac{n}{n+1}u_{n+1} \leq nW_n^2 \leq u_n$, et (u_n) étant constante, $\frac{n}{n+1}u_1 \leq nW_n^2 \leq u_1$.
D'où $\lim_{n\to+\infty} nW_n^2 = u_1$, $\lim_{n\to+\infty} \sqrt{n}W_n = \sqrt{u_1}$, soit $W_n \sim \sqrt{\frac{u_1}{n}}$. Un calcul simple donne $u_1 = 1.W_1W_0 = 1\frac{\pi}{2}.1 = \frac{\pi}{2}$, d'où le résultat.

Solution 59 1. On a $J_n = \int_{[0,+\infty[} (1-\frac{t^2}{n})^n \mathbb{1}_{[0,\sqrt{n}]}(t) \, d\lambda(t)$. Soit $t \geq 0$. Pour $n \geq t^2$, on a $(1-\frac{t^2}{n})^n \mathbb{1}_{[0,\sqrt{n}]}(t) = (1-\frac{t^2}{n})^n$. Or on sait que

$$\log\left((1-\frac{t^2}{n})^n\right) = n\log\left(1-\frac{t^2}{n}\right) \sim n\left(-\frac{t^2}{n}\right) = -t^2,$$

donc $\lim_{n\to+\infty} \log((1-\frac{t^2}{n})^n) = -t^2$, d'où $\lim_{n\to+\infty} (1-\frac{t^2}{n})^n = \exp(-t^2/2)$.
Par ailleurs, on a

$$\forall n \geq 1 \quad t \geq 0 \quad (1-\frac{t^2}{n})^n \mathbb{1}_{[0,\sqrt{n}]}(t) \leq \exp(-t^2/2).$$

En effet, l'inégalité est évidente pour $t \geq \sqrt{n}$, tandis que pour t dans $[0, \sqrt{n}[$, elle découle de l'inégalité :

$$\forall x \in [0,1[\quad -\log(1-x) \geq x,$$

que l'on applique à $x = t^2/n$. Comme e^{-t^2} est intégrable sur $[0,+\infty[$, il suffit alors d'appliquer le théorème de convergence dominée.

2. Avec le changement de variable $u = t/\sqrt{n}$, on obtient l'égalité $J_n = \sqrt{n}\int_0^1 (1-u^2)^n\, du$. Posant encore $u = \sin\theta$, on a

$$\begin{aligned}\int_0^1 (1-u^2)^n\, du &= \int_0^{\pi/2} (1-\sin^2\theta)^n \cos\theta\, d\theta \\ &= \int_0^{\pi/2} (\cos\theta)^{2n+1}\, d\theta = W_{2n+1},\end{aligned}$$

soit $J_n = \sqrt{n} W_{2n+1}$.
Vu le résultat de l'exercice précédent, $J_n \sim \sqrt{n}\sqrt{\frac{\pi}{2(2n+1)}} \sim \frac{\sqrt{\pi}}{2}$, soit $\lim\limits_{n\to+\infty} J_n = \frac{\sqrt{\pi}}{2}$, d'où $\int_0^{+\infty} \exp(-t^2)\, dt = \frac{\sqrt{\pi}}{2}$ et, avec un changement de variable $\int_0^{+\infty} \exp(-t^2/2)\, dt = \sqrt{\pi/2}$.

Solution 60 Il suffit de faire le changement de variable $t^2/2 = u$ dans la formule précédente et de relire la définition de la fonction Γ.

Solution 61 1. On sait que pour tout $u \in]-1,1[$, $\frac{1}{1-u} = \sum\limits_{n=0}^{+\infty} u^n$. En particulier, pour $x \in [0,1[$, $-x^3 \in]-1,1[$ et on a $\frac{1}{1+x^3} = \sum\limits_{n=0}^{+\infty} (-1)^n x^{3n}$. Posons pour $x \in [0,1]$ et $N \geq 1$: $f_N(x) = \sum\limits_{n=0}^{N} (-1)^n x^{3n}$. On a $f_N(x) = \frac{1-(-x^3)^{N+1}}{1+x^3}$, d'où $|f_N(x)| \leq \frac{|1|+|x|^{3(N+1)}}{1+x^3} \leq \frac{2}{1+x^3} = g(x)$. Comme g est bornée sur $[0,1]$, g est intégrable sur $[0,1]$. On sait que pour $x \in [0,1[$, $\lim\limits_{n\to+\infty} f_N(x) = \frac{1}{1+x^3}$. Ainsi sur $[0,1]$ $f_N(x)$ converge λ-presque sûrement vers $\frac{1}{1+x^3}$ lorsque N tend vers l'infini. D'après le théorème de convergence dominée, on a donc

$$\int_{[0,1]} \frac{1}{x^3+1}\, d\lambda(x) = \lim_{N\to+\infty} \int_{[0,1]} f_N(x)\, d\lambda(x).$$

Comme les fonctions considérées sont toutes continues sur un compact, on peut réécrire cela en

$$\int_0^1 \frac{1}{x^3+1}\, dx = \lim_{N\to+\infty} \int_0^1 f_N(x)\, dx.$$

Un calcul facile donne $\int_0^1 f_N(x)\,dx = \sum\limits_{n=0}^{N} \frac{(-1)^n}{3n+1}$, d'où

$$\sum_{n=0}^{+\infty} \frac{(-1)^n}{3n+1} = \int_0^1 \frac{dx}{x^3+1}.$$

2. On admet l'égalité donnée dans l'énoncé. De plus, faisant apparaître la dérivée du dénominateur au numérateur, notons que

$$\frac{x-2}{x^2-x+1} = \frac{1}{2}\left(\frac{2x-4}{x^2-x+1}\right) = \frac{1}{2}\left(\frac{2x-1-3}{x^2-x+1}\right).$$

Ainsi,

$$\begin{aligned}
\frac{1}{1+x} - \frac{x-2}{x^2-x+1} &= \frac{1}{1+x} - \frac{1}{2}\frac{2x-1}{x^2-x+1} + \frac{3}{2}\frac{1}{x^2-x+1} \\
&= \frac{1}{1+x} - \frac{1}{2}\frac{2x-1}{x^2-x+1} + \frac{3}{2}\frac{1}{(x-1/2)^2+3/4} \\
&= \frac{1}{1+x} - \frac{1}{2}\frac{2x-1}{x^2-x+1} + \sqrt{3}\frac{2}{\sqrt{3}}\frac{1}{(\frac{2}{\sqrt{3}}x - \frac{1}{\sqrt{3}})^2+1}.
\end{aligned}$$

Une primitive de $\frac{1}{1+x} - \frac{x-2}{x^2-x+1}$ est alors

$$F(x) = \log(1+x) - \frac{1}{2}\log(x^2-x+1) + \sqrt{3}\arctan\left(\frac{2}{\sqrt{3}}x - \frac{1}{\sqrt{3}}\right),$$

d'où

$$\begin{aligned}
\int_0^1 \left(\frac{1}{1+x} - \frac{x-2}{x^2-x+1}\right)\,dx &= F(1) - F(0) = \log 2 + 2\sqrt{3}\arctan\frac{1}{\sqrt{3}} \\
&= \log 2 + 2\sqrt{3}\frac{\pi}{6} = \log 2 + \frac{\pi\sqrt{3}}{3}.
\end{aligned}$$

En effet $\tan(\pi/6) = \frac{\sin(\pi/6)}{\cos(\pi/6)} = \frac{1/2}{\sqrt{3}/2} = \frac{1}{\sqrt{3}}$. Et finalement on trouve

$$\sum_{n=0}^{+\infty} \frac{(-1)^n}{3n+1} = \frac{1}{3}\int_0^1 \left(\frac{1}{1+x} - \frac{x-2}{x^2-x+1}\right)\,dx = \frac{3\log 2 + \sqrt{3}\pi}{9}.$$

Solution 62 D'après le théorème de Tonelli, on a

$$\int_{[0,a]\times[0,+\infty[} |f(x,y)|\,d(\lambda\otimes\lambda)(x,y) = \int_{[0,a]}\left(\int_{[0,+\infty[} |f(x,y)|\,d\lambda(y)\right)d\lambda(x).$$

Pour $x > 0$, on a

$$\int_{[0,+\infty[} |f(x,y)|\,d\lambda(y) = |\sin x|\int_{[0,+\infty[} e^{-xy}\,d\lambda(y) = \left|\frac{\sin x}{x}\right| \le 1,$$

d'où

$$\int_{[0,a]\times[0,+\infty[} |f(x,y)| \, d(\lambda\otimes\lambda)(x,y) \leq a < +\infty.$$

Ainsi f est intégrable sur $[0,a]\times[0,+\infty[$. On peut donc appliquer le théorème de Fubini. Dans le premier sens (intégration en y, puis en x), on a

$$I_a = \int_{[0,a]} \left(\int_{[0,+\infty[} |f(x,y)| \, d\lambda(y) \right) d\lambda(x) = \int_{[0,a]} \frac{\sin x}{x} \, d\lambda(x) = \int_0^a \frac{\sin x}{x} \, dx.$$

Dans l'autre sens (intégration en x, puis en y), on a pour tout $y > 0$

$$\begin{aligned}
\int_{[0,a]} f(x,y) \, d\lambda(x) &= \int_{[0,a]} \operatorname{Im} e^{x(i-y)} \, d\lambda(x) = \operatorname{Im} \int_{[0,a]} e^{x(i-y)} \, d\lambda(x) \\
&= \operatorname{Im} \frac{1-e^{a(i-y)}}{y-i} = \operatorname{Im} \frac{1}{y-i} - \operatorname{Im} \frac{e^{a(i-y)}}{y-i} \\
&= \operatorname{Im} \frac{y+i}{y^2+1} - \operatorname{Im} \frac{e^{a(i-y)}}{y-i} \\
&= \frac{1}{1+y^2} - \operatorname{Im} \frac{e^{a(i-y)}}{y-i},
\end{aligned}$$

d'où

$$I_a = \int_0^{+\infty} \frac{dy}{1+y^2} - \int_{[0,+\infty[} \operatorname{Im} \frac{e^{a(i-y)}}{y-i} \, d\lambda(y).$$

Or, pour $y \in \mathbb{R}$, on a $\frac{1}{|y-i|} \leq 1$, ce qui donne

$$\left| \int_{[0,+\infty[} \operatorname{Im} \frac{e^{a(i-y)}}{y-i} \, d\lambda(y) \right| \leq \int_{[0,+\infty[} \left| \frac{e^{a(i-y)}}{y-i} \right| \, d\lambda(y) \leq \int_0^{+\infty} \frac{e^{-ay}}{1} \, dy = \frac{1}{a},$$

d'où $\lim\limits_{a\to+\infty} I_a = \int_0^{+\infty} \frac{dy}{1+y^2} = \frac{\pi}{2}$. On en déduit que $\lim\limits_{a\to+\infty} \int_0^a \frac{\sin x}{x} \, dx = \frac{\pi}{2}$, soit $\int_0^{+\infty} \frac{\sin x}{x} \, dx = \frac{\pi}{2}$. (Mais $\frac{\sin x}{x}$ n'est toujours pas intégrable sur $\mathbb{R}_+$, qu'on se le dise !)

Solution 63 1. On a $\phi(\lambda) = R(\lambda) + S(\lambda)$, avec

$$R(\lambda) = \int_0^1 e^{-\lambda u} e^{iu} u^{\alpha-1} du \quad \text{et} \quad S(\lambda) = \int_1^{+\infty} e^{-\lambda u} e^{iu} u^{\alpha-1} du.$$

De plus,

$$\forall \lambda \geq 0 \quad \forall u \in [0,1] \quad |e^{-\lambda u} e^{iu} u^{\alpha-1}| \leq u^{\alpha-1}.$$

Or, la fonction $u \mapsto u^{\alpha-1}$ est intégrable sur $[0,1]$, ce qui nous donne donc la continuité de R sur $\mathbb{R}_+$ grâce au théorème de continuité d'une intégrale dépendant d'un paramètre. Par ailleurs, par définition d'une intégrale de Riemann impropre, on a

$$S(\lambda) = \lim_{M \to +\infty} \int_1^M e^{-\lambda u} e^{iu} u^{\alpha-1} du.$$

On va faire une intégration par parties pour faire apparaître une intégrale absolument convergente. On a

$$\int_1^M e^{-\lambda u} e^{iu} u^{\alpha-1} du = \left[\frac{e^{(i-\lambda)u} u^{\alpha-1}}{i-\lambda} \right]_1^M - \int_1^M \frac{e^{(i-\lambda)u}}{i-\lambda} (\alpha-1) u^{\alpha-2}\, du.$$

La majoration

$$\left| \frac{e^{(i-\lambda)u}}{i-\lambda} (\alpha-1) u^{\alpha-2} \right| \leq (\alpha-1) u^{\alpha-2} \tag{D.2}$$

donne la convergence absolue de l'intégrale de $\frac{e^{(i-\lambda)u}}{i-\lambda}(\alpha-1)u^{\alpha-2}$ entre 1 et l'infini, ce qui permet d'écrire, en faisant tendre M vers $+\infty$:

$$S(\lambda) = \frac{e^{i-\lambda}}{\lambda - i} - \int_1^{+\infty} \frac{e^{(i-\lambda)u}}{i-\lambda} (\alpha-1) u^{\alpha-2}\, du,$$

ce qui prouve au passage que $S(\lambda)$ a bien un sens.
La fonction $\lambda \mapsto \frac{e^{i-\lambda}}{\lambda-i}$ est évidemment continue ; par ailleurs la continuité de $\lambda \mapsto \int_1^{+\infty} \frac{e^{(i-\lambda)u}}{i-\lambda}(\alpha-1)u^{\alpha-2}\, du$ découle encore de la majoration (D.2) et du théorème de continuité d'une intégrale dépendant d'un paramètre. On a donc la continuité de R, de S, donc de ϕ.

2. Soit $\lambda_0 > 0$. Pour $\lambda > \lambda_0$, on a

$$\forall u > 0 \quad |e^{-\lambda u} e^{iu} u^{\alpha-1}| \leq e^{-\lambda_0 u} u^{\alpha-1}.$$

Vérifions l'intégrabilité de $e^{-\lambda_0 u} u^{\alpha-1}$ sur $\mathbb{R}_+$: elle est continue, donc localement intégrable. En 0, on a $e^{-\lambda_0 u} u^{\alpha-1} \sim u^{\alpha-1}$ qui est bien convergente car $\alpha > 0$, et en l'infini, comme $u^{\alpha-1} = o(e^{\frac{\lambda_0}{2}u})$, on obtient $e^{-\lambda_0 u} u^{\alpha-1} = o(e^{-\frac{\lambda_0}{2}u})$, ce qui donne l'intégrabilité. On a donc

$$\forall \lambda_1 > \lambda_0 \quad \phi(\lambda_1) = \int_{]0,+\infty[} e^{-\lambda_1 u} e^{iu} u^{\alpha-1}\, d\lambda(u),$$

ce qui n'était pas évident *a priori* puisque ϕ a été définie comme une intégrale de Riemann impropre, et non pas comme une intégrale de Lebesgue. De plus

$$\left| \frac{\partial}{\partial \lambda} (e^{-\lambda u} e^{iu} u^{\alpha-1}) \right| = |-e^{-\lambda u} e^{iu} u^{\alpha}| = e^{-\lambda u} u^{\alpha} \leq e^{-\lambda_0 u} u^{\alpha},$$

qui est intégrable sur $]0, +\infty[$, par les mêmes arguments que précédemment. On peut donc appliquer le théorème de dérivation sous le signe intégrale :

$$\forall \lambda_1 > \lambda_0, \quad \phi'(\lambda_1) = \int_{]0,+\infty[} -e^{-\lambda_1 u} e^{iu} u^\alpha \, d\lambda(u).$$

Mais la dérivabilité est une propriété locale et $]0,+\infty[= \bigcup_{\lambda_0>0}]\lambda_0, +\infty[$, donc on a

$$\forall \lambda_1 > 0, \quad \phi'(\lambda_1) = \int_{]0,+\infty[} -e^{-\lambda_1 u} e^{iu} u^\alpha \, d\lambda(u).$$

On fait une intégration par parties :

$$-\int_0^M e^{u(i-\lambda)} u^\alpha \, du \;=\; -\left[\frac{e^{u(i-\lambda)}}{i-\lambda} u^\alpha\right]_0^M + \alpha \int_0^M \frac{e^{u(i-\lambda)}}{i-\lambda} u^{\alpha-1} \, du.$$

En faisant tendre M vers $+\infty$, on obtient : $\phi'(\lambda) = \frac{\alpha}{i-\lambda}\phi(\lambda)$. Posons $F(\lambda) = \phi(\lambda)\exp(-\int_0^\lambda \frac{\alpha}{i-u} \, du)$. Pour tout $\lambda > 0$, $F'(\lambda)$ vaut

$$\phi(\lambda)\left[-\frac{\alpha}{i-\lambda}\exp\left(-\int_0^\lambda \frac{\alpha}{i-u} \, du\right)\right] + \phi'(\lambda)\exp\left(-\int_0^\lambda \frac{\alpha}{i-u} \, du\right)$$
$$= \exp\left(-\int_0^\lambda \frac{\alpha}{i-u} \, du\right)\left(-\frac{\alpha}{i-\lambda}\phi(\lambda) + \phi'(\lambda)\right) = 0.$$

Ainsi, pour λ et ε strictement positifs, on a

$$F(\lambda) = F(\varepsilon) = \phi(\varepsilon)\exp\left(-\int_0^\varepsilon \frac{\alpha}{i-u} \, du\right).$$

Comme ϕ est continue en 0, $\lim_{\varepsilon\to 0^+} \phi(\varepsilon) = \phi(0)$.

On obtient donc $F(\lambda) = \phi(0)$ en faisant tendre ε vers 0, soit $\phi(\lambda)\exp(-\int_0^\lambda \frac{\alpha}{i-u} \, du) = \phi(0)$, ou encore

$$\phi(\lambda) = \phi(0)\exp\left(-\int_0^\lambda \frac{\alpha}{u-i} \, du\right).$$

Or

$$\int_0^\lambda \frac{du}{u-i} = \int_0^\lambda \frac{u+i}{1+u^2} \, du = \frac{1}{2}\log(1+\lambda^2) + i\arctan\lambda,$$

d'où

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

3. On a

$$\phi(0) = (\lambda^2+1)^{\alpha/2}\exp(i\alpha\arctan\lambda)\phi(\lambda).$$

Or

$$\begin{aligned}\phi(\lambda) &= \frac{1}{\lambda^\alpha}\int_0^{+\infty} e^{-\lambda u}e^{\frac{i\lambda u}{\lambda}}(\lambda u)^{\alpha-1}\lambda\, du \\ &= \frac{1}{\lambda^\alpha}\int_0^{+\infty} e^{-x}e^{\frac{ix}{\lambda}}x^{\alpha-1}\, dx,\end{aligned}$$

d'où

$$\phi(0) = (1+\lambda^{-2})^{\alpha/2}\exp(i\alpha\arctan\lambda)\int_0^{+\infty} e^{-x}e^{\frac{ix}{\lambda}}x^{\alpha-1}\, dx. \qquad \text{(D.3)}$$

4. On a

$$\forall\lambda>0,\quad \forall x>0,\quad |e^{-x}e^{\frac{ix}{\lambda}}x^{\alpha-1}|\le e^{-x}x^{\alpha-1}$$

et pour tout $x>0$, $\lim\limits_{\lambda\to+\infty} e^{-x}e^{\frac{ix}{\lambda}}x^{\alpha-1} = e^{-x}x^{\alpha-1}$, donc par le théorème de convergence dominée,

$$\lim_{\lambda\to+\infty}\int_0^{+\infty} e^{-x}e^{\frac{ix}{\lambda}}x^{\alpha-1}\, dx = \int_0^{+\infty} e^{-x}x^{\alpha-1}\, dx,$$

d'où, en faisant tendre λ vers $+\infty$ dans (D.3) : $\phi(0)=\exp(i\alpha\frac{\pi}{2})\Gamma(\alpha)$, soit

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

et en particulier pour $\alpha=\frac12$,

$$\int_0^{+\infty}\frac{e^{iu}}{\sqrt{u}}\, du = \exp\left(i\frac{\pi}{4}\right)\Gamma\left(\frac12\right) = \frac{1+i}{\sqrt2}\sqrt{\pi},$$

soit en prenant la partie réelle et la partie imaginaire :

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

5. De simples changements de variable donnent

$$\int_0^{M}\cos(u^2)\, du = \frac12\int_0^{M^2}\frac{\cos t}{\sqrt t}dt$$

et

$$\int_0^{M}\sin(u^2)\, du = \frac12\int_0^{M^2}\frac{\sin t}{\sqrt t}dt.$$

En faisant tendre M vers $+\infty$, on obtient

$$\int_0^{+\infty}\sin(u^2)\, du = \int_0^{+\infty}\cos u^2\, du = \sqrt{\frac{\pi}{8}}.$$

Solution 64 La fonction $t\mapsto\frac{\sin t}{t}$ est continue sur $]0,+\infty[$ et est donc localement intégrable sur $]0,+\infty[$. Comme $\lim\limits_{t\to0}\frac{\sin t}{t}=1$, la fonction se prolonge par continuité en 0 et l'intégrale y est donc "faussement impropre". Seule l'intégrale en l'infini pose donc problème. On fait une intégration par parties :

$$\int_{\pi/2}^{M} \frac{\sin t}{t}\,dt = -\frac{\cos M}{M} + \int_{\pi/2}^{M} \frac{\cos t}{t^2}\,dt.$$

Comme $|\frac{\cos t}{t^2}| \le \frac{1}{t^2}$ et que $1/t^2$ est intégrable sur $[\pi/2, +\infty[$, alors $\frac{\cos t}{t^2}$ est absolument intégrable sur $[\pi/2, +\infty[$, et on a

$$\lim_{M\to+\infty} \int_{\pi/2}^{M} \frac{\sin t}{t}\,dt = \lim_{M\to+\infty} \int_{\pi/2}^{M} \frac{\cos t}{t^2}\,dt = \int_{\pi/2}^{+\infty} \frac{\cos t}{t^2}\,dt,$$

ce qui nous donne la convergence de l'intégrale impropre considérée. Posons à présent $u_n = \int_{n\pi}^{(n+1)\pi} \frac{|\sin t|}{t}\,dt$. On a

$$u_n = \int_0^{\pi} \frac{|\sin(n\pi + t)|}{n\pi + t}\,dt = \int_0^{\pi} \frac{|\sin(t)|}{n\pi + t}\,dt = \frac{1}{n\pi}\int_0^{\pi} |\sin t|\frac{n\pi}{n\pi + t}\,dt$$

Comme $\left|(\sin t)\frac{n\pi}{n\pi+t}\right| \le 1$, on a $\lim\limits_{n\to+\infty} \int_0^{\pi} |\sin t|\frac{n\pi}{n\pi+t}\,dt = \int_0^{\pi} |\sin t|\,dt = 2$, d'où l'équivalent $u_n \sim \frac{2}{n\pi}$. Grâce au théorème sur les équivalents de sommes partielles de séries divergentes on a l'équivalent à l'infini

$$\int_0^{n\pi} \frac{|\sin t|}{t}\,dt = \sum_{k=1}^{n-1} u_k \sim \sum_{k=1}^{n-1} \frac{2}{k\pi} \sim \sum_{k=1}^{n-1} \frac{2}{\pi}(\log(k+1) - \log k) \sim \frac{2}{\pi}\log n,$$

ce qui entraîne bien sûr la divergence de l'intégrale.

Solution 65 1. La première question découle immédiatement de la convergence de l'intégrale impropre $\int_0^{+\infty} \frac{\sin t}{t}\,dt$ et de la continuité de la transformée de Laplace (voir les exercices corrigés 64 et 54). On va néanmoins en donner une preuve autonome.
Posons $F_n(x) = \int_0^n \frac{\sin t}{t} e^{-xt}\,dt$. Comme on a :

— pour tout $t > 0$, $x \mapsto e^{-xt\frac{\sin t}{t}}$ est continue,
— pour tout $t > 0$, pour tout $x \ge 0$, $|e^{-xt\frac{\sin t}{t}}| \le 1$,
— 1 est intégrable sur $[0, n]$,

alors pour tout n, F_n est une fonction continue sur $\mathbb{R}_+$. On va montrer que la suite $(F_n(x))$ vérifie un critère de Cauchy uniforme, ce qui montrera à la fois que $F(x)$ est bien définie (critère de Cauchy), et que la limite $F(x)$ est continue, comme limite uniforme de fonctions continues.

Soient donc $x \geq 0$ et N, n, p avec $N \leq p \leq n$. On a

$$\begin{aligned}|F_n(x) - F_p(x)| &= |\int_p^n e^{-xt}\frac{\sin t}{t}\,dt| \\ &= |\mathrm{Im}\int_p^n e^{-xt}\frac{e^{it}}{t}\,dt| \\ &\leq |\int_p^n \frac{e^{-(x-i)t}}{t}\,dt| \\ &\leq \frac{1}{|x-i|}|\int_p^n (x-i)\frac{e^{-(x-i)t}}{t}\,dt| \\ &\leq |\int_p^n (x-i)\frac{e^{-(x-i)t}}{t}\,dt|.\end{aligned}$$

Une intégration par parties donne

$$\int_p^n (x-i)\frac{e^{-(x-i)t}}{t}\,dt = \frac{e^{-(p-i)t}}{p} - \frac{e^{-(n-i)t}}{n} - \int_p^n \frac{e^{-(x-i)t}}{t^2}\,dt$$

$$|\int_p^n (x-i)\frac{e^{-(x-i)t}}{t}\,dt| \leq \frac{1}{p} + \frac{1}{n} + \int_p^n \frac{dt}{t^2} = \frac{2}{p} \leq \frac{2}{N},$$

ce qui donne le critère de Cauchy uniforme voulu.

2. Soit $a > 0$. Pour tout $x \in]a, +\infty[$ et tout $t > 0$, on a

$$\left|\frac{\partial}{\partial x}\frac{\sin t}{t}e^{-xt}\right| = |-(\sin t)e^{-xt}| \leq e^{-at}.$$

On peut donc appliquer le théorème de dérivation sous le signe intégrale, et on a

$$\forall x > a, \quad F'(x) = \int_{]0,+\infty[} (-\sin t)e^{-xt}\,d\lambda(t).$$

Mais la dérivabilité est une propriété locale et tout point $x_0 > 0$ admet un voisinage de la forme $]a, +\infty[$. Ainsi F est dérivable sur $]0, +\infty[$, avec

$$\forall x > 0, \quad F'(x) = \int_{]0,+\infty[} (-\sin t)e^{-xt}\,d\lambda(t).$$

Or pour tout $A > 0$, on a

$$\int_0^A e^{-(i+x)t}\,dt = \frac{1 - e^{-(i+x)A}}{i+x},$$

donc

$$\lim_{A\to+\infty}\int_0^A e^{-(i+x)t}\,dt = \frac{1}{i+x} = \frac{-i+x}{1+x^2},$$

puis $\lim_{A\to+\infty} \operatorname{Im} \int_0^A e^{-(i+x)t}\, dt = -\frac{1}{1+x^2}$.
Or $\operatorname{Im} \int_0^A e^{-(i+x)t}\, dt = \int_{[0,A]} (-\sin t)e^{-xt}\, d\lambda(t)$, d'où

$$\forall x > 0, \quad F'(x) = \lim_{A\to+\infty} \int_{[0,A]} (-\sin t)e^{-xt}\, d\lambda(t) = -\frac{1}{1+x^2}.$$

3. En intégrant entre x et y, on obtient

$$\forall x, y > 0 \quad F(y) - F(x) = \arctan x - \arctan y.$$

Comme

$$|F(y)| = \left|\int_0^{+\infty} \frac{\sin t}{t} e^{-yt}\, dt\right| \le \int_0^{+\infty} \left|\frac{\sin t}{t} e^{-yt}\right|\, dt \le \int_0^{+\infty} e^{-yt}\, dt = \frac{1}{y},$$

la limite de $F(y)$ en $+\infty$ est 0, d'où

$$\forall x > 0 \quad F(x) = \frac{\pi}{2} - \arctan x.$$

Comme F est continue, en faisant tendre x vers 0, on obtient $F(0) = \frac{\pi}{2}$, ce qui est le résultat voulu.

Solution 66 Comme indiqué, on commence par traiter le cas $a = b = c = 1$. Par symétrie, on a $I = \frac{1}{8} \int_{B(0,1)} (x^2 + y^2 + z^2)\, d\lambda^{\otimes 3}(x, y, z)$. Or

$$\int_{B(0,1)} (x^2 + y^2 + z^2)\, d\lambda^{\otimes 3}(x, y, z) = \int_{\mathbb{R}^3} \phi(\|x\|_2)\, d\lambda^{\otimes 3}(x),$$

avec $\phi(x) = x^2 \mathbb{1}_{[0,1]}(x)$, donc avec la formule d'intégration des fonctions radiales, on a

$$I = \frac{1}{8} V \int_{\mathbb{R}^+} \phi(t) 3t^2\, dt = \frac{1}{8}.\frac{4}{3}\pi.\frac{3}{5} = \frac{\pi}{10}.$$

Passons au cas général. Par changement de variables, on a

$$\begin{aligned}
& \int_{V(a,b,c)} (x^2 + y^2 + z^2)\, d\lambda^{\otimes 3}(x, y, z) \\
= \; & \int_{\mathbb{R}^3} \mathbb{1}_{V(a,b,c)}(x, y, z)(x^2 + y^2 + z^2)\, d\lambda^{\otimes 3}(x, y, z) \\
= \; & \int_{\mathbb{R}^3} \mathbb{1}_{V(1,1,1)}\left(\frac{x}{a}, \frac{y}{b}, \frac{z}{c}\right) \left(a^2 \left(\frac{x}{a}\right)^2 + b^2 \left(\frac{y}{b}\right)^2 + c^2 \left(\frac{y}{c}\right)^2\right)\, d\lambda^{\otimes 3}(x, y, z) \\
= \; & abc \int_{\mathbb{R}^3} \mathbb{1}_{V(1,1,1)}(x, y, z)(a^2x^2 + b^2y^2 + c^2z^2)\, d\lambda^{\otimes 3}(x, y, z).
\end{aligned}$$

Par symétrie, on a

$$\begin{aligned}\int_{V(1,1,1)} x^2\, d\lambda^{\otimes 3}(x,y,z) &= \int_{V(1,1,1)} y^2\, d\lambda^{\otimes 3}(x,y,z)\\ &= \int_{V(1,1,1)} z^2\, d\lambda^{\otimes 3}(x,y,z)\\ &= \frac{1}{3}\int_{V(1,1,1)} (x^2+y^2+z^2)\, d\lambda^{\otimes 3}(x,y,z)\\ &= \frac{\pi}{30},\end{aligned}$$

d'où $I = \frac{\pi}{30} abc(a^2+b^2+c^2)$.

Solution 67 1. Le seul $s \in \mathbb{Z}^d$ tel que $(x_1, \dots, x_d) \in s+T$ est $s = (\lfloor x_1 \rfloor, \dots, \lfloor x_d \rfloor)$. $\mathbb{R}^d$ est donc la réunion disjointe des $(s+T)_{s\in\mathbb{Z}^d}$. Par suite, $\frac{1}{2}A$ est la réunion disjointe des $\left((\frac{1}{2}A) \cap (s+T)\right)_{s\in\mathbb{Z}^d}$.

2. Posons $B = \frac{1}{2}A$. D'après la question précédente, les $(B \cap (s+T))_{s\in\mathbb{Z}^d}$ forment une partition de B, et

$$\lambda(B) = \sum_{s\in\mathbb{Z}^d} \lambda(B\cap(s+T)) = \sum_{s\in\mathbb{Z}^d} \lambda((-s+B)\cap T) = \sum_{s\in\mathbb{Z}^d} \lambda((s+B)\cap T).$$

Notons que pour la deuxième égalité, on a utilisé l'invariance de la mesure de Lebesgue sur $\mathbb{R}^d$ par translation. On raisonne alors par l'absurde et on suppose que les translatés $(B+s)_{s\in\mathbb{Z}^d}$ sont disjoints. On obtient alors

$$\lambda(B) = \lambda((\mathbb{Z}^d + B)\cap T) \le \lambda(T) = 1,$$

ce qui donne une contradiction puisque $\lambda(B) = 2^{-d}\lambda(A)$.
Soient donc s_1, s_2 dans $\mathbb{Z}^d$ distincts avec $(B+s_1)\cap(B+s_2) \neq \varnothing$. On a ainsi $b_1 + s_1 = b_2 + s_2$ pour b_1 et b_2 dans B bien choisis, soit $b_1 - b_2 = s_1 - s_2 \in S$. Par ailleurs $b_1 - b_2 = \frac{1}{2}(2b_1 + (-2b_2))$. b_1 et b_2 sont dans B, donc $2b_1$ et $2b_2$ sont dans A, par symétrie $-2b_2 \in A$. Enfin par convexité $b_1 - b_2 \in A$.

3. Supposons que $\lambda^d(A) > 2^d|\det(u,v)|$. Soit M l'application linéaire envoyant e_1 sur u_1, e_2 sur u_2, ..., e_d sur u_d. Posons $A' = M^{-1}A$. A' est l'image d'un convexe de $\mathbb{R}^d$ symétrique par rapport à l'origine par une application linéaire : c'est aussi un convexe de $\mathbb{R}^d$ symétrique par rapport à l'origine. On a

$$\begin{aligned}\lambda(A') = \det M^{-1}\lambda(A) = (\det M)^{-1}\lambda(A) &= \det(u_1,\dots,u_d)^{-1}\lambda(A)\\ &> 2^d.\end{aligned}$$

Donc A' contient un point de $\mathbb{Z}^d\backslash\{(0,\dots,0)\}$. On en déduit que $M.A'$ contient un point de $M\mathbb{Z}^d\backslash\{M(0,\dots,0)\}$, c'est-à-dire que A contient un point de $S\backslash\{(0,\dots,0)\}$, ce qui est le résultat voulu.

4. (a) Le morphisme $\phi(x)=x^2$ de $(\mathbb{F}_p^*,\times)$ a comme noyau $\{-1,+1\}$, car $\mathbb{F}_p$ est un corps. On en déduit que $|\text{Im }\phi|=|\mathbb{F}_p^*|/|\text{Ker }\phi|=\frac{p-1}{2}$. $\mathbb{F}_p^*$ a $\frac{p-1}{2}$ carrés tandis que $\mathbb{F}_p$ a $\frac{p-1}{2}+1=\frac{p+1}{2}$ carrés. Soit C l'ensemble des carrés de $\mathbb{F}_p$. C et $-1-C$ ont même cardinal $\frac{p+1}{2}$, donc ne peuvent être disjoints puisque la réunion est de cardinal majoré par p. Ainsi, il existe x et y dans $\mathbb{F}_p$, avec $x^2=-y^2-1$, ce qui donne le résultat voulu en remontant aux représentants dans $\mathbb{Z}$.

   (b) Si $x\in\mathbb{Z}^4$ est l'image par m de $t_1e_1+t_2e_2+t_3e_3+t_4e_4$, alors $\|x\|_2^4=(pt_1+rt_3+st_4)^2+(pt_2+st_3-st_4)^2+t_3^2+t_4^2$, qui est congru modulo p à $t_3^2(r^2+s^2+1)+t_4^2(r^2+s^2+1)$, qui est lui-même divisible par p. Le volume de la boule de rayon $\sqrt{2p}$ dans $\mathbb{R}^4$ est $\frac{\pi^2}{2}(\sqrt{2p})^4=2\pi^2p^2>2^4p^2=2^4|\det M|$, donc d'après le théorème de Minkowski, A rencontre la boule en un point différent de 0.

   (c) Soit p un nombre premier. D'après la question précédente, il existe $x\in A$ avec $0<\|x\|_2^2<2p$ et $\|x\|_2^2$ divisible par p. On a donc $\|x\|_2^2=p$, ce qui montre que p est somme de 4 carrés. Par l'identité de Lagrange, l'ensemble des nombres qui sont somme de 4 carrés est stable par multiplication, donc comme tout entier non nul est produit de nombre premiers, tout entier est somme de 4 carrés.

Solution 68

1. On fait une intégration par parties : soit $T>0$, on a
$$\int_0^T t^xe^{-t}\,dt=[-t^xe^{-t}]_0^T+\int_0^T xt^{x-1}e^{-t}\,dt$$
soit
$$\int_0^T t^xe^{-t}\,dt=-T^xe^{-T}+x\int_0^T t^{x-1}e^{-t}\,dt$$
En faisant tendre T vers l'infini, on obtient $\Gamma(x+1)=x\Gamma(x)$.

   Les formules se montrent alors aisément par récurrence, en utilisant $\Gamma(1)=\int_{\mathbb{R}_+}e^t\,dt=1$.

2. Avec un changement de variables, on a
$$\begin{aligned}\int_0^{+\infty} t^{2n}e^{-t^2/2}\,dt&=2^{n-1/2}\int_0^{+\infty}(\frac{t^2}{2})^{n-1/2}e^{-t^2/2}\,t\,dt\\&=2^{n-1/2}\Gamma(n+1/2),\end{aligned}$$

d'où

$$\begin{aligned}\int_{\mathbb{R}} t^{2n}e^{-t^2/2}\,d\lambda(t) &= 2\int_0^{+\infty} t^{2n}e^{-t^2/2}\,dt\\ &= 2^{n+1/2}\Gamma(n+1/2) = \sqrt{2}.2^n\Gamma(n+1/2)\\ &= \sqrt{2}\Gamma(1/2)\frac{(2n)!}{2^n n!} = \sqrt{2\pi}\frac{(2n)!}{2^n n!}.\end{aligned}$$

3. $e^{itx}G(x)$ est intégrable car $|e^{itx}G(x)| = G(x)$; l'intégrabilité de G a déjà été démontrée (intégrale de Gauss).
   Comme $e^{itx}G(x) = (\cos(tx) + i\sin(tx))G(x)$, l'imparité de $\sin(t\cdot)G(\cdot)$ donne la première égalité. On a de plus le développement en série entière $\cos(tx)G(x) = \sum_{n=0}^{+\infty}(-1)^n\frac{(tx)^{2n}}{(2n)!}G(x) = \lim_{N\to+\infty} f_N(x)$, avec

$$f_N(x) = \sum_{n=0}^{N}(-1)^n\frac{(tx)^{2n}}{(2n)!}\exp(-x^2/2).$$

$$\text{Or}\quad |f_N(x)| \le \sum_{n=0}^{N}\frac{(tx)^{2n}}{(2n)!}G(x) \le \sum_{n=0}^{+\infty}\frac{(tx)^{2n}}{(2n)!}G(x) = \mathrm{ch}(tx)G(x).$$

   On a

$$\begin{aligned}\int_{\mathbb{R}}\mathrm{ch}(tx)G(x)\,dx &= \frac{1}{2}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{x^2}{2}+tx}\,dx + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{x^2}{2}-tx}\,dx\\ &= \frac{1}{2}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{(x+t)^2-t^2}{2}}\,dx + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{(x-t)^2-t^2}{2}}\,dx\\ &= \exp(t^2/2)\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{u^2}{2}}\,du < +\infty,\end{aligned}$$

   donc on peut appliquer le théorème de convergence dominée : l'intégrale cherchée vaut $\lim_{N\to+\infty}\int_{\mathbb{R}} f_N(x)\,dx$. Or

$$\begin{aligned}\int_{\mathbb{R}} f_N(x)\,d\lambda(x) &= \sum_{n=0}^{N}(-1)^n\frac{t^{2n}}{(2n)!}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} x^{2n}e^{-\frac{x^2}{2}}\,d\lambda(x)\\ &= \sum_{n=0}^{N}(-1)^n\frac{t^{2n}}{(2n)!}\frac{(2n)!}{2^n n!},\end{aligned}$$

   qui converge, lorsque N tend vers l'infini, vers $\exp(-\frac{t^2}{2})$. On reconnaît le développement en série de l'exponentielle. C'est le résultat voulu.

4. Si $X \sim \mathcal{N}(0,1)$, on a calculé les moments d'ordre pair de X qui valent $\mathbb{E}(X^{2n}) = \frac{(2n)!}{2^n n!}$ ainsi que la fonction caractéristique de X.

Solution 69 Comme $\frac{y}{1-y} = \sum_{p\geq 1} y^p$ pour $0 < y < 1$, on a pour tout $t > 0$:

$$t^{s-1}\frac{e^{-t}}{1-e^{-t}} = \sum_{p=1}^{+\infty} t^{s-1}e^{-pt}.$$

Comme la série est à termes positifs, on peut intervertir la série et l'intégrale :

$$\int_0^{+\infty} t^{s-1}\frac{e^{-t}}{1-e^{-t}} = \sum_{p=1}^{+\infty} \int_0^{+\infty} t^{s-1}e^{-pt}\,dt.$$

Cependant

$$\begin{aligned}\int_0^{+\infty} t^{s-1}e^{-pt}\,dt &= \frac{1}{p^s}\int_0^{+\infty}(pt)^{s-1}e^{-pt}\,pdt\\ &= \frac{1}{p^s}\int_0^{+\infty} u^{s-1}e^{-u}\,du\\ &= \frac{1}{p^s}\Gamma(s).\end{aligned}$$

On a donc

$$\int_0^{+\infty} t^{s-1}\frac{e^{-t}}{1-e^{-t}}\,dt = \sum_{p=1}^{+\infty}\frac{1}{p^s}\Gamma(s) = \Gamma(s)\zeta(s).$$

Solution 70 On peut remarquer que

$$\begin{aligned}\int_{2n}^{+\infty} x^n e^{-x}\,dx = \int_n^{+\infty}(x+n)^n e^{-(n+x)}\,dx &\leq \int_n^{+\infty}(2x)^n e^{-(n+x)}\,dx\\ &\leq \left(\frac{2}{e}\right)^n\Gamma(n+1).\end{aligned}$$

Ainsi, $\Gamma(n+1) \sim F(n) = \int_0^{2n} x^n e^{-x}\,dx$. On fait les transformations affines :

$$\begin{aligned}F(n) &= \int_0^{2n} x^n e^{-x}\,dx = \int_{-n}^{n}(x+n)^n e^{-(x+n)}\,dx\\ &= e^{-n}n^n\int_{-n}^{n}(1+\frac{x}{n})^n e^{-x}\,dx\\ &= e^{-n}n^{n+1/2}\int_{-\sqrt{n}}^{+\sqrt{n}}(1+\frac{x}{\sqrt{n}})^n e^{-\sqrt{n}x}\,dx\end{aligned}$$

puis

$$\begin{aligned}F(n) &= e^{-n}n^{n+1/2}\int_{-\sqrt{n}}^{\sqrt{n}} e^{n(\ln(1+\frac{x}{\sqrt{n}})-\frac{x}{\sqrt{n}})}\,dx\\ &= e^{-n}n^{n+1/2}\int_{-\sqrt{n}}^{\sqrt{n}} e^{x^2(\sqrt{n}/x)^2(\ln(1+\frac{x}{\sqrt{n}})-\frac{x}{\sqrt{n}})}\,dx\\ &= e^{-n}n^{n+1/2}\int_{-\sqrt{n}}^{\sqrt{n}} e^{-x^2\phi(\frac{x}{\sqrt{n}})}\,dx,\end{aligned}$$

avec $\phi(x) = \frac{1}{x^2}(x - \log(1+x))$.

Or $x \mapsto \log(1+x) - x$ se développe en série entière sur $]-1,1[$:

$$\log(1+x) - x = \sum_{k=2}^{+\infty} \frac{(-1)^{n+1}}{n} x^n.$$

Pour $x \in [0,1[$, on a $\frac{x^{n+1}}{n+1} \le \frac{x^n}{n}$ pour tout n, donc le critère spécial des séries alternées donne

$$\log(1+x) - x \le -\frac{x^2}{2} + \frac{x^3}{3} \le -\frac{x^2}{2} + \frac{x^2}{3} = -\frac{x^2}{6}.$$

Pour $x \in]-1,0]$, tous les termes sont négatifs et $\log(1+x) - x \le -\frac{x^2}{2}$. Finalement, on a toujours $\log(1+x) - x \le -\frac{x^2}{6}$, d'où on déduit que $\phi(x) \ge \frac{1}{6}$ sur $]-1,1[$. Le même développement permet de prolonger ϕ par continuité en 0 avec $\phi(0) = 1/2$. On pose alors $f_n(y) = \mathbb{1}_{]-\sqrt{n},\sqrt{n}[}(y)e^{-y^2\phi(y/\sqrt{n})}$ et de la sorte

$$F(n) = e^{-n}n^{n+1/2}\int_{\mathbb{R}} f_n(y)\,dy.$$

On a $0 \le f_n(y) \le \exp(-y^2/6)$ et $\lim_{n\to+\infty} f_n(y) = e^{-y^2\phi(0)}$, ce qui nous donne par convergence dominée

$$\lim_{n\to+\infty}\int_{\mathbb{R}} f_n(y)\,dy = \int_{\mathbb{R}} e^{-y^2/2}\,dy = \sqrt{2\pi}.$$

D'où le résultat $\Gamma(n+1) \sim F(n) \sim \sqrt{2\pi}e^{-n}n^{n+1/2}$.

Solution 71 1. Soit $z \in \mathbb{R}$. On a $(S \cap H^-)_z = S_z \cap H_z^-$, et on voit facilement que $S_z = \{(x,y) \in \mathbb{R}^2; x^2 + y^2 \le 4 - z^2\}$ tandis que $H_z = \{(x,y) \in \mathbb{R}^2; 3z \le x^2 + y^2\}$. Pour $z < 0$, on a pour tout $(x,y) \in \mathbb{R}^2$, $3z \le 0 \le x^2 + y^2$, donc $H_z = \mathbb{R}^2$, ce qui entraîne $(S \cap H^-)_z = S_z \cap H_z^- = S_z$. Soit $z \ge 0$. On a

$$S_z \cap H_z^- = \{(x,y) \in \mathbb{R}^2; 3z \le x^2 + y^2 \le 4 - z^2\}.$$

Si $3z > 4 - z^2$, cet ensemble est vide, sinon $S_z \cap H_z^-$ est la couronne située entre les cercles centrés à l'origine de rayons respectifs $\sqrt{3z}$ et

$\sqrt{4-z^2}$. Or $z^2-3z-4=(z-1)(z+4)$ est du signe de $z-1$. L'intersection est vide si $z>1$, sinon c'est une couronne non vide, éventuellement réduite à un point.

2. Avec le théorème de Tonelli, on a

$$\begin{aligned}&\lambda^{\otimes 3}(S\cap H^-)\\&=\int_{\mathbb{R}}\lambda^2((S\cap H^-)_z)\,d\lambda(z)\\&=\int_{[-\infty,0[}\lambda^2(S_z)\,d\lambda(z)+\int_{[0,1]}\lambda^2(C(\sqrt{3z},\sqrt{4-z^2}))\,d\lambda(z)\\&=\lambda^{\otimes 3}(S^-)+\int_{[0,1]}\lambda^2(C(\sqrt{3z},\sqrt{4-z^2}))\,d\lambda(z),\end{aligned}$$

où $S^-=\{(x,y,z)\in S; z<0\}$. S^- est une demi-boule, donc

$$\lambda^{\otimes 3}(S^-)=\frac{1}{2}\lambda^{\otimes 3}(S)=\frac{1}{2}\frac{4}{3}\pi.2^3=\frac{16}{3}\pi.$$

D'autre part $C(\sqrt{3z},\sqrt{4-z^2})=B(0,\sqrt{4-z^2})\backslash B(0,\sqrt{3z})$, donc

$$\lambda^{\otimes 3}(C(\sqrt{3z},\sqrt{4-z^2}))=\pi(4-z^2)-\pi(3z)=\pi(4-z^2-3z),$$

ce qui donne

$$\int_{[0,1]}\lambda^2(C(\sqrt{3z},\sqrt{4-z^2}))\,d\lambda(z)=\pi\int_0^1(4-z^2-3z)\,dz=\pi(4-\frac{1}{3}-\frac{3}{2}),$$

d'où enfin $\lambda(S\cap H^-)=\frac{15}{2}\pi$.

3. Posons $s(x,y,z)=(-x,-y,z)$, $A=S\cap H^-$, et

$$v=(v_x,v_y,v_z)=\int_A(x,y,z)\,d\lambda^{\otimes 3}(x,y,z).$$

Par linéarité, et comme $s(A)=A$, on a

$$\begin{aligned}s(v)&=\int_A s(x,y,z)\,d\lambda^{\otimes 3}(x,y,z)=\int_{\mathbb{R}^3}\mathbb{1}_A(x,y,z)s(x,y,z)\,d\lambda^{\otimes 3}(x,y,z)\\&=\int_{\mathbb{R}^3}\mathbb{1}_A(s(x,y,z))s(x,y,z)\,d\lambda^{\otimes 3}(x,y,z)\\&=\int_{\mathbb{R}^3}\mathbb{1}_A(x',y',z')(x',y',z')\,d\lambda^{\otimes 3}(x',y',z')=v\end{aligned}$$

ce qui entraîne que le centre de gravité est sur l'axe de la symétrie s. Reste à calculer sa coordonnée sur l'axe des z, soit $\frac{v_z}{\lambda^{\otimes 3}(A)}$. Comme A est borné (inclus dans une boule), $(x,y,z)\mapsto z\mathbb{1}_A(x,y,z)$ est intégrable et on peut appliquer le théorème de Fubini :

$$\begin{aligned}\int_A z\, d\lambda^{\otimes 3}(x,y,z) &= \int_{\mathbb{R}^3} z\mathbb{1}_A(x,y,z)\, d\lambda^{\otimes 3}(x,y,z)\\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}^2} z\mathbb{1}_A(x,y,z)\, d\lambda^2(x,y)\right)\, d\lambda(z)\\ &= \int_{\mathbb{R}} z \left(\int_{\mathbb{R}^2} \mathbb{1}_A(x,y,z)\, d\lambda^2(x,y)\right)\, d\lambda(z)\end{aligned}$$

puis

$$\begin{aligned}&\int_A z\, d\lambda^{\otimes 3}(x,y,z) = \int_{\mathbb{R}} z\lambda^2(A_z)\, d\lambda(z)\\ &= \int_{-2}^0 z\lambda^2(B(0,\sqrt{4-z^2})) + \int_0^1 z\lambda^2(C(\sqrt{3z},\sqrt{4-z^2}))\\ &= \int_{-2}^0 \pi z(4-z^2)\, dz + \int_0^1 \pi z(4-z^2-3z)\, dz\\ &= \pi(-\int_0^2 (4z-z^3)\, dz + \int_0^1 (4z-z^3-3z^2)\, dz)\\ &= -4\pi + \frac{3}{4}\pi = -\frac{13}{4}\pi,\end{aligned}$$

donc l'altitude du centre de masse est finalement $\frac{-\frac{13}{4}\pi}{\frac{15}{2}\pi} = -\frac{13}{30}$.

Le centre de gravité d'un solide homogène borné A est le point

$$\frac{1}{\lambda(A)} \int_A (x,y,z)\, d\lambda^{\otimes 3}(x,y,z).$$

On a donc montré que le centre de gravité de $S \cap H^-$ est le point de coordonnées $(0,0,-\frac{13}{30})$.

4. Comme dans la question précédente, un argument de symétrie donne $\int_S (x,y,z)\, d\lambda^{\otimes 3}(x,y,z) = 0$. Or S est la réunion disjointe de $S \cap H^+$ et $S \cap H^-$, donc

$$\begin{aligned}\int_{S\cap H^+} (x,y,z)\, d\lambda^{\otimes 3}(x,y,z) &= -\int_{S\cap H^-} (x,y,z)\, d\lambda^{\otimes 3}(x,y,z)\\ &= \left(\frac{13}{4}\pi, 0, 0\right)\end{aligned}$$

et $\lambda^{\otimes 3}(S\cap H^+) = \lambda^{\otimes 3}(S) - \lambda^{\otimes 3}(S \cap H^-) = \frac{32}{3}\pi - \frac{15}{2}\pi = \frac{19}{6}\pi$, et donc finalement le centre de gravité de $S \cap H^+$ est

$$\frac{1}{\frac{19}{6}\pi}\left(\frac{13}{4}\pi, 0, 0\right) = \left(\frac{39}{38}, 0, 0\right).$$

Solution 72 1. On a $T(x,y,\theta) = \begin{pmatrix} x\cos\theta \\ y \\ x\sin\theta \end{pmatrix}$. Les composantes de T ont des dérivées partielles continues par rapport à chaque variable : on a

$$DT_{(x,y\theta)} = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -x\sin\theta & 0 & x\cos\theta \end{pmatrix}.$$

L'application T est donc de classe C^1.
Si $T(x,y,\theta) = T(x',y',\theta')$ il vient aisément $y = y'$ et $x^2 = x'^2$, donc $x = x'$, puis $(\cos\theta, \sin\theta) = (\cos\theta', \sin\theta')$, ce qui entraîne $\theta = \theta'$. Vu les hypothèses faites sur α et β, T est donc injective.
Comme $\det DT_{(x,y\theta)} = x \neq 0$, $DT_{(x,y\theta)}$ est inversible et T réalise bien un C^1-difféomorphisme de $]0,+\infty[\times\mathbb{R}\times]\alpha,\beta[$ sur son image.

2. En appliquant le théorème de changement de variable C^1, on a

$$\begin{aligned}\lambda^3(R) &= \int_{V\times]\alpha,\beta[} |\det DT_{(x,y\theta)}|\, d\lambda^3(x,y,\theta) \\ &= \int_{V\times]\alpha,\beta[} x\, d\lambda^3(x,y,\theta) \\ &= \int_V \left(\int_{]\alpha,\beta[} x\, d\lambda(\theta)\right)\, d\lambda^2(x,y) \\ &= \int_V x(\beta-\alpha)\, d\lambda^2(x,y),\end{aligned}$$

ce qui est le résultat voulu.

3. Le centre de gravité de $V\times\{0\}$ est le point g_0 de coordonnées

$$\left(\frac{1}{\lambda^2(V)}\int_V x\, d\lambda^2(x,y), \frac{1}{\lambda^2(V)}\int_V y\, d\lambda^2(x,y), 0\right).$$

La rotation M_θ l'envoie sur

$$g_\theta = M_\theta g_0 = \frac{1}{\lambda^2(V)}\left(\cos\theta\int_V x\, d\lambda^2(x,y), \int_V y\, d\lambda^2(x,y), \sin\theta\int_V x\, dx\right).$$

Ainsi, g_θ décrit, dans le plan $y = \frac{\int_V y\, d\lambda^2(x,y)}{\lambda^2(V)}$, un arc de cercle de rayon $r_G(V) = \frac{\int_V x\, d\lambda^2(x,y)}{\lambda^2(V)}$ et d'angle $\beta-\alpha$: la longueur de la courbe est $L(V) = (\beta-\alpha)r_G(V) = \frac{(\beta-\alpha)\int_V x\, d\lambda^2(x,y)}{\lambda^2(V)}$. On a donc

$$L(V)\times\lambda^2(V) = (\beta-\alpha)\int_V x\, d\lambda^2(x,y) = \lambda^3(R)$$

avec la question précédente. On a ainsi démontré le théorème de Guldin.

Prenons $V = \{(x,y) \in]0,+\infty^2[; x^2+y^2 \leq R^2\}$. Par symétrie, son centre de symétrie g_0 est de la forme (r,r). Lorsque θ varie de 0 à 2π, g_θ décrit un cercle de longueur $L = 2\pi r$. Le volume de révolution engendré est une demi-boule privée d'un demi-plan fermé. Ce demi-plan est de mesure nulle, donc $\lambda^3(R) = \frac{1}{2}\frac{4}{3}\pi R^3$. Comme $\lambda^2(V) = \pi R^2/4$, on a

$$r = \frac{L}{2\pi} = \frac{\lambda^3(R)}{2\pi\lambda^2(V)} = \frac{\frac{1}{2}\frac{4}{3}\pi R^3}{2\pi\pi R^2/4} = \frac{4R}{3\pi}.$$

4. Par linéarité de l'intégrale, l'image du centre de gravité d'un volume par une similitude est le centre de gravité de l'image. En particulier, comme $T(V \times [0,2\pi])$ est laissé stable par toutes les rotations d'axe y, il en est de même de son centre de gravité, qui est donc sur l'axe des y. Bien sûr, il a le même centre de gravité que $T(V \times]0,2\pi])$: les deux ensembles ne diffèrent que d'un ensemble inclus dans un hyperplan, donc de mesure nulle. On calcule encore avec le changement de variable C^1 :

$$\begin{aligned}
&\int_{T(V\times]0,2\pi[)} y \, d\lambda(x,y,z) \\
&= \int_{V\times]0,2\pi[} Y(T(x,y,\theta)) |\det T_{(x,y,\theta)}| \, d\lambda(x,y,z) \\
&= \int_{V\times]0,2\pi[} yx \, d\lambda(x,y,\theta) \\
&= 2\pi \int_V yx \, d\lambda(x,y) \text{ avec Fubini.}
\end{aligned}$$

Comme

$$\begin{aligned}
\lambda^3(T(]0,+\infty[\times\mathbb{R}\times[0,2\pi]) &= \lambda^3(T(]0,+\infty[\times\mathbb{R}\times]0,2\pi[) \\
&= 2\pi \int_V x \, d\lambda(x),
\end{aligned}$$

le résultat demandé s'ensuit en faisant le quotient.

Solution 73 1. Pour $x \geq \delta$ et $t > 1$, on a

$$th(x) = h(x) + (t-1)h(x) \leq h(x) + (t-1)h(\delta),$$

d'où

$$\begin{aligned}
R_\delta(t) &= |\int_{\mathbb{R}_+} \mathbb{1}_{[\delta,+\infty]}(x) g(x) e^{th(x)} \, d\lambda(x)| \\
&\leq \int_{\mathbb{R}_+} |g(x)| e^{h(x)+(t-1)h(\delta)} \, d\lambda(x)| \\
&\leq Ce^{(t-1)h(\delta)} \text{ avec } C = \int_{\mathbb{R}_+} |g(x)| e^{h(x)} \, d\lambda(x)|
\end{aligned}$$

On a donc $\frac{R_\delta(t)}{e^{h(0)t}t^{-\frac{\alpha+1}{\beta}}} \leq \frac{C}{e}e^{-(h(0)-h(\delta))t}t^{\frac{\alpha+1}{\beta}}$ qui tend vers 0 par croissance comparée des fonctions puissance et exponentielle.

2. On a

$$\begin{aligned}\int_{\mathbb{R}_+} u^\alpha e^{-u^\beta}\, d\lambda(u) &= \frac{1}{\beta}\int_{\mathbb{R}_+} u^{\alpha-\beta+1}e^{-u^\beta}\beta u^{\beta-1}\, d\lambda(u)\\ &= \frac{1}{\beta}\int_{\mathbb{R}_+} (u^\beta)^{\frac{\alpha+1}{\beta}-1}e^{-u^\beta}\beta u^{\beta-1}\, d\lambda(u)\\ &= \frac{1}{\beta}\int_{\mathbb{R}_+} x^{\frac{\alpha+1}{\beta}-1}e^{-x}\, d\lambda(x) = \frac{1}{\beta}\Gamma(\frac{\alpha+1}{\beta}) = K_{\alpha,\beta}.\end{aligned}$$

3. Comme $\frac{g(x)}{Ax^\alpha}$ et $\frac{h(0)-h(x)}{cx^\beta}$ ont une limite 1 en 0, on a tout de suite l'existence de δ tel que $|q_\delta(x)| \leq 2$ et $r(x) \geq 1/2$ pour $x \in]0,\delta]$. On réécrit alors simplement

$$\begin{aligned}I_\delta(t) &= \int_{\mathbb{R}_+} \mathbb{1}_{[0,\delta]}(x)Ax^\alpha\frac{g(x)}{Ax^\alpha}e^{t(h(0)-cx^\beta r(x))}\, d\lambda(x)\\ &= Ae^{h(0)t}\int_{\mathbb{R}_+} x^\alpha q_\delta(x)e^{-ctx^\beta r(x)}\, d\lambda(x)\\ &= Ae^{h(0)t}(ct)^{-\frac{\alpha+1}{\beta}}\int_{\mathbb{R}_+} q_\delta\left(\frac{u}{(ct)^{1/\beta}}\right)u^\alpha e^{-r\left(\frac{u}{(ct)^{1/\beta}}\right)u^\beta}\, d\lambda(u)\end{aligned}$$

avec le changement de variable affine $\frac{u}{(ct)^{1/\beta}} = x$, et donc

$$\frac{I_\delta(t)}{Ae^{h(0)t}(ct)^{-\frac{\alpha+1}{\beta}}} = \int_{\mathbb{R}_+} q_\delta\left(\frac{u}{(ct)^{1/\beta}}\right)u^\alpha e^{-r\left(\frac{u}{(ct)^{1/\beta}}\right)u^\beta}\, d\lambda(u).$$

4. Comme

$$\left|q_\delta\left(\frac{u}{(ct)^{1/\beta}}\right)u^\alpha e^{-r\left(\frac{u}{(ct)^{1/\beta}}\right)u^\beta}\right| \leq 2u^\alpha e^{-u^\beta/2}$$

qui est intégrable. De plus, pour tout $u > 0$, on a

$$\lim_{t\to+\infty} q_\delta\left(\frac{u}{(ct)^{1/\beta}}\right)u^\alpha e^{-r\left(\frac{u}{(ct)^{1/\beta}}\right)u^\beta} = u^\alpha e^{-u^\beta},$$

et le théorème de convergence dominée nous donne

$$\lim_{t\to+\infty}\frac{I_\delta(t)}{Ae^{h(0)t}(ct)^{-\frac{\alpha+1}{\beta}}} = K_{\alpha,\beta}.$$

Comme $I(t) = I_\delta(t) + R_\delta(t)$, cela donne le résultat voulu avec la première question.

Solution 74 On note f une densité de ν par rapport à Lebesgue. Soit g une fonction positive mesurable. Le théorème d'intégration par rapport à une mesure image donne

$$\int_{\mathbb{R}^n} g(v)\ d(\mu*\nu)(v)=\int_{\mathbb{R}^n\times\mathbb{R}^n} g(x+y)\ d(\mu\otimes\nu)(x,y).$$

En appliquant deux fois le théorème de Tonelli et une fois l'invariance de la mesure de Lebesgue par translation, on a

$$\begin{aligned}\int_{\mathbb{R}^n} g(v)\ d(\mu*\nu)(v) &= \int_{\mathbb{R}^n\times\mathbb{R}^n} g(x+y)\ d(\mu\otimes\nu)(x,y)\\ &= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} g(x+y)d\nu(y)\right)\ d\mu(x)\\ &= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} g(x+y)f(y)d\lambda(y)\right)\ d\mu(x)\end{aligned}$$

puis

$$\begin{aligned}\int_{\mathbb{R}^n} g(v)\ d(\mu*\nu)(v) &= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} g(u)f(u-x)d\lambda(u)\right)\ d\mu(x)\\ &= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} g(u)f(u-x)d\mu(x)\right)\ d\lambda(u),\end{aligned}$$

ce qui montre que $u\mapsto\int_{\mathbb{R}^n} f(u-x)d\mu(x)$ est une densité de $\mu*\nu$ par rapport à la mesure de Lebesgue.

Solution 75 1. Comme F_n est $2\pi-$périodique et impaire, on peut se ramener au cas où $t\in[0,\pi]$. Par continuité de F_n, on peut encore se ramener au cas où $t\in]0,\pi[$. On a alors

$$\begin{aligned}F_n(t) &= \sum_{k=1}^n\int_0^t\cos(ku)\ du=\int_0^t\sum_{k=1}^n\cos(ku)\ du\\ &= \int_0^t\left(-\frac12+\frac12\sum_{k=-n}^n e^{iku}\right)\ du=\int_0^t\left(-\frac12+\frac{\sin(n+1/2)u}{\sin(u/2)}\right)\ du\\ &= \int_0^t\left(-\frac12+\frac{\sin(n+1/2)u}{2}\left(\frac{1}{\sin(u/2)}-\frac{1}{u/2}\right)\right)\ du\\ &\quad+\int_0^t\frac{\sin(n+1/2)u}{u}\ du\end{aligned}$$

La fonction $x\mapsto\frac{1}{\sin x}-\frac1x$ est continue sur $]0,\pi/2]$, équivalente à $\frac{x-\sin x}{x\sin x}\sim\frac{x^3/6}{x^2}=\frac{x}{6}$ en zéro : elle se prolonge par continuité sur $[0,\pi/2]$ et est donc bornée par une constante K. La première intégrale est donc de norme inférieure à $\frac{K+1}{2}t\le\frac{(K+1)\pi}{2}$. La deuxième vaut $\int_0^{(n+1/2)t}\frac{\sin u}{u}\ du$. Or la fonction $x\mapsto\int_0^x\frac{\sin u}{u}\ du$ est continue sur $\mathbb{R}_+$, avec une limite finie quand $x\to+\infty$ (c'est ce que dit la convergence de l'intégrale de Dirichlet, voir par exemple l'exercice 64) : elle est donc uniformément bornée par une constante C, et on a ainsi $|F_n(t)| \leq M = \frac{(K+1)\pi}{2} + C$ pour tout n et tout t.

2. Une transformation d'Abel donne

$$\begin{aligned}\sum_{k=n}^{p} \varepsilon_k \frac{\sin(kt)}{k} &= \sum_{k=n}^{p} \varepsilon_k (F_k(t) - F_{k-1}(t)) \\ &= \sum_{k=n}^{p} \varepsilon_k F_k(t) - \sum_{k=n-1}^{p-1} \varepsilon_{k+1} F_k(t) \\ &= \varepsilon_p F_p(t) - \varepsilon_n F_{n-1}(t) + \sum_{k=n}^{p-1} (\varepsilon_k - \varepsilon_{k+1}) F_k(t).\end{aligned}$$

D'où comme la suite $(\varepsilon_n)_n$ décroît vers 0, elle est positive et donc

$$\begin{aligned}|\sum_{k=n}^{p} \varepsilon_k \frac{\sin(kt)}{k}| &\leq |\varepsilon_p F_p(t)| + |\varepsilon_n F_{n-1}(t)| + \sum_{k=n}^{p-1} |\varepsilon_k - \varepsilon_{k+1}|.|F_k(t)| \\ &\leq M|\varepsilon_p| + M|\varepsilon_n| + M \sum_{k=n}^{p-1} |\varepsilon_k - \varepsilon_{k+1}| \\ &= M(\varepsilon_p + \varepsilon_n + \sum_{k=n}^{p-1} (\varepsilon_k - \varepsilon_{k+1})) = 2M\varepsilon_n.\end{aligned}$$

Comme ε_n est de limite nulle, il s'ensuit que la série de fonctions considérée satisfait un critère de Cauchy uniforme, et donc elle converge uniformément sur $\mathbb{R}$.

3. Il suffit de prendre $\varepsilon_k = \frac{1}{1+\log k}$.

D.5 Exercices sur les lois

Solution 97 X_1 et X_2 prennent des valeurs dans $\{-1, 1\}$, il en est donc de même du produit $X_3 = X_1 X_2$.
On a $\mathbb{P}(X_3 = 1) = \mathbb{P}(X_1 = 1, X_2 = 1) + \mathbb{P}(X_1 = -1, X_2 = -1) = 1/2$. Considérons $\varepsilon = \pm 1$ et $\varepsilon' = \pm 1$. On a alors :

$$\begin{aligned}\mathbb{P}(X_1 = \varepsilon, X_3 = \varepsilon') &= \mathbb{P}(X_1 = \varepsilon, X_2 = \varepsilon'/\varepsilon) \\ &= (1/2)^2 = \mathbb{P}(X_1 = \varepsilon)\mathbb{P}(X_3 = \varepsilon').\end{aligned}$$

Donc les variables aléatoires X_1 et X_3 sont indépendantes. Le même calcul donne aussi

$$\mathbb{P}(X_2 = \varepsilon, X_3 = \varepsilon') = (1/2)^2 = \mathbb{P}(X_2 = \varepsilon)\mathbb{P}(X_3 = \varepsilon').$$

Donc les variables aléatoires X_2 et X_3 sont indépendantes.

En revanche, les variables aléatoires X_1, X_2 et X_3 ne sont pas indépendantes, car on a $\mathbb{P}(X_1 = 1, X_2 = -1, X_3 = 1) = 0$, alors que le produit $\mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = -1)\mathbb{P}(X_3 = 1) = 1/8$ n'est pas nul.

Solution 98 On sait que les masses des points sont données par les sauts de discontinuité : $\mathbb{P}(X = i) = F(i) - F(i-)$. On obtient alors :

$$\begin{aligned}
\mathbb{P}(X=1) &= F(1) - F(1-) = \frac{1+1}{4} - \frac{1}{4} = \frac{1}{4},\\
\mathbb{P}(X=2) &= F(2) - F(2-) = \frac{11}{12} - \frac{3}{4} = \frac{1}{6},\\
\mathbb{P}(X=3) &= F(3) - F(3-) = 1 - \frac{11}{12} = \frac{1}{12}.
\end{aligned}$$

Pour calculer $\mathbb{P}\left(\frac{1}{2} < X < \frac{3}{2}\right)$, il suffit de remarquer que

$$\mathbb{P}\left(\frac{1}{2} < X < \frac{3}{2}\right) = \mathbb{P}\left(X < \frac{3}{2}\right) - \mathbb{P}\left(X \le \frac{1}{2}\right) = F\left(\frac{3}{2}-\right) - F\left(\frac{1}{2}\right) = \frac{5}{8} - \frac{1}{8} = \frac{1}{2}.$$

Solution 99 Pour montrer que les F_i sont des fonctions de répartition, il faut vérifier qu'elles sont continues à droites (c'est-à-dire, pour toute suite (t_n) décroissant vers t telle que $\lim_{n\to+\infty} F_i(t_n) = F_i(t)$), croissantes et satisfont $\lim_{n\to-\infty} F_i(n) = 0$ et $\lim_{n\to+\infty} F_i(n) = 1$. Cela se fait sans difficulté dans le cas de F_2. F_1 demande un peu plus de travail. D'abord, il est clair que F_2 est croissante (somme de fonctions croissantes), que $F_2(x) = 0$ pour $x \le 0$ (toutes les indicatrices de la somme sont nulles) et $F_2(x) = 1$ pour $x \ge 1$ (toutes les indicatrices de la somme sont nulles). Reste à montrer que F_2 est continue à droite. Si on pose $S_n(x) = \sum_{i=1}^n 2^{-i} \mathbb{1}_{[\frac{1}{i},\infty[}(x)$, S_n est continue à droite, comme somme de fonctions continues à droites. De plus, pour tout x réel, on a

$$0 \le F_2(x) - S_n(x) = \sum_{i>n} 2^{-i} \mathbb{1}_{[\frac{1}{i},\infty[}(x) \le \sum_{i>n} 2^{-i} = 2^{-n},$$

ce qui montre que la suite de fonctions $S_n(x)$ converge uniformément vers $F_2(x)$. Il s'ensuit que F_2 est continue à droite, comme limite uniforme de fonctions continues à droite. Calculons maintenant les probabilités voulues, en remarquant que la fonction F_2 est continue, alors que F_1 ne l'est pas.

— $[1, +\infty[$: $\mathbb{P}(X_1 \ge 1) = 1 - F_1(0.5) = 1 - 2 \times 2^{-2} = 0.5$;
  $\mathbb{P}(X_2 \ge 1) = 1 - F_2(1) = 1 - e^{-e^{-1}}$.

— $[1/10, +\infty[$: $\mathbb{P}(X_1 \ge 1/10) = 1 - F_1(1/11) = 1 - 2 \times 2^{-11} = 1 - 2^{-10}$;
  $\mathbb{P}(X_2 \ge 1/10) = 1 - F_2(1/10) = 1 - e^{-e^{-10}}$.

— $\{0\}$: $\mathbb{P}(X_1 = 0) = 0$; $\mathbb{P}(X_2 = 0) = 0$.

— $[0,1/2[$: $\mathbb{P}(0 \leq X_1 < 1/2) = F_1(1/3) = 2 \times 2^{-3} = 1/4$;
$\mathbb{P}(0 \leq X_2 < 1/2) = F_2(1/2) - F_2(0) = e^{-e^{-1/2}} - e^{-1}$.
— $]-\infty,0[$: $\mathbb{P}(X_1 < 0) = 0$; $\mathbb{P}(X_2 < 0) = e^{-1}$.
— $]0,+\infty[$: $\mathbb{P}(X_1 > 0) = 1 - F_1(0) = 1$; $\mathbb{P}(X_2 > 0) = 1 - F_2(0) = 1 - e^{-1}$.

Solution 100 Nous cherchons la fonction de répartition de la variable aléatoire $Y = \tan\Phi$, où Φ est un angle uniformément distribué dans l'intervalle $\left[0, \frac{\pi}{2}\right[$. On a alors :

$$F_Y(y) := \mathbb{P}(Y \leq y) = \mathbb{P}(\tan\Phi \leq y) = \mathbb{P}(\Phi \leq \arctan y) = F_\Phi(\arctan y)$$

avec

$$F_\Phi(x) = \begin{cases} 0 \text{ si } x < 0, \\ \frac{x+\pi/2}{\pi} \text{ si } 0 \leq x \leq \frac{\pi}{2}, \\ 1 \text{ si } x \geq \frac{\pi}{2} \end{cases}$$

On obtient donc $F_Y(y) = \frac{2\arctan y}{\pi}$ si $y \geq 0$ et vaut zéro sinon.
De manière tout à fait similaire, on obtient lorsque Φ est uniformément distribué dans $]-\frac{\pi}{2}, \frac{\pi}{2}[$:

$$F_\Phi(x) = \begin{cases} 0 \text{ si } x < -\frac{\pi}{2}, \\ \frac{2x}{\pi} \text{ si } -\frac{\pi}{2} \leq x \leq \frac{\pi}{2}, \\ 1 \text{ si } x \geq \frac{\pi}{2} \end{cases}$$

On trouve alors $F_Y(y) = \frac{\arctan y}{\pi} + \frac{1}{2}$ pour tout y.

Solution 101 On cherche la constante c telle que cx^{-3} soit une densité. On veut trouver c telle que $\int_1^{+\infty} cx^{-3}\ dx = 1$ et $cx^{-3} \geq 0$ pour tout $x \geq 1$.
1) On se place sur l'intervalle $[1,+\infty[$. Or par un calcul direct, on obtient $\int_1^{+\infty} cx^{-3}\ dx = \frac{c}{2}$. Il faut donc que $c = 2$.
2) On se place sur l'intervalle $[-2,-1]$. On obtient de manière tout à fait similaire : $c = -\frac{8}{3}$.

Solution 102 Comme X est p.s. supérieure à 1, on a
$Y = \frac{1}{2}(|X-1| + |X|) = X - \frac{1}{2}$. Ainsi Y suit la loi uniforme sur $[\frac{1}{2}, \frac{5}{2}]$.

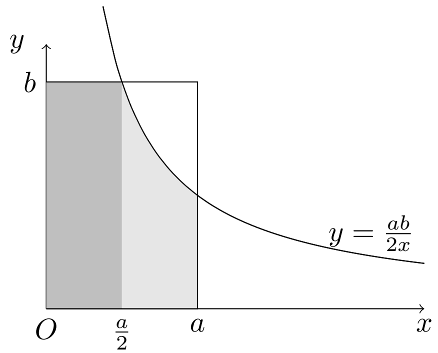
Solution 103 Le discriminant de l'équation du second degré
$4x^2 + 4xY + Y + 4 = 0$ est $\Delta = 16(Y^2 - Y - 4)$. On cherche les valeurs de Y pour lesquelles $\Delta \geq 0$.
L'équation $Y^2 - Y - 4 = 0$ admet deux racines : $y_1 = \frac{1-\sqrt{17}}{2}$ et $y_2 = \frac{1+\sqrt{17}}{2}$.
$Y^2 - Y - 4$ est de signe positif sauf entre les deux racines précédentes. On obtient alors :

$$\mathbb{P}(\text{racines réelles}) = 1 - \mathbb{P}(y_1 \leq Y \leq y_2) = 1 - \frac{\lambda([0,5] \cap [y_1,y_2])}{\lambda([0,5])} = \frac{9-\sqrt{17}}{10}.$$

Solution 104 X suit une loi uniforme sur $[0, a]$, où $a = AB$ et Y une loi uniforme sur $[0, b]$, où $b = BC$. Notons α l'angle formé par les droites (AB) et (BC). On a alors : $S_{ABC} = \frac{1}{2}ab\sin(\alpha)$ et $S_{XBY} = \frac{1}{2}XY\sin(\alpha)$. Le couple (X, Y) est uniformément réparti sur le rectangle $[0, a] \times [0, b]$.
On cherche donc

$$\begin{aligned}\mathbb{P}(S_{ABC} > 2S_{XBY}) &= \mathbb{P}(ab\sin(\alpha) > 2XY\sin(\alpha)) \\ &= \mathbb{P}(ab > 2XY) = \mathbb{P}\left(\frac{ab}{2X} > Y\right).\end{aligned}$$

Cela revient à calculer la fraction de l'aire en-dessous de l'hyperbole dans le rectangle de côtés a et b (zone gris clair du dessin ci-dessous).



Or la fraction de l'aire du rectangle gris foncé, noté S_1, vaut $\frac{ab}{2}\frac{1}{ab} = 1/2$, où $ab/2$ est l'aire du rectangle S_1 et ab celle du rectangle initial. De plus la fraction de l'aire du domaine gris clair vaut $\int_{a/2}^{a} \frac{1}{ab}\frac{ab}{2x}\,dx = \frac{1}{2}\log 2$.
On obtient donc finalement : $\mathbb{P}(S_{ABC} > 2S_{XBY}) = \frac{1}{2} + \frac{\log 2}{2}$.

Solution 105 Nous nous intéressons à une variable aléatoire discrète, donnant le nombre de succès dans une série de n expériences. On sait que la probabilité d'avoir un succès est p, où p est compris strictement entre 0 et 1. X suit une loi binomiale de paramètres n et p, soit pour tout entier $0 \leq k \leq n$

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = p_k.$$

Pour trouver la valeur de X la plus probable, il suffit alors de regarder la décroissance de p_k. Or $\frac{p_k}{p_{k+1}} = \frac{1-p}{p}\ \frac{k+1}{n-k} \geq 1$ si $k \geq p(n+1) - 1$. Comme X est à valeurs entières, la valeur la plus probable de X est $\lfloor p(n+1) \rfloor - 1$, où $\lfloor . \rfloor$ désigne la partie entière.

Solution 106 1. On remarque que $(*)$ est équivalente à

$$\mathbb{P}(X > s + t \text{ et } X > t) = \mathbb{P}(X > t) \cdot \mathbb{P}(X > s) > 0.$$

Cela revient à dire qu'en conditionnant par $\{X > s\}$, la variable aléatoire suit la même loi que si on partait de 0. C'est la loi de “non vieillissement”.

2. Nous allons démontrer qu'une variable aléatoire suivant une loi exponentielle de paramètre λ vérifie l'égalité $(*)$. On a en effet
$$\begin{aligned}\mathbb{P}(X > s+t \text{ et } X > t) &= \mathbb{P}(X > s+t) = e^{-\lambda(s+t)} = e^{-\lambda s}e^{-\lambda t}\\ &= \mathbb{P}(X > s)\mathbb{P}(X > t).\end{aligned}$$
Donc X est sans mémoire.

3. Soit $g : \mathbb{R}_+ \to]0,1]$, définie par $g(t) = \mathbb{P}(X > t)$. g est décroissante et $g(t+s) = g(t)g(s)$, ce qui entraîne en particulier que $g(0) = 1$. On veut d'abord déterminer g sur les rationnels. Remarquons que $g(1)$ est indéterminée, on sait juste que $0 < g(1) \leq 1$.

   Soient $t \geq 0$ et $k \in \mathbb{N}$. Par une récurrence facile sur k, on montre que
$$g(kt) = g(t + \cdots + t) = g(t)^k.$$
   Soit $n \in \mathbb{N}^*$. On a grâce à l'égalité ci-dessus $g(1) = g\left(n\frac{1}{n}\right) = g\left(\frac{1}{n}\right)^n$, d'où $g\left(\frac{1}{n}\right) = g(1)^{1/n}$.
   Cela donne donc que si $r = \frac{p}{q} \in \mathbb{Q}_+$, alors
$$g(r) = g(1)^{\frac{p}{q}} = g(1)^r.$$
   Puisque g a une limite nulle en l'infini, $g(1) < 1$. Soient $x \geq 0$ quelconque et (r_n) une suite de rationnels tendant vers x par valeurs supérieures. Comme $g(x) = 1 - \mathbb{P}(X \leq x)$, g est continue à droite. On a donc $g(x) = \lim_{n\to+\infty} g(r_n) = \lim_{n\to+\infty} g(1)^{r_n} = g(1)^x$. Cela implique alors $g(x) = e^{-\lambda x}$ où $\lambda = -\log(g(1))$ (car $0 < g(1) < 1$).

Solution 107 La variable $X - Y$ prend ses valeurs dans $\mathbb{Z}$. La loi de $X - Y$ est la loi image de $\mathbb{P}_{(X,Y)}$ par l'application $(a,b) \mapsto a - b$, tandis que la loi de $Y - X$ est la loi image de $\mathbb{P}_{(Y,X)}$ par l'application $(a,b) \mapsto a - b$. Comme $\mathbb{P}_{(X,Y)} = \mathcal{G}(1/2) \otimes \mathcal{G}(1/2) = \mathbb{P}_{(Y,X)}$, il s'ensuit que $X - Y$ et $Y - X$ ont même loi : la loi de $X - Y$ est symétrique. Prenons donc $\ell \in \mathbb{N}$. On a, avec le principe de partition :
$$\begin{aligned}\mathbb{P}(X - Y = \ell) &= \sum_{k\geq 1} \mathbb{P}(X = k, X - Y = \ell) = \sum_{k\geq 1} \mathbb{P}(X = k, Y = \ell + k)\\ &= \sum_{k\geq 1} \mathbb{P}(X = k)\mathbb{P}(Y = \ell + k) \text{ par indépendance}\\ &= \sum_{k\geq 1} (1/2)^k(1/2)^{k+\ell} = (1/2)^\ell \sum_{k\geq 1} (1/4)^k = \frac{1}{3}(1/2)^k,\end{aligned}$$

et par symétrie $\mathbb{P}(X - Y = k) = \frac{1}{3}(1/2)^{|k|}$ pour tout $k \in \mathbb{Z}$. Par ailleurs, XZ est une variable aléatoire à valeurs dans $\mathbb{Z}$. Comme $X \neq 0$, on trouve que $\mathbb{P}(XZ = 0) = \mathbb{P}(Z = 0) = \frac{1}{3}$. Pour $k \in \mathbb{N}^*$, on a

$$\mathbb{P}(XZ = k) = \mathbb{P}(X = k, Z = 1) = \mathbb{P}(X = k)\mathbb{P}(Z = 1) = (1/2)^k.(1/3),$$

et de même

$$\mathbb{P}(XZ = -k) = \mathbb{P}(X = k, Z = -1) = \mathbb{P}(X = k)\mathbb{P}(Z = -1) = (1/2)^k.(1/3).$$

On a donc pour tout $k \in \mathbb{Z}$, $\mathbb{P}(XZ = k) = \mathbb{P}(X - Y = k)$, ce qui montre que $X - Y$ et XZ ont même loi.

Solution 108 1. X et Y admettent la densité $f_X(t) = f_Y(t) = \frac{1}{2}\mathbb{1}_{[-1,1]}(t)$. Comme X et Y sont indépendantes, on a alors $f_{(X,Y)}(x, y) = f_X(x)f_Y(y) = \frac{1}{4}\mathbb{1}_{[-1,1]\times[-1,1]}(x, y)$, ce qui montre que (X, Y) suit la loi uniforme sur le carré $C = [-1, 1] \times [-1, 1]$

2. On a $\{1/2 < X^2 + Y^2 < 1\} = \{(X, Y) \in D\}$, où D est la couronne $D = \{(x, y) \in \mathbb{R}^2 : 1/2 < x^2 + y^2 < 1\}$. Comme (X, Y) suit la loi uniforme sur C, on a

$$\mathbb{P}((X, Y) \in D) = \frac{\lambda^2(C \cap D)}{\lambda^2(C)} = \frac{\lambda^2(D)}{\lambda^2(C)} = \frac{\pi(1 - 1/2)}{4} = \frac{\pi}{8}.$$

Solution 109 Les fonction $(x, y) \mapsto x^2$, $(x, y) \mapsto y$ et $(x, y) \mapsto x^2 + y$ sont positives sur $C = [0, 1]^2$. On peut appliquer le théorème de Tonelli.

$$\begin{aligned}\int_C (x^2 + y)\, d\lambda^2(x, y) &= \int_C x^2\, d\lambda^2(x, y) + \int_C y\, d\lambda^2(x, y) \\ &= \int_{[0,1]} x^2\, d\lambda(x) + \int_{[0,1]} y\, d\lambda(y) = \frac{4}{3}.\end{aligned}$$

La fonction $(x, y) \mapsto \frac{3}{4}(x^2 + y)\mathbb{1}_{[0,1]^2}(x, y)$ est mesurable positive, d'intégrale un : c'est la densité d'un vecteur aléatoire sur $\mathbb{R}^2$. On a maintenant

$$\begin{aligned}\mathbb{P}(X > Y) &= \int_{\mathbb{R}^2} \mathbb{1}_{\{x>y\}}\, d\mathbb{P}_{(X,Y)}(x, y) \\ &= \int_{\mathbb{R}^2} \mathbb{1}_{\{x>y\}} \frac{3}{4}(x^2 + y)\mathbb{1}_{[0,1]^2}(x, y)\, d\lambda^2(x, y) \\ &= \int_{[0,1]^2} \mathbb{1}_{\{x>y\}} \frac{3}{4}(x^2 + y)\, d\lambda^2(x, y)\end{aligned}$$

Maintenant, avec Tonelli

$$\begin{aligned}\mathbb{P}(X>Y) &= \frac{3}{4}\int_{[0,1]}\left(\int_{[0,1]}\mathbb{1}_{\{x>y\}}(x^2+y)\,d\lambda(y)\right)\,d\lambda(x)\\ &= \frac{3}{4}\int_{[0,1]}\left(\int_0^x x^2+y\,dy\right)\,d\lambda(x)\\ &= \frac{3}{4}\int_{[0,1]} x^3+\frac{x^2}{2}\,d\lambda(x) = \frac{3}{4}(\frac{1}{4}+\frac{1}{6}) = \frac{15}{48}.\end{aligned}$$

Solution 110 Comme les variables $X_1,\dots,X_n$ sont indépendantes, on a

$$\mathbb{P}(T>t) = \mathbb{P}(\bigcap_{k=1}^{n}\{X_k>t\}) = \prod_{k=1}^{n}\mathbb{P}(X_k>t).$$

Comme X_k suit la loi exponentielle de paramètre λ_k, on a $\mathbb{P}(X_k>t)=1$ pour $t<0$ et $\mathbb{P}(X_k>t)=e^{-\lambda_k t}$ sinon. Finalement

$$\mathbb{P}(T>t) = \begin{cases}\exp(-(\lambda_1+\dots+\lambda_n)t) & \text{pour } t\ge 0\\ 1 & \text{pour } t<0\end{cases}.$$

Ainsi, T a même fonction de queue (et donc même fonction de répartition) qu'une variable aléatoire de la loi exponentielle de paramètre $\lambda_1+\dots+\lambda_n$. Comme la fonction de répartition caractérise la loi, le résultat voulu s'ensuit.

Solution 111

1. Pour $a\in\mathbb{R}^d$ et $r>0$, on note $B_F(a,r)$ la boule fermée de centre a et de rayon r, $B_O(a,r)$ la boule ouverte de centre a et de rayon r. D'après le théorème de continuité séquentielle croissante, $\lim_{n\to+\infty}B_F(0,n)=1$, donc il existe $n\ge 1$ tel que $\mu(B_F(0,n))>0$. Posons $K=B_F(0,n)$. Montrons qu'il existe $a\in K$ avec $\mu(\{a\})=1$. On raisonne par l'absurde et on suppose que pour tout $a\in K$, $\mu(\{a\})<1$. D'après le théorème de continuité séquentielle décroissante, il existe $n_a\in\mathbb{N}^*$ tel que $\mu(B_O(a,1/n_a))<1$, d'où $\mu(B_O(a,1/n_a))=0$ puisque $\mu(B_O(a,r_a))\in\{0,1\}$. Les boules $(B_O(a,1/n_a))_{a\in K}$ forment un recouvrement de K par des ouverts. Comme K est compact, on peut en extraire un recouvrement fini : il existe B fini avec $B\subset K$ et

   $$K\subset\cup_{a\in B}B_O(a,1/n_a).$$

   Par sous-additivité finie,

   $$\mu(K)\le\sum_{a\in B}\mu(B_O(a,1/n_a))=\sum_{a\in B}0=0.$$

   On a une contradiction.

2. On commence par le cas où (X_n) prend ses valeurs dans $[-\pi/2,\pi/2]$. On pose $\mu=\mathbb{P}_X$. Soit $A\in\mathcal{B}(\mathbb{R})$. Pour tout $n\ge 1$, l'événement $\{X\in A\}$ est dans la tribu $\mathcal{Q}_n=\sigma(X_k,k\ge n)$ puisque X est la limite

supérieure des variables indexées à partir de n, et on sait que la limite supérieure d'une famille d'applications mesurables est mesurable. Par suite $\{X \in A\} \in \cap_{n\ge 1}\mathcal{Q}_n$, et d'après la loi 0–1 de Kolmogorov, $\mathbb{P}(X \in A) \in \{0,1\}$, soit $\mu(A) \in \{0,1\}$. Comme μ est une mesure sur $\mathbb{R}$, la question précédente montre qu'il existe $a \in \mathbb{R}$ tel que $\mu(\{a\}) = 1$, soit $\mathbb{P}(X = a) = 1$. Comme X prend ses valeurs dans $[-\pi/2, \pi/2]$, $a \in [-\pi/2, \pi/2]$. Dans le cas général, notons $\psi(x) = \arctan x$ pour x réel. On a $\psi(+\infty) = \pi/2$ et $\psi(-\infty) = -\pi/2$. Posons $Y_n = \psi(Y_n)$ et $Y = \overline{\lim\limits_{n\to+\infty}}\ Y_n$. On en déduit l'existence de $b \in [-\pi/2, \pi/2]$ tel que $\mathbb{P}(Y = b) = 1$. Comme $X = \psi^{-1}(Y)$, on a enfin $\mathbb{P}(X = \psi^{-1}(b)) = 1$.

Solution 112 Pour a réel, posons $F(a) = \mathbb{P}(X \le a)$. On a "$\mathbb{P}(X \le a) > 0$ et $\mathbb{P}(X > a) > 0$" si et seulement si $F(a) \in]0,1[$. Comme F est croissante, avec une limite 0 en $-\infty$ et une limite 1 en $+\infty$, si l'on suppose que $F(a) \in \{0,1\}$ pour tout a, alors il existe c tel que $F(x) = 0$ pour $x < c$ et $F(x) = 1$ pour $x > c$. Ainsi, pour tout $n \ge 1$,

$$\mathbb{P}(X \in [c - 1/n, c + 1/n[) = F(c + 1/n) - F(c - 1/n) = 1 - 0 = 1.$$

D'après le théorème de continuité séquentielle décroissante, on obtient que $\mathbb{P}(X = c) = 1$ et X est constante. Par contraposée, si X n'est pas constante, il existe un réel a tel que $F(a) \in]0,1[$, soit donc $\mathbb{P}(X \le a) > 0$ et $\mathbb{P}(X > a) > 0$.

On peut noter que réciproquement, si X est constante, on a pour tout a réel $\mathbb{P}(X \le a)\mathbb{P}(X > a) = 0$.

Maintenant, si X et Y ne sont pas constantes, on peut donc trouver a et b avec $\mathbb{P}(X \le a) > 0$, $\mathbb{P}(X > a) > 0$, $\mathbb{P}(Y \le b) > 0$, $\mathbb{P}(Y > b) > 0$. On a $\mathbb{P}(X + Y \le a + b) \ge \mathbb{P}(X \le a, Y \le b) = \mathbb{P}(X \le a)\mathbb{P}(Y \le b) > 0$ et $\mathbb{P}(X + Y > a + b) \ge \mathbb{P}(X > a, Y > b) = \mathbb{P}(X > a)\mathbb{P}(Y > b) > 0$. Ainsi $\mathbb{P}(X + Y \le a + b)\mathbb{P}(X + Y > a + b) > 0$ et $X + Y$ n'est pas constante.

Autre solution : Si je note $\Delta = \{(x,y) \in \mathbb{R}^2; x + y = \alpha\}$, le théorème de Tonelli me dit que $y \mapsto \mu(\{x \in \mathbb{R}; (x,y) \in \Delta\})$ est une application mesurable. Cette application est majorée par 1, d'intégrale 1, donc l'image réciproque de 1 (qui est bien mesurable) est de masse 1. En particulier, c'est un ensemble non vide et c'est fini.

Solution 113 1. Pour n, m entiers non nuls, notant $n \vee m$ le plus petit commun multiple, on a $(n\mathbb{N}^*) \cap (m\mathbb{N}^*) = (n \vee m)\mathbb{N}^*$. Ainsi les ensembles considérés forment un π-système. Pour montrer que deux probabilités qui coïncident sur ces parties sont égales, il suffit donc de montrer qu'elles engendrent $\mathcal{P}(\mathbb{N}^*)$. Comme chaque élément de $\mathcal{P}(\mathbb{N}^*)$ est réunion dénombrable de ses parties, il suffit de montrer qu'elles engendrent les singletons. Or, si $n = \prod_{i=1}^{+\infty} p_i^{\alpha_i}$, où (p_i) est la suite des nombres premiers et α_i le plus grand entier α tel que p_i^α divise n, on a

$$\{n\} = \bigcap_{i\geq 1} (p_i^{\alpha_i}\mathbb{N}^*)\backslash(p_i^{\alpha_i+1}\mathbb{N}^*),$$

ce qui donne le résultat voulu.

2. On a déjà vu que si X suit la loi Zêta de paramètre s, $\mathbb{P}(n|X) = n^{-s}$. On a donc, en utilisant ce calcul et l'indépendance

$$\begin{aligned}\mathbb{P}_{X\wedge X'}(n\mathbb{N}^*) &= \mathbb{P}(n|X\wedge X') = \mathbb{P}(n|X, n|X')\\ &= \mathbb{P}(n|X)\mathbb{P}(n|X') = n^{-s}n^{-t} = n^{-(s+t)}.\end{aligned}$$

On conclut avec la question précédente.

D.6 Exercices sur les espérances

Solution 129 On remarque que $\{X+\alpha\} = \{X+\{\alpha\}\}$, donc on peut se ramener au cas où $\alpha \in [0,1[$.
Méthode 1 : calcul de la fonction de répartition. Y est à valeurs dans $[0,1[$, donc $\mathbb{P}(Y\leq t) = 0$ pour $t \leq 0$ et $\mathbb{P}(Y\leq t) = 1$ pour $t\geq 1$. Soit donc $t\in]0,1[$.

$$\begin{aligned}\mathbb{P}(Y\leq t) &= \mathbb{P}(\{X+\alpha\}\in[0,t])\\ &= \mathbb{P}(X+\alpha\in[0,t]) + \mathbb{P}(X+\alpha\in[1,1+t])\\ &= \mathbb{P}(X\in[-\alpha,t-\alpha]) + \mathbb{P}(X\in[1-\alpha,1+t-\alpha])\\ &= \lambda([0,1]\cap[-\alpha,t-\alpha]) + \lambda([0,1]\cap[1-\alpha,1+t-\alpha]).\end{aligned}$$

Si $\alpha\geq t$, on a $[0,1]\cap[-\alpha,t-\alpha] = \varnothing$ et $[0,1]\cap[1-\alpha,1+t-\alpha] = [1-\alpha,1+t-\alpha]$, d'où une probabilité $0+t=t$. Si $\alpha<t$, on a $[0,1]\cap[-\alpha,t-\alpha] = [0,t-\alpha]$ et $[0,1]\cap[1-\alpha,1+t-\alpha] = [1-\alpha,1]$, d'où une probabilité $(t-\alpha)+\alpha = t$. Finalement, Y a la même fonction de répartition que X : les deux variables aléatoires ont la même loi.
Méthode 2 : fonction test. Soit f une fonction mesurable bornée. On a

$$\begin{aligned}\mathbb{E}(f(Y)) = \mathbb{E}(f(\{X+\alpha\})) &= \int_{\mathbb{R}} f(\{x+\alpha\})\, d\mathbb{P}_X(x)\\ &= \int_{[0,1[} f(\{x+\alpha\})\, d\lambda(x)\\ &= \int_{[0,1-\alpha[} f(x+\alpha)\, d\lambda(x) + \int_{[1-\alpha,1[} f(x+\alpha-1)\, d\lambda(x)\\ \mathbb{E}(f(Y)) &= \int_{[\alpha,1[} f(x)\, d\lambda(x) + \int_{[0,\alpha[} f(x)\, d\lambda(x)\\ &= \int_{[0,1[} f(x)\, d\lambda(x) = \mathbb{E}(f(X)),\end{aligned}$$

ce qui montre que X et Y ont même loi.

Solution 130 1. Considérons la fonction

$$\varphi : \mathbb{R} \to \mathbb{R}_+,\ x \mapsto \varphi(x) = \mathbb{E}[(X-x)^2].$$

On a $\varphi(x) = \mathbb{E}[X^2 - 2xX + x^2] = x^2 - 2x\mathbb{E}[X] + \mathbb{E}(X^2)$. C'est un polynôme du second degré de coefficient dominant 1, minimal au point $x = \mathbb{E}(X)$. Ainsi, $m = \mathbb{E}(X)$ minimise φ.

2. Notons F_X la fonction de répartition de X. L'ensemble des t tels que $F_X(t) \geq 1/2$ est un intervalle I non-vide (car F_X a une limite 1 en $+\infty$) et minoré (car F_X a une limite 0 en $-\infty$). Notons λ sa borne inférieure. Comme F_X est continue à droite, $F_X(\lambda) \geq 1/2$, soit $\mathbb{P}(X \leq \lambda) \geq 1/2$. Pour tout $n \geq 1$, $\lambda - 1/n$ n'est pas dans I, donc $\mathbb{P}(X \leq \lambda - 1/n) < 1/2$ et $\mathbb{P}(X > \lambda - 1/n) > 1/2$. On en déduit que $\mathbb{P}(X \geq \lambda) = \mathbb{P}\left(\bigcap_{n=1}^{+\infty} \{X > \lambda - 1/n\}\right) \geq 1/2$. Ainsi, λ est une médiane.

   Soit X suivant une loi de Bernoulli de paramètre $1/2$. Remarquons tout de suite que la médiane de X, notée λ, n'appartient ni à l'intervalle $]-\infty, 0[$ (car alors $\mathbb{P}(X \leq \lambda) = 0$), ni à l'intervalle $]1, +\infty[$ (car alors $\mathbb{P}(X \geq \lambda) = 0$).
   Soit $\lambda \in [0,1]$. On voit alors que $\mathbb{P}(X \geq \lambda) \geq \mathbb{P}(X = 1) = 1/2$ et $\mathbb{P}(X \leq \lambda) \geq \mathbb{P}(X = 0) = 1/2$ et donc tout $\lambda \in [0,1]$ est médiane de X.

   Soit X suivant une loi de Bernoulli de paramètre $p > 1/2$. Comme ci-dessus, on voit qu'une médiane ne peut être ni strictement négative, ni strictement supérieure à 1. Commençons par le cas $0 \leq \lambda < 1$. On a alors $\mathbb{P}(X \leq \lambda) = \mathbb{P}(X = 0) = 1 - p < 1/2$ et donc $\lambda < 1$ n'est pas une médiane.
   En revanche, si $\lambda = 1$, on a alors $\mathbb{P}(X \leq \lambda) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = 1$ et $\mathbb{P}(X \geq \lambda) = \mathbb{P}(X = 1) = p > 1/2$ et donc l'unique médiane de X est $\lambda = 1$.

   Bien sûr, lorsque X suit une loi de Bernoulli de paramètre $p < 1/2$, on trouve que l'unique médiane est $\lambda = 0$ et nous laissons le soin au lecteur de le démontrer.

Solution 131 La fonction $\Phi :]0,1] \to]0,1],\ x \mapsto 1/x - \lfloor 1/x \rfloor$ est inversible sur chaque intervalle $]1/(k+1), 1/k]$, où $k \in \mathbb{N}^*$ et son inverse est $\Psi_k :]0,1] \to]1/(k+1), 1/k],\ t \mapsto 1/(t+k)$. Ainsi pour toute fonction H mesurable bornée, la formule de changement de variable donne

$$\begin{aligned}\mathbb{E}(H(\Phi(X)) &= \int_0^1 H(\Phi(x))f(x)\,dx = \sum_{k\geq 1}\int_{1/(k+1)}^{1/k} H(\Phi(x))f(x)\,dx \\ &= \sum_{k\geq 1}\int_0^1 H(y)f(\Psi_k(y))\Psi_k'(y)\,dy \\ &= \frac{1}{\log 2}\sum_{k\geq 1}\int_0^1 H(y)\frac{1}{(y+k+1)(y+k)}\,dy.\end{aligned}$$

D'après le théorème de Fubini, on a

$$\mathbb{E}(H(\Phi(X)) = \frac{1}{\log 2}\int_0^1 H(y)\left(\sum_{k\geq 1}\frac{1}{(y+k+1)(y+k)}\right)\,dy.$$

Comme $\frac{1}{(y+k+1)(y+k)} = \frac{1}{y+k} - \frac{1}{y+k+1}$, la série est télescopique, de somme $\frac{1}{y+1}$.
Ainsi f apparaît comme la densité de la loi de $\{\frac{1}{X}\}$, qui a même loi que X.
Solution alternative : on pose $Y = \{\frac{1}{X}\}$ et on compare les fonctions de répartition de X et Y sur leur support commun $[0,1]$. Pour $t \in [0,1]$, on a aisément $F_X(t) = \frac{\log(1+t)}{\log 2}$, tandis que

$$\begin{aligned}F_Y(t) &= \sum_{k=1}^{+\infty}\mathbb{P}\left(Y\leq t, \lfloor\frac{1}{X}\rfloor = k\right) = \sum_{k=1}^{+\infty}\mathbb{P}\left(k\leq\frac{1}{X}\leq k+t\right) \\ &= \sum_{k=1}^{+\infty}\mathbb{P}\left(X\in[\frac{1}{k+t},\frac{1}{k}]\right) = \sum_{k=1}^{+\infty}\frac{1}{\log 2}\left(\log(1+\frac{1}{k}) - \log(1+\frac{1}{k+t})\right) \\ &= \frac{1}{\log 2}\sum_{k=1}^{+\infty}\log(1+k) - \log(k) + \log(k+t) - \log(k+t+1) \\ &= \sum_{k=1}^{+\infty} g_t(k) - g_t(k+1),\end{aligned}$$

avec $g_t(k) = \frac{\log(k+t)-\log(k)}{\log 2}$. On obtient finalement $F_Y(t) = g_t(1) = F_X(t)$.

Solution 132 On a

$$\mathbb{E}(Y) = \int_{\mathbb{R}}|4x^2-1|f_X(x)\,dx = \int_{-1}^1\frac{|4x^2-1|}{2}\,dx$$

$$= \int_{-1}^{-1/2}(2x^2-1/2)\,dx + \int_{-1/2}^{1/2}(1/2-2x^2)\,dx + \int_{1/2}^1(2x^2-1/2)\,dx = 1$$

et $\mathbb{E}(X+Y) = \mathbb{E}(X)+\mathbb{E}(Y) = \int_{-1}^1 x/2dx + 1 = 1$.
On trouve après un fastidieux calcul : $\mathbb{E}(Y+4X^2)^3 = 3/4 - 63/14$.

Solution 133 1. Comme les variables $(X_n)_{n\geq 1}$ sont indépendantes identiquement distribuées, on a

$$\mathbb{P}(\max(X_1,\dots,X_n)\le x)=\mathbb{P}(\bigcap_{k=1}^{n}\{X_k\le x\})=\mathbb{P}(X_1\le x)^n.$$

On connait la fonction de répartition de X_1, qui suit la loi uniforme sur $[0,1]$. On en déduit

$$\mathbb{P}(\max(X_1,\dots,X_n)\le x)=\begin{cases}0 & \text{pour } x<0\\ x^n & \text{pour } x\in[0,1]\\ 1 & \text{pour } x>1\end{cases}$$

Comme la fonction de répartition trouvée est continue et C^1 par morceaux, on obtient en dérivant la densité $f_{\max(X_i)}(x)=nx^{n-1}\mathbb{1}_{[0,1]}(x)$. Un calcul direct de l'espérance permet alors de trouver

$$\mathbb{E}(\max(X_1,\dots,X_n))=\int_0^1 x(nx^{n-1})\,dx=\frac{n}{n+1}.$$

2. On considère maintenant une suite $(Y_n)_n$ de variables indépendantes identiquement distribuées de loi exponentielle de paramètre 1. On a, pour $x>0$, $\mathbb{P}(\min(Y_1,\dots,Y_n)>x)=\mathbb{P}(Y_1>x)^n=e^{-nx}$, ce qui nous donne alors

$$\begin{aligned}\mathbb{E}(\min(Y_1,\dots,Y_n))&=\int_{\mathbb{R}_+}\mathbb{P}(\min(Y_1,\dots,Y_n)>t)\,d\lambda(t)\\&=\int_0^{+\infty}e^{-nt}\,dt=\frac{1}{n}.\end{aligned}$$

Solution 134 1. On va montrer par l'absurde le résultat plus fort : il existe une infinité de nombres premiers congrus à -1 modulo 6. Notons N le produit de tous les nombres premiers congrus à -1 modulo 6. Décomposons $6N-1$ en produit de facteurs premiers : tous les termes du produit sont congrus à 1 ou -1 modulo 6 ; il y en a au moins 1 congru à -1 modulo 6 (sinon $6N-1$ serait congru à 1 modulo 6, ce qui est absurde). C'est un des facteurs de N, donc il divise $6N$ et $6N-1$: contradiction.

2. Si un ensemble E d'un anneau est sans-somme et que x est inversible, alors $x.E$ est sans-somme. Cela se voit aisément par contraposée : si $x.E$ n'est pas sans somme, alors on peut trouver a,b,c dans E vérifiant $x.a=x.b+x.c$. En multipliant par x^{-1}, on obtient $a=b+c$, ce qui montre que E n'est pas sans somme. Ainsi $X.B_0$ est sans-somme, donc $A'\cap(X.B_0)$ aussi.

3. Soient a',b' dans B_0. Si a et b sont dans $[k+1,\dots,2k+1]$ des représentants respectivement de a' et b', on trouve l'inégalité suivante : $2k+2\le a+b\le 4k+2=(3k+2)+k=p+k$, donc

$a+b \in [2k+2,\dots,p-1] \cup [p,\dots,p+k]$, ce qui fait que $a+b \notin B_0$: B_0 est sans-somme.

4. On a

$$\begin{aligned}
\mathbb{E}[|B'|] &= \sum_{b\in\mathbb{F}_p} \mathbb{P}(b \in B') = \sum_{b\in\mathbb{F}_p} \mathbb{P}(b\in A', b \in X.B_0)\\
&= \sum_{b\in A'} \mathbb{P}(b \in X.B_0) = \sum_{b\in A'} \mathbb{P}(\exists y \in B_0; b = X.y)\\
&= \sum_{b\in A'} \mathbb{P}(\exists y \in B_0; X = b.y^{-1}) = \sum_{b\in A'} \mathbb{P}(X \in b.B_0^{-1})\\
&= \sum_{b\in A'} \frac{|B_0|}{p-1} = |A'|.\frac{k+1}{3k+1} > |A| \times \frac{1}{3}.
\end{aligned}$$

5. Par contraposée : si on avait $|B'| \le |A|/3$ presque sûrement, on aurait $\mathbb{E}[|B'|] \le |A| \times \frac{1}{3}$.
6. Si $\mathbb{P}(|B'| > |A|/3) > 0$, alors $\{\omega \in \Omega; |B'(\omega)| > |A|/3\}$ n'est pas vide. Pour un tel ω, $B(\omega) = \pi_A^{-1}(B'(\omega))$ est un ensemble sans-somme inclus dans A et de cardinal strictement supérieur à $|A|/3$.

Solution 135 On rappelle deux résultats à connaître :

— Formule de Poincaré (aussi appelée formule du crible) : Quels que soient les ensembles finis $A_1, A_2, \dots, A_n$, on a

$$|\bigcup_{i=1}^n A_i| = \sum_{B\in\mathcal{P}(\{1,\dots,n\})\setminus\varnothing} (-1)^{1+|B|} |\bigcap_{j\in B} A_j|.$$

Si les (A_k) sont interchangeables (c'est-à-dire le cardinal de l'intersection de i éléments quelconques distincts pris dans les A_k ne dépend que de i), on a

$$|\bigcup_{i=1}^n A_i| = \sum_{k=1}^n (-1)^{k-1}\binom{n}{k}|A_1 \cap A_2 \dots \cap A_k|.$$

— Le nombre de solutions dans $\mathbb{N}^n$ de $x_1 + \dots + x_n = p$ est $\binom{n+p-1}{n-1}$ (voir l'annexe des rappels de dénombrements).

On note A_0 les solutions de $x_1 + \dots + x_8 = 40$ et A_i les éléments de A_0 avec $x_i \ge 10$. Le nombre de solutions admissibles est donc $|A_0| - |\bigcup_i A_i|$, que l'on doit calculer. On a $|A_0| = \binom{40+8-1}{8-1}$ et le cardinal de l'union s'évalue à l'aide de la formule du crible.

Soit $(x_1,\dots,x_8) \in A_1 \cap A_2 \cap \dots \cap A_i$. On pose $y_k = x_k$ pour $k > i$ et $y_k = x_k - 10$ pour $k \le i$. On définit ainsi une bijection entre $A_1 \cap A_2 \cap \dots \cap A_i$ et les solutions de $y_1 + \dots + y_8 = 40 - 10i$ dans $\mathbb{N}^8$. Si $i > 4$, il n'y a pas de solution. Ainsi $|A_1 \cap A_2 \cap \dots \cap A_i| = \binom{40-10i+8-1}{8-1} = \binom{47-10i}{7}$, d'où finalement le résultat final

$$\sum_{i=0}^{4}(-1)^i\binom{47-10i}{7}\binom{8}{i} = 4303545.$$

Solution 136 1. On a

$$\mathbb{P}(|X-Y|\le 2) \le \mathbb{P}(|\lfloor X\rfloor - \lfloor Y\rfloor| \le 2) = \sum_{\ell=-2}^{2} \mathbb{P}(\lfloor X\rfloor - \lfloor Y\rfloor = \ell).$$

Pour $k\in\mathbb{Z}$, posons $p_k = \mathbb{P}(\lfloor X\rfloor = k) = \mathbb{P}(\lfloor Y\rfloor = k)$. On a pour tout $\ell\in\mathbb{Z}$

$$\begin{aligned}
\mathbb{P}(\lfloor X\rfloor - \lfloor Y\rfloor = \ell) &= \sum_{k\in\mathbb{Z}} \mathbb{P}(\lfloor X\rfloor - \lfloor Y\rfloor = \ell, \lfloor Y\rfloor = k) = \sum_{k\in\mathbb{Z}} p_k p_{k+\ell} \\
&\le \left(\sum_{k\in\mathbb{Z}} p_k^2\right)^{1/2} \left(\sum_{k\in\mathbb{Z}} p_{k+\ell}^2\right)^{1/2} \\
&= \sum_{k\in\mathbb{Z}} p_k^2 = \mathbb{P}(\lfloor X\rfloor = \lfloor Y\rfloor),
\end{aligned}$$

par application de l'inégalité de Cauchy–Schwarz. D'où

$$\mathbb{P}(|X-Y|\le 2) \le 5\mathbb{P}(\lfloor X\rfloor = \lfloor Y\rfloor) \le 5\mathbb{P}(|X-Y|\le 1).$$

2. Notons que dans cet exemple, le couple (X,Y) suit la loi uniforme sur $\{2,4,6,\dots,2N\}^2$. On est donc ramené à un dénombrement : on a

$$\mathbb{P}(|X-Y|\le 1) = \mathbb{P}(X=Y) = \frac{N}{N^2} = \frac{1}{N}$$

et

$$\mathbb{P}(|X-Y|\le 2) = \frac{N+2(N-1)}{N^2} = \frac{3}{N} - \frac{2}{N^2}.$$

Ainsi $\frac{3}{N} - \frac{2}{N^2} \le \frac{C}{N}$, soit $C \ge 3 - \frac{2}{N}$, et donc $C\ge 3$ en faisant tendre N vers l'infini.

Solution 137 1. Comme $s_1^{Z^1} s_2^{Z^2} \dots s_k^{Z^k}$ est une fonction de X, le théorème de transfert s'applique et

$$\mathbb{E}\left[s_1^{Z^1} s_2^{Z^2} \dots s_k^{Z^k}\right] = \sum_{i=1}^{k} p_i s_i.$$

2. Soit $Z_i^j = \mathbb{1}_{\{X_i=j\}}$, avec $i = 1, \dots, n$ et $j = 1, \dots, k$. Alors

$$\begin{aligned}
\mathbb{E}\left[s_1^{N^1} s_2^{N^2} \dots s_k^{N^k}\right] &= \mathbb{E}\left[s_1^{\sum_{i=1}^n Z_i^1} \dots s_k^{\sum_{i=1}^n Z_i^k} \right] \\
&= \mathbb{E}\left[\prod_{i=1}^n (s_1^{Z_i^1} \dots s_k^{Z_i^k}) \right] = \prod_{i=1}^n \mathbb{E}[s_1^{Z_i^1} \dots s_k^{Z_i^k}]
\end{aligned}$$

car les vecteurs $(Z_i)_{1 \le i \le n}$ sont indépendants. Ainsi avec la formule du multinôme

$$\mathbb{E}\left[s_1^{N^1} s_2^{N^2} \dots s_k^{N^k}\right] = \left(\sum_{i=1}^k p_i s_i \right)^n$$

$$= \sum \binom{n}{a_1, a_2, \dots, a_k} p_1^{a_1} p_k^{a_k} (s_1^{a_1} \dots s_k^{a_k}).$$

où la sommation s'effectue sur toutes les familles $(a_1, \dots, a_k)$ d'entiers naturels dont la somme est n. Par ailleurs, le théorème de transfert donne $\mathbb{E}\left[s_1^{N^1} s_2^{N^2} \dots s_k^{N^k}\right] = \sum \mathbb{P}(N^1 = a_1, \dots, N^k = a_k)(s_1^{a_1} \dots s_k^{a_k})$. Comme deux polynômes qui coïncident sur $\mathbb{R}^k$ sont identiques, cela permet d'identifier les coefficients.

Solution 138 Si X et Y sont indépendantes alors $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. Supposons que $X \in \{x_1, x_2\}$ et $Y \in \{y_1, y_2\}$. Posons $X' = \frac{X - x_1}{x_2 - x_1}$ et $Y' = \frac{Y - y_1}{y_2 - y_1}$ de sorte que $X', Y' \in \{0, 1\}$. Soient $p(a, b) = \mathbb{P}(X' = a, Y' = b)$, $p = \mathbb{P}(X' = 1)$ et $q = \mathbb{P}(Y' = 1)$. Par hypothèse, on a

$$p(1, 1) = \mathbb{E}(X'Y') = \mathbb{E}(X')\mathbb{E}(Y') = pq.$$

D'où $p(0, 1) = q - p(1, 1) = q - pq$, $p(1, 0) = p - qp$ et

$$p(0, 0) = 1 - p - p(0, 1) = 1 - p - q + pq = (1 - p)(1 - q).$$

Montrons que ce résultat est faux avec plus de deux valeurs. Soient X et Z des variables aléatoires indépendantes.

On suppose que $\mathbb{P}_X = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$ et que Z est une variable aléatoire discrète centrée, mais de distribution non symétrique (c'est-à-dire que $\mathbb{E}Z = 0$, mais qu'il existe a tel que $\mathbb{P}(Z = a) \neq \mathbb{P}(Z = -a)$). On peut prendre par exemple $S = \{-1, 0, 2\}$ avec $\mathbb{P}(Z = 1) = 2\mathbb{P}(Z = 2)$.

Soit $Y = XZ$. Alors $XY = X^2 Z = Z$ car $\mathbb{P}(X^2 = 1) = 1$ et donc

$$\mathbb{E}(XY) = \mathbb{E}(Z) = 0 = \mathbb{E}(X)\mathbb{E}(Y).$$

Par ailleurs,

$$\mathbb{P}(X = 1, Y = a) = \frac{1}{2}\mathbb{P}(Z = a)$$

et

$$\begin{aligned}\mathbb{P}(X=1)\mathbb{P}(Y=a) &= \frac{1}{2}\left(\mathbb{P}(Z=a,X=1)+\mathbb{P}(Z=-a,X=-1)\right)\\ &= \frac{1}{4}\left(\mathbb{P}(Z=a)+\mathbb{P}(Z=-a)\right) \neq \mathbb{P}(X=1,Y=a).\end{aligned}$$

Solution 139 On remarque que

$$S_n^2 = ((S_n - S_k) + S_k)^2 \geq S_k^2 + 2S_k(S_n - S_k)$$

d'où

$$\mathbb{E}(S_n^2) = \mathbb{E}\left(\sum_{k=1}^n \mathbb{1}_{A_k} S_n^2\right) \geq \mathbb{E}\left(\sum_{k=1}^n \mathbb{1}_{A_k}(S_k^2 + 2S_k(S_n - S_k))\right)$$

Cependant, la variable $\mathbb{1}_{A_k} S_k$ est $\sigma(X_1, \dots, X_k)$-mesurable tandis que $S_n - S_k$ est $\sigma(X_{k+1}, \dots, X_n)$-mesurable. Elles sont donc indépendantes et on a l'égalité $\mathbb{E}(\mathbb{1}_{A_k} S_k(S_n - S_k)) = \mathbb{E}(\mathbb{1}_{A_k} S_k)\mathbb{E}(S_n - S_k) = 0$. On en déduit

$$\mathbb{E}(S_n^2) \geq \mathbb{E}\left(\sum_{k=1}^n \mathbb{1}_{A_k} S_k^2\right) \geq \mathbb{E}\left(\sum_{k=1}^n \mathbb{1}_{A_k} x^2\right) = x^2 \sum_{k=1}^n \mathbb{P}(A_k).$$

Par ailleurs

$$\sum_{k=1}^n \mathbb{P}(A_k) = \mathbb{P}\left(\max_{1\leq k\leq n} |S_k| \geq x\right).$$

Solution 140 Soit $f(y) = \mathbb{E}(|X-y|)$. Soient x, y réels et $\theta \in [0,1]$. On a $X - (\theta x + (1-\theta)y) = \theta(X-x) + (1-\theta)(X-y)$, donc

$$|X - (\theta x + (1-\theta)y)| \leq \theta|X-x| + (1-\theta)|X-y|,$$

d'où en intégrant $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$. L'application f est donc convexe. Or, avec Tonelli, on a

$$\begin{aligned}\mathbb{E}|X-Y| &= \int_{\mathbb{R}^2} |x-y|\, d\mathbb{P}_{(X,Y)}(x,y) = \int_{\mathbb{R}^2} |x-y|\, d(\mathbb{P}_X \otimes \mathbb{P}_Y)(x,y)\\ &= \int_{\mathbb{R}}\left(\int_{\mathbb{R}} |x-y|\, d\mathbb{P}_X(x)\right)\, d\mathbb{P}_Y(y) = \int_{\mathbb{R}} f(y)\, d\mathbb{P}_Y(y)\\ &= \mathbb{E}[f(Y)].\end{aligned}$$

Mais, d'après l'inégalité de Jensen, on a

$$\mathbb{E}(f(Y)) \geq f(\mathbb{E}(Y)) = f(0) = \mathbb{E}(|X|).$$

On obtient bien l'inégalité voulue.

Solution 141 Comme $\min(X,Y)$ est à valeurs positives, on a l'écriture intégrale $\mathbb{E}(\min(X,Y)) = \int_{\mathbb{R}_+} \mathbb{P}(\min(X,Y) > t)\, d\lambda(t)$.
Or $\{\min(X,Y) > t\} = \{X > t, Y > t\}$, et comme X et Y sont indépendantes,

on a alors

$$\mathbb{P}(\min(X,Y) > t) = \mathbb{P}(X > t)\mathbb{P}(Y > t) = e^{-at}e^{-at} = e^{-2at},$$

ce qui nous donne $\mathbb{E}(\min(X,Y)) = \int_0^{+\infty} e^{-2at}\, dt = \frac{1}{2a}$.

D'autre part $\max(X,Y) + \min(X,Y) = X + Y$, donc, par linéarité,

$$\mathbb{E}(\max(X,Y)) = \mathbb{E}(X) + \mathbb{E}(Y) - \mathbb{E}(\min(X,Y)) = \frac{1}{a} + \frac{1}{a} - \frac{1}{2a} = \frac{3}{2a}.$$

Solution 142

1. En procédant comme dans l'exercice précédent, l'indépendance des X_i nous donne $\mathbb{P}(M \geq x) = \prod_{i=1}^{n} \mathbb{P}(X_i \geq x) = e^{-anx}$ pour $x \geq 0$ et vaut 1 sinon. La loi de M est donc une loi exponentielle de paramètre $an > 0$.

2. Remarquons tout d'abord que $N_t = \sum_{i=1}^{n} \mathbb{1}_{\{X_i \geq t\}}$. Comme les X_i sont indépendantes et de même loi, les variables de la somme sont des variables de Bernoulli indépendantes de paramètre $\mathbb{P}(X_1 \geq t)$. Pour $m \in \{0, \ldots, n\}$, N_t suit donc une loi binomiale $\mathcal{B}(n,p)$ avec $p = e^{-at}$. D'où l'espérance et la variance de N_t sont

$$\mathbb{E}(N_t) = np = ne^{-at} \text{ et } \operatorname{Var}(N_t) = np(1-p) = ne^{-at}(1 - e^{-at}).$$

Solution 143 Soient $X \sim \mathcal{U}([0,3])$ et $Y = \max\{2, X\}$. Si $y \leq 2$, on a alors $\{Y \leq y\} = \varnothing$ tandis que si $y \geq 2$, alors $\{Y \leq y\} = \{X \leq y\}$. On trouve ainsi la fonction de répartition de Y :

$$F_Y(y) = \mathbb{P}(Y \leq y) = 0 \text{ si } y < 2, \; \frac{y}{3} \text{ si } 2 \leq y \leq 3, \; 1 \text{ sinon.}$$

Y est une fonction de X dont la densité est simple. Le plus facile est d'appliquer le théorème de transfert : on a

$$\mathbb{E}(Y) = \mathbb{E}[\max(2,X)] = \int_0^3 \frac{1}{3} \max\{2, x\} dx = \frac{13}{6}.$$

Il y a un saut au point 2 dans la fonction de répartition de Y, donc cette variable aléatoire n'admet pas de densité.

Solution 144 Rappelons la formule de changement de variables. Soient U, V des ouverts de $\mathbb{R}^n$, $\Phi : U \to V$ un difféomorphisme et Z un vecteur aléatoire de densité $h(x)\mathbb{1}_{\{x \in U\}}$. Alors la variable aléatoire $\Phi(Z)$ admet pour densité

$$y \mapsto h(\Phi^{-1}(y)) |\det(D\Phi^{-1}(y))| \mathbb{1}_V(y).$$

Soit $Z = (X,Y)$, Z a pour densité

$$h(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}.$$

Soient $U = \mathbb{R} \times \mathbb{R}^*$ et $\Phi : U \to U$, $(x, y) \mapsto (x/y, y)$ le difféomorphisme d'inverse $\Phi^{-1}(u, v) = (uv, v)$. Par application de la formule de changement de variables, $\Phi(Z)$ admet pour densité

$$f(u,v) = h(uv, v)|v|\mathbb{1}_{\{(u,v)\in U\}} = \frac{1}{2\pi} e^{-(v^2(1+u^2))/2}|v|\mathbb{1}_{\{(u,v)\in U\}}.$$

Pour calculer la densité de X/Y, on intégre celle de $(X/Y, Y)$ par rapport à v. Soit, en notant f_1 cette densité

$$f_1(u) = \int_0^{+\infty} f(u,v)\ dv = 2\int_0^{+\infty} \frac{1}{2\pi} e^{-(v^2(1+u^2))/2} v\ dv = \frac{1}{\pi(u^2+1)}.$$

On reconnaît ici la loi de Cauchy. Si X/Y et Y étaient indépendantes, $(X/Y)^2$ et Y^2 le seraient aussi, et on aurait $\mathbb{E}(X^2) = (\mathbb{E}(X/Y)^2)\mathbb{E}(Y^2)$, ce qui est faux, car $\mathbb{E}(X^2) = \mathbb{E}(Y^2) = 1$, tandis que $\mathbb{E}((X/Y)^2) = \int_{\mathbb{R}} u^2 f_1(u)\ d\lambda(u) = +\infty$.

Solution 145 La variable aléatoire G_k s'écrit en fonction de N sous la forme

$$G_k = \begin{cases} aN - b(k-N) = (a+b)N - bk & \text{si } N \le k \\ ak - c(N-k) = (a+c)k - cN & \text{si } N \ge k \end{cases}$$

L'expression de $\mathbb{E}(G_k)$ en découle. Mais il est plus astucieux de considérer

$$G_{k+1} - G_k = -b + (a+b+c)\mathbb{1}_{\{N>k\}} = \begin{cases} -b & \text{si } N \le k \\ a+c & \text{si } N > k \end{cases}$$

Finalement, la différence

$$\mathbb{E}(G_{k+1}) - \mathbb{E}(G_k) = -b + (a+b+c)\mathbb{P}(N > k)$$

est positive pour $k < k' = \min\left\{\ell; \mathbb{P}(N > \ell) \le \frac{b}{a+b+c}\right\}$, puis change de signe. Le choix optimal est obtenu pour $k = k'$, soit approximativement

$$\mathbb{P}(N > k) \simeq \frac{b}{a+b+c}.$$

Solution 146 1. Pour calculer la loi de X_1, il suffit d'appliquer le théorème de calcul d'une mesure image par un C^1-difféomorphisme.
Ici $y \mapsto y^{1/p}$ réalise un C^1 difféomorphisme de $]0, +\infty[$ dans lui-même.
On a $(T^{-1})'(x) = px^{p-1}$ et la densité de X_1 est donc :

$$\begin{aligned} px^{p-1}.\frac{\mathbb{1}_{]0,+\infty[}(x)}{\Gamma(1/p)}(x^p)^{\frac{1}{p}-1}e^{-x^p} &= \frac{\mathbb{1}_{]0,+\infty[}(x)}{\frac{1}{p}\Gamma(1/p)}e^{-x^p} \\ &= \frac{e^{-x^p}}{\Gamma(\frac{1}{p}+1)}\mathbb{1}_{]0,+\infty[}(x). \end{aligned}$$

On a de plus $S = Y_1 + \cdots + Y_n$. Comme pour tout k, $Y_k \sim \Gamma(1/p, 1)$, la propriété de convolution des lois Gamma de même échelle donne
$S \sim \Gamma(n/p, 1)$.

2. On a
$$\int_{\mathbb{R}_+^n} \phi(\|x\|_p^p)\, d\lambda^{\otimes n}(x)$$
$$= \int_{\mathbb{R}_+^n} \left(\phi(\|x\|_p^p)\Gamma(1/p+1)^n e^{\|x\|_p^p}\right) \frac{1}{\Gamma(1/p+1)^n} e^{-\|x\|_p^p}\, d\lambda^{\otimes n}(x)$$
puis
$$\begin{aligned}\int_{\mathbb{R}_+^n} \phi(\|x\|_p^p)\, d\lambda^{\otimes n}(x) &= \int_{\mathbb{R}_+^n} \psi(\|x\|_p^p)\frac{1}{\Gamma(1/p+1)^n} e^{-\|x\|_p^p}\, d\lambda^{\otimes n}(x)\\ &= \int_{\mathbb{R}_+^n} \psi(\|x\|_p^p)\, d\mathbb{P}_{(X_1,\dots,X_n)}(x).\end{aligned}$$
En effet, les X_i étant indépendantes à densité sur $\mathbb{R}_+$, la loi du n-uplet $(X_1,\dots,X_n)$ a pour densité sur $\mathbb{R}_+^n$ le produit des densités. S étant l'image de $X = (X_1,\dots,X_n)$ par $x \mapsto \|x\|_p^p$, la loi de S est la loi image de $\mathbb{P}_X$ par $x \mapsto \|x\|_p^p$, soit, d'après le théorème de transfert
$$\begin{aligned}\int_{\mathbb{R}_+^n} \phi(\|x\|_p^p)\, d\lambda^{\otimes n}(x) &= \int_{\mathbb{R}_+^n} \psi(\|x\|_p^p)\, d\mathbb{P}_{(X_1,\dots,X_n)}(x)\\ &= \int_{\mathbb{R}_+^n} \psi(s)\, d\mathbb{P}_S(x) = \mathbb{E}[\psi(S)].\end{aligned}$$
Par symétrie
$$\begin{aligned}\int_{\mathbb{R}^n} \phi(\|x\|_p^p)\, d\lambda^{\otimes n}(x) &= 2^n \int_{\mathbb{R}_+^n} \phi(\|x\|_p^p)\, d\lambda^{\otimes n}(x) = 2^n\mathbb{E}[\psi(S)]\\ &= 2^n \int_{[0,+\infty[} \frac{1}{\Gamma(n/p)} u^{n/p-1}e^{-u}\psi(u)\, d\lambda(u)\\ &= \frac{2^n\Gamma(\frac{1}{p}+1)^n}{\Gamma(n/p)} \int_{\mathbb{R}_+} u^{\frac{n}{p}-1}\phi(u)\, du.\end{aligned}$$

3. On applique la formule précédente à $\phi(x) = \mathbb{1}_{[0,1]}(x)$. Le résultat est alors immédiat, avec la simplification $\frac{n}{p}\Gamma\left(\frac{n}{p}\right) = \Gamma\left(\frac{n}{p}+1\right)$.

   Pour $p = 2$, on retrouve la formule du volume de la boule euclidienne, qui avait été démontrée en cours à l'aide de la formule d'intégration des fonctions radiales.

4. On applique encore la formule avec $p = 1$ et $\phi(x) = \frac{1}{x}\mathbb{1}_{[0,1]}(x)$
$$\begin{aligned}\int_{T_n} \frac{d\lambda^{\otimes n}(x)}{x_1+\dots+x_n} &= \frac{1}{2^n}\int_{\mathbb{R}^n} \phi(\|x\|_1^1)\, d\lambda^{\otimes n}(x)\\ &= \frac{\Gamma(\frac{1}{1}+1)^n}{\Gamma(n/1)}\int_{\mathbb{R}_+} u^{\frac{n}{1}-1}\phi(u)\, du = \frac{1}{(n-1)!(n-1)}.\end{aligned}$$

Solution 147 1. Soient

$$\begin{aligned} U &= \{x=(x_1,\dots,x_n)\in\mathbb{R}^n : x_i>0 \text{ pour tout } i\},\\ V &= \{(u,s)\in\mathbb{R}^{n-1}\times\mathbb{R} : u_i>0 \text{ pour tout } i,\ \sum_{i=1}^{n-1} u_i>1,\ s>0\}\end{aligned}$$

et $\Phi : U \to V$, $(x_1,\dots,x_n) \mapsto (u_1,\dots,u_{n-1},s)$ le difféomorphisme défini par

$$u_i=\frac{x_i}{s} \text{ pour } i=1,\dots,n-1 \quad \text{et} \quad s=\sum_{i=1}^{n-1} x_i.$$

L'inverse de Φ est le difféomorphisme $\Psi : V \to U$ défini par

$$x_i=u_is,\ i=1,\dots,n-1 \text{ et } x_n=s\left(1-\sum_{i=1}^{n-1}u_i\right).$$

$$\text{Ainsi } D\Psi(u,s)=\begin{pmatrix} s & 0 & \dots & 0 & u_1\\ 0 & s & \dots & 0 & u_2\\ \vdots & & \ddots & & \vdots\\ 0 & & 0 & s & u_{n-1}\\ -s & -s & & -s & (1-\sum u_i)\end{pmatrix}.$$

Pour calculer le déterminant Jacobien $J=\det(D\Psi(u,s))$ il suffit d'ajouter chacune des $n-1$ premières lignes à la dernière ligne de $D\Psi(u,s)$. Ainsi

$$J=\begin{vmatrix} s & 0 & \dots & 0 & u_1\\ 0 & s & \dots & 0 & u_2\\ \vdots & & \ddots & & \vdots\\ 0 & & 0 & s & u_{n-1}\\ 0 & 0 & & 0 & 1\end{vmatrix}=s^{n-1}.$$

Par application de la formule du changement de variables, la densité de (U,S_n) est égale à

$$e^{-s}s^{n-1}\mathbb{1}_V.$$

Le vecteur $U=(U_1,\dots,U_n)$ suit donc une loi uniforme sur l'ensemble

$$\Sigma^{n-1}=\{u=(u_1,\dots,u_{n-1})\in\mathbb{R}^{n-1} :\ u_i>0 \text{ pour tout } i,\sum_{i=1}^{n-1}u_i<1\}.$$

Soit

$$\Delta^{n-1}=\{x=(x_1,\dots,x_n)\in\mathbb{R}^n :\ x_i>0 \text{ pour tout } i,\sum_{i=1}^{n}x_i=1\}.$$

Le vecteur $(U_1,\dots,U_n)$ est l'image du vecteur $(U_1,\dots,U_{n-1})$ par l'application

$$P : \Sigma^{n-1} \to \Delta^{n-1}, \, (u_1, \ldots, u_{n-1}) \mapsto \left(u_1, \ldots, u_{n-1}, 1 - \sum_{i=1}^{n-1} u_i \right).$$

La loi de $(U_1, \ldots, U_n)$ est donc la loi uniforme sur Δ^{n-1}. On peut remarquer qu'on retrouve le résultat de la section 6.12.1 : cette loi est aussi la loi de Dirichlet de paramètre $(1, \ldots, 1)$.

2. Identifions T à Δ^2, et a, b, c aux points $(1,0,0), (0,1,0)$ et $(0,0,1)$. Alors, pour $V = (U_1, U_2, U_3) \in \Delta^2$ la surface du triangle (a, b, V) est proportionnelle au déterminant $[\overrightarrow{Va}, \overrightarrow{Vb}] = U_3$. Si V suit une loi uniforme sur T, la première question montre que la loi du rapport de cette surface à la surface de T est égale à la loi de la variable

$$U_3 = X_1/(X_1 + X_2 + X_3)$$

où X_1, X_2, X_3 sont des variables indépendantes identiquement distribuées de loi exponentielle $\mathcal{E}(1)$. D'où, pour tout $a \in]0,1[$,

$$\mathbb{P}(U_3 \leq a) = \mathbb{P}(X_1 \leq \frac{a}{1-a}(X_2 + X_3)).$$

Posons $t = \frac{a}{1-a}$ et $B_t = \{(x_1, x_2, x_3) \in \mathbb{R}_+^3 ; x_1 \leq t(x_2 + x_3)\}$. On a

$$\begin{aligned} \mathbb{P}(U_3 \leq a) &= \int_{\mathbb{R}_+^3} \mathbb{1}_{B_t}(x_1, x_2, x_3) \, d\mathbb{P}_{(X_1, X_2, X_3)}(x_1, x_2, x_3) \\ &= \int_{\mathbb{R}_+^3} \mathbb{1}_{B_t}(x_1, x_2, x_3) \, d(\mathbb{P}_{X_1} \otimes \mathbb{P}_{(X_2, X_3)})(x_1, x_2, x_3). \end{aligned}$$

Avec Tonelli, on a alors

$$\begin{aligned} \mathbb{P}(U_3 \leq a) &= \int_{\mathbb{R}_+^2} \mathbb{P}((X_1, x_2, x_3) \in B_t) \, d\mathbb{P}_{(X_2, X_3)}(x_2, x_3) \\ &= \int_{\mathbb{R}_+^2} (1 - e^{-(x_2 + x_3)t}) \, d\mathbb{P}_{(X_2, X_3)}(x_2, x_3) \\ &= \mathbb{E}(1 - e^{-t(X_2 + X_3)}) = 1 - \mathbb{E}(e^{-tX_2} e^{-tX_3}) \\ &= 1 - \mathbb{E}(e^{-tX_2}) \mathbb{E}(e^{-tX_3}) = 1 - (\mathbb{E}(e^{-tX_2}))^2. \end{aligned}$$

Un calcul simple donne $\mathbb{E}(e^{-tX_2}) = \frac{1}{1+t} = 1 - a$. On en déduit que $\mathbb{P}(U_3 \leq a) = 1 - (1-a)^2 = a(2-a)$.

Solution 148 Les variables X et Y ne sont pas indépendantes car le support de la loi conjointe du couple (X, Y) n'est pas un produit cartésien.
On a par exemple $\mathbb{P}(X > \sqrt{2}, Y > \sqrt{2}) \leq \mathbb{P}(X^2 + Y^2 > 4) = 0$ tandis que $\mathbb{P}(X > \sqrt{2})$ et $\mathbb{P}(Y > \sqrt{2})$ sont non-nuls car g n'est pas presque partout nulle

sur $]\sqrt{2},+\infty[\times\mathbb{R}$ ou $\mathbb{R}\times]\sqrt{2},+\infty[$. L'application

$$\Phi : D \to]0,2[\times]0,\pi/2[, \quad (x,y) \mapsto (\rho,\theta)$$

définie par

$$x = \rho\cos\theta, \quad y = \rho\sin\theta$$

est un C^1 difféomorphisme. D'après la formule de changement de variables, le couple $(R = \sqrt{(X^2+Y^2)}, \ \Theta = \arctan(Y/X))$ admet pour densité la fonction

$$f(\rho,\theta) = \frac{1}{2}\rho^3\cos\theta\sin\theta \mathbb{1}_{]0,2[\times]0,\pi/2[}(\rho,\theta).$$

On en déduit que les variables R et Θ sont indépendantes de densité respective $\frac{\rho^3}{4}\mathbb{1}_{]0,2[}(\rho)$ et $2\cos\theta\sin\theta\mathbb{1}_{]0,\frac{\pi}{2}[}(\theta) = \sin(2\theta)$.
Soit $Z = R^2$. Alors pour $t \in [0,4]$

$$\mathbb{P}(Z \le t) = \mathbb{P}(R \le \sqrt{t}) = \int_0^{\sqrt{t}} \frac{\rho^3}{4} d\rho = \frac{t^2}{16}.$$

Bien sûr, $\mathbb{P}(Z \le t)$ vaut 0 pour $t < 0$ et 1 pour $t \ge 4$. Comme la fonction de répartition est continue et C^1 par morceaux, on en déduit par dérivation la densité de Z : $f_Z(t) = \frac{t}{8}\mathbb{1}_{[0,4]}(t)$.

Solution 149

1. L'application qui échange la coordonnée 1 et la coordonnée i, la coordonnée 2 et la coordonnée j, réalise évidemment une bijection entre $\Omega_{1,2}^{a,b}$ et $\Omega_{i,j}^{a,b}$. On en déduit
$$\mathbb{P}(X_i = a, X_j = b) = \frac{|\Omega_{i,j}^{a,b}|}{|\Omega|} = \frac{|\Omega_{1,1}^{a,b}|}{|\Omega|} = \mathbb{P}(X_1 = a, X_2 = b).$$
Comme l'égalité a lieu pour toutes les valeurs possibles du couple (a,b), on en déduit que les vecteurs (X_i,X_j) et (X_1,X_2) ont même loi. En particulier, la loi de la première marginale est commune à X_1 et X_i qui ont donc même loi.

2. X_i prend ses valeurs dans $\{0,1\}$. C'est donc une variable de Bernoulli. Comme $k = \sum_{i=1}^n X_i$, on a par linéarité de l'espérance $k = \sum_{i=1}^n \mathbb{E}(X_i)$. Comme les X_i ont toutes la même loi, elles ont la même espérance : $k = n\mathbb{E}(X_1)$, soit $\mathbb{E}(X_1) = k/n$. Ainsi, pour tout i, X_i suit la loi de Bernoulli de paramètre k/n. Sa variance est $\frac{k}{n}(1-\frac{k}{n})$.

3. Par linéarité $\mathbb{E}(S_r) = \sum_{i=1}^r \mathbb{E}(X_i) = r\mathbb{E}(X_1) = \frac{kr}{n}$. Pour tous les couples (i,j), avec i et j distincts, (X_1,X_2) a même loi que (X_i,X_j) donc $\operatorname{Covar}(X_i,X_j) = \operatorname{Covar}(X_1,X_2)$. De même $\operatorname{Var} X_i = \operatorname{Var} X_1$,

donc par bilinéarité,

$$\begin{aligned}\operatorname{Var}(X_1+\dots+X_r) &= \sum_{i=1}^{r}\operatorname{Var}X_i + 2\sum_{1\le i<j\le r}\operatorname{Covar}(X_i,X_j)\\ &= r\operatorname{Var}X_1 + r(r-1)\operatorname{Covar}(X_1,X_2).\end{aligned}$$

Il suffit donc de poser $P_n(X) = X\operatorname{Var}X_1 + X(X-1)\operatorname{Covar}(X_1,X_2)$ et on a le résultat voulu.

4. Comme la variable $X_1+\dots+X_n$ est constante égale à k, sa variance est nulle. Le polynôme P_n est un polynôme de degré 2, nul en 0 et en n : il s'écrit donc sous la forme $P_n(X) = AX(n-X)$. On a

$$A.1.(n-1) = P_n(1) = \operatorname{Var}S_1 = \operatorname{Var}X_1,$$

donc $A = \frac{\operatorname{Var}X_1}{n-1}$, d'où la formule

$$\operatorname{Var}S_r = P_n(r) = \frac{r(n-r)}{n-1}\operatorname{Var}X_1 = \frac{r(n-r)}{n-1}\frac{k}{n}\left(1-\frac{k}{n}\right).$$

5. On prend sans remise k boules indiscernables au toucher dans une urne contenant n boules, r d'entre elles ayant été marquées. Vu les symétries du problème, il est naturel de modéliser le vecteur de présence des différentes boules dans le tirage par la loi uniforme sur Ω. Il suffit alors de donner les indices $1,\dots,r$ aux boules marquées, et on voit que S_r modélise le nombre de boules marquées dans le tirage sans remise.

Solution 150 1. Le plus simple est d'appliquer la technique de la fonction test. Prenons ϕ mesurable de $[0,+\infty[$ dans lui-même : en appliquant plusieurs fois le théorème de Tonelli, on a

$$\begin{aligned}\mathbb{E}\left(\phi(\frac{X}{Y})\right) &= \int_{]0,+\infty[^2} f_{(X,Y)}(x,y)\phi(\frac{x}{y})\,d\lambda^2(x,y)\\ &= \int_{]0,+\infty[^2} f_X(x)f_Y(y)\phi(\frac{x}{y})\,d\lambda^2(x,y)\\ &= \int_{]0,+\infty[} f_Y(y)\left(\int_{]0,+\infty[} f_X(x)\phi(\frac{x}{y})\,d\lambda(x)\right)\,d\lambda(y)\\ &= \int_{]0,+\infty[} f_Y(y)\left(\int_{]0,+\infty[} f_X(yv)\phi(v)y\,d\lambda(v)\right)\,d\lambda(y)\\ &= \int_{]0,+\infty[} \phi(v)\left(\int_{]0,+\infty[} f_Y(y)f_X(yv)y\,d\lambda(y)\right)\,d\lambda(v).\end{aligned}$$

Cela nous donne la densité de V :

$$f_V(v) = \int_{]0,+\infty[} f_X(yv)f_Y(y)y\,d\lambda(y)\mathbb{1}_{\{v>0\}}.$$

2. Dans le cas où $X\sim\Gamma(a,\gamma)$ et $Y\sim\Gamma(b,\gamma)$, avec $a,b>0$, on a pour

$x, y > 0$, $f_X(x) = c_1 x^{a-1}e^{-\gamma x}$, $f_Y(y) = c_2 y^{b-1}e^{-\gamma y}$, et on obtient, pour $v > 0$

$$\begin{aligned} f_V(v) &= \int_{]0,+\infty[} c_1(uv)^{a-1}e^{-\gamma uv}c_2 u^{b-1}e^{-\gamma u}u\ d\lambda(u) \\ &= c_1c_2v^{a-1}\int_{]0,+\infty[} u^{a+b-1}e^{-\gamma(v+1)u}\ d\lambda(u) \\ &= c_1c_2v^{a-1}(\gamma(v+1))^{-(a+b)}\Gamma(a+b) \\ &= K_{a,b}\frac{v^{a-1}}{(1+v)^{a+b}}, \end{aligned}$$

avec

$$\begin{aligned} K_{a,b} = c_1c_2\gamma^{-(a+b)}\Gamma(a+b) &= \frac{\gamma^a}{\Gamma(a)}\frac{\gamma^b}{\Gamma(b)}\gamma^{-(a+b)}\Gamma(a+b) \\ = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} &= \frac{1}{B(a,b)}. \end{aligned}$$

Comme l'intégrale d'une densité fait toujours 1, on en déduit que

$$B(a,b) = \int_0^{+\infty} \frac{u^{a-1}}{(1+u)^{a+b}}\ du.$$

On notera que la loi trouvée ne dépend pas du paramètre γ. C'était prévisible vu les propriétés d'échelle des lois γ. En effet si X et Y sont indépendantes de lois respectives $\Gamma(a,\gamma)$ et $\Gamma(b,\gamma)$, la loi de $\frac{X}{Y}$ est la loi de $\frac{\gamma X}{\gamma Y}$. Mais γX et γY sont des variables indépendantes de lois respectives $\Gamma(a,1)$ et $\Gamma(b,1)$: la loi de leur quotient ne dépend pas de γ.

3. Par la question précédente,

$$B(1/n, 1-1/n) = \int_0^{+\infty} \frac{v^{1/n-1}}{1+v}\ dv = n\int_0^{+\infty} \frac{dx}{1+x^n}\ dx,$$

avec le changement de variables $x^n = u$. Cette dernière intégrale se calcule facilement avec le théorème des résidus : on intègre le long du chemin qui va de 0 à $R > 1$ puis, en tournant, de R à $R\omega^2$, puis revient de $R\omega^2$ à 0. Posant $f(z) = \frac{1}{1+z^n}$, on a

$$\int_0^R f(r)\ dr - \int_0^R f(r\omega^2)\omega^2\ dr + \int_0^{2\pi/n} f(Re^{i\theta})iRe^{i\theta}\ d\theta = 2i\pi Res(f(z)dz, \omega).$$

Comme $f(\omega^2 r) = f(r)$, on a

$$(1-\omega^2)\int_0^R f(r)\ dr + \int_0^{2\pi/n} f(Re^{i\theta})iRe^{i\theta}\ d\theta = 2i\pi Res(f(z)dz, \omega).$$

Comme ω est un pôle simple de f, le résidu est égal à

$$\lim_{z\to\omega}(z-\omega)f(z) = (n\omega^{n-1})^{-1} = \frac{1}{n}\omega^{1-n} = -\frac{1}{n}\omega.$$

Comme $|f(Re^{i\theta})| \leq \frac{1}{R^n-1}$, en faisant tendre R vers l'infini, on obtient finalement le résultat voulu, à savoir

$$\int_0^{+\infty} \frac{dx}{1+x^n} = \frac{1}{1-\omega^2}(-\frac{1}{n}\omega)\times 2i\pi = \frac{\pi}{n}\frac{2i}{\omega-\omega^{-1}} = \frac{\pi/n}{\sin(\pi/n)}.$$

Multipliant les deux membres de l'équation par $1/n$, on obtient $\frac{1}{n}B\left(\frac{1}{n}, 1-\frac{1}{n}\right) = \frac{\pi/n}{\sin(\pi/n)}$, soit $\frac{1}{n}\Gamma(1/n)\Gamma(1-1/n) = \frac{\pi/n}{\sin(\pi/n)}$ et enfin $\Gamma(1+1/n)\Gamma(1-1/n) = \frac{\pi/n}{\sin(\pi/n)}$.

La fonction $z \mapsto \Gamma(1+z)\Gamma(1-z) - \frac{\pi z}{\sin(\pi z)}$ est holomorphe sur l'ensemble $\{z \in \mathbb{C} : -1 < \text{Re(z)} < 1\}$. Sur cet ensemble, 0 est un point d'accumulation des zéros, qui sont les points $1/n$. La fonction est donc identiquement nulle et on a donc sur $\{z \in \mathbb{C} : -1 < \text{Re(z)} < 1\}$: $\Gamma(1+z)\Gamma(1-z) - \frac{\pi z}{\sin(\pi z)}$. En divisant par z, on obtient la formule voulue.

Solution 151 1. Soient $\Pi = (A_1, \dots, A_N)$ une partition de $\{1, \dots, n\}$ et $x_1 \in A_1, \dots, x_n \in A_n$. L'application qui à f associe sa restriction à $\{x_1, \dots, x_N\}$ réalise une bijection entre l'ensemble des fonctions f telles que $P(f) = \Pi$ et l'ensemble des injections de $\{x_1, \dots, x_N\}$ dans $\{1, \dots, p\}$. Ainsi, il y a $p(p-1)\dots(p-N+1)$ telles fonctions.

2. Pour tout entier naturel p, on peut écrire

$$\begin{aligned} p^n &= |\{1, \dots, p\}^n| \\ &= \sum_{\Pi \text{ Partition de } \{1,\dots,n\}} |\{f \in \{1, \dots, p\}^n ; P(f) = \Pi\}| \\ &= \sum_{\Pi \text{ Partition de } \{1,\dots,n\}} P_{N(\Pi)}(p). \end{aligned}$$

Deux polynômes qui coïncident en un nombre infini de points sont égaux ; d'où le résultat.

3. Avec le théorème de transfert, il vient
$$\begin{aligned}\mathbb{E}(P_p(X)) &= \sum_{k=0}^{+\infty} \mathbb{P}(X=k)P_p(k) = \sum_{k=p}^{+\infty} \mathbb{P}(X=k)P_p(k) \\ &= \sum_{i=0}^{+\infty} \mathbb{P}(X=p+i)P_p(p+i) \\ &= \sum_{i=0}^{+\infty} \frac{1}{e}\frac{1}{(p+i)!}\frac{(p+i)!}{i!} = 1.\end{aligned}$$

4. En prenant l'espérance de la relation de 2., on a par linéarité
$$\begin{aligned}\mathbb{E}(X^n) &= \sum_{\Pi \text{ Partition de } \{1,\dots,n\}} \mathbb{E}(P_{N(\Pi)}(X)) \\ &= \sum_{\Pi \text{ Partition de } \{1,\dots,n\}} 1 = B_n.\end{aligned}$$
Maintenant, avec le théorème de transfert, on a
$$\begin{aligned}B_n = \mathbb{E}(X^n) &= \sum_{k=0}^{+\infty} k^n \mathbb{P}(X=k) \\ &= \sum_{k=0}^{+\infty} k^n e^{-1}\frac{1^k}{k!} = \frac{1}{e}\sum_{k=0}^{+\infty}\frac{k^n}{k!}.\end{aligned}$$

5. Soit $u>0$. On a
$$\begin{aligned}\phi(u) &= \sum_{n\geq 0}\frac{u^n}{n!}B_n = \sum_{n\geq 0}\frac{u^n}{n!}\mathbb{E}(X^n) = \mathbb{E}(\sum_{n\geq 0}\frac{u^n}{n!}X^n) \text{ avec Tonelli} \\ &= \mathbb{E}\exp(uX) = \sum_{k=0}^{+\infty}\exp(-1)\frac{1}{k!}e^{uk} = \exp(e^u-1).\end{aligned}$$
On en déduit que le rayon de convergence de la série définissant $\phi(x)$ est infini. D'après le théorème des zéros isolés, l'identité entre $\phi(x)$ et $\exp(\exp(x)-1)$ se prolonge à $\mathbb{C}$ tout entier.

6. Avec la formule de Cauchy, pour toute courbe simple entourant l'origine, on a
$$\frac{B_n}{n!} = \text{Res}_0 \frac{\phi(z)}{z^{n+1}}\,dz = \frac{1}{2i\pi}\int_\gamma \frac{\phi(z)}{z^{n+1}}\,dz.$$
En particulier, en prenant pour γ le cercle de centre 0 et de rayon r, il vient
$$\frac{B_n}{n!} = \frac{1}{2\pi r^n}\int_{-\pi}^{+\pi} e^{\phi(re^{i\theta})}e^{-in\theta}\,d\theta \leq \frac{M(r)}{r^n},$$

où $M(r) = \sup\{|\phi(z); |z| = r\}$. Cette dernière inégalité est parfois appelée inégalité de Cauchy. Comme les coefficients définissant $\phi(z)$ sont positifs, on a facilement avec l'inégalité triangulaire $|\phi(z)| \le \phi(|z|)$, d'où $M(r) = \phi(r)$. Il s'agit donc de minimiser $\frac{\phi(r)}{r^n}$.
Comme $\frac{\phi(r)}{r^n} = \exp(e^r - 1 - n\log r)$, on voit facilement que $\frac{\phi(r)}{r^n}$ est minimal au point r_n tel que $e^{r_n} - \frac{n}{r_n} = 0$, soit $r_n e^{r_n} = n$. La quantité $\log n$ est probablement une bonne estimation de r_n [1]donc on va utiliser l'inégalité en $\log n$: il vient alors

$$\frac{B_n}{n!} \le \frac{\phi(\log n)}{(\log n)^n} = \frac{e^{n-1}}{(\log n)^n},$$

ce qui donne l'inégalité voulue.

Solution 152 1. On pose $X' = \log X$, $Y' = \log Y$. X' et Y' sont indépendantes et de même loi. Comme X' et Y' sont dans L^2, $Z' = X' - Y'$ aussi et on a $\operatorname{Var} Z' = \operatorname{Var} X' + \operatorname{Var} Y' = 2\operatorname{Var} X'$. Comme différence de deux variables de même espérance, elle est centrée, donc $\mathbb{E}[(Z')^2] = \operatorname{Var}(Z') = 2\operatorname{Var} X'$. On a donc bien $\operatorname{Var} X' = \frac{1}{2}\mathbb{E}[(Z')^2]$. Comme $Z' = \log(X/Y)$ et $-Z' = \log(Y/X)$, Z' et $-Z'$ ont même loi, qui est la loi image de $\mu^{\otimes 2}$ par $(x, y) \mapsto \log(x/y)$. Donc on a maintenant

$$\begin{aligned}
\mathbb{E}(Z'^2) &= \mathbb{E}((Z')^2 \mathbb{1}_{\{Z'>0\}}) + \mathbb{E}((Z')^2 \mathbb{1}_{\{Z'<0\}}) + \mathbb{E}((Z')^2 \mathbb{1}_{\{Z'=0\}}) \\
&= \mathbb{E}((Z')^2 \mathbb{1}_{\{Z'>0\}}) + \mathbb{E}((-Z')^2 \mathbb{1}_{\{-Z'>0\}}) \\
&= 2\mathbb{E}((Z')^2 \mathbb{1}_{\{Z'>0\}}) \\
&= 2\int_0^{+\infty} \mathbb{P}((Z')^2 \mathbb{1}_{\{Z'>0\}} > u)\, du \\
&= 2\int_0^{+\infty} \mathbb{P}((Z')^2 \mathbb{1}_{\{Z'>0\}} > t^2) 2t\, dt \\
&= 4\int_0^{+\infty} \mathbb{P}(Z' > t) t\, dt,
\end{aligned}$$

ce qui donne les dernières égalités.

1. La solution de $re^r = x$, notée $W(x)$, est appelée *fonction de Lambert*. On peut montrer en l'infini le développement $W(x) = \log x + O(\log\log x)$.

2. Pour $t > 0$, on a

$$\begin{aligned}\mathbb{P}(Z > t) = \mathbb{P}(X/Y > t) &= \mathbb{P}_{(X,Y)}(\{(x,y) \in \mathbb{R}_+^2 ; x > yt > 0\}) \\ &= \int_{\mathbb{R}_+^2} e^{-x} e^{-y} \mathbb{1}_{\{x>yt\}} \, d\lambda^{\otimes 2}(x,y) \\ &= \int_{\mathbb{R}+} e^{-y} \left(\int e^{-y} \mathbb{1}_{\{x>yt\}} \, d\lambda(x) \right) \, d\lambda(y) \\ &= \int_{\mathbb{R}+} e^{-y} e^{-ty} \, d\lambda(y) = \frac{1}{1+t}.\end{aligned}$$

On en déduit, pour t réel $\mathbb{P}(Z' > t) = \mathbb{P}(e^{Z'} > e^t) = \mathbb{P}(Z > e^t) = \frac{1}{1+e^t}$, ce qui, avec la question précédente, nous donne

$$\mathrm{Var}(\log X) = 2 \int_0^{+\infty} t\mathbb{P}(Z' > t) \, dt = 2 \int_{]0,+\infty[} \frac{t}{1+e^t} \, d\lambda(t).$$

On peut noter que $\frac{2t}{1+e^{-t}} = \frac{2t}{e^t-1} - \frac{4t}{e^{2t}-1}$ et $\frac{t}{e^t-1} = \frac{te^{-t}}{1-e^{-t}}$. Comme pour tout $x \in]-1,1[$, on a $\frac{1}{1-x} = \sum_{k=0}^{+\infty} x^k$, alors pour tout $t > 0$, on a $\frac{1}{1-e^{-t}} = \sum_{k=0}^{+\infty} e^{-kt}$ et

$$\frac{te^{-t}}{1-e^{-t}} = \sum_{k=0}^{+\infty} te^{-(k+1)t}.$$

Pour tout $k \geq 0$, on a

$$\begin{aligned}\int_0^{+\infty} te^{-(k+1)t} \, dt &= \frac{1}{(k+1)^2} \int_0^{+\infty} ((k+1)t) e^{-(k+1)t} (k+1) \, dt \\ &= \frac{1}{(k+1)^2} \int_0^{+\infty} te^{-t} \, dt = \frac{\mathbb{E}(X)}{(k+1)^2} = \frac{1}{(k+1)^2}.\end{aligned}$$

Avec le théorème de convergence monotone, on en déduit alors que si on pose $I = \int_{[0,+\infty[} \frac{te^{-t}}{1-e^{-t}} \, d\lambda(t)$, on a

$$I = \sum_{k=0}^{+\infty} \frac{1}{(k+1)^2} = \frac{\pi^2}{6}.$$

Finalement, un changement de variable pour la deuxième intégrale nous donne

$$\begin{aligned}\mathrm{Var}(\log X) &= 2 \int_{[0,+\infty[} \frac{te^{-t}}{1-e^{-t}} \, d\lambda(t) - \int_{[0,+\infty[} \frac{4te^{-2t}}{1-e^{-2t}} \, d\lambda(t) \\ &= 2I - I = I \\ &= \frac{\pi^2}{6}.\end{aligned}$$

Solution 153 (X,Y) prend presque sûrement ses valeurs dans $\mathbb{R}^* \times \mathbb{R}_+^*$. X

et Y étant indépendantes, la loi du couple (X,Y) admet la densité

$$(x,y) \mapsto \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) \frac{1}{2^{n/2}\Gamma(n/2)} e^{-y/2} y^{n/2-1} \mathbb{1}_{\mathbb{R}_+}(y).$$

Montrons que $(x,y) \mapsto (y, \frac{x}{\sqrt{y/n}})$ réalise un C^1 difféomorphisme de $\mathbb{R}^* \times \mathbb{R}_+^*$ dans $\mathbb{R}^* \times \mathbb{R}^*$. En effet, il est clair que l'application est C^1 (ses composantes le sont) et c'est une bijection : $z = \frac{x}{\sqrt{y/n}} \iff x = z\sqrt{y/n}$ et sa réciproque est $(y,z) \mapsto (x,y) = (z\sqrt{y/n}, y)$ qui est de classe C^1. On a

$$\frac{D(x,y)}{D(y,z)} = \| \begin{pmatrix} \frac{z}{2\sqrt{ny}} & \sqrt{y/n} \\ 1 & 0 \end{pmatrix} \| = \sqrt{y/n} = y^{1/2} n^{-1/2}.$$

Donc (Y,Z) suit la loi de densité

$$f_{(Y,Z)}(y,z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} z^2 \frac{y}{n}) \frac{1}{2^{n/2}\Gamma(n/2)} e^{-y/2} y^{n/2-1} y^{1/2} n^{-1/2} \mathbb{1}_{\mathbb{R}_+}(y).$$

Par suite, Z suit la loi de densité

$$z \mapsto \int_{\mathbb{R}} f_{(Y,Z)}(y,z)\, dy = \frac{1}{\sqrt{2\pi}} \frac{n^{-1/2}}{2^{n/2}\Gamma(n/2)} \int_{\mathbb{R}_+} \exp(-\frac{1}{2}(\frac{z^2}{n}+1)y) y^{\frac{n+1}{2}-1}\, dy.$$

On fait le changement de variable $(\frac{z^2}{n}+1)y = \theta$, soit $y = \frac{n}{n+z^2}\theta$, d'où

$$\begin{aligned}\int_{\mathbb{R}_+} \exp(-\frac{1}{2}(\frac{z^2}{n}+1)y) y^{\frac{n+1}{2}-1}\, dy &= \left(\frac{n}{n+z^2}\right)^{\frac{n+1}{2}} \int_{\mathbb{R}_+} \exp(-\frac{\theta}{2}) \theta^{\frac{n+1}{2}-1}\, d\theta \\ &= \left(\frac{n}{n+z^2}\right)^{\frac{n+1}{2}} \Gamma(\frac{n+1}{2}) 2^{\frac{n+1}{2}}.\end{aligned}$$

Après simplifications, Z suit la loi de densité

$$f_Z(z) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} n^{n/2} \frac{1}{(z^2+n)^{\frac{n+1}{2}}}.$$

Solution 154 1. On a $M = \sum_{i=1}^{+\infty} \mathbb{1}_{\{N_i \neq 0\}}$. En appliquant le théorème de convergence monotone aux sommes partielles, on a

$$\begin{aligned}\mathbb{E}(M) &= \sum_{i=1}^{+\infty} \mathbb{E}(\mathbb{1}_{\{N_i \neq 0\}}) = \sum_{i=1}^{+\infty} \mathbb{P}(N_i \neq 0) \\ &= \sum_{i=1}^{+\infty} p_i^{-s} \leq \sum_{i=1}^{+\infty} i^{-s} < +\infty.\end{aligned}$$

En particulier $\mathbb{P}(M < +\infty) = 1$, ce qui entraîne que la variable X, qui prend *a priori* ses valeurs dans $\mathbb{N}^* \cup \{+\infty\}$, est en fait presque sûrement

à valeurs dans $\mathbb{N}^*$.
Si $n = \prod_{i=1}^{+\infty} p_i^{\alpha_i}$, où (p_i) est la suite des nombres premiers et α_i le plus grand entier α tel que p_i^α divise n, alors

$$\{X = n\} = \bigcap_{i\geq 1} \{N_i = \alpha_i\},$$

et avec le théorème de continuité séquentielle décroissante

$$\begin{aligned}
\mathbb{P}(X = n) &= \lim_{k\to+\infty} \mathbb{P}(\bigcap_{i=1}^{k} \{N_i = \alpha_i\}) = \lim_{k\to+\infty} \prod_{i=1}^{k} \mathbb{P}(N_i = \alpha_i) \\
&= \lim_{k\to+\infty} \prod_{i=1}^{k} (1 - p_i^{-s}) p_i^{-s\alpha_i} \\
&= \left(\prod_{i\geq 1} (1 - p_i^{-s})\right) \prod_{i\geq 1} p_i^{-s\alpha_i} = C_s n^{-s},
\end{aligned}$$

avec $C_s = \prod_{i\geq 1} (1 - p_i^{-s})$. Comme $\sum_{n\geq 1} \mathbb{P}(X = n) = 1$, on a $C_s = \frac{1}{\zeta(s)}$, et X suit la loi Zêta de paramètre s.

2. Comme on a l'identité $X = \prod_{i=1}^{n} p_i^{N_i}$ pour n (aléatoire) assez grand, on trouve $\phi(X) = \lim_{n\to+\infty} \phi(\prod_{i=1}^{n} p_i^{N_i})$. Posons $Y_n = \phi(\prod_{i=1}^{n} p_i^{N_i})$. Comme $|Y_n| \leq M$, le théorème de convergence dominée nous donne $\mathbb{E}[\phi(X)] = \lim_{n\to+\infty} \mathbb{E}[Y_n]$. Par la propriété de multiplicativité, on a $Y_n = \prod_{i=1}^{n} \phi(p_i^{N_i})$. De plus, l'indépendance nous donne

$$\mathbb{E}[Y_n] = \prod_{i=1}^{n} \mathbb{E}[\phi(p_i^{N_i})].$$

Il reste à calculer $\mathbb{E}[\phi(p_i^{N_i})]$. Par le théorème de transfert, on a

$$\mathbb{E}[\phi(p_i^{N_i})] = \sum_{m=0}^{+\infty} \mathbb{P}(N_i = m)\phi(p_i^m) = \sum_{m=0}^{+\infty} p_i^{-ms}(1 - p_i^{-1})\phi(p_i^m),$$

d'où

$$\begin{aligned}
\mathbb{E}(\phi(X)) &= \prod_{i=1}^{+\infty} \left(\sum_{m=0}^{+\infty} p_i^{-ms}(1 - p_i^{-1})\phi(p_i^m)\right) \\
&= C_s \prod_{i=1}^{+\infty} \left(\sum_{m=0}^{+\infty} p_i^{-ms}\phi(p_i^m)\right).
\end{aligned}$$

Comme X suit la loi Zêta de paramètre s, le théorème de transfert donne simplement $\mathbb{E}(\phi(X)) = \frac{1}{\zeta(s)} \sum_{n=1}^{+\infty} \frac{\phi(n)}{n^s}$. En multipliant les deux expressions de $\mathbb{E}(\phi(X))$ par $\zeta(s)$, on obtient l'égalité voulue. Le cas où ϕ est complètement multiplicative se déduit d'un simple calcul de série géométrique.

Solution 155 1. Comme $\mathcal{E}(\lambda) = \Gamma(1, \lambda)$, on voit que pour tout n, X_n suit la loi $\Gamma(1, \lambda)$. On sait que si deux variables aléatoires indépendantes X et Y vérifient $X \sim \Gamma(a, \lambda)$ et $Y \sim \Gamma(a, \lambda)$, alors $X + Y \sim \Gamma(a + b, \lambda)$. Or, S_n est clairement mesurable par rapport à la tribu $\sigma(X_1, \ldots, X_n)$. D'après le théorème d'associativité de l'indépendance, cette tribu est indépendante de la tribu engendrée par X_{n+1}, donc les variables aléatoires S_n et X_{n+1} sont indépendantes. Il vient alors aisément par récurrence sur n que pour tout entier n strictement positif, la variable aléatoire S_n suit une loi gamma $\Gamma(n, \lambda)$.

2. Comme S_n suit la loi $\Gamma(n, \lambda)$, S_n a la densité $x \mapsto \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} \mathbb{1}_{\mathbb{R}_+}(x)$. On a donc
$$\begin{aligned}\mathbb{P}(S_n < t) &= \int_0^t \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} \, dx = \int_0^t \frac{1}{\Gamma(n)} (\lambda x)^{n-1} e^{-\lambda x} \, \lambda dx \\ &= \int_0^{\lambda t} \frac{1}{\Gamma(n)} x^{n-1} e^{-x} \, dx =: g_n(\lambda t).\end{aligned}$$
Une intégration par parties donne alors
$$\begin{aligned}g_{n+1}(t) &= \frac{1}{\Gamma(n+1)} \int_0^t x^n e^{-x} \, dx \\ &= \Big[\frac{1}{\Gamma(n+1)} x^n (-e^{-x})\Big)_0^t + \frac{1}{\Gamma(n+1)} \int_0^t n x^{n-1} e^{-x} \\ &= \frac{1}{\Gamma(n+1)} t^n e^{-t} + \frac{1}{n\Gamma(n)} \int_0^t n x^{n-1} e^{-x} = \frac{1}{\Gamma(n+1)} e^{-t} t^n + g_n(t).\end{aligned}$$

3. Comme les X_n sont positifs, la suite $(S_n)_{n \geq 0}$ est croissante. Il s'ensuit que $\{N_t = n\} = \{S_n < t \leq S_{n+1}\} = \{S_n < t\} \backslash \{S_{n+1} < t\}$. Comme il y a inclusion entre ces deux ensembles, il vient
$\mathbb{P}(N_t = n) = \mathbb{P}(S_n < t) - \mathbb{P}(S_{n+1} < t)$ et donc pour tout $n \geq 1$:
$$\begin{aligned}\mathbb{P}(N_t = n) &= \mathbb{P}(S_n < t) - \mathbb{P}(S_{n+1} < t) = g_n(\lambda t) - g_{n+1}(\lambda t) \\ &= \frac{1}{\Gamma(n+1)} e^{\lambda t} (\lambda t)^n = \frac{1}{n!} e^{-\lambda t} (\lambda t)^n.\end{aligned}$$
Pour $n = 0$, on a $\mathbb{P}(N_t = 0) = \mathbb{P}(S_1 \geq t) = \mathbb{P}(X_1 \geq t) = e^{-\lambda t}$. On a bien montré que N_t suit une loi de Poisson de paramètre λt.

D.7 Exercices sur les espaces L^p

Solution 176 1. Supposons $0 < p < p' \le p_0$. Soient α, β avec $\alpha > 1$ et $\frac{1}{\alpha} + \frac{1}{\beta} = 1$. La variable X étant positive, on a d'après l'inégalité de Hölder

$$\mathbb{E}[X^p.1] \le \mathbb{E}[X^{p\alpha}]^{1/\alpha}\mathbb{E}[1^\beta]^{1/\beta},$$

soit $\mathbb{E}[X^p] \le \mathbb{E}[X^{p\alpha}]^{1/\alpha}$, ou encore $(\mathbb{E}[X^p])^{1/p} \le \mathbb{E}[X^{p\alpha}]^{1/(p\alpha)}$. Il suffit de prendre α tel que $p\alpha = p'$, soit $\alpha = p'/p > 1$, et on obtient le résultat voulu, à savoir $N(p) \le N(p')$.

2. $p \mapsto N(p)$ est croissante, minorée par 0, donc admet une limite lorsque p tend vers 0 par valeurs supérieures. Dans le cas où X suit la loi uniforme sur $[0,1]$, on a pour $p > 0$, avec le théorème de transfert : $\mathbb{E}[X^p] = \int_{\mathbb{R}} \mathbb{1}_{[0,1]}(x) x^p \, d\lambda(x) = \int_0^1 x^p \, dx = \frac{1}{p+1}$, d'où
$$N(p) = (1/(p+1))^{1/p} \text{ et } \log N(p) = \frac{1}{p}\log(\frac{1}{p+1}) = -\frac{1}{p}\log(1+p).$$
Or on a l'équivalent en 0 : $\log(1+p) \sim p$, donc $\lim\limits_{p\to 0^+} \log N(p) = -1$,

d'où par continuité de l'exponentielle $\lim\limits_{p\to 0^+} N(p) = e^{-1}$.

3. F est $(\mathbb{R}_+^2, \mathcal{B}(\mathbb{R}_+^2)) - (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable, comme produit de deux fonctions mesurables. Comme F est mesurable positive, on peut appliquer le théorème de Tonelli :
$$\int_{\mathbb{R}_+\times\mathbb{R}_+} F(x,t)\, d(\mathbb{P}_X \otimes \lambda)(x,t) = \int_{\mathbb{R}_+} \left(\int_{\mathbb{R}_+} F(x,t)\, d\lambda(t)\right) d\mathbb{P}_X(x).$$
Or,
$$\begin{aligned}\int_{\mathbb{R}_+} F(x,t)\, d\lambda(t) &= \int_{\mathbb{R}_+} pt^{p-1}\mathbb{1}_{\{t\le x\}}\, d\lambda(t)\\ &= \int_0^x pt^{p-1}\, dt = x^p.\end{aligned}$$
D'où
$$\begin{aligned}\int_{\mathbb{R}_+} \left(\int_{\mathbb{R}_+} F(x,t)\, d\lambda(t)\right) d\mathbb{P}_X(x) &= \int_{\mathbb{R}_+} x^p \, d\mathbb{P}_X(x)\\ &= \int_\Omega X^p \, d\mathbb{P}(\omega)\\ &= \mathbb{E}[X^p] \le 1 + \mathbb{E}[X^{p_0}] < +\infty,\end{aligned}$$
ce qui donne la première assertion. D'un autre côté, toujours d'après

Tonelli, on a

$$\int_{\mathbb{R}_+\times\mathbb{R}_+} F(x,t)\ d(\mathbb{P}_X\otimes\lambda)(x,t) \quad = \quad \int_{\mathbb{R}_+}\left(\int_{\mathbb{R}_+} F(x,t)\ d\mathbb{P}_X(x)\right)\ d\lambda(t).$$

Or,

$$\begin{aligned}\int_{\mathbb{R}_+} F(x,t)\ d\mathbb{P}_X(x) &= \int_{\mathbb{R}_+} pt^{p-1}\mathbb{1}_{\{t\le x\}}\ d\mathbb{P}_X(x)\\ &= pt^{p-1}\int_{\mathbb{R}_+}\mathbb{1}_{\{t\le x\}}\ d\mathbb{P}_X(x)\\ &= pt^{p-1}\int_{\mathbb{R}_+}\mathbb{1}_{[t,+\infty[}(x)\ d\mathbb{P}_X(x)\\ &= pt^{p-1}\mathbb{P}_X([t,+\infty[) = pt^{p-1}\mathbb{P}(X\ge t),\end{aligned}$$

ce qui donne $\mathbb{E}[X^p] = \int_{[0,+\infty[} pt^{p-1}\mathbb{P}(X\ge t)\ d\lambda(t)$.

4. (a) Par définition de $\|X\|_{\infty,\text{ess}}$, on a $\mathbb{P}(X\ge M)=0$, d'où $0\le X\le M$ $\mathbb{P}$-presque sûrement, et $0\le X^p\le M^p$ $\mathbb{P}$-presque sûrement. On obtient donc $0\le\mathbb{E}[X^p]\le M^p$, et en élevant à la puissance $1/p$, $N(p)\le M<+\infty$ ce qui entraîne que X est dans L^p.

   (b) Soit $M>\|X\|_{\infty,\text{ess}}$. Comme pour tout $p>0$, on a $N(p)\le M$, il s'ensuit que $\overline{\lim\limits_{p\to+\infty}}\ N(p)\le M$. En passant à la borne inférieure sur tous les $M>\|X\|_{\infty,\text{ess}}$, on obtient $\overline{\lim\limits_{p\to+\infty}}\ N(p)\le\|X\|_{\infty,\text{ess}}$.

   (c) On suppose $M<\|X\|_{\infty,\text{ess}}$. Par définition de $\|X\|_{\infty,\text{ess}}$, il existe un $M'\ge M$ tel que $\mathbb{P}(X\ge M')>0$, ce qui entraîne $\mathbb{P}(X\ge M)>0$. D'après l'inégalité de Markov, on a

$$\mathbb{P}(X\ge M)=\mathbb{P}(X^p\ge M^p)\le\frac{\mathbb{E}[X^p]}{M^p},$$

   soit $\mathbb{E}[X^p]\ge M^p\mathbb{P}(X\ge M)$, soit encore $N(p)\ge M\mathbb{P}(X\ge M)^{1/p}$. On en déduit

$$\underline{\lim}_{p\to+\infty}\ N(p)\ge\underline{\lim}_{p\to+\infty}\ M\mathbb{P}(X\ge M)^{1/p}.$$

   Or $\lim\limits_{p\to+\infty} M\mathbb{P}(X\ge M)^{1/p}=M$ (en effet pour tout $\alpha>0$, on a $\lim\limits_{p\to+\infty}\alpha^{1/p}=1$), d'où $\underline{\lim}_{p\to+\infty}\ N(p)\ge M$. Ainsi, pour tout $M<\|X\|_{\infty,\text{ess}}$, on a $\underline{\lim}_{p\to+\infty}\ N(p)\ge M$. En passant à la borne supérieure sur tous les $M<\|X\|_{\infty,\text{ess}}$, on obtient

$$\underline{\lim}_{p\to+\infty} N(p) \geq \|X\|_{\infty,\text{ess}}.$$

(d) On a $\|X\|_{\infty,\text{ess}} \leq \underline{\lim}_{p\to+\infty} N(p) \leq \overline{\lim}_{p\to+\infty} N(p) \leq \|X\|_{\infty,\text{ess}}$, d'où

$$\|X\|_{\infty,\text{ess}} = \underline{\lim}_{p\to+\infty} N(p) = \overline{\lim}_{p\to+\infty} N(p),$$

soit encore $\lim_{p\to+\infty} N(p) = \|X\|_{\infty,\text{ess}}$.

5. a) Traitons d'abord le cas où $\log X$ est intégrable.
On a $\log N(p) = \frac{1}{p}\log \mathbb{E}[X^p] = \frac{1}{p}\log(1+p\phi(p))$, avec

$$\phi(p) = \mathbb{E}\left[\frac{X^p-1}{p}\right] = \int_{]0,+\infty[} \frac{\exp(p\log x)-1}{p}\, d\mathbb{P}_X(x).$$

Pour tout x, on a $|e^x-1| \leq |x|\max(1,e^x)$ grâce à l'inégalité des accroissements finis, d'où $|\exp(p\log x)-1| \leq p|\log x|\max(x^p,1)$. Ainsi, pour $p \leq p_0/2$, comme $|\log x| \leq \frac{2}{p_0}x^{p_0/2}$ pour $x \geq 1$, on a

$$\begin{aligned}
\left|\frac{\exp(p\log x)-1}{p}\right| &\leq |\log x|\max(x^{p_0/2},1)\\
&\leq |\log x|\mathbb{1}_{\{x\leq 1\}} + \frac{2x^{p_0}\mathbb{1}_{\{x>1\}}}{p_0}.
\end{aligned}$$

Or $\lim_{p\to 0^+} \frac{\exp(p\log x)-1}{p} = \log x$, donc d'après le théorème de convergence dominée

$$\lim_{p\to 0^+}\phi(p) = \int \log x\, d\mathbb{P}_X(x) = \mathbb{E}[\log X],$$

d'où, avec un développement limité, $\lim_{p\to 0^+} \log N(p) = \mathbb{E}[\log X]$, ce qui entraîne

$$\lim_{p\to 0^+} N(p) = \exp(\mathbb{E}[\log X]).$$

b) Passons au cas où $\log X$ n'est pas intégrable. Comme X^{p_0} est intégrable, $(\log X)^+$ est intégrable ; c'est donc que $\mathbb{E}[(\log X)^-] = +\infty$.
Soit $N \geq 1$. On remarque que

$$0 \leq (\mathbb{E}[X^p])^{1/p} \leq (\mathbb{E}[\max(X,1/N)^p])^{1/p},$$

donc en faisant tendre p vers 0, on a

$$0 \leq \overline{\lim}_{p\to 0^+} N(p) \leq \exp(\mathbb{E}[\log(\max(X,1/N))]).$$

On a

$$\begin{aligned} & \mathbb{E}[\log(\max(X, 1/N))] \\ = \ & \mathbb{E}[\log(\max(X, 1/N))^+] - \mathbb{E}[\log(\max(X, 1/N))^-] \\ = \ & \mathbb{E}[(\log X)^+] - \mathbb{E}[\log(\max(X, 1/N))^-]. \end{aligned}$$

D'après le théorème de convergence monotone, on a

$$\lim_{N\to+\infty} \mathbb{E}[\log(\max(X, 1/N))^-] = \mathbb{E}[(\log X)^-] = +\infty,$$

ce qui nous donne $\lim\limits_{N\to+\infty} \mathbb{E}[\log(\max(X, 1/N))] = -\infty$, d'où

$\lim\limits_{p\to 0^+} N(p) = 0.$

Solution 177 Soit $x \in]0,1[$. Pour montrer que $x \in \overline{\lim\limits_{n\to+\infty}} A_n$, il faut montrer que pour tout $N \in \mathbb{N}$, il existe $n \geq N$ avec $x \in A_n$. Soit donc $N \in \mathbb{N}$. Soit $N' \geq N$ tel que $n \geq N$ implique $x_n < \min(1-x, x)$. La première question de l'exercice 6 nous dit qu'il existe $n \geq N'$ avec $\{s_n\} \leq x \leq \{s_{n+1}\}$, ce qui dit exactement que $x \in A_n$. Ainsi pour tout $x \in]0,1[$, la limite supérieure de $f_n(x)$ est 1. Soit N tel que $x_n < 1$ pour $n \geq N$.
Si $\{s_n\} \leq \{s_{n+1}\}$, alors $s_n - \lfloor s_n \rfloor \leq s_n + x_{n+1} - \lfloor s_{n+1} \rfloor$, ce qui implique $\lfloor s_{n+1} \rfloor - \lfloor s_n \rfloor \leq x_{n+1}$, d'où $\lfloor s_{n+1} \rfloor - \lfloor s_n \rfloor = 0$, ce qui entraîne alors l'égalité $\lambda(A_n) = \{s_{n+1}\} - \{s_n\} = s_{n+1} - s_n = x_{n+1}$. Ainsi pour tout $n \geq N$, $\int |f_n| \, d\lambda = \lambda(A_n) \leq x_{n+1}$, ce qui montre que f_n tend dans L^1 vers zéro.

Soit $x \in [0,1[$. Comme $x \in \overline{\lim\limits_{n\to+\infty}} A_n$, il existe une suite (n_k) avec $\{s_{n_k}\} \leq x \leq \{s_{n_{k+1}}\}$, donc $|x - s_{n_k}| = x - s_{n_k} \leq s_{n_{k+1}} - s_{n_k} = x_{n_{k+1}}$, ce qui montre que $\{s_{n_k}\}$ tend vers x : x est valeur d'adhérence de la suite. Finalement l'ensemble des valeurs d'adhérence contient $]0,1[$. Mais l'ensemble des valeurs d'adhérence est un fermé donc l'ensemble des valeurs d'adhérence contient $[0,1]$. Comme il est évidemment inclus dans $[0,1]$, c'est $[0,1]$.

Solution 178 1. La suite $(B_{k,n})_{n\geq 1}$ est décroissante. Son intersection est l'ensemble des points x où $f_n(x)$ ne converge pas vers $f(x)$, qui est de mesure nulle, puisque f_n converge presque partout vers f : d'après le théorème de continuité séquentielle décroissante, il existe n_k tel que $\mu(B_{k,n}) \leq \frac{\varepsilon}{2^k}$.
En posant $A = \bigcup\limits_{k=1}^{+\infty} B_{k,n_k}$, on a alors $\mu(A) \leq \sum\limits_{k=1}^{+\infty} \mu(B_{k,n_k}) \leq \varepsilon$. Soit A^c le complémentaire de A. Soient $\varepsilon > 0$ et $k > 1/\varepsilon$. Pour $i \geq n_k$,

on a $|f_i(x) - f(x)| \le \frac{1}{k} \le \varepsilon$ pour tout $x \in A^c$. Ainsi, f_n converge uniformément vers f sur A^c.

2. Soit f_n une suite de fonctions continues sur $[a,b]$ convergeant vers f dans $L^1([a,b])$. On peut en extraire une sous-suite (f_{n_k}) qui converge presque partout vers f. D'après la question précédente, il existe un ensemble E dont le complémentaire a une mesure qui ne dépasse pas ε sur lequel f_{n_k} converge uniformément. Comme limite uniforme de fonctions continues, la limite des f_{n_k} sur E, soit f, est continue.

3. Pour n assez grand, la mesure des points x tels que $|f| > n$ est inférieure à ε d'après le théorème de continuité séquentielle décroissante. On applique alors la question précédente à $|f|\mathbb{1}_{\{|f|\le n\}}$, qui est bien sûr intégrable.

Solution 179 1. Sur $]0,+\infty[$, $\frac{\Gamma'}{\Gamma}$ est la dérivée de $\log\Gamma$. On peut se souvenir qu'on a vu en 7.1.1, que $\log\Gamma$ est convexe, donc sa dérivée est positive. Cependant, il est peut-être plus naturel d'observer le signe de la dérivée de $\frac{\Gamma'}{\Gamma}$. On a

$$(\frac{\Gamma'}{\Gamma})' = \frac{\Gamma''\Gamma - (\Gamma')^2}{\Gamma^2}.$$

Or l'inégalité de Cauchy-Schwarz donne

$$\begin{aligned}
|\Gamma'(x)|^2 &= \left(\int_{]0,+\infty[}(-\log t)t^{x-1}e^{-t}\,d\lambda(t)\right)^2\\
&= \left(\int_{]0,+\infty[}(-\log t)t^{\frac{x-1}{2}}e^{-t/2}.t^{\frac{x-1}{2}}e^{-t/2}\,d\lambda(t)\right)^2\\
&\le \left(\int_{]0,+\infty[}(-\log t)^2t^{x-1}e^{-t}\,d\lambda(t)\right)\left(\int_{]0,+\infty[}t^{x-1}e^{-t}\,d\lambda(t)\right)\\
&= \Gamma''(x)\Gamma(x).
\end{aligned}$$

La fonction $\frac{\Gamma'}{\Gamma}$ est donc bien croissante sur $]0,+\infty[$.

2. En dérivant l'identité $\Gamma(x+1) = x\Gamma(x)$, on obtient

$$\Gamma'(x+1) = \Gamma(x) + x\Gamma'(x),$$

et en divisant la deuxième identité par la première, on a

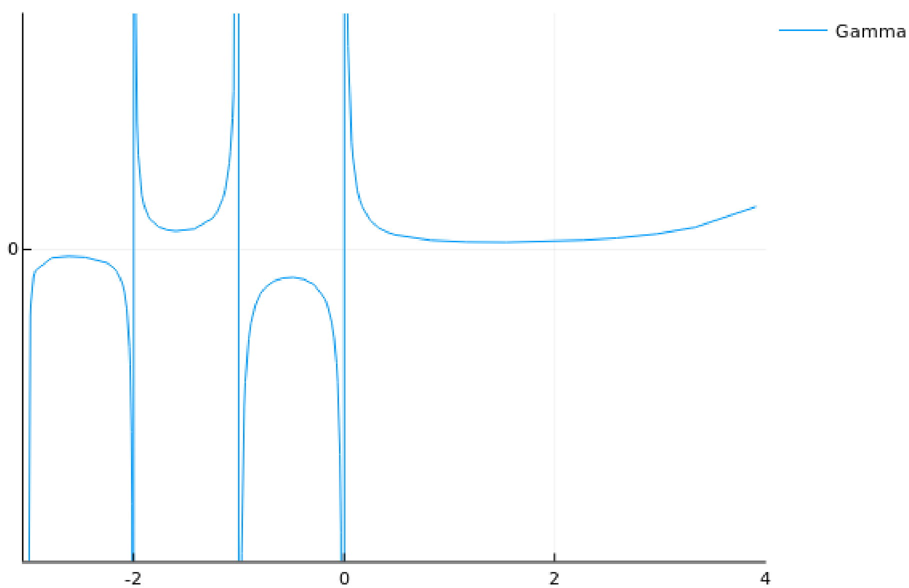
$$\frac{\Gamma'(x+1)}{\Gamma(x+1)} = \frac{1}{x} + \frac{\Gamma'(x)}{\Gamma(x)},$$

que l'on peut réécrire

$$\frac{\Gamma'(x)}{\Gamma(x)} = -\frac{1}{x} + \frac{\Gamma'(x+1)}{\Gamma(x+1)}.$$

Comme la fonction $x \mapsto -\frac{1}{x}$ est croissante, on en déduit par récurrence que pour tout $n \geq -1$, $\frac{\Gamma'}{\Gamma}$ est croissante sur $]-(n+1), -n[$.

3. Comme $\Gamma(x) > 0$ sur $]0, +\infty[$, l'identité $\Gamma(x+1) = x\Gamma(x)$ permet de montrer par récurrence que pour tout $n \geq -1$, Γ a le signe de $(-1)^{n+1}$ sur $]-(n+1), -n[$. Comme $\frac{\Gamma'}{\Gamma}$ est croissante sur $]-(n+1), -n[$, sa dérivée est positive. Or elle a le signe de $\Gamma\Gamma'' - (\Gamma')^2$. Ainsi on trouve que $\Gamma\Gamma'' > 0$ sur $]-(n+1), -n[$. La fonction Γ est donc concave sur $]-(n+1), -n[$ quand n est pair, et convexe sur $]-(n+1), -n[$ quand n est impair. On sait déjà que $\lim_{x\to 0^+} \Gamma(x) = +\infty$. Au voisinage de 0 à gauche, comme $\Gamma(x) = \frac{\Gamma(x+1)}{x} \sim \frac{1}{x}$, $\lim\limits_{x\to 0^-} \Gamma(x) = -\infty$. Par récurrence, comme $\Gamma(x) = \frac{\Gamma(x+1)}{x}$, il s'ensuit que pour n pair, $\lim\limits_{x\to n^+} \Gamma(x) = +\infty$ et $\lim\limits_{x\to n^-} \Gamma(x) = -\infty$, tandis que pour n impair, $\lim\limits_{x\to n^+} \Gamma(x) = -\infty$ et $\lim\limits_{x\to n^-} \Gamma(x) = +\infty$. En mettant ensemble les deux résultats, on voit que pour n pair, Γ est strictement négative et concave sur $]-(n+1), n[$, d'abord croissante, puis décroissante, avec une limite $-\infty$ à la fois en $-(n+1)$ à droite et en $-n$ à gauche. Quand n est impair, c'est l'inverse : Γ est strictement positive et convexe sur $]-(n+1), n[$, d'abord décroissante, puis croissante, avec une limite $+\infty$ à la fois en $-(n+1)$ à droite et en $-n$ à gauche.



D.8 Exercices sur la convolution et Fourier

Solution 193 1. — Première solution. Il s'agit de calculer

$$f * g(x) = \int_{\mathbb{R}} \exp\left\{-\frac{(x-t)^2}{2a^2}\right\} \exp\left\{-\frac{t^2}{2b^2}\right\}\, d\lambda(t)$$
$$= \int_{\mathbb{R}} \exp\left\{-\left(\frac{(x-t)^2}{2a^2} + \frac{t^2}{2b^2}\right)\right\}\, d\lambda(t).$$

On commence par exprimer $\frac{(x-t)^2}{2a^2} + \frac{t^2}{2b^2}$, vu comme un polynôme du second degré t, sous forme canonique. Après calculs, on trouve

$$\frac{(x-t)^2}{2a^2} + \frac{t^2}{2b^2} = \frac{a^2+b^2}{2a^2b^2}\left(t - \frac{b^2}{a^2+b^2}x\right)^2 + \frac{x^2}{2(a^2+b^2)},$$

ce qui nous donne

$$\begin{aligned}
&(f*g)(x)\\
&= \exp\left\{-\frac{x^2}{2(a^2+b^2)}\right\} \int_{\mathbb{R}} \exp\left\{-\frac{a^2+b^2}{2a^2b^2}\left(t - \frac{b^2}{a^2+b^2}x\right)^2\right\}\, d\lambda(t)\\
&= \exp\left\{-\frac{x^2}{2(a^2+b^2)}\right\} \int_{\mathbb{R}} \exp\left\{-\frac{a^2+b^2}{2a^2b^2}t^2\right\}\, d\lambda(t)\\
&= \exp\left\{-\frac{x^2}{2(a^2+b^2)}\right\} \frac{ab}{\sqrt{a^2+b^2}} \int_{\mathbb{R}} \exp\left\{-\frac{t^2}{2}\right\}\, d\lambda(t)\\
&= \exp\left\{-\frac{x^2}{2(a^2+b^2)}\right\} \frac{ab\sqrt{2\pi}}{\sqrt{a^2+b^2}}.
\end{aligned}$$

— Deuxième solution. On remarque que $f(x) = \sqrt{2\pi a^2} f_X(x)$ et $g(x) = \sqrt{2\pi b^2} f_Y(x)$, où f_X et f_Y sont les densités de deux variables aléatoires indépendantes X et Y suivant respectivement les lois $\mathcal{N}(0, a^2)$ et $\mathcal{N}(0, b^2)$ On a alors

$$(f*g)(x) = \sqrt{2\pi a^2}\sqrt{2\pi b^2} f_X * f_Y(x) = \sqrt{2\pi a^2}\sqrt{2\pi b^2} f_{X+Y}(x),$$

où f_{X+Y} est la densité de $X+Y$. Ceci vient de la formule de la densité de la somme de deux variables aléatoires indépendantes à densité. Mais on sait que la loi de $X+Y$ est $\mathcal{N}(0, a^2+b^2)$, donc

$$f_{X+Y}(x) = \frac{1}{\sqrt{2\pi(a^2+b^2)}} \exp\left(-\frac{x^2}{2(a^2+b^2)}\right),$$

ce qui redonne le résultat.

2. On a

$$\begin{aligned}
(f*g)(x) &= \int_{\mathbb{R}} \mathbb{1}_{[-a,a]}(x-t)\mathbb{1}_{[-b,b]}(t) = \int_{\mathbb{R}} \mathbb{1}_{[-a,a]}(t-x)\mathbb{1}_{[-b,b]}(t) \\
&= \int_{\mathbb{R}} \mathbb{1}_{[-a+x,a+x]}(t)\mathbb{1}_{[-b,b]}(t) = \lambda([-a+x,a+x]\cap[-b,b]) \\
&= \lambda([-a+x,+\infty[\cap]-\infty,a+x]\cap[-b,+\infty[\cap]-\infty,b]) \\
&= \lambda([\max(-a+x,-b),+\infty[\cap]-\infty,\min(a+x,b)]) \\
&= (\min(a+x,b)-\max(-a+x,-b))^+ \\
&= (a+\min(x,b-a)-(-a+\max(x,a-b)))^+ \\
&= (2a+\min(x,b-a)-\max(x,a-b)))^+.
\end{aligned}$$

Le produit de convolution commutant, on peut supposer que $b-a\geq 0$. Si $x\geq b-a\geq 0$, on a donc

$$(2a+\min(x,b-a)-\max(x,a-b)))^+ = (2a+(b-a)-x)^+ = (a+b-x)^+.$$

De même, si $x\leq -(b-a)\leq 0$, on a

$$(2a+\min(x,b-a)-\max(x,a-b)))^+ = (2a+x-(a-b))^+ = (a+b+x)^+,$$

tandis que pour $x\in]-(b-a),b-a[$,

$$(2a+\min(x,b-a)-\max(x,a-b)))^+ = (2a+x-x)^+ = 2a.$$

Finalement on obtient

$$(f*g)(x) = \begin{cases} 2a & \text{si } |x|\leq b-a \\ a+b-|x| & \text{si } b-a<|x|<a+b \\ 0 & \text{sinon} \end{cases}$$

Solution 194 1. On a $0\leq g_n(x)\leq \mathbb{1}_{[-1,1]}(x)$ et $\lim_{n\to+\infty} g_n(x)=0$ pour tout $x\neq 0$. Comme $x\mapsto \mathbb{1}_{[-1,1]}(x)$ est évidemment intégrable, la convergence de (a_n) vers 0 découle immédiatement du théorème de convergence dominée. On a

$$\begin{aligned}
a_n = \int_{-1}^{1}(1-x^2)^n\,dx = 2\int_0^1 (1-x^2)^n\,dx \geq \quad & 2\int_0^1 (1-x)^n\,dx \\
&= 2\int_0^1 x^n\,dx = \frac{2}{n+1}.
\end{aligned}$$

2. Comme $\int_{\mathbb{R}} k_n(t)\,d\lambda(t)=1$, on a

$$\begin{aligned}
f*k_n(x)-f(x) &= \Big(\int_{\mathbb{R}} f(x-t)k_n(t)\,d\lambda(t)\Big) - \Big(\int_{\mathbb{R}} k_n(t)\,d\lambda(t)\Big)f(x) \\
&= \int_{\mathbb{R}} (f(x-t)-f(x))k_n(t)\,d\lambda(t).
\end{aligned}$$

Notons $\omega_f(\delta) = \sup\{|f(x)-f(y)|; |x-y| \le \delta\}$ le module de continuité de f. Soit $\delta \in]0,1[$. On a pour tout $x \in \mathbb{R}$

$$\begin{aligned}
|f * k_n(x) - f(x)| &\le \int_{\mathbb{R}} |f(x-t) - f(x)| k_n(t)\, d\lambda(t) \\
&\le \int_{\mathbb{R}} \left(\omega_f(\delta) + 2\|f\|_\infty \mathbb{1}_{\{|t|>\delta\}}\right) k_n(t)\, d\lambda(t) \\
&\le \omega_f(\delta) + 4\|f\|_\infty \int_\delta^1 \frac{(1-t^2)^n}{a_n}\, dt \\
&\le \omega_f(\delta) + 4\|f\|_\infty \frac{(1-\delta^2)^n}{a_n} \\
&\le \omega_f(\delta) + 8(n+1)\|f\|_\infty (1-\delta^2)^n.
\end{aligned}$$

En passant au supremum en x, on a
$\|f * k_n - f\|_\infty \le \omega_f(\delta) + 8(n+1)\|f\|_\infty (1-\delta^2)^n$, puis en passant à la limite supérieure en n :

$$\overline{\lim_{n\to+\infty}} \; \|f * k_n - f\|_\infty \le \omega_f(\delta).$$

Comme f est uniformément continue, $\lim_{\delta \to 0} \omega_f(\delta) = 0$, ce qui donne le résultat voulu.

3. Pour tout $x \in \mathbb{R}$, on obtient

$$\begin{aligned}
(f * k_n)(x) &= \int_{\mathbb{R}} f(t) k_n(x-t)\, d\lambda(t) = \frac{1}{a_n} \int_{\mathbb{R}} f(t) g_n(x-t)\, d\lambda(t) \\
&= \frac{1}{a_n} \int_{[-1/2,1/2]} f(t) g_n(x-t)\, d\lambda(t).
\end{aligned}$$

Lorsque $|x| \le 1/2$, on a $|x-t| \le 1$ pour tout $t \in [-1/2, 1/2]$, donc

$$(f * k_n)(x) = \frac{1}{a_n} \int_{[-1/2,1/2]} f(t)(1-(x-t)^2)^n\, d\lambda(t).$$

Or $(1-(x-t)^2)^n$ se développe en

$$\sum_{k=0}^n (-1)^k \binom{n}{k} (x-t)^{2k} = \sum_{k=0}^n (-1)^k \binom{n}{k} \sum_{i=0}^{2k} \binom{2k}{i} (-t)^{2k-i} x^i,$$

d'où en posant $c_k = \int_{[-1/2,1/2]} t^k f(t)\, d\lambda(t)$:

$$(f * k_n)(x) = \sum_{k=0}^n (-1)^k \binom{n}{k} \sum_{i=0}^{2k} \binom{2k}{i} (-1)^{2k-i} \frac{c_{2k-1}}{a_n} x^i,$$

qui est bien un polynôme en x de degré $2n$.

4. Soit f une fonction continue sur $[-1/2, 1/2]$. On peut la prolonger de

manière classique en une fonction $\tilde{f}$ continue, nulle à l'extérieur de $[-1,1]$, affine sur $[-1,-1/2]$ et $[1/2,1]$. $\tilde{f}$ est uniformément continue sur $\mathbb{R}$, avec $\omega_{\tilde{f}}(\delta) \leq \max(\omega_f(\delta), 2\delta|f(1/2)|, 2\delta|f(-1/2)|)$. D'après ce qui précède, $(\tilde{f} * k_n)(x)$ est une suite de polynômes qui converge uniformément vers f sur $[-1/2,1/2]$. Passons au cas général : soit f une fonction continue sur $[a,b]$, et notons $g(\theta) = (b-a)\theta + \frac{a+b}{2}$. $f \circ g$ est une fonction continue sur $[-1/2,1/2]$, donc il existe une suite (P_n) de polynômes qui converge uniformément vers $f \circ g$ sur $[-1/2,1/2]$. Posons alors $Q_n = P_n \circ g^{-1}$. Q_n est un polynôme avec

$$\begin{aligned}\sup_{x \in [a,b]} |Q_n(x) - f(x)| &= \sup_{x \in [a,b]} |P_n(g^{-1}(x)) - (f \circ g)(g^{-1}(x))| \\ &= \sup_{y \in [-1/2,1/2]} |P_n(y) - (f \circ g)(y)|,\end{aligned}$$

qui tend bien vers 0 lorsque n tend vers l'infini.

Solution 195 1. Posons $f(x) = \mathbb{1}_E * \mathbb{1}_{-E}$. On a $\forall x, y \in \mathbb{R}$

$$f(x) - f(y) = \int_{\mathbb{R}} (\mathbb{1}_{-E}(x-t) - \mathbb{1}_{-E}(y-t))\mathbb{1}_E(t)\, d\lambda(t),$$

donc quels que soient x et y réels, on a

$$|f(x) - f(y)| \leq \int_{\mathbb{R}} |\mathbb{1}_E(t-x) - \mathbb{1}_E(t-y)|\, d\lambda(t) = \|T_x\mathbb{1}_E - T_y\mathbb{1}_E\|_1.$$

Or pour $g \in L^1$ l'application $x \mapsto T_x g$ est continue de $\mathbb{R}$ dans L^1, ce qui donne la continuité de f.

2. $f(0) = \int_{\mathbb{R}} \mathbb{1}_E(t)\mathbb{1}_{-E}(0-t)\, d\lambda(t) = \int_{\mathbb{R}} \mathbb{1}_E(t)^2\, d\lambda(t) = \lambda(E) > 0$. Par continuité de f, il existe α tel que $f(x) > 0$ pour $x \in [-\alpha, \alpha]$. Mais si $f(x) > 0$, la mesure de Lebesgue des t tels que

$$\mathbb{1}_E(x-t)\mathbb{1}_{-E}(t) = \mathbb{1}_E(x-t)\mathbb{1}_E(-t) > 0$$

est strictement positive. En particulier, il existe t tel que $x - t \in E$ et $-t \in E$, donc $x = x - t - (-t) \in E - E$.

D.9 Exercices sur les fonctions caractéristiques

Solution 203 1. Pour tout $s \geq 0$, on a

$$\mathcal{L}_{\Gamma(r,\lambda)}(t) = \int_{-\infty}^{+\infty} e^{-tx}\gamma_{r,\lambda}(x)\, dx = \frac{\lambda^r}{(\lambda+t)^r}\int_0^{+\infty}\gamma_{r+t,\lambda}(x)\, dx = \frac{\lambda^r}{(\lambda+t)^r}.$$

2. Comme X et Y sont indépendantes, on a l'égalité

$$\mathcal{L}_{X+Y}(t) = \mathcal{L}_X(t)\mathcal{L}_Y(t) = \frac{\lambda^{r+s}}{(\lambda+t)^{r+s}} = \mathcal{L}_{\Gamma(r+s,\lambda)}(t),$$

ce qui permet de conclure puisque la transformée de Laplace caractérise la loi des variables aléatoires positives.

3. Avec la question précédente, une récurrence immédiate donne que $X_1 + \cdots + X_n$ suit la loi $\Gamma(n, \lambda)$.

4. On a

$$\begin{aligned}\mathcal{L}_{X^2}(t) &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{-tx^2}\, d\lambda(x) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2(1+2t)}{2}}\, d\lambda(x)\\ &= \frac{1}{\sqrt{1+2t}} = \mathcal{L}_{1/2,1/2}(t),\end{aligned}$$

donc la loi de X_1^2 est $\Gamma(1/2, 1/2)$. Avec la question précédente, on conclut que la loi de $X_1^2 + \cdots + X_n^2$ est donc la loi $\Gamma(n/2, 1)$.

Solution 204 1. On a $t^S\, \mathbb{1}_{\{T=n\}} = t^{S_n}\, \mathbb{1}_{\{T=n\}}$, donc on obtient par indépendance $\mathbb{E}[t^S\, \mathbb{1}_{\{T=n\}}] = \mathbb{E}[t^{S_n}\, \mathbb{1}_{\{T=n\}}] = \mathbb{E}[t^{S_n}]\mathbb{E}[\mathbb{1}_{\{T=n\}}]$.
Comme les X_i sont indépendantes identiquement distribuées, on a $\mathbb{E}[t^{S_n}] = G_{S_n}(t) = G_{X_1}(t)^n$. Finalement,

$$\begin{aligned}\mathbb{E}[t^S\, \mathbb{1}_{\{T=n\}}] &= G_{X_1}(t)^n \mathbb{E}[\, \mathbb{1}_{\{T=n\}}]\\ &= \mathbb{E}[G_{X_1}(t)^n\, \mathbb{1}_{\{T=n\}}] = \mathbb{E}[G_{X_1}(t)^T\, \mathbb{1}_{\{T=n\}}].\end{aligned}$$

En sommant sur $n \in \mathbb{N}$ les deux extrémités, le théorème de Tonelli nous donne $\mathbb{E}[t^S] = \mathbb{E}[G_{X_1}(t)^T] = G_T(G_{X_1}(t))$, soit $G_S = G_T \circ G_{X_1}$.

2. Comme X_1 et T ont un moment d'ordre 1, G_{X_1} et G_T sont dérivables à gauche en 1, avec $G'_{X_1}(1) = \mathbb{E}[X_1]$ et $G'_T(1) = \mathbb{E}[T]$. Or $G_{X_1}(1) = 1$, donc la formule précédente entraîne que G_S est dérivable à gauche au point 1, avec $G'_S(1) = G'_T(1)G'_{X_1}(1) = \mathbb{E}[T]\mathbb{E}[X_1]$. Cela entraîne que S est intégrable, avec $\mathbb{E}[S] = \mathbb{E}[T]\mathbb{E}[X_1]$.

Solution 205 1. Soit ψ une fonction continue bornée. On a par indépendance de ε et X

$$\begin{aligned}\mathbb{E}\psi(\varepsilon X) &= \mathbb{E}[\, \mathbb{1}_{\{\varepsilon=1\}}\psi(X) + \mathbb{1}_{\{\varepsilon=-1\}}\psi(-X)]\\ &= \mathbb{E}[\, \mathbb{1}_{\{\varepsilon=1\}}]\mathbb{E}[\psi(X)] + \mathbb{E}[\, \mathbb{1}_{\{\varepsilon=-1\}}]\mathbb{E}[\psi(-X)]\\ &= \frac{1}{2}\mathbb{E}[\psi(X)] + \frac{1}{2}\mathbb{E}[\psi(-X)]\\ &= \frac{1}{2}\int_{\mathbb{R}_+} e^{-x}\psi(x)\, d\lambda(x) + \frac{1}{2}\int_{\mathbb{R}_+} e^{-x}\psi(-x)\, d\lambda(x)\\ &= \frac{1}{2}\int_{\mathbb{R}_+} e^{-x}\psi(x)\, d\lambda(x) + \frac{1}{2}\int_{\mathbb{R}_-} e^{x}\psi(x)\, d\lambda(x)\\ &= \frac{1}{2}\int_{\mathbb{R}} e^{-|x|}\psi(x)\, d\lambda(x).\end{aligned}$$

Cela entraîne que εX admet la densité $x \mapsto \frac{1}{2}e^{-|x|}$.

2. On prend $\psi(x) = e^{itx}$. Avec la formule de la troisième ligne, on voit que $\phi_{\varepsilon X}(t) = \frac{1}{2}\left(\phi_X(t) + \phi_X(-t)\right)$. Or
$$\phi_X(t) = \int_{\mathbb{R}_+} e^{itx}e^{-x}\, d\lambda(x) = \frac{1}{1-it},$$
d'où
$$\phi_{\varepsilon X}(t) = \frac{1}{2}\left(\frac{1}{1-it} + \frac{1}{1+it}\right) = \frac{1}{1+t^2}.$$

Solution 206 1. La fonction caractéristique de U est donnée par
$$\begin{aligned}
\phi_U(t) &= \mathbb{E}[e^{itX_{1,1}X_{1,2}}] = \int_{\mathbb{R}^2} e^{itxy}\, d\mathbb{P}_{(X_{1,1},X_{1,2})}(x,y) \\
&= \int_{\mathbb{R}^2} e^{itxy}\, d(\mathbb{P}_{X_{1,1}} \otimes \mathbb{P}_{X_{1,2}})(x,y) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{itxy}\, d\mathbb{P}_{X_{1,1}}(x)\right)\, d\mathbb{P}_{X_{1,2}}(y) \\
&= \int_{\mathbb{R}} \phi_{X_{1,1}}(ty)\, d\mathbb{P}_{X_{1,2}}(y) \\
&= \int_{\mathbb{R}} \exp(-\frac{1}{2}t^2y^2)\, d\mathbb{P}_{X_{1,2}}(y) \\
&= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}(t^2+1)\right)\, d\lambda(y) \\
&= \frac{1}{\sqrt{t^2+1}}.
\end{aligned}$$
Comme U et V ont même loi (qui est la loi image de $\mathcal{N}(0,1)^{\otimes 2}$ par $(x,y) \mapsto xy$), elles ont donc même fonction caractéristique.

2. Le déterminant est $U - V$. Comme U et V sont indépendantes, on a $\phi_{U-V}(t) = \mathbb{E}[e^{itU}e^{-itV}] = \mathbb{E}[e^{itU}]\mathbb{E}[e^{-itV}] = \phi_U(t)\phi_V(-t) = \frac{1}{t^2+1}$. On reconnaît la fonction caractéristique de la loi de Laplace. Comme la fonction caractéristique caractérise la loi, on en déduit que le déterminant suit la loi de Laplace.

Solution 207 1. On peut remarquer que cette première question n'est autre qu'un cas particulier du lemme 9.24. On va toutefois le rédiger de manière un peu différente. D'après le théorème de transfert discret, on a
$$\phi_X(t) = \sum_{n\in\mathbb{Z}} \mathbb{P}(X=n)e^{int}.$$
Ainsi, $\phi_X(t)$ est la somme d'une série trigonométrie normalement convergente, et il est bien connu que dans ce cas, les coefficients apparais-

sant en face des $e^{in\cdot}$ ne sont autres que les coefficients de Fourier de la fonction somme. Montrons-le.
Si on pose $f_N(t) = \sum_{n=-N}^{N} \mathbb{P}(X=n)e^{int}$, on a clairement

$\lim_{N\to+\infty} f_N(t)e^{-ikt} = \phi_X(t)e^{-ikt}$. Or

$$|f_N(t)e^{-ikt}| \le \sum_{k=-N}^{N} \mathbb{P}(X=n) \le 1,$$

et la fonction 1 étant intégrable sur $[0,2\pi]$, on peut appliquer le théorème de convergence dominée et il vient :

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \phi_X(t)e^{-ikt}\, dt = \lim_{N\to+\infty}\int_{-\pi}^{\pi} \frac{1}{2\pi} f_N(t)e^{-ikt}\, dt.$$

Or l'identité $\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i(n-k)t}\, dt = \delta_{k,n}$ entraîne que pour $N \ge |k|$, on a $\frac{1}{2\pi}\int_{-\pi}^{\pi} f_N(t)(t)e^{-ikt}\, dt = \mathbb{P}(X=k)$, et donc

$$\mathbb{P}(X=k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \phi_X(t)e^{-ikt}\, dt.$$

2. (a) Grâce à la formule de la question précédente, on a

$$\mathbb{P}(S_{2n}=0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \phi_{S_{2n}}(\theta)\, d\theta.$$

Comme S_{2n} est la somme de $2n$ variables aléatoires indépendantes ayant la même loi que X_1, on a $\phi_{S_{2n}}(\theta) = \phi_{X_1}(\theta)^{2n}$. On a de plus l'identité

$$\phi_{X_1}(\theta) = \mathbb{E}[e^{i\theta X_1}] = \mathbb{P}(X_1=-1)e^{-i\theta} + \mathbb{P}(X_1=1)e^{i\theta} = \cos\theta,$$

donc

$$\mathbb{P}(S_{2n}=0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos^{2n}(\theta)\, d\theta.$$

On a de plus

$$\sum_{n=0}^{+\infty} \mathbb{P}(S_{2n=0}) = \lim_{N\to+\infty}\sum_{n=0}^{N} \mathbb{P}(S_{2n=0}) = \lim_{N\to+\infty}\frac{1}{2\pi}\int_{-\pi}^{\pi} f_N(t)\, dt$$

avec $f_N(t) = \sum_{k=0}^{N} (\cos t)^{2n}$. Avec le théorème de convergence monotone, il vient

$$\sum_{n=0}^{+\infty} \mathbb{P}(S_{2n=0}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{dt}{1-\cos^2 t}\, dt = +\infty,$$

puisqu'on a en 0 l'équivalent positif : $\frac{1}{1-\cos^2 t} = \frac{1}{\sin^2 t} \sim \frac{1}{t^2}$.

(b) Par symétrie du cosinus, on a $\int_{-\pi}^{\pi} \cos^{2n}(\theta)\ d\theta = 4\int_0^{\pi/2} \cos^{2n}(\theta)\ d\theta$.
Il y a alors deux approches possibles.
Méthode 1 : on reconnait une intégrale de Wallis (voir exercice 58), d'où $\mathbb{P}(S_{2n} = 0) = \frac{2}{\pi} W_{2n}$. En injectant dans l'équivalent de Wallis, on obtient le résultat voulu.
Méthode 2 : on applique la méthode de Laplace (voir exercice 73) sur $[0, \pi/2[$, avec $h(x) = \log(\cos^2 x)$ et $g(x) = 1$.

Solution 208 1. La fonction $(x,t) \mapsto e^{it(x-a)}$ est bornée, donc intégrable par rapport à la mesure de probabilité $\mu \otimes \mathcal{U}([-T,T])$. Notons I_a son intégrale. Avec le théorème de Fubini, on calcule I_a de deux manières : pour tout $t \in [-T,T]$, on a

$$\int_{\mathbb{R}} e^{it(x-a)}\ d\mu(x) = e^{-ita}\phi(t),$$

d'où, en réintégrant par rapport à $\mathcal{U}([-T,T])$:

$$I_a = \frac{1}{2T}\int_{[-T,T]} e^{-ita}\phi(t)\ d\lambda(t).$$

Quant à l'intégrale de $t \mapsto e^{it(x-a)}$ par rapport à $\mathcal{U}([-T,T])$, elle vaut 1 pour $x = a$, et $\frac{\sin(T(x-a)}{T(x-a)}$ sinon. En réintégrant par rapport à μ, on a

$$\begin{aligned} I_a &= \int_{\mathbb{R}} \delta_a(x) + \mathbb{1}_{\mathbb{R}\backslash\{a\}}(x)\frac{\sin(T(x-a))}{T(x-a)}\ d\mu(x) \\ &= \mu(\{a\}) + \int_{\mathbb{R}\backslash\{a\}} \frac{\sin(T(x-a))}{T(x-a)}\ d\mu(x), \end{aligned}$$

d'où l'égalité voulue.

2. Soit a un réel. Pour $x \in \mathbb{R}\backslash a$, la majoration $|\frac{\sin(T(x-a))}{T(x-a)}| \leq \frac{1}{T|x-a|}$ donne $\lim\limits_{T\to+\infty} \frac{\sin(T(x-a))}{T(x-a)} = 0$. Par ailleurs, on a la majoration, uniforme en T : $|\frac{\sin(T(x-a))}{T(x-a)}| \leq 1$, et la fonction 1 est intégrable par rapport à μ. On peut donc appliquer le théorème de convergence dominée et on trouve ainsi que $\lim\limits_{T\to+\infty} \int_{\mathbb{R}\backslash\{a\}} \frac{\sin(T(x-a))}{T(x-a)}\ d\mu(x) = 0$, ce qui donne bien le résultat voulu avec la première question.
3. Soit T_0 tel que $|\phi(t)| \leq \varepsilon$ pour $t \geq T_0$.
On a la majoration $|e^{-ita}\phi(t)| \leq \varepsilon + \mathbb{1}_{[-T_0,T_0]}(t)$, d'où

$$\frac{1}{2T}\int_{[-T,T]} e^{-ita}\phi(t)\ d\lambda(t) \leq \varepsilon + \frac{T_0}{T}.$$

En faisant tendre T vers $+\infty$, on obtient l'inégalité $\mu(\{a\}) \leq \varepsilon$. Comme ε peut être pris arbitrairement petit, on a $\mu(\{a\}) = 0$.

D.10 Exercices sur la convergence presque sûre

Solution 221 1. Les variables aléatoires $(X_k)_{k\geq 1}$ sont centrées, avec $\operatorname{Var} X_k = k$. Comme elles sont indépendantes, on a

$$\operatorname{Var} S_n = \sum_{k=1}^{n} \operatorname{Var} X_k = \sum_{k=1}^{n} k = \frac{n(n+1)}{2}.$$

Maintenant, d'après l'inégalité de Chebychev, on a pour tout $\varepsilon > 0$,

$$\mathbb{P}(|S_n/n^{3/2}| \geq \varepsilon) \leq \frac{\operatorname{Var}(S_n/n^{3/2})}{\varepsilon^2} = \frac{\operatorname{Var}(S_n)}{n^3\varepsilon^2} = \frac{n(n+1)/2}{n^3\varepsilon^2} \leq \frac{1}{\varepsilon^2 n}.$$

Cela implique que $S_n/n^{3/2}$ converge en probabilité vers 0.

2. On fait la même majoration que précédemment :

$$\mathbb{P}(|S_{n^2}/n^3| \geq \varepsilon) \leq \frac{1}{n^2\varepsilon^2}.$$

Or la série de terme général $\frac{1}{n^2\varepsilon^2}$ converge, donc la série de terme général $\mathbb{P}\left(\left|\frac{S_{n^2}}{n^3}\right| \geq \varepsilon\right)$ aussi et d'après Borel-Cantelli, on obtient que $\mathbb{P}\left(\overline{\lim_{n\to+\infty}} \left|\frac{S_{n^2}}{n^3}\right| \geq \varepsilon\right) = 0$. Comme $\varepsilon > 0$ est quelconque, le critère fondamental de convergence presque sûre nous dit que $\frac{S_{n^2}}{n^3}$ converge presque sûrement vers 0.

3. Comme suggéré, on s'inspire de la preuve du théorème 10.21.

Soit maintenant $n \geq 4$ et notons $p = p(n) = \lfloor\sqrt{n}\rfloor^2$. On peut noter que $n - p \leq 2\lfloor\sqrt{n}\rfloor \leq 2\sqrt{n} \leq \frac{n}{2}$, ce qui entraîne également $n - p \leq p$. Notons $D_{n,p} = X_{p+1} + \cdots + X_n$. On a

$$\frac{S_p}{p^{3/2}} - \frac{S_n}{n^{3/2}} = \frac{S_p}{p^{3/2}} - \frac{S_p + D_{n,p}}{n^{3/2}} = \left(\frac{1}{p^{3/2}} - \frac{1}{n^{3/2}}\right) S_p - \frac{1}{n^{3/2}} D_{n,p}.$$

Comme S_p et $D_{n,p}$ ne sont pas corrélés, on a

$$\operatorname{Var}\left(\frac{S_p}{p^{3/2}} - \frac{S_n}{n^{3/2}}\right) = \left(\frac{1}{p^{3/2}} - \frac{1}{n^{3/2}}\right)^2 \operatorname{Var} S_p + \frac{1}{n^3} \operatorname{Var} D_{n,p}.$$

Or on a les majorations suivantes $\operatorname{Var} S_p = 1 + 2 + \cdots + p \leq p^2$ et $\operatorname{Var} D_{n,p} = (p+1) + \cdots + n \leq (n-p)n$. D'autre part, l'inégalité des accroissements finis donne

$$\left(\frac{1}{p^{3/2}} - \frac{1}{n^{3/2}}\right)^2 \leq \left(\frac{3(n-p)}{2p^{5/2}}\right)^2 \leq \frac{9(n-p)}{4p^4} \leq \frac{9(n-p)}{n^2p^2}.$$

On en déduit que $\operatorname{Var}\left(\frac{S_p}{p^{3/2}} - \frac{S_n}{n^{3/2}}\right) \leq \frac{10(n-p)}{n^2} \leq \frac{20}{n^{3/2}}$. Argumentant comme précédemment, $\frac{S_n}{n^{3/2}} - \frac{S_{p(n)}}{p(n)^{3/2}}$ tend presque sûrement vers 0.

Comme $\lim_{n\to+\infty} p(n) = +\infty$ et que $\frac{S_{p(n)}}{p(n)^{3/2}}$ contient les mêmes termes que $\frac{S_{n^2}}{n^3}$, $\frac{S_{p(n)}}{p(n)^{3/2}}$ tend également presque sûrement vers 0, d'où on déduit que $\frac{S_n}{n^{3/2}}$ tend presque sûrement vers 0.

Solution 222 1. Le plus simple ici est de raisonner ω par ω. Pour tout $n > X(\omega)$, on a $Y_n(\omega) = X(\omega)$. Ainsi, Y_n converge presque sûrement vers X (et donc aussi en probabilité).

2. $\mathbb{E}(Y_n) = \int_0^n x f_X(x)\, dx + e^{2n}\mathbb{P}(X \geq n) = 1 - (n+1)e^{-n} + e^n$ (et cette espérance tend vers l'infini avec n).

3. Non, car $\lim_{n\to+\infty} \mathbb{E}(Y_n) = +\infty$, tandis que $\mathbb{E}(X) = 1$.

Solution 223 Il suffit de remarquer que la série de terme général $\mathbb{P}(X_n \neq 0)$ converge. En effet, d'après le lemme de Borel-Cantelli, X_n est p.s. nulle à partir d'un certain rang, donc S_n est p.s. constante à partir d'un certain rang et S_n/n converge évidemment presque sûrement vers 0.

Solution 224 1. Calculons l'espérance de la variable aléatoire S_n. Par définition de S_n et par linéarité, on a $\mathbb{E}(S_n) = \mathbb{E}(Y_1) + \cdots + \mathbb{E}(Y_n)$. Comme $Y_k = \mathbb{1}_{\{X_k \neq X_{k+1}\}}$, Y_k suit une loi de Bernoulli de paramètre $q = \mathbb{P}(X_k \neq X_{k+1})$ et son espérance vaut

$$\begin{aligned}\mathbb{E}(Y_k) &= \mathbb{P}(X_k \neq X_{k+1})\\ &= \mathbb{P}(X_k = 0, X_{k+1} = 1) + \mathbb{P}(X_k = 1, X_{k+1} = 0).\end{aligned}$$

Comme X_k et X_{k+1} sont indépendantes, on obtient $\mathbb{E}(Y_k) = 2p(1-p)$, d'où $\mathbb{E}(S_n) = n\mathbb{E}(Y_1) = 2np(1-p)$. La variance de Y_k vaut

$$\mathrm{Var}(Y_k) = q(1-q) = 2p(1-p)[1 - 2p(1-p)].$$

Calculons maintenant la variance de S_n :

$$\mathrm{Var}(S_n) = \sum_{k=1}^{n} \mathrm{Var}(Y_k) + 2 \sum_{1\leq i<j\leq n} \mathrm{Cov}(Y_i, Y_j).$$

Les X_k étant indépendantes, Y_k est indépendante de toutes les Y_{k+i} avec $i \geq 2$. Il ne nous reste donc plus qu'à calculer la covariance

$$\mathrm{Covar}(Y_k, Y_{k+1}) = \mathbb{E}(Y_k Y_{k+1}) - \mathbb{E}(X_k)\mathbb{E}(X_{k+1}).$$

On a de plus

$$\begin{aligned}
&\mathbb{E}(Y_k Y_{k+1})\\
&= \mathbb{E}[\mathbb{1}_{\{X_k \neq X_{k+1}\}} \mathbb{1}_{\{X_{k+1} \neq X_{k+2}\}}]\\
&= \mathbb{P}(X_k \neq X_{k+1}, X_{k+1} \neq X_{k+2})\\
&= \mathbb{P}(X_k = 1, X_{k+1} = 0, X_{k+2} = 1) + \mathbb{P}(X_k = 0, X_{k+1} = 1, X_{k+2} = 0)\\
&= p^2(1-p) + (1-p)^2 p = p(1-p).
\end{aligned}$$

Cela donne donc $\mathrm{Covar}(Y_k, Y_{k+1}) = p(1-p) - 4p^2(1-p)^2$, ce qui implique

$$\begin{aligned}
&\mathrm{Var}(S_n)\\
&= n \,\mathrm{Var}\, X_1 + 2(n-1)\,\mathrm{Covar}(X_1, X_2)\\
&= 2np(1-p)[1-2p(1-p)] + 2(n-1)\left(p(1-p) - 4p^2(1-p)^2\right).
\end{aligned}$$

2. On a $\|\frac{S_n}{n} - 2p(1-p)\|_2 = \|\frac{S_n - \mathbb{E}S_n}{n}\|_2$. Il suffit donc de montrer que $\|\frac{S_n - \mathbb{E}S_n}{n}\|_2^2$ tend vers 0. Or $\|\frac{S_n - \mathbb{E}S_n}{n}\|_2^2 = \frac{1}{n^2} \mathrm{Var}\, S_n$ et $\mathrm{Var}\, S_n = O(n)$, ce qui donne le résultat voulu.

3. Posons $I_n = Y_1 + \cdots + Y_{2n-1}$ et $P_n = Y_2 + \cdots + Y_{2n}$. Les variables $Y_1, Y_3, \ldots, Y_{2n-1}$ étant indépendantes, d'après la loi forte des grands nombres, I_n/n converge presque sûrement vers $\mathbb{E}[Y_1] = 2p(1-p)$. De même, les variables $Y_2, Y_4, \ldots, Y_{2n}$ sont indépendantes, donc, d'après la loi forte des grands nombres P_n/n converge presque sûrement vers $\mathbb{E}[Y_1] = 2p(1-p)$, et par suite $\frac{S_{2n}}{2n} = \frac{1}{2}(\frac{P_n}{n} + \frac{I_n}{n})$ converge presque sûrement vers $\mathbb{E}[Y_1] = 2p(1-p)$. Comme $|S_{2n+1} - S_{2n}| \leq 1$, $\frac{S_{2n+1}}{2n}$ converge également vers $2p(1-p)$ et donc $\frac{S_{2n+1}}{2n+1}$ aussi. Ainsi, S_n/n converge presque sûrement vers $2p(1-p)$.

Solution 225 Considérons les événements $A_n =$ “obtenir face au n-ième lancer” (qui sont indépendants, de probabilité $1/2$) et $B_n =$ “obtenir face aux lancers n et $n+1$”. Posons $B = \overline{\lim}_{n \to +\infty} B_n$.

La difficulté pour calculer $\mathbb{P}(B)$ est que les événements B_n ne sont pas indépendants. On considère alors les événements $C_n = B_{2n}$ qui sont, eux, indépendants. Si on a une infinité de B_{2n} qui sont réalisés, on a bien sûr une infinité de B_n qui sont réalisés, soit $B \supset \overline{\lim}_{n \to +\infty} C_n$. Or $\mathbb{P}(C_k) = 1/4$ et la série de terme général $\mathbb{P}(C_k)$ diverge. D'après le second lemme de Borel-Cantelli, on a $\mathbb{P}\left(\overline{\lim}_{n \to +\infty} C_n\right) = 1$. On en conclut que $\mathbb{P}(B) = \mathbb{P}\left(\overline{\lim}_{n \to +\infty} B_n\right) = 1$.

Solution 226 1. $\left\{ \overline{\lim\limits_{n\to+\infty}} \frac{|X_n|}{n} \geq a \right\} \supset \left\{ \frac{|X_n|}{n} \geq a \quad \text{i.s.} \right\}$. Il suffit donc de montrer que ce dernier événement est de probabilité 1. Pour ce faire, d'après le second lemme le Borel-Cantelli et l'indépendance des X_n, il suffit de montrer que la série de terme général $\mathbb{P}\left(\frac{|X_n|}{n} \geq a\right)$ diverge. Or

$$\begin{aligned}\sum_{n=1}^{+\infty} \mathbb{P}\left(\frac{|X_n|}{n} \geq a\right) &\geq \sum_{n=1}^{+\infty} \mathbb{P}\left(\lfloor\frac{|X_n|}{a}\rfloor \geq n\right) \\ &= \sum_{n=1}^{+\infty} \mathbb{P}\left(\lfloor\frac{|X_1|}{a}\rfloor \geq n\right) = \mathbb{E}(\lfloor\frac{|X_1|}{a}\rfloor)\end{aligned}$$

ce qui donne le résultat puisque $\lfloor\frac{|X_1|}{a}\rfloor \geq \frac{|X_1|}{a} - 1$ et que $\mathbb{E}[|X_1|] = +\infty$.

2. Pour tout entier $N \geq 1$, l'événement $\left\{ \overline{\lim\limits_{n\to+\infty}} \frac{|X_n|}{n} \geq N \right\}$ est de probabilité 1. Comme une intersection dénombrable d'événements de probabilité 1 est de probabilité 1, on a $\overline{\lim\limits_{n\to+\infty}} \frac{|X_n|}{n} = +\infty$ p.s., ce qui entraîne que $\sup\limits_{n\geq 1} \frac{|X_n|}{n} = +\infty$ p.s. Cependant on a l'inclusion

$$\left\{ \sup_{n\geq 1} \frac{|S_n|}{n} < +\infty \right\} \subset \left\{ \sup_{n\geq 1} \frac{|X_n|}{n} < +\infty \right\}.$$

En effet, si $|S_n/n| \leq M$ pour tout n, on a alors pour tout n :
$|\frac{X_n}{n}| = \frac{|S_n - S_{n-1}|}{n} \leq \frac{|S_n|}{n} + \frac{|S_{n-1}|}{n} \leq 2M$.
Ainsi $\mathbb{P}\left(\sup\limits_{n\geq 1} \frac{|S_n|}{n} < +\infty \right) \leq \mathbb{P}\left(\sup\limits_{n\geq 1} |X_n| < +\infty \right) = 0$, donc
$\mathbb{P}\left(\sup\limits_{n\geq 1} \frac{|S_n|}{n} < +\infty \right) = 0$ et donc $\mathbb{P}\left(\sup\limits_{n\geq 1} \frac{|S_n|}{n} = +\infty \right) = 1$.

Solution 227 1. (a) Soit $\varepsilon > 0$. On a

$$\mathbb{P}\left(|\frac{X_n}{n}| \geq \varepsilon\right) = \mathbb{P}\left(\frac{|X_n|}{\varepsilon} \geq n\right) = \mathbb{P}\left(\frac{|X_1|}{\varepsilon} \geq n\right).$$

Comme $\frac{|X_1|}{\varepsilon}$ est intégrable, la série de terme général $\mathbb{P}\left(\frac{|X_1|}{\varepsilon} \geq n\right)$ converge. Donc d'après le lemme de Borel-Cantelli, on obtient $\mathbb{P}\left(\overline{\lim\limits_{n\to+\infty}} \{|\frac{X_n}{n}| \geq \varepsilon\}\right) = 0$ et le critère fondamental de convergence presque sûre implique que X_n/n tend p.s. vers 0.

(b) Posons $S_n = \sum\limits_{k=1}^{n} X_k$. On a $\frac{S_{Mn}}{Mn} = \frac{1}{M} \sum\limits_{i=1}^{M} \frac{1}{n} \sum\limits_{j=0}^{n-1} X_{Mj+i}$.
Pour tout i compris entre 1 et M, les variables $(X_{Mj+i})_{j\geq 0}$ sont des

variables deux-à-deux indépendantes, identiquement distribuées, intégrables, donc d'après la loi forte des grands nombres $\frac{1}{n}\sum_{j=0}^{n-1} X_{Mj+i}$ tend presque sûrement vers $\mathbb{E}[X_i]$. On en déduit que $\frac{S_{nM}}{nM}$ converge presque sûrement vers $\frac{1}{M}\sum_{i=1}^{M} \mathbb{E}[X_i] = \mathbb{E}[X_1]$. Posons $k(n)=M\lfloor \frac{n}{M} \rfloor$. On a $|S_n - S_{k(n)}| \le |X_n| + |X_{n-1}| + \dots + |X_{n-M+1}|$. Grâce à la première question $\frac{S_n - S_{k(n)}}{n}$ tend presque sûrement vers 0 (c'est une somme finie de termes tendant vers 0). Comme $\frac{S_{k(n)}}{k(n)}$ tend presque sûrement vers $\mathbb{E}[X_1]$ et que $k(n)/n$ tend vers 1, $\frac{S_{k(n)}}{n}$, tend presque sûrement vers $\mathbb{E}[X_1]$. On en déduit que S_n/n tend presque sûrement vers $\mathbb{E}[X_1]$.

2. Posons $Y_k = \mathbb{1}_{\{X_{k+1}=a_1,\dots,X_{k+\ell}=a_\ell\}}$. Les (Y_k) sont des variables de Bernoulli, d'espérance $\mathbb{P}(X_{k+1} = a_1, \dots, X_{k+\ell} = a_\ell) = \prod_{i=1}^{n} p_{a_i}$. Il est aisé de voir que Y_i et Y_j sont indépendantes dès que $|i-j| \ge \ell$. Ainsi, la question précédente donne immédiatement le résultat voulu.

3. Si on note $a_1, \dots, a_\ell$ la suite des caractères des livres des bibliothèques, la question précédente montre que S_n/n a une limite non nulle, en particulier S_n est non-nul pour n assez grand, ce qui atteste de l'apparition de la suite cherchée. Le texte de Borel met en exergue le fait que le grand nombre d'expériences indépendantes contrebalance la très faible probabilité de l'événement considéré.

Solution 228 1. Par définition de la partie fractionnaire, il est immédiat que la suite des X_i^g prend ses valeurs dans $[0,1[$. Comme $0 \le gX_i^g < g$, il est également clair que A_i^g prend ses valeurs dans $\{0,\dots,g-1\}$. On a $gX_i = \lfloor gX_i \rfloor + \{gX_i\}$, soit $gX_i = A_i + X_{i+1}$, ce qui s'écrit encore $\frac{X_i}{g^i} = \frac{A_i}{g^{i+1}} + \frac{X_{i+1}}{g^{i+1}}$. Ainsi

$$\sum_{i=j}^{n} \frac{A_i}{g^{i+1}} = \sum_{i=j}^{n} \left(\frac{X_i}{g^i} - \frac{X_{i+1}}{g^{i+1}} \right) = \frac{X_j}{g^j} - \frac{X_{n+1}}{g^{n+1}}.$$

Soit en faisant tendre n vers l'infini : $\frac{X_j}{g^j} = \sum_{i=j}^{+\infty} \frac{A_i}{g^{i+1}}$. En particulier, pour $j=0$, on obtient l'écriture voulue. Reste à voir que A_i ne peut être constamment égal à $j-1$ à partir d'un certain rang. En effet, si on avait $A_i = g-1$ pour $i > j$, on aurait $X_j = g^j \sum_{i=j}^{+\infty} \frac{g-1}{g^{i+1}} = 1$, ce qui est exclu, car $X_j \in [0,1[$. On a donc l'existence de la décomposition. Passons à l'unicité. Supposons qu'on ait deux écritures propres distinctes :

$$\omega = \sum_{i=1}^{+\infty} a_i g^{-i} = \sum_{i=1}^{+\infty} b_i g^{-i}.$$

Soit i_0 le plus petit entier tel que $a_i \neq b_i$: on a $\sum_{i=i_0}^{+\infty} a_i g^{-i} = \sum_{i=i_0}^{+\infty} b_i g^{-i}$.
On peut supposer que $b_i < a_i$. On a alors

$$\begin{aligned}\frac{a_{i_0}}{g^{i_0}} \leq \sum_{i=i_0}^{+\infty} a_i g^{-i} = \sum_{i=i_0}^{+\infty} b_i g^{-i} &= \frac{b_{i_0}}{g^{i_0+1}} + \sum_{i=i_0+1}^{+\infty} b_i g^{-i} \\ &< \frac{b_{i_0}}{g^{i_0}} + \sum_{i=i_0+1}^{+\infty} (g-1) g^{-i},\end{aligned}$$

où l'inégalité stricte vient du fait qu'au moins un des b_i est strictement inférieur à $g-1$. On a alors $\frac{a_{i_0}}{g^{i_0}} < \frac{b_{i_0}+1}{g^{i_0}}$, soit $a_{i_0} \leq b_{i_0}$: contradiction.

2. Note : dans la suite, comme g est le plus souvent fixé, on omet l'exposant g afin de simplifier les écritures.
On a $\{G(X_0) \in \{k\} \times [0,x]\} = \{gX_0 \in [k, k+x]\} = \{X_0 \in [\frac{k}{g}, \frac{k}{g} + \frac{x}{g}]\}$, donc $\mathbb{P}(G(X_0) \in \{k\} \times [0,x]\}) = \mathbb{P}(X_0 \in [\frac{k}{g}, \frac{k}{g} + \frac{x}{g}]) = \frac{x}{g}$. Par addition, on en déduit que pour tout $k \in \{0, \dots, g-1\}$, on a

$$\mathbb{P}(G(X_0) \in [0, k[\times[0,x]) = \frac{k}{g} x,$$

soit

$$\mathbb{P}_G([0, k[\times[0,x]) = (\mathcal{U}(\{0, \dots, g-1\}) \otimes \mathbb{P})([0, k[\times[0,x]).$$

Comme on a ici un π-système qui engendre $\mathcal{B}(\{0, \dots, g-1\} \times [0,1])$, cela permet d'identifier les deux probabilités.

3. Soit p_2 la projection sur la seconde coordonnée. On a $p_2 \circ G = \{g.\}$. D'après la question 2., la loi image de $\mathbb{P}$ par $x \mapsto \{gx\}$ est $\mathbb{P}$ elle-même, ce qui donne par récurrence que tous les X_i ont $\mathbb{P}$ comme loi commune. D'autre part $A_i = (p_1 \circ G)(X_{i-1})$. Comme X_i suit la loi $\mathbb{P}$, la question 2. nous dit que A_i suit la loi uniforme sur $\{0, \dots, g-1\}$.

4. Il suffit de montrer que pour tout n, les variables $A_0, \dots, A_n$ sont indépendantes. Pour cela, il suffit encore de montrer que pour tout entier $k \in \{0, \dots, n-1\}$, A_k est indépendant du vecteur $(A_{k+1}, \dots, A_n)$. Or pour tout $i > k$, A_i est $\sigma(X_i)$-mesurable. Comme, grâce aux relations de récurrence, on a $\sigma(X_i) \subset \sigma(X_{i-1}) \subset \dots \subset \sigma(X_{k+1})$, le vecteur $(A_{k+1}, \dots, A_n)$ est $\sigma(X_{k+1})$-mesurable. Or $(X_{k+1}, A_k) = G(X_k)$. Comme X_k suit la loi $\mathbb{P}$, d'après la question 2., la loi de (X_{k+1}, A_k) est une loi produit, ainsi X_{k+1} et A_k sont indépendants, autrement dit que

les tribus $\sigma(X_{k+1})$ et $\sigma(A_k)$ sont indépendantes. Comme $(A_{k+1},\dots,A_n)$ est $\sigma(X_{k+1})$-mesurable et A_k est bien sûr $\sigma(A_k)$-mesurable, on a bien l'indépendance voulue.

5. Comme les variables $A_i^g(\omega)$ sont indépendantes, suivent la loi uniforme sur $\{0,\dots,g-1\}$ et représentent le développement de ω en base g, le résultat voulu découle immédiatement de l'exercice 227 et de la définition de la g-normalité.

6. On sait que pour tout $g \geq 2$, $\omega \in [0,1[$ est $\mathbb{P}$-presque sûrement g-normal. Comme une intersection dénombrable d'événements de probabilité 1 est encore de probabilité 1, $\mathbb{P}$-presque tous les $\omega \in [0,1[$ sont normaux.

Solution 229 1. On a $2^n S = 2^n S_n + R_n$, avec $R_n = \sum_{k=1}^{+\infty} \frac{X_{n+k}}{2^k}$. $2^n S_n$ est à valeurs entières et R_n est à valeurs dans $[0,1]$. Comme $R_n \geq 0$, on a $\{2^n S < k\} \subset \{2^n S_n < k\}$, mais comme k est entier, on a l'inclusion $\{2^n S_n < k\} \cap \{R_n < 1\} \subset \{2^n S < k\}$.

Ainsi $\{S_n < \frac{k}{2^n}\} \setminus \{S < \frac{k}{2^n}\} \subset \{R_n = 1\} = \bigcap_{k=n+1}^{+\infty} \{X_k = 1\}$, d'où

$$\mathbb{P}\left(S_n < \frac{k}{2^n}\right) - \mathbb{P}\left(S < \frac{k}{2^n}\right) \leq \mathbb{P}\left(\bigcap_{k=n+1}^{+\infty} \{X_k = 1\}\right) = 0.$$

Or, si $i \in \{0,\dots,2^n-1\}$ s'écrit $i = \sum_{k=0}^{n-1} a_k 2^{n-k}$, on a

$$\mathbb{P}(2^n S_n = i) = \mathbb{P}\left(\bigcap_{k=1}^{n} \{X_k = a_k\}\right) = \prod_{k=1}^{n} \mathbb{P}(X_k = a_k) = \frac{1}{2^n},$$

d'où $\mathbb{P}(2^n S_n < k) = \frac{k}{2^n}$, ce qui donne $\mathbb{P}\left(S < \frac{k}{2^n}\right) = \frac{k}{2^n}$.

2. Pour tout $a \in [0,1[$, on a $[0,a] = \bigcap_{n=1}^{+\infty} [0, \frac{[2^n a]}{2^n}[$, donc la tribu engendrée par les ensembles de la forme $\left[0, \frac{k}{2^n}\right[$ contient la tribu engendrée par les ensembles de la forme $[0,a]$, c'est-à-dire la tribu borélienne de $[0,1[$. Comme les ensembles $\left[0, \frac{k}{2^n}\right[$ sont eux-mêmes des boréliens, ils engendrent donc exactement la tribu borélienne de $[0,1[$. $\mathbb{P}_S$ et $\lambda_{|[0,1[}$ sont deux probabilités qui coïncident sur les ensembles de la forme $[0, \frac{k}{2^n}[$, qui forment un π-système engendrant $\mathcal{B}([0,1[)$, donc elles sont égales : $\mathbb{P}_S = \lambda_{|[0,1[}$. Autrement dit, S suit la loi uniforme sur $[0,1[$ (ou sur $[0,1]$: c'est la même chose).

3. À i fixé, les variables aléatoires $X_{2^j(2i+1)}$ sont des variables aléatoires indépendantes suivant la loi de Bernoulli de paramètre $1/2$, donc d'après la question précédente, sous $\mathbb{P}$, U_i suit la loi uniforme.

Posons $\mathcal{T}_i = \sigma(X_{2^j(2i+1)}; j \geq 1)$. U_i est clairement $\mathcal{T}_i$-mesurable. Or, par associativité de l'indépendance, les tribus $\mathcal{T}_i$ sont indépendantes, donc les U_i forment une suite de variables indépendantes de loi uniforme sur $[0, 1[$.

4. On a vu dans l'exercice précédent que sur l'espace probabilisé $([0, 1[, \mathcal{B}([0, 1[), \lambda_{|[0,1[})$, on peut définir une suite de variables aléatoires $(X_n)_{n\geq 1}$ indépendantes de loi commune $\mathrm{Ber}(1/2)$. D'après la question qui précède, on peut alors construire une suite $(U_i)_{i\geq 1}$ de variables indépendantes de loi uniforme sur $[0, 1[$. Soit F_i la fonction de répartition de μ_i. Posons
$$\forall u \in [0, 1[\quad Q_i^*(u) = \inf\{x \in \mathbb{R}; 1 - F_i(x) \leq u\}.$$
D'après le théorème 5.36, $Z_i = Q_i^*(U_i)$ est une variable aléatoire réelle de fonction de répartition F_i, donc de loi μ_i. Par construction, les Z_i sont indépendantes et elles répondent donc à la question posée.

Solution 230 Soit $a > 0$. Comme X_n admet la densité $f(x) = e^{-x}\mathbb{1}_{\mathbb{R}_+}(x)$, on a
$$\mathbb{P}\left(\frac{X_n}{\log n} \geq a\right) = \mathbb{P}(X_n \geq a \log n) = \int_{[a\log n, +\infty[} e^{-x}\, d\lambda(x) = e^{-a\log n} = \frac{1}{n^a}.$$
Lorsque $a > 1$, la série de terme général $\mathbb{P}\left(\frac{X_n}{\log n} \geq a\right)$ converge, donc le premier lemme de Borel-Cantelli assure que $\mathbb{P}\left(\overline{\lim_{n\to+\infty}} \{\frac{X_n}{\log n} \geq a\}\right) = 0$.
Lorsque $a \leq 1$, la série de terme général $\mathbb{P}\left(\frac{X_n}{\log n} \geq a\right)$ diverge. Or les événements $\{\frac{X_n}{\log n} \geq a\}$ sont indépendants, donc le deuxième lemme de Borel-Cantelli nous permet d'affirmer que $\mathbb{P}\left(\overline{\lim_{n\to+\infty}} \{\frac{X_n}{\log n} \geq a\}\right) = 1$. Il y a deux manières de conclure :

— Si on se souvient du théorème 10.3, on peut conclure que
$\overline{\lim_{n\to+\infty}} \frac{X_n}{\log n} = 1$ presque sûrement.

— Sinon, on refait partiellement le raisonnement permettant de montrer ce théorème. Soit $k \in \mathbb{N}^*$. On a $\mathbb{P}\left(\overline{\lim_{n\to+\infty}} \{\frac{X_n}{\log n} \geq 1 + \frac{1}{k}\}\right) = 0$, et donc prenant la réunion (dénombrable) sur $k \in \mathbb{N}^*$,
$\mathbb{P}\left(\overline{\lim_{n\to+\infty}} \frac{X_n}{\log n} > 1\right) = 0$. D'autre part $\mathbb{P}\left(\overline{\lim_{n\to+\infty}} \{\frac{X_n}{\log n} \geq 1\}\right) = 1$, donc $\mathbb{P}\left(\overline{\lim_{n\to+\infty}} \frac{X_n}{\log n} \geq 1\right) = 1$. Les deux arguments rassemblés im-

pliquent que $\overline{\lim}_{n\to+\infty} \frac{X_n}{\log n} = 1$ presque sûrement.

Nous proposons ici une seconde démonstration (plus courte, mais plus astucieuse) de ce résultat. On sait que $\mathbb{P}(X_n \geq \log n) = \frac{1}{n}$ et donc le (second) lemme de Borel-Cantelli implique que $\mathbb{P}\left(\overline{\lim}_{n\to+\infty} \frac{X_n}{\log n} \geq 1 \right) = 1$. Comme on sait de plus que $\mathbb{P}(X_n \geq \log n + 2\log\log n) = \frac{1}{n(\log n)^2}$, le (premier) lemme de Borel-Cantelli donne

$$\mathbb{P}\left(\overline{\lim}_{n\to+\infty} \frac{X_n}{\log n + 2\log\log n} \leq 1 \right) = 1.$$

Puisque $\frac{1}{n(\log n)^2}$ est le terme général d'une série convergente, on utilise ensuite l'équivalent à l'infini $\log n \sim \log n + 2\log\log n$ pour conclure.

Solution 231 Une suite à valeurs entières convergente est nécessairement constante à partir d'un certain rang (critère de Cauchy), et la limite est entière. Par suite, les valeurs d'adhérence d'une suite entière sont exactement les entiers qui sont pris une infinité de fois par la suite. Ainsi, pour tout a entier, on a

$$V_a = \{a \text{ est valeur d'adhérence de } X_n\} = \overline{\lim}_{n\to+\infty} \{X_n = a\}.$$

D'après la loi du $0-1$ de Borel, a est donc presque sûrement valeur d'adhérence de X_n si la série de terme général $\mathbb{P}(X_n = a)$ diverge. Inversement, si la série de terme général $\mathbb{P}(X_n = a)$ converge, a n'est presque sûrement pas valeur d'adhérence. Posons $\alpha = 0,01789$ et $p_n = \frac{1}{n^{1+\alpha}}$. On a

$$\mathbb{P}(X_n = a) = \binom{n}{a} p_n^a (1-p_n)^{n-a}.$$

En l'infini, on a $\binom{n}{a} \sim \frac{n^a}{a!}$ et $(1-p_n)^a \sim 1$, donc

$$\mathbb{P}(X_n = a) = \binom{n}{a} p_n^a (1-p_n)^{n-a} \sim \frac{n^a}{a!} \frac{1}{n^{a(1+\alpha)}} (1-p_n)^n$$

$$\text{où} \quad \log((1-p_n)^n) = n\log(1-p_n) \sim n(-p_n) = -\frac{1}{n^\alpha}$$

tend vers 0 lorsque n tend vers l'infini. Donc $(1-p_n)^n = \exp(\log((1-p_n)^n))$ tend vers 1 lorsque n tend vers l'infini, ce qui nous donne

$$\mathbb{P}(X_n = a) \sim \frac{n^a}{a!} \frac{1}{n^{a(1+\alpha)}} = \frac{1}{a!} \frac{1}{n^{a\alpha}}.$$

Ainsi, la série diverge si et seulement si $a\alpha < 1$, soit $a < 1/\alpha$.
Comme $55 < \frac{1}{\alpha} < 56$, on obtient que la série diverge pour $a \in \{0, \ldots, 55\}$ et

converge pour $a > 56$. L'événement

$$\left(\bigcap_{a=1}^{55} V_a\right) \cap \left(\bigcap_{a=56}^{+\infty} V_a^c\right)$$

est de probabilité 1 comme intersection dénombrable d'événements de probabilité 1. Donc avec probabilité 1, l'ensemble des valeurs d'adhérence de (X_n) est exactement $\{0, \dots, 55\}$.

Solution 232 1. Quitte à remplacer X par $X' = \frac{X}{\mathbb{E}[X]}$, on peut supposer que $\mathbb{E}[X] = 1$. On a, d'une part,

$$\begin{aligned}\mathbb{E}[X\mathbb{1}_{\{X\geq\lambda\}}] &= \mathbb{E}[X] - \mathbb{E}[X\mathbb{1}_{\{X<\lambda\}}] \\ &\geq 1 - \mathbb{E}[\lambda\mathbb{1}_{\{X<\lambda\}}] \geq 1 - \lambda.\end{aligned}$$

D'autre part, d'après l'inégalité de Cauchy–Schwarz, on a

$$(\mathbb{E}[X\mathbb{1}_{\{X\geq\lambda\}}])^2 \leq \mathbb{E}[X^2]\mathbb{E}[\mathbb{1}^2_{\{X\geq\lambda\}}] = \mathbb{E}[X^2]\mathbb{P}(X \geq \lambda),$$

d'où

$$\mathbb{P}(X \geq \lambda) \geq \frac{(\mathbb{E}[X\mathbb{1}_{\{X\geq\lambda\}}])^2}{\mathbb{E}[X^2]} \geq \frac{(1-\lambda)^2}{\mathbb{E}[X^2]}.$$

2. Posons $f_n = 1 - \mathbb{1}_{B_n}$. Les f_n sont des fonctions mesurables positives, donc d'après le lemme de Fatou, $\mathbb{E}\left[\underline{\lim}_{n\to+\infty} f_n\right] \leq \underline{\lim}_{n\to+\infty} \mathbb{E}[f_n]$, puis

$$\begin{aligned}\mathbb{E}\left[\overline{\lim}_{n\to+\infty} -f_n\right] &\geq \overline{\lim}_{n\to+\infty} \mathbb{E}[-f_n] \\ \mathbb{E}\left[\overline{\lim}_{n\to+\infty} 1-f_n\right] &\geq \overline{\lim}_{n\to+\infty} \mathbb{E}[1-f_n] \\ \mathbb{E}\left[\overline{\lim}_{n\to+\infty} \mathbb{1}_{B_n}\right] &\geq \overline{\lim}_{n\to+\infty} \mathbb{E}[\mathbb{1}_{B_n}].\end{aligned}$$

Comme $\overline{\lim}_{n\to+\infty} \mathbb{1}_{B_n}$ est l'indicatrice de $\overline{\lim}_{n\to+\infty} B_n$, on obtient l'inégalité voulue.

3. Soit $\lambda \in]0, 1[$. On a

$$\overline{\lim}_{n\to+\infty} B_n \supset \{N_n \geq \lambda\mathbb{E}[N_n] \text{ i.s.}\},$$

où i.s. signifie “infiniment souvent”, donc

$$\begin{aligned}\mathbb{P}\left(\overline{\lim_{n\to+\infty}} B_n\right) &\geq \mathbb{P}(N_n \geq \lambda\mathbb{E}[N_n] \text{ i.s.}) \geq \overline{\lim_{n\to+\infty}} \mathbb{P}(N_n \geq \lambda\mathbb{E}[N_n]) \\ &\geq \overline{\lim_{n\to+\infty}} (1-\lambda)^2\frac{(\mathbb{E}N_n)^2}{\mathbb{E}[N_n^2]} \\ &= (1-\lambda)^2 \overline{\lim_{n\to+\infty}} \frac{(\mathbb{E}N_n)^2}{\mathbb{E}[N_n^2]}.\end{aligned}$$

En faisant tendre λ vers 0 dans la dernière inégalité, on obtient

$$\mathbb{P}\left(\overline{\lim_{n\to+\infty}} B_n\right) \geq \overline{\lim_{n\to+\infty}} \frac{(\mathbb{E}N_n)^2}{\mathbb{E}[N_n^2]}.$$

4. On a $\frac{(\mathbb{E}N_n)^2}{\mathbb{E}[N_n^2]} = \frac{(\mathbb{E}N_n)^2}{(\mathbb{E}[N_n])^2+\operatorname{Var} N_n} = 1 - \frac{\operatorname{Var} N_n}{(\mathbb{E}[N_n])^2+\operatorname{Var} N_n}$. Or, on sait que $\mathbb{E}N_n = \sum_{k=1}^{n} \mathbb{P}(B_k) \geq \sum_{k=1}^{n} \mathbb{P}(B_k)(1-\mathbb{P}(B_k)) = \operatorname{Var} N_n$.
Ainsi $\frac{\operatorname{Var} N_n}{(\mathbb{E}[N_n])^2+\operatorname{Var} N_n} \leq \frac{\mathbb{E}[N_n]}{(\mathbb{E}[N_n])^2} = \frac{1}{\mathbb{E}(N_n)}$ qui tend vers 0, donc $\frac{(\mathbb{E}N_n)^2}{\mathbb{E}[N_n^2]}$ tend vers 1. On déduit alors de la question précédente que $\mathbb{P}\left(\overline{\lim_{n\to+\infty}} B_n\right) = 1$: on a redémontré le deuxième lemme de Borel–Cantelli.

Solution 233 1. On a

$$\mathbb{E}[N_n^2] = \mathbb{E}\left[\sum_{k=1}^{n} \mathbb{1}_{B_k}^2 + 2\sum_{k=2}^{n}\sum_{p=1}^{k-1} \mathbb{1}_{B_k}\mathbb{1}_{B_p}\right].$$

Or $\sum_{k=1}^{n} \mathbb{1}_{B_k}^2 = \sum_{k=1}^{n} \mathbb{1}_{B_k} = N_n$ d'où en prenant l'espérance

$$\mathbb{E}[N_n^2] = \mathbb{E}[N_n] + 2\sum_{k=2}^{n}\sum_{p=1}^{k-1} \mathbb{P}(B_k \cap B_p).$$

Comme $p < k$, on a $B_k \cap B_p = B_p \cap \{X_{2p^2+1} + \dots + X_{2k^2} = 0\}$. Puisque B_p et $\{X_{2p^2+1} + \dots + X_{2k^2} = 0\}$ sont indépendants, on a

$$\mathbb{P}(B_k\cap B_p) = \mathbb{P}(B_p)\mathbb{P}(X_{2p^2+1}+\dots+X_{2k^2} = 0) = \mathbb{P}(B_p)\mathbb{P}(S_{2k^2-2p^2} = 0),$$

ce qui donne l'identité voulue.
Rappelons l'équivalent $\mathbb{P}(S_{2n} = 0) \sim \frac{1}{\sqrt{\pi n}}$ et que $\mathbb{P}(S_{2n} = 0) > 0$. On en déduit l'existence de constantes c_1 et c_2 telles que

$$\forall n \geq 0, \quad \frac{c_1}{\sqrt{n}} \leq \mathbb{P}(S_{2n} = 0) \leq \frac{c_2}{\sqrt{n}},$$

ce qui nous donne $\mathbb{P}(S_{2p^2}=0)\mathbb{P}(S_{2(k^2-p^2)}=0)\le \frac{c_2^2}{p\sqrt{k^2-p^2}}$, d'où la majoration demandée avec $C=2c_2^2$.

2. Soit $p\le k/2$. Les inégalités suivantes sont équivalentes

$$\begin{aligned} c_{p_k}\ge c_{k-p,k} &\iff p^2(k^2-p^2)\le (k-p)^2(k^2-(k-p)^2)\\ &\iff p^2(k^2-p^2)\le (k-p)^2p(2k-p)\\ &\iff p(k+p)\le (k-p)(2k-p), \end{aligned}$$

ce qui est vrai car $p\le k-p$ et $k+p\le 2k-p$. Comme les termes sont positifs, on a alors, pour tout $k\ge 2$:

$$\begin{aligned} s_k &= \sum_{p=1}^{k-1} c_{p,k}\le \sum_{1\le p\le k/2}(c_{p,k}+c_{k-p,k})\\ &\le 2\sum_{1\le p\le k/2} c_{p,k}=2\sum_{1\le p\le k/2}\frac{1}{p\sqrt{k^2-p^2}}\\ &\le 2\sum_{1\le p\le k/2}\frac{1}{p\sqrt{k^2-\frac{k^2}{4}}}=\frac{4}{\sqrt{3}}\frac{1}{k}\sum_{1\le p\le k/2}\frac{1}{p}\\ &\le \frac{4}{\sqrt{3}}\frac{1}{k}\left(1+\sum_{2\le p\le k/2}\int_{k-1}^{k}\frac{dx}{x}\right)\\ &\le \frac{4}{\sqrt{3}}\frac{1}{k}(1+\log(k/2))\le \frac{4}{\sqrt{3}}\frac{1}{k}(\log 2+\log(k))\\ &\le \frac{8}{\sqrt{3}}\frac{\log k}{k}. \end{aligned}$$

3. On déduit des questions précédentes que

$$\frac{\mathbb{E}[N_n^2]}{(\mathbb{E}[N_n])^2}\le \frac{1}{\mathbb{E}[N_n]}+\frac{CD}{(\mathbb{E}[N_n])^2}\sum_{k=2}^{n}\frac{\log k}{k}.$$

On a d'une part

$$\mathbb{E}[N_n]=\sum_{k=1}^{n}\mathbb{P}(B_k)\ge \sum_{k=1}^{n}\frac{c_1}{k}\ge c_1\sum_{k=1}^{n}\int_{k}^{k+1}\frac{dx}{x}=c_1\log(n+1)$$

et d'autre part, comme $x\mapsto \frac{\log x}{x}$ est décroissante sur $[e,+\infty[$,

$$\begin{aligned} \sum_{k=2}^{n}\frac{\log k}{k} &\le \frac{\log 2}{2}+\frac{\log 3}{3}+\sum_{k=4}^{n}\int_{k-1}^{k}\frac{\log x}{x}\,dx\\ &\le \frac{\log 2}{2}+\frac{\log 3}{3}+\frac{1}{2}(\log n)^2. \end{aligned}$$

En mettant tout ensemble, on obtient

$$\overline{\lim_{n\to+\infty}} \frac{\mathbb{E}[N_n^2]}{(\mathbb{E}[N_n])^2} \leq \frac{CD}{2c_1^2} = \frac{8c_2^2}{c_1^2\sqrt{3}} < +\infty.$$

4. Passant à l'inverse, on a

$$\underline{\lim_{n\to+\infty}} \frac{(\mathbb{E}[N_n])^2}{\mathbb{E}[N_n^2]} > 0,$$

et donc, d'après le lemme de Kochen–Stone, $\mathbb{P}\left(\overline{\lim_{n\to+\infty}} B_n\right) > 0$. Or l'événement $\overline{\lim_{n\to+\infty}} B_n$ est dans $\mathcal{Q}$, donc d'après la loi du 0–1 de Hewitt et Savage, $\mathbb{P}\left(\overline{\lim_{n\to+\infty}} B_n\right) = 1$.

Solution 234 On raisonne par l'absurde : si a_n ne tendait pas vers 0, il existerait $\alpha > 0$ et une suite strictement croissante d'indices $(n_k)_{k\geq 1}$ avec $|a_{n_k}| \geq \alpha$ pour tout k. Comme $|\alpha X_{n_k}| \leq |a_{n_k} X_{n_k}|$, $(\alpha X_{n_k})_{k\geq 1}$ convergerait en probabilité, donc en loi vers 0. Or la suite $(\alpha X_{n_k})_{k\geq 1}$ est constante en loi. La loi de αX_1 serait donc identiquement nulle, ce qu'on a exclu.
La réciproque est simple. En effet, si (a_n) converge vers 0, alors $a_n X_n$ converge en loi vers 0 (car les X_n ont toutes même loi). Comme 0 est une constante, cela implique la convergence en probabilité de X_n vers 0.

Solution 235 1. Comme X_n suit une loi de Poisson de paramètre λ_n, on a $\mathbb{P}(X_n = 1) = e^{-\lambda_n}\lambda_n = f(\lambda_n)$ avec $f(x) = xe^{-x}$.
Comme $f'(x) = (1-x)e^{-x}$, il est facile de voir que f est maximale en 1, donc $\mathbb{P}(X_n = 1) \leq e^{-1}$ et $\mathbb{P}(X_n \neq 1) \geq 1 - e^{-1}$. La série de terme général $\mathbb{P}(X_n \neq 1)$ diverge donc trivialement. Comme les événements $\{X_n \neq 1\}$ sont indépendants, le deuxième lemme de Borel-Cantelli entraîne que $\mathbb{P}(X_n \neq 1 \text{ pour une infinité de } n) = 1$

2. À ω fixé, La suite $(Y_n(\omega))$ est une suite d'entiers : elle ne peut converger que si elle est constante à partir d'un certain rang. Un produit d'entiers non nuls étant non nul, $(Y_n(\omega))$ converge donc vers 0 si et seulement si il existe n tel que $X_n(\omega) = 0$. Comme $Y_{n+1} - Y_n = (X_{n+1} - 1)X_n$, la suite $(Y_n(\omega))$ converge alors vers une limite non-nulle si et seulement si $(X_n(\omega))$ est constamment égale à 1 à partir d'un certain rang. Maintenant si $X_n(\omega)$ n'est pas constamment égale à 1 à partir d'un certain rang et ne contient pas de 0, la suite $(Y_n(\omega))$ est croissante et non majorée (sinon elle convergerait, ce qui n'est pas possible car $X_n(\omega)$ n'est pas constamment égale à 1 à partir d'un certain rang) : elle tend vers l'infini.

On a donc $\Omega\backslash\{Y_n \to 0\} = B \cup \{Y_n \to +\infty\}$, avec

$$B = \{X_n = 1 \text{ à partir d'un certain rang}\} = \underline{\lim}_{n\geq 1} \{X_n = 1\}.$$

On a

$$\begin{aligned}\mathbb{P}(B) &= 1 - \mathbb{P}(B^c) = 1 - \mathbb{P}(\overline{\lim}_{n\geq 1}\{X_n \neq 1\}) \\ &= 1 - \mathbb{P}(X_n \neq 1 \text{ i.s}) = 1 - 1 = 0.\end{aligned}$$

Donc

$$\begin{aligned}\mathbb{P}(Y_n \to +\infty) &= \mathbb{P}(B \cup \{Y_n \to +\infty\}) \\ &= \mathbb{P}(\Omega\backslash\{Y_n \to 0\}) = \mathbb{P}(\forall n \geq 1; X_n \neq 0).\end{aligned}$$

Si on pose $A_N = \cap_{n=1}^N \{X_n \neq 0\}$, on a $\{\forall n \geq 1; X_n \neq 0\} = \cap_{N\geq 1} A_N$, donc avec le théorème de continuité séquentielle croissante, on a

$$\mathbb{P}(\forall n \geq 1; X_n \neq 0) = \lim_{N\to+\infty} \mathbb{P}(A_N).$$

Or, par indépendance

$$\mathbb{P}(A_N) = \prod_{n=1}^N \mathbb{P}(X_n \neq 0) = \prod_{n=1}^N (1 - \mathbb{P}(X_n = 0)),$$

d'où

$$\mathbb{P}(\forall n \geq 1; X_n \neq 0) = \lim_{N\to+\infty} \prod_{n=1}^N (1 - e^{-\lambda_n}) = \prod_{n=1}^{+\infty} (1 - e^{-\lambda_n}).$$

Comme $\mathbb{P}(\Omega\backslash\{Y_n \to 0\}) = 1 - \mathbb{P}(Y_n \to 0)$, on a l'égalité voulue.

3. (a) Soit N_1 tel que $\lambda_n \leq \log n$ pour $n \geq N_1$. Pour $N \geq N_1$, on a

$$\mathbb{P}(A_N) \leq \prod_{n=N_1}^N (1 - 1/n) = \prod_{n=N_1}^N \frac{n-1}{n} = \frac{N_1 - 1}{N},$$

ce qui donne $\mathbb{P}(Y_n \to +\infty) = 0$ en faisant tendre N vers l'infini.

(b) On a ici $e^{-\lambda_n} = \frac{1}{(n+1)^2}$ et

$$1 - e^{-\lambda_n} = \frac{(n+1)^2 - 1}{(n+1)^2} = \frac{n(n+2)}{(n+1)^2} = \frac{q_n}{q_{n+1}}$$

avec $q_n = \frac{n}{n+1}$. De la sorte

$$\mathbb{P}(A_N) = \prod_{n=1}^N \frac{q_n}{q_{n+1}} = \frac{q_1}{q_{N+1}}.$$

Comme $\lim_{n\to+\infty} q_n = 1$, on a $\mathbb{P}(Y_n \to +\infty) = q_1 = \frac{1}{2}$.

Solution 236 1. Pour tout entier naturel k , on a $\mathbb{1}_{\{k\geq 2\}} \leq \frac{1}{2}k(k-1)$ (les deux membres de l'inégalité sont nuls pour $k=0,1$ et pour $k\geq 2$, on a $k-1\geq 1$ d'où en faisant le produit $k(k-1)\geq 2$. On a donc $\mathbb{1}_{\{X_n\geq 2\}} \leq \frac{1}{2}X_n(X_n-1)$. On en déduit que

$$\mathbb{P}(X_n\geq 2)=\mathbb{E}(\mathbb{1}_{\{X_n\geq 2\}})\leq \frac{1}{2}\mathbb{E}(X_n(X_n-1))\leq \frac{1}{2}(\mathbb{E}(X_n^2)-\mathbb{E}(X_n))$$
$$\leq \frac{1}{2}(\mathrm{Var}(X_n)+(\mathbb{E}X_n)^2-\mathbb{E}(X_n))=\frac{1}{2}\lambda_n^2.$$

Ainsi, avec le lemme de Borel-Cantelli, $\mathbb{P}(X_n\geq 2 \text{ i.s})=0$, donc presque sûrement, à partir d'un certain rang $X_n<2$, soit $X_n\leq 1$, ce qui entraîne $\overline{\lim_{n\to+\infty}}\ X_n\leq 1$.

2. Comme précédemment, on a

$$\mathbb{1}_{\{X_n\geq k\}}\leq \frac{X_n(X_n-1)\dots(X_n-k+1)}{k!},$$

d'où

$$\mathbb{P}(X_n\geq k)\leq \frac{\mathbb{E}(X_n(X_n-1)\dots(X_n-k+1))}{k!}$$
$$=\sum_{i=k}^{+\infty}e^{-\lambda_n}\frac{\lambda_n^i}{i!}i(i-1)\dots(i-k+1)$$
$$=\sum_{i=k}^{+\infty}e^{-\lambda_n}\frac{\lambda_n^i}{(i-k)!}=\lambda_n^k.$$

Comme précédemment, on obtient avec le lemme de Borel-Cantelli $\mathbb{P}(X_n\geq k \text{ i.s.})=0$, puis $\overline{\lim_{n\to+\infty}}\ X_n\leq k-1$ presque sûrement.

3. Posons $k=\lfloor 1/\alpha\rfloor+1$. On a $k>1/\alpha$ donc $\alpha k>1$. La série de terme général $\lambda_n^k=(n^{-\alpha})^k=n^{-\alpha k}$ converge donc d'après la question précédente $\overline{\lim}X_n\leq k-1=\lfloor 1/\alpha\rfloor$ presque sûrement, ce qui entraîne que presque sûrement, l'ensemble des valeurs d'adhérence de (X_n) est inclus dans $\{0,\dots,\lfloor 1/\alpha\rfloor\}$, puisque les valeurs d'adhérence d'une suite d'entiers sont toujours des entiers. Prenons maintenant $i\in\{0,\dots,\lfloor 1/\alpha\rfloor\}$. On a $0\leq i\leq \lfloor 1/\alpha\rfloor\leq 1/\alpha$, donc $i\alpha\in[0,1]$. On obtient ainsi

$$\sum_{n=0}^{+\infty}\mathbb{P}(X_n=i)=\sum_{n=0}^{+\infty}\frac{\lambda_n^i}{i!}e^{-\lambda_n}.$$

Or $\frac{\lambda_n^i}{i!}e^{-\lambda_n}\sim\frac{\lambda_n^i}{i!}=\frac{1}{i!}\frac{1}{n^{\alpha i}}$, donc d'après le théorème sur les équivalents de séries à termes positifs, la série diverge. Comme les événements sont indépendants, le deuxième lemme de Borel-Cantelli implique que $\mathbb{P}(X_n = i \text{ i.s}) = 1$, ce qui entraîne que presque sûrement, i est une valeur d'adhérence de (X_n). Comme l'intersection d'un nombre fini d'ensembles de probabilité 1 est encore de probabilité 1, on peut dire que presque sûrement, l'ensemble des valeurs d'adhérence de (X_n) est l'ensemble $\{0, \dots, \lfloor 1/\alpha \rfloor\}$.

Solution 237 1. Soit L_r l'ensemble des familles libres de cardinal r. On a $L_1 = E\backslash\{0\}$, donc $|L_1| = q^n - 1$. Pour $r \geq 1$, notons ψ_r la projection canonique de L_{r+1} dans L_r : $\psi_r(u_1, \dots, u_{r+1}) = (u_1, \dots, u_r)$. Pour tout $u \in L_r$, on a

$$\begin{aligned} \psi_r^{-1}(u) &= \{(u, x); x \in E\backslash\text{Vect}(u_1, \dots, u_r)\} \\ |\psi_r^{-1}(u)| &= |E| - |\text{Vect}(u_1, \dots, u_r)| = q^n - q^r, \end{aligned}$$

car la famille $(u_1, \dots, u_r)$ est libre. Le lemme des bergers nous donne alors $|L_{r+1}| = (q^n - q^r)$, ce qui donne par récurrence

$$|L_r| = \prod_{i=0}^{r-1} (q^n - q^i).$$

Considérons l'application ϕ de L_r dans $\text{Gr}_r(E)$ qui à $(u_1, \dots, u_r) \in L_r$ associe $\text{Vect}(u_1, \dots, u_r)$. Pour tout $V \in \text{Gr}_r(E)$, on a $\phi^{-1}(V) = L_r(V)$ et

$$|\phi^{-1}(V)| = \prod_{i=0}^{r-1} (q^r - q^i).$$

En appliquant encore le lemme des bergers, il vient

$$\begin{aligned} |\text{Gr}_r(E)| &= \frac{|L_r|}{\prod_{i=0}^{r-1}(q^r - q^i)} = \frac{\prod_{i=0}^{r-1}(q^n - q^i)}{\prod_{i=0}^{r-1}(q^r - q^i)} \\ &= \frac{\prod_{i=0}^{r-1}(q^{n-i} - 1)}{\prod_{i=0}^{r-1}(q^{r-i} - 1)} = \frac{\prod_{i=n-r+1}^{n}(q^i - 1)}{\prod_{i=1}^{r}(q^i - 1)} \\ &= \frac{Q(n)}{Q(r)Q(n-r)}. \end{aligned}$$

2. (a) Soit $\varepsilon > 0$. On veut montrer que $\lim_{n\to+\infty} \mathbb{P}(|\frac{R_n}{n} - 1| > \varepsilon) = 0$. Comme $R_n \leq n$, on a $\mathbb{P}(|\frac{R_n}{n} - 1| > \varepsilon) = \mathbb{P}(R_n < (1-\varepsilon)n)$. Soit r_n le plus petit entier dépassant $n(1-\varepsilon)$. Bien sûr, $r_n \sim n(1-\varepsilon)$. On veut montrer que $\mathbb{P}(R_n \leq r_n)$ tend vers 0. Le nombre de sous-espaces vectoriels de dimension r_n de $\mathbb{F}_q^n$ vaut

$$|\text{Gr}_{r_n}(\mathbb{F}_q^n)| = \frac{Q(n)}{Q(r_n)Q(n-r_n)},$$

avec

$$Q(n) = \prod_{i=1}^{n}(q^i - 1) = q^{n(n+1)/2}\prod_{i=1}^{n}(1 - q^{-i}).$$

Comme le produit $P = \prod_{i=1}^{+\infty}(1 - q^{-i})$ converge, on a l'équivalent $Q(n) \sim Pq^{n(n+1)/2}$. Après simplifications, on trouve que

$$|\mathrm{Gr}_{r_n}(\mathbb{F}_q^n)| \sim \frac{q^{nr_n}}{Pq^{r_n^2}}.$$

Alors

$$\begin{aligned}\mathbb{P}(R_n \leq r_n) &= \mathbb{P}(\exists V \in \mathrm{Gr}_{r_n}(\mathbb{F}_q^n); \mathrm{Im}\, M_n \subset E)\\ &\leq \sum_{V\in\mathrm{Gr}_{r_n}(\mathbb{F}_q^n)} \mathbb{P}(\mathrm{Im}\, M_n \subset E) = \sum_{V\in\mathrm{Gr}_{r_n}(\mathbb{F}_q^n)} (\frac{q^{r_n}}{q^n})^n\\ &\leq |\mathrm{Gr}_{r_n}(\mathbb{F}_q^n)|(\frac{q^{r_n}}{q^n})^n.\end{aligned}$$

Finalement $\mathbb{P}(R_n \leq r_n) = O(q^{-(n-r_n)^2})$, qui est bien de limite nulle quand n tend vers l'infini : ceci donne la convergence en probabilité.

Comme la suite $\frac{R_n}{n}$ est bornée par une constante, la convergence en probabilité entraîne la convergence dans L^1, et donc celle des espérances.

En effet, pour tout ε positif, on a

$$\begin{aligned}|\frac{R_n}{n} - 1| &\leq \varepsilon + \mathbb{1}_{\{|\frac{R_n}{n}-1|>\varepsilon\}}\\ \mathbb{E}|\frac{R_n}{n} - 1| &\leq \varepsilon + \mathbb{P}(|\frac{R_n}{n} - 1| > \varepsilon),\end{aligned}$$

d'où $\overline{\lim}_{n\to+\infty} \mathbb{E}(|\frac{M_n}{n}-1|) \leq \varepsilon$, ce qui donne le résultat voulu puisque ε peut être pris arbitrairement petit.

(b) Si $n - r_n \sim c\sqrt{\log n}$, on a

$$\log \mathbb{P}(R_n \leq r_n) \leq (\log q)(-c^2 \log n) + O(1).$$

Cela donne la convergence de la série de terme général $\mathbb{P}(R_n \leq r_n)$ si $c^2 \log q > 1$. On conclut avec le lemme de Borel–Cantelli.

D.11 Exercices sur la convergence en loi

Solution 247 1. Soit $n \geq 1$. Utilisons la fonction caractéristique pour déterminer la loi de X_n. D'après l'indépendance des variables U_k,

$$\begin{aligned}\phi_{X_n}(t) &= \mathbb{E}e^{itX_n} = \mathbb{E}e^{it(\theta X_{n-1}+U_n)} = \mathbb{E}e^{itU_n+it\theta U_{n-1}+\ldots+it\theta^n X_0}\\ &= \prod_{k=0}^{n-1}\mathbb{E}(e^{it\theta^k U_{n-k}}) = \prod_{k=0}^{n-1}\phi_{U_{n-k}}(t\theta^k).\end{aligned}$$

Comme les variables U_k sont identiquement distribuées, elles ont même fonction caractéristique et donc

$$\begin{aligned}\phi_{X_n}(t) &= \prod_{k=0}^{n-1}\phi_{U_n}(t\theta^k) = \prod_{k=0}^{n-1}\exp\left\{-\frac{t^2}{2}\theta^{2k}\right\}\\ &= \exp\left\{-\frac{t^2}{2}(1+\theta^2+\theta^4+\ldots+\theta^{2n-2})\right\}\end{aligned}$$

et donc X_n est une variable gaussienne centrée de variance

$$1+\theta^2+\ldots+\theta^{2(n-1)} = \frac{1-\theta^{2n}}{1-\theta^2}.$$

2. Si $|\theta| < 1$, la fonction caractéristique de X_n converge vers la fonction $t \mapsto \exp\{-\frac{t^2}{2(1-\theta^2)}\}$. Ainsi, X_n converge en loi vers une variable gaussienne centrée de variance $\frac{1}{1-\theta^2}$.
Si $|\theta| \geq 1$, la série de terme général θ^{2k} diverge. Ainsi, la suite des fonctions caractéristiques converge vers l'indicatrice du point 0, qui n'est pas une fonction continue. Donc la suite $(X_n)_{n\geq 1}$ ne converge pas en loi.

Solution 248 Soit θ non nul. La fonction caractéristique de X_n est équivalente en l'infini à $\phi_{X_n}(\theta) = \mathbb{E}[e^{i\theta X_n/n}] = \frac{1}{n}\sum_{k=0}^{n-1} e^{ik\theta/n} = \frac{1}{n}\frac{1-e^{i\theta}}{1-e^{i\theta/n}} \sim \frac{1-e^{i\theta}}{-i\theta}$. En $\theta = 0$, la suite est identiquement égale à 1. Or si U suit la loi uniforme sur $[0,1]$, $\phi_U(0) = 1$ tandis que $\phi_U(\theta) = \int_0^1 e^{i\theta u}\,du = \frac{e^{i\theta}-1}{i\theta}$. Avec le théorème de Lévy, ceci montre que X_n/n converge en loi vers la loi uniforme sur $[0,1]$.

Solution 249 1. Soit $\varepsilon > 0$. Pour $n \geq 1$, on a

$$\begin{aligned}\mathbb{P}(|X_n| > \varepsilon) &= \int_\varepsilon^{+\infty} n^2 x e^{-nx}\,dx\\ &\leq e^{-(n-1)\varepsilon)}\int_\varepsilon^{+\infty} n^2 x e^{-x}\,dx \leq n^2 e^{-(n-1)\varepsilon},\end{aligned}$$

qui tend vers 0 quand n tend vers l'infini. Cela montre que X_n converge en probabilité, et donc en loi vers 0.

2. Posons $\phi(x) = |\frac{\sin x}{x}|$ pour $x \neq 0$, et $\phi(0) = 1$. La fonction ϕ est continue sur $\mathbb{R}$ car $\sin x \sim x$ en 0, et elle est bornée par 1. Par définition de la convergence en loi, on a alors $\lim\limits_{n\to+\infty} \mathbb{E}(\phi(X_n)) = \phi(0) = 1$. Or, le théorème de transfert donne
$$\mathbb{E}(\phi(X_n)) = \int_0^{+\infty} \phi(x)n^2xe^{-nx}\,dx = n^2\int_0^{+\infty} |\sin x|e^{-nx}\,dx,$$
d'où $\int_0^{+\infty} |\sin t|e^{-nt}\,dt \sim \frac{1}{n^2}$. Cet équivalent peut se retrouver de manière indépendante à l'aide de la méthode de Laplace, ou plus simplement encore par convergence dominée après le changement de variable $t = u/n$.

Solution 250 1. On trouve que $\mathbb{P}(M_n \leq x) = 1$ si $x \geq 1$, $\mathbb{P}(M_n \leq x) = x^n$ si $0 \leq x \leq 1$ et $\mathbb{P}(M_n \leq x) = 0$ si $x \leq 0$.

2. La limite de la fonction de répartition de M_n vaut :
$$\lim_{n\to+\infty} \mathbb{P}(M_n \leq x) = \begin{cases} 1 & \text{si } x \geq 1 \\ 0 & \text{si } x < 1 \end{cases}$$
On reconnaît la fonction de répartition de la variable aléatoire constante valant 1 partout, donc M_n converge en loi vers 1.
Soit $x \in [0,1[$. Posons $A_n = \{M_n \leq x\}$. La série de terme général $\mathbb{P}(A_n) = x^n$ converge, donc le premier lemme de Borel–Cantelli implique $\mathbb{P}\left(\overline{\lim\limits_{n\to+\infty}}\, A_n\right) = 0$, ce qui entraîne $\mathbb{P}\left(\underline{\lim\limits_{n\to+\infty}}\, M_n \geq x\right) = 1$.
En prenant $x = 1 - 1/k$ et en procédant par intersection dénombrable, on obtient que p.s. $\underline{\lim\limits_{n\to+\infty}}\, M_n \geq 1$; comme de plus $M_n \leq 1$, M_n converge presque sûrement vers 1.

3. Étudions la fonction de répartition de $n(1 - M_n)$. On sait que, pour tout $x \in \mathbb{R}$, on a
$$\mathbb{P}(n(1 - M_n) \leq x) = \mathbb{P}(M_n \geq 1 - x/n) = 1 - \mathbb{P}(M_n < 1 - x/n).$$
Or nous savons ce que vaut cette dernière probabilité. En effet, on a vu que $\mathbb{P}(M_n < 1 - x/n) = 1$ si $1 - x/n \geq 1$, soit $x \leq 0$. De plus, on sait que $\mathbb{P}(M_n < 1 - x/n) = \left(1 - \frac{x}{n}\right)^n$ si $0 \leq 1 - \frac{x}{n} \leq 1$, soit $0 \leq x \leq n$, tandis que cette probabilité vaut 1 dès que $x \geq n$. Nous faisons maintenant tendre n vers $+\infty$. Dans ce cas, on trouve que la limite en l'infini de $\left(1 - \frac{x}{n}\right)^n = \exp\{n \log(1 - x/n)\}$ est e^{-x}. Ainsi, on obtient
$$\lim_{n\to+\infty} \mathbb{P}(n(1 - M_n) \leq x) = (1 - e^{-x})\mathbb{1}_{\{x>0\}}.$$

On reconnaît la fonction de répartition d'une loi exponentielle de paramètre 1. Donc, $n(1-M_n)$ converge en loi vers une variable de loi exponentielle de paramètre 1.

Solution 251 1. Il suffit d'appliquer le théorème de Slutsky. On sait en effet que la suite $(a_n)_n$ converge vers a (donc la convergence a aussi lieu en probabilité) et donc a_nY_n converge en loi vers aY. De même comme b_n est déterministe, on en déduit que $a_nY_n + b_n$ converge en loi vers $aY + b$.

2. D'après le théorème central limite, $\frac{S_n - \mathbb{E}S_n}{\sqrt{\mathrm{Var}(S_n)}}$ converge en loi vers une variable gaussienne $\mathcal{N}(0,1)$. Il nous faut donc déterminer les valeurs de $\mathbb{E}S_n$ et $\mathrm{Var}\, S_n$. Par linéarité de l'espérance, on obtient

$$\mathbb{E}S_n = \sum_{k=0}^{n} \mathbb{E}X_k = (n+1)\mathbb{E}X_1 = n+1.$$

De plus, l'indépendance des variables X_k implique

$$\mathrm{Var}(S_n) = \sum_{k=0}^{n} \mathrm{Var}\, X_k = (n+1)\,\mathrm{Var}\, X_1 = n+1.$$

Ainsi, $\frac{S_n-(n+1)}{\sqrt{n+1}}$ converge en loi vers une variable gaussienne $\mathcal{N}(0,1)$. Il suffit d'utiliser la question précédente avec $Y_n = \frac{S_n-(n+1)}{\sqrt{n+1}}$, $a_n = \sqrt{\frac{n+1}{n}}$ (qui converge vers 1) et $b_n = 1/\sqrt{n}$ (qui converge vers 0) pour obtenir le résultat voulu.

3. Montrons par récurrence que S_n suit la loi $\Gamma(n+1,1)$. C'est clair pour $n=0$, puisque $\mathcal{E}(1) = \Gamma(1,1)$. Comme S_n est $\sigma(X_0, \dots X_n)$-mesurable et que les X_i sont indépendants, S_n est indépendant de X_{n+1}, ce qui nous donne $\mathbb{P}_{S_{n+1}} = \mathbb{P}_{S_n} * \Gamma(1,1)$. L'égalité $\mathbb{P}_{S_n} = \Gamma(n+1,1)$ vient alors aisément par récurrence, car $\Gamma(a,\lambda) * \Gamma(b,\lambda) = \Gamma(a+b,\lambda)$ [2]. Par ailleurs, pour $x \in \mathbb{R}$:

$$\mathbb{P}\left(\frac{S_n - n}{\sqrt{n}} \leq x\right) = \mathbb{P}(S_n \leq n + x\sqrt{n}).$$

2. Cela a été démontré par convolution des densités dans le cours et avec la transformée de Laplace dans l'exercice 203.

Exprimant la densité de $\frac{S_n-n}{\sqrt{n}}$ en fonction de celle de S_n, on obtient

$$\begin{aligned} f_{\frac{S_n-n}{\sqrt{n}}}(x) &= \sqrt{n} f_{S_n}(n+x\sqrt{n}) \\ &= \sqrt{n}\frac{1}{\Gamma(n+1)}(n+x\sqrt{n})^n e^{-(n+x\sqrt{n})}\ \mathbb{1}_{\mathbb{R}_+^*}(n+x\sqrt{n}) \\ &= \frac{\sqrt{n}}{\Gamma(n+1)} n^n(1+x/\sqrt{n})^n e^{-(n+x\sqrt{n})}\ \mathbb{1}_{[-\sqrt{n},+\infty[}(x), \end{aligned}$$

ce qui donne le résultat voulu par identification des fonctions h_n et a_n.

4. Comme $\frac{S_n-n}{\sqrt{n}}$ converge en loi vers la variable gaussienne standard N qui ne charge ni 0 ni 1, on obtient
$$\int_0^1 g_n(x)\,dx = \mathbb{P}\left(\frac{S_n-n}{\sqrt{n}} \in [0,1]\right) \underset{n\to+\infty}{\to} \int_0^1 \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx.$$
5. Remarquons que
$$\left(1+\frac{x}{\sqrt{n}}\right)^n = \exp\left\{n\log(1+x/\sqrt{n})\right\} = \exp\left\{n\left(\frac{x}{\sqrt{n}}-\frac{x^2}{2n}\right)+o(1)\right\}.$$
Ainsi, h_n converge ponctuellement vers $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. De plus, pour n suffisamment grand, h_n est dominée par $h(x) = e^{-x^2/3}$, qui est une fonction intégrable sur $\mathbb{R}$. Ainsi, on conclut par le théorème de convergence dominée. Autre solution :
On peut noter que $h_n(x) = \frac{1}{\sqrt{2\pi}}\mathbb{1}_{]-1,+\infty[}(\frac{x}{\sqrt{n}})\exp(-x^2\psi(\frac{x}{\sqrt{n}}))$ avec, pour $x>-1$, $\psi(x) = \frac{x-\log(1+x)}{x^2}$. Un développement limité donne classiquement $\lim_{x\to 0}\psi(x) = \frac{1}{2}$, donc h_n converge ponctuellement vers $x \mapsto \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. De plus, comme $\psi(x)\geq 0$ pour tout $x>-1$, h_n est dominée par $h(x) = \frac{1}{\sqrt{2\pi}}$. Ainsi, on conclut par le théorème de convergence dominée.
6. Pour finir, on écrit $\lim\limits_{n\to+\infty} a_n = \dfrac{\lim\limits_{n\to+\infty}\int_0^1 g_n(x)\,dx}{\lim\limits_{n\to+\infty}\int_0^1 h_n(x)\,dx} = 1$. De plus, on sait que $\Gamma(n+1) = n!$ et par définition de a_n, on a
$$1 = \lim_{n\to+\infty} a_n = \lim_{n\to+\infty}\frac{n^{n+1/2}e^{-n}\sqrt{2\pi}}{\Gamma(n+1)} = \lim_{n\to+\infty}\frac{n^{n+1/2}e^{-n}\sqrt{2\pi}}{n!}.$$
On obtient ainsi la formule de Stirling.

Solution 252 Dans toute cette correction, on suppose $\theta > 0$. Les résultats cherchés s'en déduisent aisément pour $\theta < 0$.

1. Soit donc $\theta > 0$. Les variables $Y_{n,k}$ étant indépendantes, la fonction

caractéristique de Z_n est

$$\begin{aligned}\phi_{Z_n}(\theta) &= \mathbb{E}e^{i\theta Z_n} = \mathbb{E}\exp\left\{i\theta\frac{1}{n}\sum_{k=-n^2}^{n^2} kY_{n,k}\right\}\\ &= \prod_{k=-n^2}^{n^2}\mathbb{E}\exp\left\{i\theta\frac{k}{n}Y_{n,k}\right\}.\end{aligned}$$

De plus, on sait que $Y_{n,k}$ suit une loi de Poisson, ce qui donne par le théorème du transfert

$$\begin{aligned}\mathbb{E}\exp\left\{i\theta\frac{k}{n}Y_{n,k}\right\} &= \sum_{j=0}^{+\infty}\exp\left\{i\theta\frac{k}{n}j\right\}\mathbb{P}(Y_{n,k}=j)\\ &= e^{-cn^\alpha/|k|^{1+\alpha}}\sum_{j=0}^{+\infty}\left(\exp\left\{i\theta\frac{k}{n}\right\}\frac{cn^\alpha}{|k|^{1+\alpha}}\right)^j\frac{1}{j!}\\ &= \exp\left\{-c\frac{n^\alpha}{|k|^{1+\alpha}} + e^{i\theta\frac{k}{n}}c\frac{n^\alpha}{|k|^{1+\alpha}}\right\}\\ &= \exp\left\{-c\frac{n^\alpha}{|k|^{1+\alpha}}\left(1-e^{i\theta\frac{k}{n}}\right)\right\}.\end{aligned}$$

On déduit de ce résultat l'expression de la fonction caractéristique de Z_n (en remarquant que pour $k=0$, le terme multiplicatif vaut 1) :

$$\begin{aligned}\phi_{Z_n}(\theta) &= \prod_{k=-n^2}^{n^2}\exp\left\{-c\frac{n^\alpha}{|k|^{1+\alpha}}\left(1-e^{i\theta\frac{k}{n}}\right)\right\}\\ &= \prod_{k=1}^{n^2}\exp\left\{-c\frac{n^\alpha}{k^{1+\alpha}}\left(2-e^{i\theta\frac{k}{n}}-e^{-i\theta\frac{k}{n}}\right)\right\}\end{aligned}$$

puis

$$\begin{aligned}\phi_{Z_n} &= \prod_{k=1}^{n^2}\exp\left\{-c\frac{n^\alpha}{k^{1+\alpha}}\left(2-2\cos(\theta\frac{k}{n})\right)\right\}\\ &= \exp\left\{-2c\theta^\alpha\sum_{k=1}^{n^2}\frac{n^\alpha}{\theta^\alpha k^{1+\alpha}}\left(1-\cos(\theta\frac{k}{n})\right)\right\}.\end{aligned}$$

Pour mettre en lumière l'expression souhaitée, il faut utiliser le “lien série-intégrale”. On sait en effet que pour toute fonction f continue sur $[0,\theta]$, on a

$$\frac{\theta}{n}\sum_{k=1}^{n} f\left(\frac{k}{n}\theta\right) = \int_0^\theta f\left(\lfloor\frac{nx}{\theta}\rfloor\frac{\theta}{n}\right)\,dx. \tag{D.4}$$

On décompose alors la série en

$$\sum_{k=1}^{n^2} \frac{n^\alpha}{\theta^\alpha k^{1+\alpha}} \left(1-\cos(\theta\frac{k}{n})\right) = \theta \sum_{j=0}^{n-1} \frac{1}{n} \sum_{k=jn+1}^{(j+1)n} \frac{\left(1-\cos(j\theta+\theta\frac{k}{n})\right)}{\left(j\theta+\theta\frac{k}{n}\right)^{1+\alpha}}.$$

D'après l'équation (D.4), pour tout $j \geq 0$, on a

$$\frac{\theta}{n} \sum_{k=jn+1}^{(j+1)n} f\left(j\theta+\theta\frac{k}{n}\right) = \int_{j\theta}^{(j+1)\theta} f\left(\lfloor\frac{nx}{\theta}\rfloor\frac{\theta}{n}\right)\, dx.$$

On obtient le résultat voulu en sommant sur tous les $0 \leq j \leq n-1$.

2. Rappelons que $0 < \alpha < 2$. Soit $x > 0$. Justifions d'abord l'inégalité $f(x) \leq \min\left(\frac{1}{x^\alpha}, \frac{2}{x^{1+\alpha}}, \frac{1}{2x^{\alpha-1}}\right)$. En effet, il suffit tour à tour de voir que $|\cos(x)| \leq 1$, $|1-\cos x| \leq x^2/2$ et $|1-\cos x| \leq x$. Ainsi, en majorant par $x^{1-\alpha}\mathbb{1}_{]0,1]}(x) + 2x^{-(1+\alpha)}\mathbb{1}_{]1,+\infty[}(x)$, le théorème de convergence dominée donne

$$\begin{aligned} \lim_{n\to+\infty} u_n(\theta) &= \lim_{n\to+\infty} \int_0^{+\infty} f\left(\frac{\theta}{n}\lfloor\frac{nx}{\theta}\rfloor\right) \mathbb{1}_{[0,n\theta]}(x)\, dx \\ &= \int_0^{+\infty} f(x)\, dx. \end{aligned}$$

On choisit alors $c = \frac{1}{2\int_0^{+\infty} f(x)\ dx}$ pour obtenir le résultat voulu.

3. On sait que la fonction $f(\theta) = e^{-|\theta|^\alpha}$ est continue en 0. Le résultat découle alors du théorème de continuité de Lévy.

4. On sait que la transformée de Fourier de la mesure m_α est $e^{-|\theta|^\alpha}$. Les variables X_k étant indépendantes, on a

$$\mathbb{E}\exp\{i\lambda_n\theta(X_1+\cdots+X_n)\} = \prod_{k=1}^n \mathbb{E}e^{i\lambda_n\theta X_k} = \prod_{k=1}^n e^{-|\theta\lambda_n|^\alpha} = e^{-|\theta|^\alpha\lambda_n^\alpha n}.$$

Choisissant λ_n tel que $\lambda_n^\alpha n = 1$, c'est-à-dire $\lambda_n = n^{-1/\alpha}$, on en déduit que $\lambda_n(X_1+\cdots+X_n)$ suit la loi m_α, puisque la transformée de Fourier caractérise la loi.

Solution 253 1. Soit $\varepsilon > 0$. Il existe $A > 0$ tel que $\mathbb{P}(|X| \geq A) \leq \varepsilon/2$. D'après le théorème de Portmanteau, on a

$$\overline{\lim_{n\to+\infty}}\ \mathbb{P}(|X_n| \geq A) \leq \mathbb{P}(|X| \geq A) \leq \varepsilon/2,$$

donc il existe n_0 tel que $\mathbb{P}(|X_n| \geq A) \leq \varepsilon$ pour $n \geq n_0$. On a ainsi $\mathbb{P}_{X_n}([-A,A]) \geq 1-\varepsilon$. Comme la famille finie $(X_i)_{1\leq i<n_0}$ est tendue, il existe un réel $A' \geq A$ tel que pour tout $i \in \{1,\ldots,n_0-1\}$, on a

$\mathbb{P}_{X_i}([-A', A']) \geq 1 - \varepsilon$. On a ainsi $\mathbb{P}_{X_i}([-A', A']) \geq 1 - \varepsilon$ pour tout $i \geq 1$: $(X_n)_{n\geq 1}$ est tendue.

2. Soit $\varepsilon > 0$. D'après la question précédente, il existe A tel que $\mathbb{P}(|X_n| > A) \leq \varepsilon/4$ pour tout $n \geq 1$.
   Comme $|e^{ixs} - e^{ixt}| \leq \min(|s-t||x|, 2)$, on obtient
   $$\begin{aligned} |e^{iX_n s} - e^{iX_n t}| &\leq |s-t||X_n| \, \mathbb{1}_{\{|X_n|\leq A\}} + 2 \, \mathbb{1}_{\{|X_n|>A\}} \\ &\leq A|s-t| + 2 \, \mathbb{1}_{\{|X_n|>A\}}, \end{aligned}$$
   d'où
   $$\begin{aligned} |\phi_{X_n}(s) - \phi_{X_n}(s)| &= |\mathbb{E}[e^{iX_n s} - e^{iX_n t}]| \leq \mathbb{E}[|e^{iX_n s} - e^{iX_n t}|] \\ &\leq A|s-t| + 2\mathbb{P}(|X_n| > A). \end{aligned}$$
   On en déduit que
   $$\forall n \geq 1, \quad \forall (s,t) \in \mathbb{R}^2, \quad |s-t| \leq \frac{\varepsilon}{2A} \Longrightarrow |\phi_{X_n}(s) - \phi_{X_n}(t)| \leq \varepsilon,$$
   ce qui est le résultat voulu.

Solution 254 Comme la suite (S_n) est croissante, on a pour tout $t \geq 0$ et tout réel a
$$\begin{aligned} \left\{ \frac{N_t - t}{\sqrt{t}} \leq a \right\} &= \{N_t \leq t + a\sqrt{t}\} = \{S_{\lfloor t+a\sqrt{t}\rfloor} \geq t\} \\ &= \left\{ \frac{S_{\lfloor t+a\sqrt{t}\rfloor} - (t + a\sqrt{t})}{\sqrt{t}} \geq -a \right\}. \end{aligned}$$
Soient (t_n) une suite tendant vers l'infini et a un réel. Le théorème central limite donne la convergence de $\frac{S_{\lfloor t_n+a\sqrt{t_n}\rfloor} - \lfloor t_n+a\sqrt{t_n}\rfloor}{\sqrt{\lfloor t_n+a\sqrt{t_n}\rfloor}}$ vers $X \sim \mathcal{N}(0,1)$. Avec le lemme de Slutsky, on en déduit la convergence de $\frac{S_{\lfloor t_n+a\sqrt{t_n}\rfloor} - (t_n+a\sqrt{t_n})}{\sqrt{t_n}}$ vers X, et donc, puisque la loi gaussienne ne charge pas $-a$, la convergence de
$$\mathbb{P}\left(\frac{N_{t_n} - t_n}{\sqrt{t_n}} \leq a \right) = \mathbb{P}\left(\frac{S_{\lfloor t+a\sqrt{t_n}\rfloor} - (t_n + a\sqrt{t_n})}{\sqrt{t_n}} \geq -a \right)$$
vers $\mathbb{P}(X \geq -a) = \mathbb{P}(-X \leq a) = \mathbb{P}(X \leq a)$. Comme a est pris quelconque, ceci montre que $\frac{N_{t_n} - t_n}{\sqrt{t_n}}$ converge en loi vers X, ce qui est le résultat voulu.

On a ainsi montré une convergence en loi à l'aide d'une autre convergence en loi en procédant à une réécriture d'événements : ce n'est pas un cas isolé.

Solution 255 1. Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes de loi de Poisson de paramètre 1. On pose $S_n = \sum_{k=1}^{n} X_k$. D'après les propriétés de convolution des lois de Poisson, S_n suit la loi de Poisson de paramètre n. La quantité $e^{-n}\sum\limits_{k=0}^{n}\frac{n^k}{k!}$ est égale à $\mathbb{P}(S_n\le n)$. Comme $\mathbb{E}(S_n)=n$, et que la loi de Poisson a un moment d'ordre deux, on a encore

$$\mathbb{P}\left(\frac{S_n-\mathbb{E}(S_n)}{\sqrt{\operatorname{Var} S_n}}\le 0\right)=e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}.$$

Comme la loi $\mathcal{N}(0,1)$ ne charge aucun point, le TCL entraîne que $\mathbb{P}(\frac{S_n-\mathbb{E}(S_n)}{\sqrt{\operatorname{Var} S_n}}\le 0)$ tend vers $\mathbb{P}(X\le 0)$ où $X\sim\mathcal{N}(0,1)=\frac{1}{2}$, ce qui donne le résultat voulu.

2. La suite $(Y_k)_{k\ge 1}$ des couleurs tirées se modélise par une suite de variables aléatoires indépendantes suivant la loi uniforme sur $\{1,\dots,n\}$. Ainsi

$$X_n=\sup\{k:(Y_i)_{1\le i\le k}\text{ injective}\}.$$

X_n prend des valeurs entières comprises entre 1 et n. Prenons k entre 0 et $n-1$. On a

$$\{X_n>k\}=\{(Y_i)_{1\le i\le k+1}\text{ injective}\}.$$

Comme $(Y_1,\dots,Y_{k+1})$ suit la loi uniforme sur $\{1,\dots,n\}^{k+1}$, on a

$$\mathbb{P}(X_n>k)=\frac{n(n-1)\dots(n-k)}{n^{k+1}}=\frac{n!}{(n-(k+1))!n^{k+1}},$$

d'où

$$\begin{aligned}\mathbb{E}[X_n]&=\sum_{k=0}^{n-1}\mathbb{P}(X_n>k)=\sum_{k=1}^{n}\frac{n!}{(n-k)!n^k}\\&=\frac{n!}{n^n}\sum_{k=0}^{n-1}\frac{n^k}{k!}.\end{aligned}$$

On a vu à la première question que $e^{-n}\sum\limits_{k=0}^{n-1}\frac{n^k}{k!}$ tend vers $1/2$. On en déduit l'équivalent $\mathbb{E}[X_n]\sim\frac{n!}{n^n}e^n/2\sim\sqrt{2\pi n}/2$, avec l'aide de la formule de Stirling.

Solution 256 Si c'était vrai, $Z_{2n}-Z_n$ convergerait en probabilité vers 0. La convergence en probabilité vers une constante entraîne la convergence en loi vers cette constante. Or le TCL permet de montrer assez simplement que $Z_{2n}-Z_n$ converge en loi vers $\mathcal{N}(0,2-\sqrt{2})$. En effet

$$Z_{2n}-Z_n=\frac{1}{\sqrt{2}}\frac{X_1+\dots+X_n}{\sqrt{n}}+(\frac{1}{\sqrt{2}}-1)\frac{X_{n+1}+\dots+X_{2n}}{\sqrt{n}}.$$

Comme le vecteur
$$\left(\frac{X_1+\dots+X_n}{\sqrt{n}}, \frac{X_{n+1}+\dots+X_{2n}}{\sqrt{n}}\right)$$
converge en loi vers $\mathcal{N}(0,1)\otimes\mathcal{N}(0,1)$, la convergence vers $\mathcal{N}(0,\frac{1}{2}+(\frac{1}{\sqrt{2}}-1)^2)$ s'ensuit.

Une autre façon de le montrer est de dire que s'il y avait convergence en probabilité, la variable aléatoire limite serait mesurable par rapport à la tribu terminale, et donc égale presque sûrement à une constante en vertu de la loi du 0-1 de Kolmogorov (voir par exemple l'exercice 111).

Solution 257 On a
$$\frac{\mathbb{E}[|S_n|]}{\sqrt{n}} = \int_0^{+\infty} \mathbb{P}\left(\frac{|S_n|}{\sqrt{n}} > t\right)\, dt.$$
Les X_i sont i.i.d., centrées, de variance 1, donc d'après le théorème central limite, $\frac{|S_n|}{\sqrt{n}}$ converge en loi vers $X \sim \mathcal{N}(0,1)$.
Comme $\mathbb{P}(\frac{|S_n|}{\sqrt{n}} > t) = \mathbb{P}(\frac{|S_n|}{\sqrt{n}} \in]-t,t[)$ et que $\mathcal{N}(0,1)$ ne charge ni $-t$, ni t, $\mathbb{P}(\frac{|S_n|}{\sqrt{n}} > t)$ converge vers $\mathbb{P}(X \in]-t,t[) = \mathbb{P}(|X|>t)$. D'autre part, l'inégalité de Chebychev entraîne que
$$\mathbb{P}\left(\frac{|S_n|}{\sqrt{n}} > t\right) = \mathbb{P}(|S_n - \mathbb{E}(S_n)| > \sqrt{n}t) \le \frac{\operatorname{Var} S_n}{nt^2} = \frac{1}{t^2},$$
d'où $\mathbb{P}(\frac{|S_n|}{\sqrt{n}} > t) \le \min(1, \frac{1}{t^2})$. Comme cette fonction est intégrable sur $\mathbb{R}_+$, on en déduit avec le théorème de convergence dominée que $\frac{\mathbb{E}[|S_n|]}{\sqrt{n}}$ converge vers
$$\begin{aligned}\int_0^{+\infty} \mathbb{P}(|X|>t)\, dt = \mathbb{E}(|X|) &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} |x| e^{-x^2/2}\, d\lambda(x)\\ &= \frac{2}{\sqrt{2\pi}}[-e^{-x^2/2}]_0^{+\infty} = \sqrt{2/\pi},\end{aligned}$$
ce qui donne le résultat avec $A = \sqrt{2/\pi}$.

Solution 258 1. (a) Si $a_1,\dots,a_n$ sont des entiers naturels non nuls, on a pour toute variable Y à valeurs dans $\mathbb{N}^*$.
$$\{\nu_{p_1}(Y) \ge a_1, \dots, \nu_{p_N}(Y) \ge a_N\} = \{\prod_{k=1}^N p_k^{a_k} | Y\},$$
donc la convergence des probabilités $\mathbb{P}(m|X_n)$ vers $\mathbb{P}(m|X)$ entraîne celle des quantités du type
$$\mathbb{P}(\nu_{p_1}(X_n) \ge a_1, \dots, \nu_{p_N}(X_n) \ge a_N)$$

vers $\mathbb{P}(\nu_{p_1}(X) \geq a_1, \ldots, \nu_{p_N}(X) \geq a_N)$. Cela suffit à entraîner la convergence en loi. En effet, comme

$$\begin{aligned}
&\mathbb{1}_{\{Y^1=a_1,\ldots,Y^N=a_N\}} = \prod_{k=1}^N (\mathbb{1}_{\{Y^k \geq a_k\}} - \mathbb{1}_{\{Y^k \geq a_k+1\}}) \\
&= \prod_{k=1}^N \sum_{j=0}^1 (-1)^j \mathbb{1}_{\{Y^k \geq a_k+j\}} \\
&= \sum_{(b_1,\ldots,b_N)\in\{0,1\}^N} \prod_{k=1}^N (-1)^{b_1+\cdots+b_N} \mathbb{1}_{\{\forall k\in\{1,\ldots,n\}, Y^k \geq a_k+b_k\}}
\end{aligned}$$

on a

$$\begin{aligned}
&\mathbb{P}(Y^1 = a_1, \ldots, Y^N = a_N) \\
&= \sum_{(b_1,\ldots,b_N)\in\{0,1\}^N} \prod_{k=1}^N (-1)^{b_1+\cdots+b_N} \mathbb{P}(\forall k \in \{1, \ldots, n\}, Y^k \geq a_k + b_k),
\end{aligned}$$

ce qui permet de déduire la convergence en loi de celle des fonctions de queues jointes.
Comme $\psi_N = g \circ (\nu_{p_1}, \ldots, \nu_{p_N})$ avec $g(a_1, \ldots, a_N) = \prod_{k=1}^N p_k^{a_k}$, et que g est continue, on en déduit la convergence en loi de

$$\psi_N(X_n) = g(\nu_{p_1}(X_n), \ldots, \nu_{p_N}(X_n)).$$

(b) La famille $(X, (X_n)_{n\geq 1})$ étant tendue, il existe une constante A telle que $\mathbb{P}(X > A) \leq \varepsilon/3$ et $\mathbb{P}(X_n > A) \leq \varepsilon/3$ pour tout n. On prend alors N tel que $p_N \geq A$. On a les inclusions

$$\begin{aligned}
\{Y = i\}\Delta\{\psi_N(Y) = i\} &\subset \{Y \neq \psi_N(Y)\} = \{\exists k > N; \nu_{p_k}(Y) > 0\} \\
&\subset \{Y > p_N\},
\end{aligned}$$

$$\begin{aligned}
\text{d'où} \quad |\mathbb{P}(Y = i) - \mathbb{P}(\psi_N(Y) = i)| &\leq \mathbb{P}(\{Y = i\}\Delta\{\psi_N(Y) = i\}) \\
&\leq \mathbb{P}(Y > p_N).
\end{aligned}$$

(c) On en déduit avec l'inégalité triangulaire

$$\begin{aligned}
|\mathbb{P}(X_n = i) - \mathbb{P}(X = i)| &\leq |\mathbb{P}(\psi_N(X_n) = i) - \mathbb{P}(\psi_N(X) = i)| \\
&\quad + \mathbb{P}(X_n > p_N) + \mathbb{P}(X > p_N) \\
&\leq |\mathbb{P}(\psi_N(X_n) = i) - \mathbb{P}(\psi_N(X) = i)| + \frac{2\varepsilon}{3}.
\end{aligned}$$

Comme $\psi_N(X_n)$ converge en loi vers $\psi_N(X)$, il existe n_0 tel que $|\mathbb{P}(\psi_N(X_n) = i) - \mathbb{P}(\psi_N(X) = i)| \leq \varepsilon/3$ pour $n \geq n_0$, ce qui donne $|\mathbb{P}(X_n = i) - \mathbb{P}(X = i)| \leq \varepsilon$ pour $n \geq n_0$: la convergence est démontrée.

2. On a
$$\mathbb{P}(N|Z_n) = \mathbb{P}(N|X_n, N|Y_n) = \mathbb{P}(N|X_n)^2 = \left(\frac{\lfloor n/N \rfloor}{n}\right)^2$$
et
$$\mathbb{P}(N|W_n) = \mathbb{P}(N^2|X_n) = \frac{\lfloor n/N^2 \rfloor}{n}.$$
Quand n tend vers l'infini, les deux quantités convergent évidemment vers $\frac{1}{N^2}$, qui est la probabilité qu'une variable suivant la loi Zêta de paramètre 2 soit divisible par N.
Reste à montrer la tension. Quels que soient les entiers naturels n et a, on a
$$\mathbb{P}(Z_n \geq a) = \sum_{i=a}^{+\infty} \mathbb{P}(Z_n = i) \leq \sum_{i=a}^{+\infty} \mathbb{P}(i|Z_n) \leq \sum_{i=a}^{+\infty} \frac{1}{i^2},$$
et de même $\mathbb{P}(W_n \geq a) \leq \sum_{i=a}^{+\infty} \frac{1}{i^2}$. Soit alors $\varepsilon > 0$. Si je prends a tel que $\sum_{i=a}^{+\infty} \frac{1}{i^2} < \varepsilon$, l'ensemble $F = \{1, \dots, a\}$ vérifie $\mathbb{P}(Z_n \in F) \geq 1 - \varepsilon$ et $\mathbb{P}(W_n \in F) \geq 1 - \varepsilon$, ce qui montre que l'hypothèse de tension est bien satisfaite et on a la convergence en loi voulue. On a donc en particulier $\lim_{n \to +\infty} \mathbb{P}(Z_n = 1) = \lim_{n \to +\infty} \mathbb{P}(W_n = 1) = \dfrac{1}{\zeta(2)} = \dfrac{6}{\pi^2}$: la proportion des couples premiers entre eux parmi les couples de $\{1, \dots, n\}^2$ et la proportion des entiers sans facteur carré entre 1 et n tendent vers $\frac{6}{\pi^2}$ quand n tend vers l'infini.

D.12 Exercices sur les vecteurs gaussiens

Solution 264 Posons $Y = r\frac{X}{\|X\|}$. Comme X est un vecteur gaussien centré, de matrice de covariance Id_n, OX suit la loi gaussienne $\mathcal{N}(0, O\mathrm{Id}_n O^*)$. Comme $OO^* = \mathrm{Id}_n$, X et OX ont même loi. On en déduit que $r\frac{X}{\|X\|_2}$ et $r\frac{OX}{\|OX\|_2}$ ont même loi. Cependant, $r\frac{OX}{\|OX\|_2} = r\frac{OX}{\|X\|_2} = OY$, donc Y et OY ont même loi, ce qui est le résultat voulu.

Solution 265 Soient $(X_i)_{i \geq 1}$ des variables indépendantes identiquement distribuées de loi $\mathcal{N}(0,1)$. Notons $X^n = (X_1, \dots, X_n)$. C'est un vecteur gaussien. D'après l'exercice 264, μ_n est la loi de la variable aléatoire
$$Y^n = \sqrt{n}\frac{X^n}{||X^n||}.$$
D'où l'on déduit que γ_n est la loi de $Y_1^n = \sqrt{n}\frac{X_1}{||X^n||}$. Par application de la loi (forte ou faible) des grands nombres, $\frac{||X^n||^2}{n}$ converge en loi vers $\mathbb{E}(X_1^2) = 1$,

par composition $\frac{\sqrt{n}}{||X^n||}$ converge aussi en loi vers 1. Avec le lemme de Slutsky, Y_1^n converge en loi vers la loi de X_1, soit $\mathcal{N}(0,1)$.

Solution 266 On suit l'indication :

$$XY = \frac{1}{4}\left((X+Y)^2 - (X-Y)^2)\right) = \frac{1}{2}(U^2 - V^2),$$

où l'on a posé

$$\begin{pmatrix} U \\ V \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

La matrice M considérée est une matrice orthogonale (une matrice de rotation), donc la loi $\mathcal{N}(0,\ \mathrm{Id}_2)$ est laissée invariante. En effet, (U,V) est un vecteur gaussien centré, de matrice de covariance $M\ \mathrm{Id}_2 M^* =\ \mathrm{Id}_2$. La loi de XY est donc la loi image de $\mathcal{N}(0,\ \mathrm{Id}_2)$ par l'application qui à (u,v) associe $\frac{1}{2}(u^2-v^2)$. Mais cette loi est aussi la loi de $\frac{1}{2}(X^2-Y^2)$, puisque (X,Y) suit la loi $\mathcal{N}(0,\ \mathrm{Id}_2)$. Ainsi, XY et $\frac{1}{2}(X^2-Y^2)$ ont même loi.

Solution 267

1. $(a) \Longrightarrow (b)$. Si f est croissante et $X \geq Y$, alors $f(X) \geq f(Y)$, d'où $\mathbb{E}[f(X)] \geq \mathbb{E}[f(Y)]$, soit $\int f\ d\mu \geq \int f\ d\nu$.
   $(b) \Longrightarrow (c)$ Il suffit de considérer la fonction croissante $f = \mathbb{1}_{]t,+\infty[}$.
   $(c) \Longrightarrow (a)$. Comme dans le théorème 5.36, on pose pour toute loi γ

   $$\forall u \in]0,1[\quad Q^*_\gamma(u) = \min\{x \in \mathbb{R} :\ \gamma(]t,+\infty]) \leq u\}.$$

   Pour tout t, on a $\mu(]t,+\infty[) \geq \nu(]t,+\infty[)$, donc $\mu(]t,+\infty[) \leq u$ entraîne $\nu(]t,+\infty[) \leq u$. Ainsi

   $$\{x \in \mathbb{R} :\ \mu(]t,+\infty]) \leq u\} \subset \{x \in \mathbb{R} :\ \nu(]t,+\infty]) \leq u\},$$

   d'où $Q^*_\mu(u) \geq Q^*_\nu(u)$. Soit U suivant la loi uniforme sur $[0,1]$: d'après le théorème 5.36, $Q^*_\mu(U)$ suit la loi μ et $Q^*_\nu(U)$ suit la loi ν. Comme $Q^*_\mu(U) \geq Q^*_\nu(U)$, on a le couplage voulu.

2. Soit f croissante mesurable positive. On a

   $$\begin{aligned} \int_{\mathbb{R}} f\ d(\mu_1 * \nu_1) &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x+y)\ d\mu_1(x)\right)\ d\nu_1(y) \\ &\geq \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x+y)\ d\mu_2(x)\right)\ d\nu_1(y) = \int_{\mathbb{R}} f\ d(\mu_2 * \nu_1). \end{aligned}$$

   En effet, pour tout y, la fonction $x \mapsto f(x+y)$ est croissante, donc $\int_{\mathbb{R}} f(x+y)\ d\mu_1(x) \geq \int_{\mathbb{R}} f(x+y)\ d\mu_2(x)$ puisque $\mu_1 \succeq \mu_2$.
   Ainsi $\mu_1 * \nu_1 \succeq \mu_2 * \nu_1$. On montre de même que $\mu_2 * \nu_1 \succeq \mu_2 * \nu_2$, d'où le résultat.

3. Soient $U_1, \dots, U_n$ des variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On a l'inégalité

$\sum_{k=1}^{m} \mathbb{1}_{\{U_k \leq q\}} \leq \sum_{k=1}^{m} \mathbb{1}_{\{U_k \leq p\}} \leq \sum_{k=1}^{n} \mathbb{1}_{\{U_k \leq p\}}$.

Les trois membres suivent respectivement la loi $\mathcal{B}(m,q)$, $\mathcal{B}(m,p)$ et $\mathcal{B}(n,p)$. Cela donne les couplages voulus.

De même si $X_1, \dots, X_n$ sont des variables aléatoires indépendantes suivant la loi gaussienne $\mathcal{N}(0,1)$, l'inégalité $\sum_{k=1}^{n} X_k^2 \geq \sum_{k=1}^{m} X_k^2$ entraîne que $\chi^2(n) \succeq \chi^2(m)$.

4. (a) La loi de $\sum_{i=1}^{n} (X_i + a_i)^2$ est la loi de $\|Y_a\|_2^2$ où Y_a est un vecteur gaussien $\mathcal{N}(a, \mathrm{Id}_n)$, avec $a = (a_1, \dots, a_n)$. Si $\|a\|_2 = \|b\|_2$, alors il existe une transformation orthogonale O qui envoie a sur b (par exemple la rotation envoyant a sur b et laissant l'orthogonal du plan contenant a et b invariant).
La loi de OY_a est $\mathcal{N}(Oa, O\mathrm{Id}_n O^*) = \mathcal{N}(b, \mathrm{Id}_n)$. Ainsi OY_a et Y_b ont même loi, donc $\|OY_a\|_2^2$ et $\|Y_b\|_2^2$ ont même loi.
Comme $\|OY_a\|_2^2 = \|Y_a\|_2^2$, le résultat s'ensuit.

(b) La question précédente entraîne que $\chi^2(n, \lambda) = \chi^2(1, \lambda/n)^{*n}$. La convolution préservant l'ordre stochastique, il suffit de montrer que $\chi^2(1, \lambda/n) \succeq \chi^2(1, \lambda'/n)$. Ainsi on est ramené à étudier le cas $n = 1$. On va utiliser la caractérisation (c) de l'ordre stochastique. On doit montrer que pour tout t, la fonction $\lambda \mapsto \chi^2(1, \lambda)(]t, +\infty))$ est décroissante en λ. Il est inutile d'étudier le cas $t \leq 0$, puisqu'alors la fonction est constante égale à 1. On posera donc $t = b^2$. Si $X \sim \mathcal{N}(0,1)$, $(X+a)^2$ suit la loi $\chi^2(1, a^2)$. Il suffit donc, à $b > 0$ fixé, d'étudier le sens de variation de $g_b(a) = \mathbb{P}((X+a)^2 \geq b^2)$ en fonction de a. Or

$$\begin{aligned} g_b(a) &= \mathbb{P}(X + a \geq b) + \mathbb{P}(X + a \leq -b) \\ &= \mathbb{P}(X \geq b - a) + \mathbb{P}(-X \geq b + a) \\ &= \mathbb{P}(X \geq b - a) + \mathbb{P}(X \geq b + a). \end{aligned}$$

La dérivée par rapport à a est $f(b-a) - f(b+a)$, où f est la densité de la gaussienne. Comme la densité de la gaussienne est paire et décroissante sur $\mathbb{R}_+$, cette différence a le même signe que la quantité $(a+b)^2 - (a-b)^2 = 4ab \geq 0$, ce qui donne le résultat voulu.

Solution 268 Réalisons M à l'aide de $X \sim \mathcal{N}(0, I_m)$. Posons $Q = \|P_F(M)\|_2^2$.

On a

$$Q = \|P_F(M)\|_2^2 = \|P_F(\frac{X}{\|X\|_2})\|^2 = \frac{\|P_F(X)\|_2^2}{\|X\|_2}$$
$$= \frac{\|P_F(X)\|_2^2}{\|P_F(X)\|_2^2 + \|P_{F^\perp}(X)\|_2^2}.$$

D'après le théorème de Cochran, $A = \|P_F(X)\|_2^2$ et $B = \|P_{F^\perp}(X)\|_2^2$ sont indépendantes, avec $A \sim \chi^2(\frac{n}{2}) = \Gamma(\frac{n}{2}, \frac{1}{2})$ et $B \sim \chi^2(\frac{m-n}{2}) = \Gamma(\frac{m-n}{2}, \frac{1}{2})$. On peut encore écrire

$$Q = \frac{A}{B} = \frac{A/2}{A/2 + B/2}.$$

$A/2$ et $B/2$ sont indépendantes, avec $A/2 \sim \Gamma(\frac{n}{2}, 1)$ et $B/2 \sim \Gamma(\frac{m-n}{2}, 1)$. Le résultat de l'exercice sur les lois de Dirichlet de la section 6.12.1 donne alors le résultat voulu.

D.13 Exercices sur les statistiques

Solution 280
1. Au niveau de confiance 95%, on obtient l'intervalle de confiance $[1/10 - \sqrt{1/400},\ 1/10 + \sqrt{1/400}] = [0,05;\ 0,15]$.
2. L'effectif doit être tel que $\sqrt{1/N} \leq 0,005$. D'où $N \geq 40000$.
3. On sait que $\sigma\sqrt{\frac{\theta(1-\theta)}{N}} \leq \frac{1}{2}\sigma\sqrt{1/N}$. Pour obtenir le même intervalle qu'à la question précédente, il faut choisir $0,5\sigma\sqrt{1/N} \leq 0,005$. Ainsi pour $N = 400$, on a $\sigma \leq 0,01$. Le niveau de confiance est alors $\alpha \approx \mathbb{P}(G \in [-0,01;\ 0,01]) \leq 0,17$ où G une variable gaussienne $\mathcal{N}(0,1)$.

Solution 281 On remarque tout d'abord que comme $g \in L^\infty([0,1]^d)$, la variable $g(X_1)$ est en particulier de carré intégrable. Pour simplifier les notations, posons pour tout $n \geq 1$, $u_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i) - \int_{[0,1]^d} g(x)\,dx$. Soient $\varepsilon > 0$ et $\alpha > 0$. On a par l'inégalité de Markov que

$$\mathbb{P}(u_n > \varepsilon) = \mathbb{P}(e^{\alpha u_n} > e^{\alpha\varepsilon}) \leq \frac{\mathbb{E}[e^{\alpha u_n}]}{e^{\alpha\varepsilon}}.$$

Calculons donc cette espérance. Par indépendance des variables X_i, on a

$$\begin{aligned}
\mathbb{E}[e^{\alpha u_n}] &= \mathbb{E}\left[\exp\left(\frac{\alpha}{n}\left(\sum_{i=1}^{n} g(X_i) - \int_{[0,1]^d} g(x)\,dx\right)\right)\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{n} \exp\left(\frac{\alpha}{n}\left(g(X_i) - \int_{[0,1]^d} g(x)\,dx\right)\right)\right] \\
&= \prod_{i=1}^{n} \mathbb{E}\left[\exp\left(\frac{\alpha}{n} g(X_i) - \int_{[0,1]^d} g(x)\,dx\right)\right] \\
&= \left(\mathbb{E}\left[\exp\left(\frac{\alpha}{n}\left(g(X_1) - \int_{[0,1]^d} g(x)\,dx\right)\right)\right]\right)^n .
\end{aligned}$$

Une rapide étude de la fonction $t \mapsto e^{-t}(1+t+t^2)$ montre que pour tout $|t| \leq 1$, on a $e^t \leq 1+t+t^2$. De plus, pour presque tout $x \in [0,1]^d$, l'inégalité triangulaire implique

$$|g(x) - \int_{[0,1]^d} g(x)\,dx| \leq |g(x)| + \left|\int_{[0,1]^d} g(x)\,dx\right| \leq 2\|g\|_\infty .$$

On applique la majoration précédente de l'exponentielle avec $\alpha \leq \frac{n}{2\|g\|_\infty}$:

$$\begin{aligned}
\mathbb{E}[e^{\alpha u_n}] &\leq \left(\mathbb{E}[1 + \frac{\alpha}{n}(X_1 - \int_{[0,1]^d} g(x)\,dx) + \frac{\alpha^2}{n^2}(g(X_1) - \int_{[0,1]^d} g(x)\,dx)^2]\right)^n \\
&\leq \left(1 + \frac{\alpha^2}{n^2}\operatorname{Var}(g(X_1))\right)^n \leq \left(1 + \frac{\alpha^2}{n^2}\|g\|_2^2\right)^n \leq \exp\left\{\frac{\alpha^2}{n}\|f\|_2^2\right\}.
\end{aligned}$$

Cette dernière inégalité est due au fait que pour tout $t \geq 0$ on a $1+t \leq e^t$. On en déduit alors que pour tout $\varepsilon > 0$ et tout $\alpha \leq \frac{n}{2\|f\|_\infty}$, on a

$$\mathbb{P}(u_n > \varepsilon) \leq \exp(-\alpha\varepsilon)\exp\left\{\frac{\alpha^2}{n}\|f\|_2^2\right\} = \exp\left\{\alpha\left(\frac{\alpha}{n}\|f\|_2^2 - \varepsilon\right)\right\}.$$

Il reste maintenant à optimiser en α pour conclure. On cherche le minimum (en α et sur l'intervalle $]0, \frac{n}{2\|f\|_\infty}[$) de la quantité $\exp\left\{\alpha\left(\frac{\alpha}{n}\|f\|_2^2 - \varepsilon\right)\right\}$, ce qui correspond au minimum de $\alpha \mapsto \alpha\left(\frac{\alpha}{n}\|f\|_2^2 - \varepsilon\right)$. On voit aisément que le minimum sur $\mathbb{R}$ est atteint en $\alpha_0 = \frac{n}{2\|f\|_2^2}\varepsilon$. On remarque enfin que

$$\alpha_0 \leq \frac{n}{2\|f\|_\infty} \iff \varepsilon \leq \frac{\|f\|_2^2}{\|f\|_\infty}.$$

On trouve donc que pour tout $0 < \varepsilon \leq \frac{\|f\|_2^2}{\|f\|_\infty}$, on a $\mathbb{P}(u_n > \varepsilon) \leq \exp(-\frac{n}{4\|f\|_2^2}\varepsilon^2)$. En appliquant cette inégalité à la fonction $-f$, on obtient enfin l'inégalité $\mathbb{P}(-u_n < -\varepsilon) \leq \exp(-\frac{n}{4\|f\|_2^2}\varepsilon^2)$, ce qui complète la preuve.

Solution 282 La fonction de vraisemblance vaut

$$L(k_1, \dots, k_n, \lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{k_i}}{k_i!} \qquad = C(k) e^{-n\lambda} \lambda^{k_1 + \cdots + k_n},$$

et la log-vraisemblance $\log C(k) - n\lambda + (k_1 + \cdots + k_n) \log(\lambda)$. Sa dérivée vaut $-n + \frac{k_1 + \cdots + k_n}{\lambda}$, donc la log-vraisemblance et la vraisemblance sont maximales pour $\lambda = \frac{k_1 + \cdots + k_n}{n}$. Ainsi, pour un échantillon $(X_1, \dots, X_n)$, l'estimateur du maximum de vraisemblance pour le modèle poissonnien est $\hat{\lambda}_n = \frac{X_1 + \cdots + X_n}{n}$. Il coïncide avec la moyenne empirique.

Solution 283 1. La moyenne est :

$$m = \sum_{i=1}^{200} \frac{n_i x_i}{N} = \frac{800}{200} = 4$$

et la variance :

$$V = \sum_{i=1}^{200} \frac{n_i x_i^2}{N} - m^2 = 3930/200 - 4^2 = 19,65 - 16 = 3,65.$$

2. m et V sont voisins donc la distribution est susceptible de suivre une distribution de Poisson. Cette première impression est renforcée par l'allure de l'histogramme des effectifs.

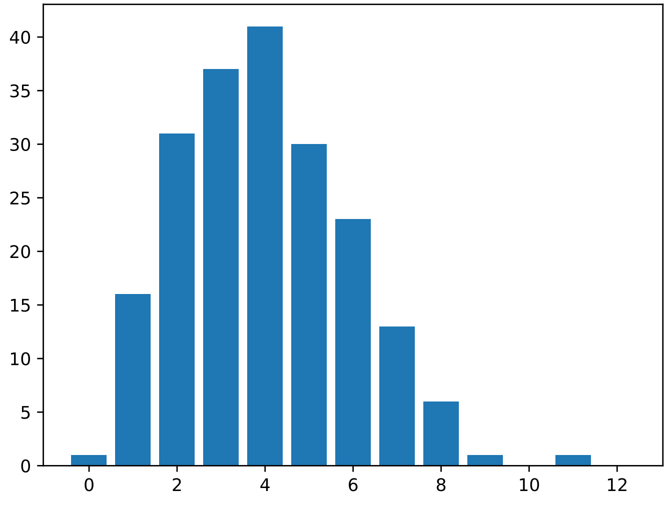


FIGURE D.1 – Histogramme des effectifs

Ci-après, le code Julia ayant permis de tracer la figure.

```
using PyPlot
c=[1,16,31,37,41,30,23,13,6,1,0,1,0]
valeurs=collect(0:12); dessin=bar(valeurs,c)
lambda=sum(c.*valeurs)/sum(c)
sdeux=sum(c.*(valeurs.^2))/sum(c)
```

Il est de plus standard de modéliser un tel problème par une loi de Poisson.

3. On fait un test de χ^2 d'adéquation à une distribution de Poisson. L'hypothèse à vérifier est

   H_0 : La distribution expérimentale suit une loi de Poisson . Le paramètre estimé est $\hat{\lambda} = 4$. Si $X \sim \mathcal{P}(\hat{\lambda})$, on a les valeurs théoriques $p_{\text{th}}(k) = \mathbb{P}(X = k) = \frac{m^k}{k!}e^{-m} = \frac{4^k}{k!}e^{-4}$ et $E_{\text{th}}(k) = 200p_k$. Afin de pouvoir faire un test du χ^2, il faut vérifier que les effectifs des classes sont tous supérieurs à 5. On fait donc un regroupement par classes extrêmes. Notons n le nombre d'appels, E_o l'effectif observé du nombre d'appels, p_{th} la probabilité théorique et $E_{\text{th}} = 200p_{\text{th}}$. On obtient alors le tableau suivant :

k	0 – 1	2	3	4	5	6	7	8 – 12
E_o	17	31	37	41	30	23	13	8
p_{th}	0,091	0,147	0,195	0,195	0,156	0,104	0,06	0,051
E_{th}	18,3	29,3	39,1	39,1	31,3	20,8	11,9	10,2

   On calcule alors la valeur de la statistique

$$\chi^2 = \sum_k \frac{(E_o(k) - E_{\text{th}}(k))^2}{E_{\text{th}}(k)}$$
$$= \frac{(17-18,3)^2}{18,3} + \frac{(31-29,3)^2}{29,3} + \frac{(37-39,1)^2}{39,1} + \ldots + \frac{(8-10,2)^2}{10,2}$$
$$= 1,26.$$

   Il faut comparer cette valeur avec la valeur théorique correspondant au risque 5% et au bon nombre de degrés de liberté. Ici, le paramètre de la loi de Poisson a été estimé à l'aide de $\hat{\lambda}$ qui est l'estimateur du maximum de vraisemblance. Il nous faut donc réduire le nombre de degrés de liberté en conséquence. Le nombre de degrés de liberté est ici $\nu = q - 1 - r$, où le nombre de classes est $q = 8$ et $r = 1$ (car la loi de Poisson a un seul paramètre). Ainsi, $\nu = 6$. Au risque 5%, la lecture de la table du χ^2_6 donne : $\chi^2_0 = 12,59$. Comme $\chi^2 \leq \chi^2_0$, au risque de 5% de se tromper, l'hypothèse de suivi d'une loi de Poisson n'est pas rejetée (elle est vraisemblable).

Ici encore, l'outil informatique peut être utile. On donne ci-après un code en Julia ; on suppose que lambda et c ont été définis comme plus haut.

```
  using Distributions
  function Calcul_chi_deux(obs,ref)
    th=ref./sum(ref)
    effectif=sum(obs)
    dobs=obs./effectif
    s=effectif*sum(((th-dobs).^2)./th)
    return(s)
   end

F=cdf.(Poisson(lambda),collect(1:7))
dth=[F;1]-[0;F]
b=[c[1]+c[2];c[3:8];sum(c[9:13])]
chi_deux=Calcul_chi_deux(b,dth)
risque=0.05
chi_deux_zero=cquantile(Chisq(length(b)-2),risque)
```

Solution 284 1. Il s'agit d'une comparaison d'un pourcentage observé avec un pourcentage théorique.

Les hypothèses à tester sont

$H_0 : \pi_H = 0,5$ où π_H est le pourcentage d'hommes dans la population de diabétiques
contre $H_1 : \pi_H \neq 0,5$ (test bilatéral).

	Hommes	Femmes	Total
Effectifs observés	510	490	1000
Effectifs calculés	500	500	1000

Les conditions d'application sont vérifiées : les effectifs calculés sont supérieurs à 5.

Le χ^2 calculé vaut donc

$$\chi^2 = \frac{(510-500)^2}{500} + \frac{(490-500)^2}{500} = 0,4.$$

La valeur absolue de cette valeur doit être comparée à la valeur seuil pour $\alpha = 5\%$ de la loi du χ^2 à 1 degré de liberté, c'est-à-dire 3,84. Comme $\chi^2 < 3,84$: on ne rejette pas l'hypothèse H_0 au risque 5%, on ne montre pas de différence significative entre les pourcentages d'hommes et de femmes dans la population de diabétiques étudiée.

2. Il s'agit d'une comparaison de deux pourcentages observés sur échantillons indépendants.

   Les hypothèses à tester sont

   $H_0 : \pi_I = \pi_{II}$ où π_I et π_{II} sont les pourcentages vrais de patients diabétiques de plus de 50 ans chez les hypertendus, respectivement les non hypertendus

   contre $H_1 : \pi_I \neq \pi_{II}$ (test bilatéral).

	Groupe I	Groupe II	Total
- de 50 ans	30 (45)	420 (405)	450
+ de 50 ans	70 (55)	480 (495)	550
Total	100	900	1000

   Les conditions d'application sont vérifiées : les effectifs calculés (entre parenthèses) sont supérieurs à 5. On calcule alors la valeur de la statistique

$$\chi^2 = \frac{(30-45)^2}{45} + \cdots + \frac{(480-495)^2}{495} = 10,1.$$

   On compare cette valeur à celle d'une loi de χ^2 à 1 degré de liberté, au risque de 5% : $\chi_0^2 = 3,84$.

   On trouve $\chi^2 \geq 3,84$: on rejette l'hypothèse H_0 au risque 5%, le pourcentage de patients de plus de 50 ans est significativement différent chez les hypertendus que chez les non hypertendus.

D.14 Exercices sur les sommes de variables aléatoires indépendantes

Solution 291 D'après le théorème des trois séries, les deux conditions sont nécessaires. Pour montrer qu'elles sont suffisantes, il suffit de montrer que la série de terme général $\mathrm{Var}(X_n \mathbb{1}_{\{X_n \leq c\}})$ converge. On pourra alors appliquer le théorème des trois séries, dans le sens de la réciproque cette fois. Or on a

$$\begin{aligned}\mathrm{Var}(X_n \mathbb{1}_{\{X_n \leq c\}}) &= \mathbb{E}[(X_n \mathbb{1}_{\{X_n \leq c\}})^2] - (\mathbb{E}[X_n \mathbb{1}_{\{X_n \leq c\}}])^2 \\ &\leq \mathbb{E}[(X_n \mathbb{1}_{\{X_n \leq c\}})^2].\end{aligned}$$

X_n étant positive, on a aussi $(X_n \mathbb{1}_{\{X_n \leq c\}})^2 = X_n^2 \mathbb{1}_{\{X_n \leq c\}} \leq cX_n \mathbb{1}_{\{X_n \leq c\}}$, d'où

$$\sum_{n=1}^{+\infty} \mathrm{Var}(X_n \mathbb{1}_{\{X_n \leq c\}}) \leq \sum_{n=1}^{+\infty} c\mathbb{E}[X_n \mathbb{1}_{\{X_n \leq c\}}] = c\sum_{n=1}^{+\infty} \mathbb{E}(X_n \mathbb{1}_{\{X_n \leq c\}}) < +\infty,$$

ce qui achève la preuve.

Solution 292 1. On procède par récurrence sur n. Pour $n = 1$, σ_1 est toujours l'application identité, ce qui montre la propriété pour $n = 1$. Soit $\gamma \in \mathfrak{S}_{n+1}$. On doit montrer que $\mathbb{P}(\sigma_{n+1} = \gamma) = \frac{1}{(n+1)!}$. On a

$$\begin{aligned}\sigma_{n+1} = \gamma &\iff \sigma_n \circ (n+1\ X_{n+1}) = \gamma \\ &\iff \sigma_n = \gamma \circ (n+1\ X_{n+1}) \quad \text{car } (n+1\ X_{n+1})^2 = \mathrm{Id} \\ &\iff X_{n+1} = \gamma^{-1}(n+1) \ \text{ et } \sigma_n = \gamma \circ (n+1\ \gamma^{-1}(n+1)).\end{aligned}$$

Comme σ_n est $\sigma(X_1, \dots, X_n)$-mesurable, on a alors, par hypothèse de récurrence :

$$\begin{aligned}\mathbb{P}(\sigma_{n+1} = \gamma) &= \mathbb{P}(X_{n+1} = \gamma^{-1}(n+1))\mathbb{P}(\sigma_n = \gamma \circ (n+1\ \gamma^{-1}(n+1))) \\ &= \frac{1}{n+1}\frac{1}{n!} = \frac{1}{(n+1)!}\end{aligned}$$

ce qui achève la démonstration.

2. On passe de σ_{n-1} à σ_n en insérant n dans une des orbites de σ_{n-1}. Or n ne peut être le plus petit élément de son orbite que s'il est seul dans son orbite, c'est-à-dire si $\sigma_n(n) = n$, c'est-à-dire si $X_n = n$. On a alors $Y_{n,n} = \mathbb{1}_{\{X_n = n\}}$. Cependant, insérer un élément plus grand que tous les éléments présents ne modifie pas le plus petit élément de chaque orbite. Ainsi, pour tout $n > k$, on garde $Y_{n,k} = Y_{k,k} = \mathbb{1}_{\{X_k = k\}}$. Donc, à n fixé, les variables aléatoires $Y_{n,1}, \dots, Y_{n,n}$ sont bien indépendantes.

3. Chaque orbite admet un seul plus petit élément, donc le nombre de plus petits éléments est bien le nombre d'orbites. Pour tout $z \in \mathbb{C}$, on a

$$z^{C_n} = \prod_{k=1}^{n} z^{Y_{n,k}},$$

et par indépendance

$$\begin{aligned}\mathbb{E}(z^{C_n}) &= \prod_{k=1}^{n} \mathbb{E}(z^{Y_{n,k}}) = \prod_{k=1}^{n} (\mathbb{P}(X_k \neq k) + \mathbb{P}(X_k = k)z) \\ &= \prod_{k=1}^{n} \frac{(k-1)+z}{k} = \frac{1}{n!}\prod_{k=1}^{n} ((k-1)+z).\end{aligned}$$

Notons $c'(n,k)$ le nombre d'éléments de $\mathfrak{S}_n$ ayant exactement k cycles. Comme σ_n suit la loi uniforme sur $\mathcal{S}_n$, on a $\mathbb{P}(C_n = k) = \frac{c'(n,k)}{n!}$. Il vient alors

$$\mathbb{E}(z^{C_n}) = \sum_{k=0}^{n} \mathbb{P}(C_n = k)z^k = \sum_{k=0}^{n} \frac{c'(n,k)}{n!} z^k.$$

En égalant les deux quantités, on obtient

$$\prod_{i=0}^{n-1}(z+i) = \sum_{k=0}^{n} c'(n,k)z^k.$$

Comme $\prod_{i=0}^{n-1}(z+i) = \sum_{k=0}^{n} c(n,k)z^k$, il suffit d'identifier les coefficients pour avoir le résultat voulu.

4. $Y_{n,k}$ est une variable de Bernoulli, d'espérance $\mathbb{E}[Y_{n,k}] = \frac{1}{k}$ et de variance $\mathrm{Var}(Y_{n,k}) = \frac{1}{k}(1-\frac{1}{k})$. Comme les variables sont indépendantes, on a

$$\mathbb{E}[C_n] = \sum_{k=1}^{n} \frac{1}{k} \quad \text{et} \quad \mathrm{Var}(C_n) = \sum_{k=1}^{n} \frac{1}{k}\left(1-\frac{1}{k}\right).$$

Or on a

$$\mathrm{Var}\, C_n = \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=1}^{n} \frac{1}{k^2} = \log(n) + O(1),$$

et de même $\mathbb{E}[C_n] = \log n + O(1)$. Les variables sont indépendantes et bornées : le théorème central limite de Lindeberg s'applique et on a

$$\frac{C_n - \mathbb{E}[C_n]}{\sqrt{C_n}} \Longrightarrow \mathcal{N}(0,1).$$

Comme $\frac{\sqrt{\mathrm{Var}\, C_n}}{\sqrt{\log n}}$ tend vers 1, on en déduit avec le lemme de Slutsky que $\frac{C_n - \mathbb{E}[C_n]}{\sqrt{\log n}} \Longrightarrow \mathcal{N}(0,1)$. De même, comme $\frac{\log n - \mathbb{E}[C_n]}{\sqrt{\log n}}$ tend vers 0, on a finalement

$$\frac{C_n - \log n}{\sqrt{\log n}} \Longrightarrow \mathcal{N}(0,1).$$

Solution 293 1. Si f est une fonction holomorphe définie sur un ouvert U, on sait que le développement en série entière de f au point $z \in U$: $h \mapsto \sum_{n=0}^{+\infty} \frac{f^n(z)}{n!}h^n$ converge sur le disque ouvert de rayon $d(z, U^c)$, et la somme coïncide avec $f(z+h)$. Le rayon de convergence de cette série est donné par la formule de Hadamard $\frac{1}{R(z)} = \overline{\lim_{n\to+\infty}} |\frac{f^n(z)}{n!}|^{1/n}$. D'après ce qui précède, $R(z) \geq d(z, U^c)$, mais si $R(z) > d(z, U^c)$, on peut prolonger f en une fonction holomorphe sur l'ouvert plus grand $U \cup B(z, R(z))$. On en déduit que $\bigcup_{\substack{r\in[0,1[\cap\mathbb{Q} \\ \theta\in[0,1]\cap\mathbb{Q}}} E(re^{i2\pi\theta}) \subset \{A_r \neq \varnothing\}$. Cependant, si u de norme 1 est un point régulier, f admet un prolongement $\tilde{f}$ holomorphe sur $B(0,1) \cup B(u, \varepsilon)$ pour un certain ε. Mais

$B(0,1) \cap B(u, \varepsilon/3)$ contient un point de la forme $re^{2i\pi\theta}$ avec r et θ rationnels de $[0,1[$. Comme $B(re^{2i\pi\theta}, 2\varepsilon/3) \subset B(u,\varepsilon)$, on a $R(re^{2i\pi\theta}) \geq 2\varepsilon/3 > \varepsilon/3 > 1 - |re^{2i\theta}|$, ce qui donne l'inclusion inverse.

2. La série se dérive terme à terme à l'intérieur du disque ouvert de convergence et on a

$$\frac{f^{(n)}(z)}{n!} = \sum_{k=0}^{+\infty} a_{n+k} e^{2i\pi X_{n+k}} \binom{n+k}{k} z^k.$$

Ainsi $\sup\limits_{n\geq N} |\frac{f^{(n)}(z)}{n}|^{1/n}$ est $\sigma(X_k, k \geq N)$-mesurable.

Les variables $\sup\limits_{n\geq N} |\frac{f^{(n)}(z)}{n}|^{1/n}$ pour $N \geq N_0$ sont $\sigma(X_k, k \geq N_0)$-mesurables, donc leur limite $\overline{\lim\limits_{n\to+\infty}} |\frac{f^{(n)}}{n}(z)|^{1/n}$ est $\sigma(X_k, k \geq N_0)$-mesurable. Comme N_0 est quelconque, $\overline{\lim\limits_{n\to+\infty}} |\frac{f^{(n)}}{n}|^{1/n}$ est mesurable par rapport à la tribu de queue, donc l'événement $\{R(z) > 1 - |z|\}$ a une probabilité 0 ou 1 d'après la loi du 0–1 de Kolmogorov. Il nous reste à montrer que $\mathbb{P}(E(r)) = \mathbb{P}(E(re^{2i\pi\theta}))$. On a

$$\frac{f^{(n)}(re^{2i\pi\theta})}{n!} = \sum_{k=0}^{+\infty} a_{n+k} e^{2i\pi(X_{n+k}+\theta k)} \binom{n+k}{k} r^k$$

et

$$\begin{aligned}\left|\frac{f^{(n)}(re^{i\theta})}{n!}\right| &= \left|e^{2i\pi n\theta}\frac{f^{(n)}(re^{i\theta})}{n!}\right| \\ &= \left|\sum_{k=0}^{+\infty} a_{n+k} e^{2i\pi(X_{n+k}+\theta(n+k))} \binom{n+k}{k} r^k\right|.\end{aligned}$$

Ainsi si l'on note Ψ l'application de $[0,1]^{\mathbb{N}}$ dans $\overline{\mathbb{R}}$ qui à $(x_n)_{n\geq 0}$ associe

$$\overline{\lim_{n\to+\infty}} \left|\sum_{k=0}^{+\infty} a_{n+k} e^{2i\pi x_{n+k}} \binom{n+k}{k} r^k\right|^{1/n},$$

on a

$$\begin{aligned} E(r) &= \left\{\psi((X_n)_{n\geq 0}) < \frac{1}{1-r}\right\} \\ \text{et} \quad E(re^{2i\pi\theta}) &= \left\{\psi((X_n + n\theta))_{n\geq 0}) < \frac{1}{1-r}\right\}. \end{aligned}$$

Posons $Y_n = \{X_n + n\theta\}$. On a $e^{2i\pi(X_n+n\theta)} = e^{2i\pi Y_n}$, donc

$$E(re^{2i\pi\theta}) = \left\{\psi((Y_n)_{n\geq 0}) < \frac{1}{1-r}\right\}.$$

Or d'après le résultat de l'exercice 129, Y_n suit, tout comme X_n, la loi uniforme sur $[0,1]$.
Ainsi les suites $(X_n)_{n\geq 0}$ et $(Y_n)_{n\geq 0}$ suivent chacune la loi $\mathcal{U}([0,1])^{\otimes\mathbb{N}}$ (voir par exemple le corollaire 5.45).
Par suite, les variables $\psi((X_n)_{n\geq 0}))$ et $\psi((Y_n)_{n\geq 0})$ ont même loi, d'où $\mathbb{P}(E(r)) = \mathbb{P}(E(re^{2i\pi\theta}))$, ce qui est le résultat voulu.

3. Raisonnons par l'absurde et supposons que $\mathbb{P}(A_r \neq \varnothing) > 0$. D'après la première question, il existe $r_0 \in [0,1[\cap\mathbb{Q}$ et $\theta_0 \in [0,1]\cap\mathbb{Q}$ tels que $\mathbb{P}(E(r_0e^{2i\pi\theta_0})) > 0$. Par construction, quel que soit θ, les points de $B(r_0e^{2i\pi\theta}, \psi((X_n+n\theta)_{n\geq 0})^{-1})$ sont des points réguliers. La variable $\psi((X_n+n\theta)_{n\geq 0})$ est mesurable par rapport à la tribu de queue des (X_n), qui est triviale d'après la loi 0–1 de Kolmogorov : elle est donc presque sûrement constante (voir par exemple l'exercice 111). L'argument de la question précédente nous dit que la valeur de cette variable ne dépend pas de θ. Soit ℓ cette valeur. On a donc $\ell < \frac{1}{1-r_0}$ et
$$\mathbb{P}\left(\bigcap_{\theta\in[0,1]\cap\mathbb{Q}} \{\psi((X_n+n\theta)_{n\geq 0}) = \ell\}\right) = 1.$$
Ainsi, avec probabilité 1, les points de $\bigcup_{\theta\in[0,1]\cap\mathbb{Q}} B(r_0e^{2i\pi\theta}, \ell^{-1})$ sont réguliers, en particulier tous les points du cercle unité sont réguliers, ce qui est impossible d'après le résultat d'analyse que nous avons rappelé.

Solution 294 On pose dans la suite
$$S_n = \sum_{k=1}^{N_n} \mathbb{1}_{A_{n,k}}, \quad m_n = \max_{k=1}(p_{n,k}, 1\leq k\leq N_n), \text{ et } s_n = \sum_{k=1}^{N_n} p_{n,k}.$$
L'idée est, comme dans la preuve du TCL de Lindeberg, d'approcher la somme de variables de Bernoulli S_n par une somme de variables indépendantes de même type que la loi limite.

La comparaison entre la somme de Bernoulli et la somme de Poisson se fait au niveau des fonctions caractéristiques. L'indépendance des variables $A_{n,k}$ donne aisément
$$\phi_{S_n}(t) = \prod_{k=1}^{N_n} \phi_{\mathbb{1}_{A_{n,k}}}(t) = \prod_{k=1}^{N_n}((1-p_{n,k}) + p_{n,k}e^{it}).$$
Mais si les $P_{n,k}$ sont des variables aléatoires indépendantes suivant les lois de

Poisson de paramètres respectifs $p_{n,k}$ et si on pose $S'_n = \sum_{k=1}^{N_n} P_{n,k}$, alors

$$\phi_{S'_n}(t) = \prod_{k=1}^{N_n} \phi_{P_{n,k}}(t) = \prod_{k=1}^{N_n} (\exp(p_{n,k}(e^{it}-1))) = \exp(s_n(e^{it}-1)).$$

On a vu, lors de la preuve du théorème central limite de Lindeberg, que le module de la différence de deux produits de nombres complexes dont les modules ne dépassent pas un, n'excède pas la somme des modules des différences. Or, une fonction caractéristique prend toujours ses valeurs dans le disque unité, donc

$$\begin{aligned}|\phi_{S_n}(t) - \phi_{S'_n}(t)| &= |\prod_{k=1}^{N_n} \phi_{\mathbb{1}_{A_{n,k}}}(t) - \prod_{k=1}^{N_n} \phi_{P_{n,k}}(t)| \\ &= \left| \prod_{k=1}^{N_n} \left((1-p_{n,k}) + p_{n,k}e^{it}\right) - \prod_{k=1}^{N_n} \left(\exp(p_{n,k}(e^{it}-1))\right) \right| \\ &\le \sum_{k=1}^{N_n} |g(p_{n,k}(e^{it}-1))|,\end{aligned}$$

avec $g(z) = e^z - 1 - z$. Mais

$$\begin{aligned}g(z) &= \sum_{k=2}^{+\infty} \frac{z^k}{k!} = \sum_{k=0}^{+\infty} \frac{z^{k+2}}{(k+2)!} \\ &= z^2 \sum_{k=0}^{+\infty} \frac{z^k}{k!} \frac{1}{(k+1)(k+2)}\end{aligned}$$

d'où

$$|g(z)| \le |z|^2 \sum_{k=0}^{+\infty} \frac{|z|^k}{k!} |\frac{1}{(k+1)(k+2)}| \le |z|^2 \frac{e^{|z|}}{2}.$$

Avec les inégalités $|p_{n,k}(e^{it}-1)| \le 2p_{n,k} \le 2$, on obtient

$$|\phi_{S_n}(t) - \phi_{S'_n}(t)| \le \sum_{k=1}^{N_n} (2p_{n,k})^2 \frac{e^2}{2} = 2e^2 \sum_{k=1}^{N_n} p_{n,k}^2 \le 2e^2 s_n m_n.$$

Comme s_n tend vers λ et que m_n tend vers zéro, $|\phi_{S_n}(t) - \phi_{S'_n}(t)|$ tend vers zéro. Maintenant,

$$\begin{aligned}|\phi_{S_n}(t) - \exp(\lambda(e^{it}-1))| &\le |\phi_{S_n}(t) - \phi_{S'_n}(t)| + |\phi_{S'_n}(t) - \exp(\lambda(e^{it}-1))| \\ &\le |\phi_{S_n}(t) - \phi_{S'_n}(t)| \\ &\quad + |\exp(s_n(e^{it}-1)) - \exp(\lambda(e^{it}-1))|,\end{aligned}$$

qui tend bien vers 0 puisque s_n tend vers λ.

On conclut avec le théorème de Lévy.

Autre méthode : On utilise l'inégalité de Le Cam démontrée dans l'exercice 175 par une méthode de couplage. Soit k un entier naturel quelconque. On a

$$|\mathbb{P}(S_n = k) - \mathcal{P}(\lambda)(k)| \leq |\mathbb{P}(S_n = k) - \mathcal{P}(s_n)(k)| + |\mathcal{P}(s_n)(k) - \mathcal{P}(\lambda)(k)|.$$

La convergence de $\mathcal{P}(s_n)(k)$ vers $\mathcal{P}(\lambda)(k)$ découle immédiatement de la convergence de s_n vers λ. Pour le premier terme, l'inégalité de Le Cam donne

$$|\mathbb{P}(S_n = k) - \mathcal{P}(s_n)(k)| \leq \sum_{k=1}^{N_n} p_{n,k}^2 \leq m_n s_n,$$

qui converge vers 0, comme vu précédemment. Cette convergence pour tout k fixé suffit à entraîner la convergence en loi (voir le corollaire 11.5).

Solution 295

1. On peut écrire
$$1 + z = e^z((1+z)e^{-z}) = e^z(1 + \psi(z))$$
avec
$$\psi(z) = e^{-z}(1+z) - 1 = z^2 \sum_{k=0}^{+\infty} (-1)^{k+1} z^k \left(\frac{1}{(k+1)!} - \frac{1}{(k+2)!} \right).$$
Notons que pour $|z| \leq 1$, on a $|\psi(z)| \leq |z|^2$. Ainsi
$$\prod_{n=1}^{N} (1 + Z_n) = \prod_{n=1}^{N} (1 + \psi(Z_n)) \exp\left(\sum_{n=1}^{N} Z_n \right).$$
Comme $|Z_n| \leq c_n < 1$, on a $|\psi(Z_n)| \leq |Z_n|^2 \leq c_n^2$ et la convergence de la série des (c_n^2) entraîne celle du produit $\prod_{n=1}^{+\infty} (1 + \psi(Z_n))$. Les (Z_n) sont des variables aléatoires indépendantes centrées dont la somme des moments d'ordre deux converge car $\mathbb{E}|Z_n|^2 \leq c_n^2$. Donc la série des Z_n converge presque sûrement. Ainsi $\exp\left\{ \sum_{n=1}^{N} X_n \right\}$ admet une limite non nulle quand N tend vers l'infini, ce qui donne le résultat voulu.
2. Si on pose $c_p = \frac{1}{p^s}$, et $Z_p = -\frac{X_p}{p^s}$, la convergence de $\sum_{p \in \mathcal{P}} \frac{1}{p^{2s}}$ entraîne celle du produit $\prod_{p \in \mathcal{P}} (1 - \frac{X_p}{p^s})$ et par suite celle du produit
$$\prod_{p \in \mathcal{P}} (1 - \frac{X_p}{p^s})^{-1}.$$

Il y avait 147 exercices corrigés.

Annexe E

Tables

Fonction de répartition de la loi normale centrée réduite

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.719	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.881	0.883
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936

ν \ p	0.005	0.01	0.025	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.975	0.99	0.995
1	0.000	0.000	0.001	0.004	0.016	0.064	0.148	0.275	0.455	0.708	1.074	1.642	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	0.446	0.713	1.022	1.386	1.833	2.408	3.219	4.605	5.991	7.378	9.21	10.6
3	0.072	0.115	0.216	0.352	0.584	1.005	1.424	1.869	2.366	2.946	3.665	4.642	6.251	7.815	9.348	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	1.649	2.195	2.753	3.357	4.045	4.878	5.989	7.779	9.488	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	1.61	2.343	3	3.655	4.351	5.132	6.064	7.289	9.236	11.07	12.83	15.09	16.75
6	0.676	0.872	1.237	1.635	2.204	3.07	3.828	4.57	5.348	6.211	7.231	8.558	10.64	12.59	14.45	16.81	18.55
7	0.989	1.239	1.69	2.167	2.833	3.822	4.671	5.493	6.346	7.283	8.383	9.803	12.02	14.07	16.01	18.48	20.28
8	1.344	1.646	2.18	2.733	3.49	4.594	5.527	6.423	7.344	8.351	9.524	11.03	13.36	15.51	17.53	20.09	21.95
9	1.735	2.088	2.7	3.325	4.168	5.38	6.393	7.357	8.34 3	9.414	10.66	12.24	14.68	16.92	19.02	21.67	23.59
10	2.156	2.558	3.247	3.94	4.865	6.179	7.267	8.295	9.342	10.47	11.78	13.44	15.99	18.31	20.48	23.21	25.19
11	2.603	3.053	3.816	4.575	5.578	6.989	8.148	9.237	10.34	11.53	12.9	14.63	17.28	19.68	21.92	24.72	26.76
12	3.074	3.571	4.404	5.226	6.304	7.807	9.034	10.18	11.34	12.58	14.01	15.81	18.55	21.03	23.34	26.22	28.3
13	3.565	4.107	5.009	5.892	7.042	8.634	9.926	11.13	12.34	13.64	15.12	16.98	19.81	22.36	24.74	27.69	29.82
14	4.075	4.66	5.629	6.571	7.79	9.467	10.82	12.08	13.34	14.69	16.22	18.15	21.06	23.68	26.12	29.14	31.32
15	4.601	5.229	6.262	7.261	8.547	10.31	11.72	13.03	14.34	15.73	17.32	19.31	22.31	25	27.49	30.58	32.8
16	5.142	5.812	6.908	7.962	9.312	11.15	12.62	13.98	15.34	16.78	18.42	20.47	23.54	26.3	28.85	32	34.27
17	5.697	6.408	7.564	8.672	10.09	12	13.53	14.94	16.34	17.82	19.51	21.61	24.77	27.59	30.19	33.41	35.72
18	6.265	7.015	8.231	9.39	10.86	12.86	14.44	15.89	17.34	18.87	20.6	22.76	25.99	28.87	31.53	34.81	37.16
19	6.844	7.633	8.907	10.12	11.65	13.72	15.35	16.85	18.34	19.91	21.69	23.9	27.2	30.14	32.85	36.19	38.58
20	7.434	8.26	9.591	10.85	12.44	14.58	16.27	17.81	19.34	20.95	22.77	25.04	28.41	31.41	34.17	37.57	40
21	8.034	8.897	10.28	11.59	13.24	15.44	17.18	18.77	20.34	21.99	23.86	26.17	29.62	32.67	35.48	38.93	41.4
22	8.643	9.542	10.98	12.34	14.04	16.31	18.1	19.73	21.34	23.03	24.94	27.3	30.81	33.92	36.78	40.29	42.8
23	9.26	10.2	11.69	13.09	14.85	17.19	19.02	20.69	22.34	24.07	26.02	28.43	32.01	35.17	38.08	41.64	44.18
24	9.886	10.86	12.4	13.85	15.66	18.06	19.94	21.65	23.34	25.11	27.1	29.55	33.2	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	18.94	20.87	22.62	24.34	26.14	28.17	30.68	34.38	37.65	40.65	44.31	46.93
26	11.16	12.2	13.84	15.38	17.29	19.82	21.79	23.58	25.34	27.18	29.25	31.79	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	20.7	22.72	24.54	26.34	28.21	30.32	32.91	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	21.59	23.65	25.51	27.34	29.25	31.39	34.03	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	22.48	24.58	26.48	28.34	30.28	32.46	35.14	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.6	23.36	25.51	27.44	29.34	31.32	33.53	36.25	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	32.34	34.87	37.13	39.34	41.62	44.16	47.27	51.81	55.76	59.34	63.69	66.77
60	35.53	37.48	40.48	43.19	46.46	50.64	53.81	56.62	59.33	62.13	65.23	68.97	74.4	79.08	83.3	88.38	91.95
80	51.17	53.54	57.15	60.39	64.28	69.21	72.92	76.19	79.33	82.57	86.12	90.41	96.58	101.9	106.6	112.3	116.3

Table du $\chi^2 : \chi^2_\nu([0,x]) = p$

Tables de Kolmogorov-Smirnov

Pour les petites valeurs de n :

n \ α	0.20	0.15	0.1	0.05	0.01		0.20	0.15	0.1	0.05	0.01
1	0.9	0.925	0.95	0.975	0.995	26	0.204	0.217	0.233	0.259	0.311
2	0.684	0.726	0.776	0.842	0.929	27	0.2	0.213	0.229	0.254	0.305
3	0.565	0.596	0.636	0.708	0.829	28	0.197	0.209	0.225	0.25	0.3
4	0.493	0.525	0.565	0.624	0.734	29	0.193	0.205	0.221	0.246	0.295
5	0.447	0.474	0.509	0.563	0.669	30	0.19	0.202	0.218	0.242	0.29
6	0.41	0.435	0.468	0.519	0.617	31	0.187	0.199	0.214	0.238	0.285
7	0.381	0.405	0.436	0.483	0.576	32	0.184	0.196	0.211	0.234	0.281
8	0.358	0.381	0.41	0.454	0.542	33	0.182	0.193	0.208	0.231	0.277
9	0.339	0.36	0.387	0.43	0.513	34	0.179	0.19	0.205	0.227	0.273
10	0.323	0.342	0.369	0.409	0.489	35	0.177	0.187	0.202	0.224	0.269
11	0.308	0.327	0.352	0.391	0.468	36	0.174	0.185	0.199	0.221	0.265
12	0.296	0.314	0.338	0.375	0.449	37	0.172	0.182	0.196	0.218	0.262
13	0.285	0.302	0.325	0.361	0.432	38	0.17	0.18	0.194	0.215	0.258
14	0.275	0.292	0.314	0.349	0.418	39	0.167	0.178	0.191	0.213	0.255
15	0.266	0.282	0.304	0.338	0.404	40	0.165	0.176	0.189	0.21	0.252
16	0.258	0.274	0.295	0.327	0.392	41	0.163	0.174	0.187	0.208	0.249
17	0.25	0.266	0.286	0.318	0.381	42	0.162	0.172	0.185	0.205	0.246
18	0.244	0.259	0.279	0.309	0.371	43	0.16	0.17	0.183	0.203	0.243
19	0.237	0.252	0.271	0.301	0.361	44	0.158	0.168	0.18	0.201	0.241
20	0.232	0.246	0.265	0.294	0.352	45	0.156	0.166	0.179	0.198	0.238
21	0.226	0.24	0.259	0.287	0.344	46	0.155	0.164	0.177	0.196	0.235
22	0.221	0.235	0.253	0.281	0.337	47	0.153	0.162	0.175	0.194	0.233
23	0.216	0.23	0.247	0.275	0.33	48	0.151	0.161	0.173	0.192	0.231
24	0.212	0.225	0.242	0.269	0.323	49	0.15	0.159	0.171	0.19	0.228
25	0.208	0.221	0.238	0.264	0.317	50	0.148	0.158	0.17	0.188	0.226

Pour les grandes valeurs de n :

n \ α	0.20	0.15	0.1	0.05	0.01
$n > 50$	$\frac{1.073}{\sqrt{n}}$	$\frac{1.138}{\sqrt{n}}$	$\frac{1.224}{\sqrt{n}}$	$\frac{1.358}{\sqrt{n}}$	$\frac{1.628}{\sqrt{n}}$

Bibliographie

[1] P. Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995.

[2] Z. W. Birnbaum and F. H. Tingey. One-sided confidence contours for probability distribution functions. *Ann. Math. Statistics*, 22 :592–596, 1951.

[3] O. Bordellès. *Thèmes d'arithmétique*. Ellipses, Paris, 2006.

[4] A. Bulinski and A. Shashkin. *Limit theorems for associated random fields and related systems*. Advanced Series on Statistical Science & Applied Probability, 10. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2007.

[5] Raphaël Cerf and Pierre Petit. A short proof of Cramér's theorem in $\mathbb{R}$. *Amer. Math. Monthly*, 118(10) :925–931, 2011.

[6] N. G. de Bruijn. *Asymptotic methods in analysis*. Dover Publications Inc., New York, third edition, 1981.

[7] Pedro Egydio de Oliveira Carvalho. On the distribution of the Kolmogorov-Smirnov D-statistic. *Ann. Math. Statist.*, 30 :173–176, 1959.

[8] E. F. Drion. Some distribution-free tests for the difference between two empirical cumulative distribution functions. *Ann. Math. Statistics*, 23 :563–574, 1952.

[9] M. J. Dubourdieu. Sur un théorème de M. S. Bernstein relatif à la transformation de Laplace-Stieltjes. *Compositio Math.*, 7 :96–111, 1939.

[10] J. Durbin. *Distribution theory for tests based on the sample distribution function*. Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1973.

[11] N. Etemadi. An elementary proof of the strong law of large numbers. *Z. Wahrsch. Verw. Gebiete*, 55(1) :119–122, 1981.

[12] N. Etemadi. On some classical results in probability theory. *Sankhyā Ser. A*, 47(2) :215–221, 1985.

[13] K. Falconer. *Techniques in fractal geometry*. John Wiley & Sons Ltd., Chichester, 1997.

[14] K. Falconer. *Fractal geometry*. John Wiley & Sons Inc., Hoboken, NJ, second edition, 2003.

[15] M. Fekete. Über die Verteilung der Würzeln bei gewissen algebraischen Gleichungen mit ganzzähligen Koeffizienten. *Math. Z.*, 17(1) :228–249, 1923.

[16] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.

[17] Olivier Garet. Les lois zêta pour l'arithmétique. *Quadrature*, (96) :10–18, 2015.

[18] Olivier Garet. *Probabilités et processus stochastiques*. auto-édité, distribué par Amazon, 2017.

[19] B. V. Gnedenko and V. S. Korolyuk. On the maximum discrepancy between two empirical distributions. *Doklady Akad. Nauk SSSR (N.S.)*, 80 :525–528, 1951.

[20] J. M. Hammersley. Generalization of the fundamental theorem on subadditive functions. *Proc. Cambridge Philos. Soc.*, 58 :235–238, 1962.

[21] J. M. Hammersley. Postulates for subadditive processes. *Ann. Probability*, 2 :652–680, 1974.

[22] Jean-Pierre Kahane. *Séries de Fourier absolument convergentes*. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 50. Springer-Verlag, Berlin-New York, 1970.

[23] J. F. C. Kingman. Subadditive processes. In *École d'Été de Probabilités de Saint-Flour, V–1975*, pages 167–223. Lecture Notes in Math., Vol. 539. Springer, Berlin, 1976.

[24] Simon Kochen and Charles Stone. A note on the Borel-Cantelli lemma. *Illinois J. Math.*, 8 :248–251, 1964.

[25] E. Lesigne. *Pile ou face*. Ellipses, Paris, 2001.

[26] P. Lévy. Sur quelques points de la théorie des probabilités dénombrables. *Ann. Inst. H. Poincaré*, 6(2) :153–184, 1936.

[27] P. Mattila. *Geometry of sets and measures in Euclidean spaces*, volume 44 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995.

[28] G. Pólya and G. Szegő. *Problems and theorems in analysis. I*. Classics in Mathematics. Springer-Verlag, Berlin, 1998.

[29] J.-P. Portmanteau. Un espoir pour l'ensemble vide. *Annales de l'Université de Felletin*, 1915.

[30] H. Queffélec. *Topologie*. Dunod, 2006.

[31] Hervé Queffélec and Claude Zuily. *Analyse pour l'agrégation*. Dunod, 2013.

[32] W. Rudin. *Analyse réelle et complexe*. Masson, Paris, 1980.

Index