Année universitaire 2019-2020

UNIVERSITÉ DE LORRAINE

Olivier GARET

Probabilités et Statistiques

Introduction

Le cours contenu dans le présent polycopié reproduit pour l'essentiel (dans les sept premiers chapitres) le contenu de divers enseignements de Licence que j'ai donnés à Orléans, puis à Nancy. Le chapitre 8 reproduit le cours de statistique donné par Aline Kurtzmann dans le cadre de la préparation à l'agrégation de mathématiques.

Le cours de ce polycopié a été un des ingrédients de base de l'ouvrage « De l'Intégration aux Probabilités » [3], que j'ai écrit avec Aline Kurtzmann et que nous avons publié aux éditions Ellipses. Vous êtes invités à vous y reporter pour compléter votre culture.

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

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

Table des matières

Notations

$\mathrm{Card}(A)$ 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

Calcul de lois

Ce premier chapitre introduit peu de notions théoriques nouvelles. Il vise essentiellement à voir comment pratiquer efficacement les techniques de calcul de loi introduites au premier semestre.

1.1 Prologue : les mathématiques, la modélisation

La théorie des probabilités, comme une grande partie des mathématiques, a pour origine des questionnements sur des problèmes issus de la vie réelle. Il convient aujourd'hui de bien faire la différence entre la modélisation, qui est la construction d'une représentation du monde réel par des objets mathématiques, et le calcul des probabilités proprement dit. On trouve encore de nombreux textes mathématiques qui entretiennent la confusion entre ces deux étapes. Cette confusion induit souvent les étudiants en erreur et est source de déconvenues pour les usagers des probabilités qui en attendent des miracles.

La modélisation, ce n'est pas des mathématiques Tout d'abord, entendons nous bien : le jardinage, ce n'est pas non plus des mathématiques, et le dire n'est pas dénigrer le jardinage. La phrase est à entendre dans un contexte de mathématiciens, pour lesquels faire des mathématiques, c'est faire des démonstrations. Ainsi la phase de modélisation est caractérisée par le fait que

— On fait des hypothèses. Ces hypothèses sont souvent guidées par le bon sens et l'expérience empirique. Cette expérience peut excéder (et parfois de beaucoup) la culture générale d'un mathématicien.

— On ne fait pas de preuves.

Ainsi, on ne peut pas démontrer que les observations d'une suite de lancers d'une pièce non truquée sont des variables aléatoires indépendantes valant "pile" ou "face" avec une même probabilité. Cependant, l'expérience nous a appris que "tout ce passe comme si" et que des choix basés sur cette hypothèse n'avaient pas eu de conséquence fâcheuse.

Les erreurs courantes

Usage inapproprié de la loi uniforme L'erreur la plus courante consiste à faire trop d'hypothèses, qui ne sont pas basées sur l'expérience. Un exemple classique est l'étude de la loi de la somme de deux dés à six faces. Une erreur courante consiste à remarquer que les valeurs possibles vont de 2 à 12, et en induire que la loi de la somme est la loi uniforme sur l'ensemble $\{2, \ldots, 12\}$. Essayons d'analyser cette erreur.

D'abord, notons que l'emploi de la loi uniforme est naturel, et semble plein de bon sens dès lors que le phénomène que l'on étudie présente d'importantes symétries. Par exemple, il est tout à fait raisonnable de penser que si on lance une pièce, les deux côtés ont même probabilité de tomber. De même si on lance un dé qui à la forme d'un tétraèdre régulier ou d'un cube, l'hypothèse d'équiprobabilité est raisonnable.

En revanche, la somme de deux dés est un phénomène qui manque de symétrie, par exemple parce que 2 ne peut s'obtenir que comme somme de deux 1 et que 7 peut s'obtenir de 6 manières différentes.

S'écarter de la réalité Cet exemple est également typique du danger qu'il y a à s'écarter de la réalité physique directement observée pour courir à l'aspect particulier de l'aspect que l'on veut étudier. Ici, la réalité physique est l'observation de deux nombres indiqués sur les dés. Si l'on s'en tient à cette observation, on peut postuler que les $6 \times 6 = 36$ observations possibles ont la même probabilité. Ce postulat conduira à des calculs qui donneront le résultat raisonnable pour la loi de la somme, le même que l'on obtient si l'on suppose que les résultats des deux dés sont indépendants et suivent la loi uniforme sur $\{1, \ldots, 6\}$.

Toujours dans le cadre de l'expérience aléatoire d'un lancer de deux dés, une autre erreur classique est d'oublier que les deux dés sont des objets que l'on peut observer séparément, d'identifier les résultats symétriques (1–5 et 5–1 par exemple) et de postuler que les 21 résultats possibles sont équiprobables. Dans ce genre de problème, il ne faut jamais perdre de vue que, si on veut construire une modélisation probabiliste qui puisse rendre compte (autant que faire se peut) de la réalité observée, il faut choisir une représentation qui puisse tirer parti de ce que la vie nous a appris. Si on représente le lancer comme un couple non ordonné, l'expérience (la vie) ne nous a rien appris du tout, c'est d'ailleurs ce qui justifie que l'on confie le soin au probabiliste de nous éclairer sur les chances des différentes issues.

Le choix des mots À la décharge des élèves et des étudiants qui font ce genre d'erreur, il faut reconnaître qu'un certain nombre de formulations codifiées que l'on trouve dans les manuels sont assez pousse-au-crime : ainsi de nombreux textes parlent de dés indiscernables, ce qui incite à une modélisation qui ne fait pas la différence entre les deux dés. Or, s'il est bien vrai que je ne peux pas faire la différence entre les deux dés lorsque je les sors de la boîte, il n'en demeure pas moins que ce sont bien deux objets distincts, ce dont une modélisation efficace tiendra compte.

Notons aussi qu'il y a, dans le cadre des problèmes discrets, un très ancien cousinage entre les problèmes de probabilités et les problèmes de dénombrement. Dans le cas où la probabilité mise sur l'espace est une loi uniforme, le calcul des probabilités peut se réduire à un problème de dénombrement. Encore faut-il que l'espace Ω des possibles ait été choisi de telle sorte que la probabilité uniforme rende compte de la réalité. Encore une fois, l'observation de la réalité pèse sur le choix du modèle.

Dans le cadre de la description du problème et du modèle, une source d'erreur assez fréquente est l'utilisation de termes mathématiques, qui ont un sens précis, dans le cadre informel de la description de l'expérience. On lit par exemple dans le sujet du capes externe 2014 :

« Soient $n \in \mathbb{N}^*$ et $N \in \mathbb{N}^*$. On dispose de N urnes $U_1, \ldots, U_N$ contenant des boules rouges et des boules blanches et telles que, pour tout $j \in \{1, \ldots, N\}$, la proportion de boules rouges dans U_j est j/N. On choisit une urne au hasard et on effectue dans cette urne n tirages indépendants d'une boule avec remise. »

Une erreur relativement fréquente est de prendre l'énoncé au pied de la lettre est de penser que les variables aléatoires $X_1, \ldots, X_n$ représentant les n tirages sont des variables indépendantes. Elles ne le sont évidemment pas : si je tire successivement 999 boules rouges, il y a fort à parier que je suis en train de piocher une urne contenant beaucoup de boules rouges, et donc que j'en tirerai encore une la millième fois.

Ici, l'énoncé, en même temps qu'il décrit l'expérience physique, donne une petite indication sur la modélisation. En réalité, les tirages sont indépendants conditionnellement au tirage de l'urne : si on note 0 les boules blanches et 1 les boules rouges, on a pour tout $(x_1, \ldots, x_n) \in \{0,1\}^n$ et tout $j \in \{1, \ldots, N\}$:

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = n | U = j) = (j/N)^{\sum x_i}(1-j/N)^{n-\sum x_i} = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i | U = j).$$

Pour terminer ces remarques sur la modélisation, enfonçons le clou : il n'est pas possible de "montrer que" des lancers de dés sont indépendants, même si on trouve encore parfois des énoncés qui perpétuent cet abus de langage. Il faut donc les interpréter comme "donner des arguments heuristiques qui permettent de penser que", même s'il n'existe le plus souvent d'autre réponse possible qu ""il est raisonnable de faire l'hypothèse que".

Le calcul des probabilités : des mathématiques Une fois qu'est défini le modèle, les lois des variables aléatoires, le lien à la vie réelle devient plus ténu, on peut même l'oublier puisqu'on travaille avec des objets mathématiques idéaux parfaitement définis. Cependant, ce lien avec le réel pourra tout de même guider l'intuition vers la formation de conjectures.

1.2 Qu'est-ce qu'une loi ?

1.2.1 Rappels

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 (ou d'un vecteur 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 écrit le plus souvent $\mathbb{P}(\{X \in A\})$ et même $\mathbb{P}(X \in A)$ pour désigner $\mathbb{P}_X(A)$. Dans le cas d'une variable aléatoire réelle, 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.

Le but de ce chapitre est de permettre à l'apprenti probabiliste de répondre à l'angoissante question : comment *calculer la loi de X* ?

1.2.2 Importance de la notion de mesure image

Une nouveauté du cours de L3 par rapport à un cours de lycée où de L2, c'est que la loi d'une variable aléatoire apparaît maintenant comme une loi image, ou une mesure image. De fait, le calcul d'une loi est toujours, d'une manière où d'une autre un calcul de loi image.

Théorème 1.1. *Soit μ une mesure de probabilité sur $\mathbb{R}^n$, ν une mesure de probabilité sur $\mathbb{R}^p$, ϕ une application mesurable de $\mathbb{R}^n$ dans $\mathbb{R}^p$. On a équivalence entre :*

— *ν est la mesure image de μ par ϕ*

— *Si X et Y sont des variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$ avec $\mathbb{P}_X = \mu$ et $Y = \phi(X)$, alors $\mathbb{P}_Y = \nu$.*

Démonstration. Soit A borélien de $\mathbb{R}^p$:

$$\begin{aligned}\mathbb{P}_Y(A) &= \mathbb{P}(Y^{-1}(A)) = \mathbb{P}((\phi \circ X)^{-1}(A)) \\ &= \mathbb{P}(X^{-1}(\phi^{-1}(A)) = \mathbb{P}_X(\phi^{-1}(A)) = \mu(\phi^{-1}(A)) = \mu_\phi(A)\end{aligned}$$

□

Ainsi, un calcul de loi est toujours d'une manière ou d'une autre un calcul de mesure image.

Il faut noter qu'un certain nombre de lois sont, par définition, des lois images. Par exemple, la loi binomiale $\mathcal{B}(n, p)$ est la loi image de $\mathrm{Ber}(p)^{\otimes n}$ par $(x_1, \ldots, x_n) \mapsto x_1 + \cdots + x_n$. Une autre loi, un peu moins classique, est la loi hypergéométrique

Une loi image : la 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 1.2. *On note $\mathcal{B}(N, k)$ est l'ensemble des parties de $\{1, \ldots, N\}$ de cardinal k. La loi hypergéométrique est la loi image de la loi uniforme sur $\Omega = \mathcal{B}(N, k)$ par l'application*

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

Ainsi, pour $i \in \{0, \ldots, \min(n, k)\}$, on a

$$\mathbb{P}(X = i) = \mathcal{H}(N, n, k)(i) = \frac{\binom{n}{i}\binom{N-n}{k-i}}{\binom{N}{k}}.$$

Démonstration. Notons $\mathbb{P}$ la loi uniforme sur Ω. On a

$$\mathcal{H}(N,n,k)(i) = \mathbb{P}(\omega \in S),$$

où $S = \{\omega \in \mathcal{B}(N,k); |\{1,\dots,n\} \cap \omega| = i\}$. L'application

$$\begin{array}{rcl} \mathcal{B}(\{1,\dots,n\},i) \times \mathcal{B}(\{n+1,\dots,N\},k-i) & \to & S \\ (A,B) & \mapsto & A \cup B \end{array}$$

est une bijection, donc

$$|S| = |\mathcal{B}(\{1,\dots,n\},i) \times \mathcal{B}(\{n+1,\dots,N\},k-i)| = \binom{n}{i}\binom{N-n}{k-i}.$$

Comme $\mathbb{P}$ est la loi uniforme sur Ω, et $|\Omega| = \binom{N}{k}$, le résultat s'ensuit. □

Il faut retenir qu'une méthode très efficace pour montrer que deux variables aléatoires ont la même loi est de les représenter comme la loi image de deux vecteurs de même loi par une même application.

Exemple : soient X, Y, Z trois variables aléatoires indépendantes suivant la loi exponentielle de paramètre 1. Alors X/Y et Y/Z ont même loi.

En effet la loi de (X/Y) est la loi image de $\mathbb{P}_{(X,Y)}$ par $(x,y) \mapsto x/y$ la loi de (Y/Z) est la loi image de $\mathbb{P}_{(Y,Z)}$ par $(x,y) \mapsto x/y$. Or $\mathbb{P}_{(X,Y)} = \mathcal{E}(1) \otimes \mathcal{E}(1) = \mathbb{P}_{(Y,Z)}$, d'où le résultat.

1.2.3 Douce loi des couples et indépendance

On sait que X et Y sont indépendants sous $\mathbb{P}$ si et seulement si $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$.

L'égalité est utile dans les deux sens : dans l'exemple précédent, l'hypothèse d'indépendance jointe à la connaissance des marginales nous a permis d'obtenir la loi d'un couple. C'est également utile dans l'autre sens : il est parfois plus facile de calculer la loi du couple (X,Y) que de chercher séparément les lois de X et Y.

Exemple : soient $(A_n)_{n\geq 1}$ des variables aléatoires suivant la loi de Bernoulli de paramètre p. On pose

$$X = \inf\{n \geq 1; A_n = 1\} \text{ et } Y = \inf\{n \geq 1; A_{X+n} = 1\}.$$

La loi de X est immédiate : c'est la loi géométrique de paramètre p. Le calcul direct de la loi de Y ne semble pas facile, mais le calcul de la loi du couple (X,Y) est assez simple : si n et ℓ sont des entiers naturels non nuls, on a

$$\begin{aligned} &\mathbb{P}(X = n, Y = \ell) \\ &= \mathbb{P}(A_1 = 0, \dots, A_{n-1} = 0, A_n = 1, A_{n+1} = 0, \dots, A_{n+\ell-1} = 0, A_{n+\ell} = 1) \\ &= (1-p)^{n-1}p(1-p)^{\ell-1}p = (\mathcal{G}(p) \otimes \mathcal{G}(p))(n,\ell), \end{aligned}$$

donc X et Y sont indépendantes et suivent la loi géométrique de paramètre p.

1.3 Identifier une loi

1.3.1 Un outil universel : la fonction test

On se souvient que les intégrales caractérisent les mesures.

Théorème 1.3. *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'application continues, et converge simplement vers l'indicatrice de K. Comme $|g_n| \le \mathbb{1}_{K+\overline{B}(0,1)}$ 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 $\mu = \nu$. $\square$

Corollaire 1.4. *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. $\square$

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 est une composante continue.

Exemple : soit X une variable suivant la loi uniforme sur $[-1, 1]$. Calculer la loi de $Y = \max(0, X)$.

On a $\phi(Y) = \phi(0)\mathbb{1}_{\{X<0\}} + \phi(X)\mathbb{1}_{\{X\ge0\}}$, donc

$$\begin{aligned}
\mathbb{E}\phi(Y) &= \mathbb{E}(\phi(0)\mathbb{1}_{\{X<0\}}) + \mathbb{E}(\phi(X)\mathbb{1}_{\{X\ge0\}}) \\
&= \phi(0)\mathbb{P}(X < 0) + \frac{1}{2}\int_{[-1,1]} \phi(x)\mathbb{1}_{\{x\ge0\}})\, d\lambda(x) \\
&= \frac{1}{2}\phi(0) + \frac{1}{2}\int_{[0,1]} \phi(x)\, d\lambda(x)
\end{aligned}$$

On reconnait là l'intégrale de ϕ par rapport à la mesure $\frac{1}{2}\delta_0 + \frac{1}{2}\lambda_{[0,1]}$.

1.3.2 Le cas discret

Rappelons qu'une variable (un vecteur) aléatoire X est discret si il existe D dénombrable avec $\mathbb{P}(X \in D) = 1$.

Dans le cas d'une variable aléatoire discrète (ou d'un vecteur discret), identifier la loi revient basiquement à

— identifier les valeurs possibles

— calculer les probabilités des différentes valeurs.

En effet, une loi discrète est complètement caractérisée dès lors que l'on connaît les masses des singletons.

Cette méthode est universelle, toutefois elle n'est pas toujours la plus rapide, en particulier lorsque la loi à identifier est une loi classique.

1.3.3 Le cas continu

La fonction de répartition de la variable aléatoire réelle X, définie par

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

ractérise sa loi. C'est une conséquence du théorème suivant :

Théorème 1.5 (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 $\mathbb{P} = \mathbb{Q}$.

Démonstration. Voir Garet–Kurtzmann. □

On va voir comment la fonction de répartition permet parfois de retrouver la loi d'une variable aléatoire réelle.

Théorème 1.6. *Soit F la fonction de répartition associée à la loi μ. On suppose que F est de classe C^1 par morceaux, avec les points de discontinuité $a_1, \dots, a_n$. Alors μ se décompose en la somme d'une partie à densité, f qui est la dérivée de F là où F est dérivable, et d'une partie discrète qui est $\nu = \sum_{i=1}^{n} \mu(a_i)\delta_{a_i}$.*

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

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

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

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

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

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

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

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

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

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

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

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

□

Remarque : dans le cours de deuxième partie de semestre, vous verrez les fonctions génératrices et les fonctions caractéristiques, qui sont des outils plus élaborés d'identification des lois.

1.4 Transformations

On a bien compris que le travail sur les lois revenait à trouver des lois images par des transformations. Voyons concrètement comment on procède.

1.4.1 Le cas discret

Si X est une variable aléatoire discrète à valeurs dans D, la variable aléatoire $Y = \phi(X)$ est encore une variable aléatoire discrète, à valeurs dans $F = \phi(D)$, caractérisée par la valeur, pour $a \in F$ de

$$\mathbb{P}(Y = a) = \mathbb{P}(\phi(X) = a) = \mathbb{P}(X \in \phi^{-1}(a)) = \sum_{x \in \phi^{-1}(a)} \mathbb{P}(X = x).$$

Exemple : soit X suivant la loi binomiale $\mathcal{B}(2n, \frac{1}{2})$. On pose $Y = |X - n|$. Ici X est à valeurs dans $D = \{0, \ldots, 2n\}$, $f(x) = |x - n|$ et $F = f(D) = \{0, \ldots, n\}$. Pour tout $k \in \{1, \ldots, n\}$, on a $f^{-1}(k) = \{n - k; n + k\}$, tandis que $f^{-1}(0) = \{n\}$. Ainsi $\mathbb{P}(Y = 0) = \mathbb{P}(X = n) = \frac{1}{2^{2n}}\binom{2n}{n} = \frac{1}{2^{2n}}\frac{(2n)!}{n!^2}$ et pour $k \in \{1, \ldots, n\}$, on a $\mathbb{P}(Y = k) = \mathbb{P}(X = n - k) + \mathbb{P}(X = n + k) = \frac{1}{2^{2n}}\binom{2n}{n-k} + \frac{1}{2^{2n}}\binom{2n}{n+k} = \frac{1}{2^{2n-1}}\frac{(2n)!}{(n+k)!(n-k)!}$.

1.4.2 Le cas continu

Changement de variables C^1

Théorème 1.7. *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 dans ce cas*

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

Remarque 1.8. *La quantité $\det D_x\ \phi$ est appelée déterminant jacobien (ou plus simplement Jacobien) de ϕ au point x.*

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 1.9. *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*

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

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,

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

ce qui donne le résultat voulu. □

Exemple : loi Beta, loi Gamma, et loi de Dirichlet
Quelques rappels :

— Soient a, b des réels strictement positifs. La densité de probabilité de la loi Beta 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 Beta, fonction d'Euler de première espèce qui peut s'exprimer comme

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

— 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

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

Théorème 1.10. *Soient $a_1, a_2, \lambda > 0$. Soit X_1 et X_2 indépendantes avec $X_1 \sim \Gamma(a_1, \lambda)$ et $X_2 \sim \Gamma(a_1, \lambda)$. On pose*

$$Y_1 = X_1/(X_1 + X_2) \text{ et } Y_2 = X_2/(X_1 + X_2).$$

On dit que (Y_1, Y_2) suit la loi de Dirichlet de paramètres (a_1, a_2).

Alors Y_1 suit la loi Bêta de paramètres (a_1, a_2), $X_1 + X_2$ suit la loi $\Gamma(a_1 + a_2, \lambda)$, et Y_1 et $X_1 + X_2$ sont indépendantes.

De plus $\beta(a_1, a_2) = \frac{\Gamma(a_1)\Gamma(a_2)}{\Gamma(a_1+a_2)}$.

Démonstration. L'application $T : (x, y) \mapsto (\frac{x}{x+y}, x+y)$ réalise un C^1-difféomorphisme de $(\mathbb{R}_+^*)^2$ dans $]0, 1[\times\mathbb{R}_+^*$, dont la réciproque est $T^{-1}(\theta, s) = (\theta s, (1 - \theta)s)$. On a $|DT^{-1}_{(\theta,s)}| = \begin{vmatrix} s & \theta \\ -s & 1-\theta \end{vmatrix} = s$. Comme la densité de (X_1, X_2) est

$$(x, y) \mapsto \frac{\lambda^{a_1+a_2}}{\Gamma(a_1)\Gamma(a_2)} x^{a_1-1}y^{a_2}e^{-\lambda(x+y)}\mathbb{1}_{\mathbb{R}_+^*}(x)\mathbb{1}_{\mathbb{R}_+^*}(y),$$

on obtient alors la densité de $(Y_1, X_1 + X_2)$:

$$\begin{aligned}(\theta, s) \quad \mapsto \quad & \frac{\lambda^{a_1+a_2}}{\Gamma(a_1)\Gamma(a_2)} s(\theta s)^{a_1-1}((1-\theta)s)^{a_2-1}e^{-\lambda s}\mathbb{1}_{]0,1[}(\theta)\mathbb{1}_{\mathbb{R}_+^*}(s) \\ & = K\frac{\theta^{a_1-1}(1-\theta)^{a_2-1}\mathbb{1}_{]0,1[}(\theta)}{\beta(a_1, a_2)} \frac{\lambda^{a_1+a_2}}{\Gamma(a_1 + a_2)} s^{a_1+a_2-1}e^{-\lambda s}\mathbb{1}_{\mathbb{R}_+^*}(s),\end{aligned}$$

où l'on a posé $K = \frac{\Gamma(a_1+a_2)\beta(a_1,a_2)}{\Gamma(a_1)\Gamma(a_2)}$. On reconnait ainsi

$$\mathbb{P}_{(Y_1,X_1+X_2)} = K\,(\text{Bêta}(a_1, a_2) \otimes \Gamma(a_1 + a_2, \lambda))$$

En évaluant cette égalité de mesures en Ω, on obtient $1 = K.1$, d'où

$$\mathbb{P}_{B(a_1,a_2)} = \text{Bêta}(a_1, a_2) \otimes \Gamma(a_1 + a_2, \lambda),$$

ce qui nous dit que Y_1 et $S_1 + S_2$ sont indépendantes et suivent respectivement les lois Bêta(a_1, a_2) et $\Gamma(a_1 + a_2, \lambda)$. □

Lorsque ϕ est une application de classe C^1 de $\mathbb{R}^n$ dans $\mathbb{R}^p$ avec $p < n$, une technique classique est de construire $\psi; \mathbb{R}^n \to \mathbb{R}^{n-p}$ telle que $x \mapsto (\phi(x), \psi(x))$ réalise un C^1-difféomorphisme. On applique alors le théorème de changement de variable pour trouver la densité de $(\phi(X), \psi(X))$ et en réintégrant, on obtient la densité de $\phi(X)$.

Exemple : soient X, Y indépendantes suivant la loi uniforme sur $[0, 1]$. On veut calculer la loi de XY. Il suffit de compléter XY en (XY, X) Si $T(x, y) = (xy, x)$, T est un C^1-difféomorphisme de $]0, 1[^2$ dans $O = \{(u, v) \in \mathbb{R}^2; 0 < u < v < 1\}$. Si $(u, v) = T(x, y)$ On a alors

$$|\det DT_{(x,y)}| = \left|\det \begin{pmatrix} y & x \\ 1 & 0 \end{pmatrix}\right| = x = v,$$

ce qui nous donne $|\det DT^{-1}_{(u,v)}| = v^{-1}$. Avec le théorème de C^1 difféomorphisme, le couple $(U, V) = (XY, X)$ a la densité $(u, v) \mapsto \frac{1}{v}\mathbb{1}_O(u, v)$. La densité de $U = XY$ est donc $u \mapsto \int_{\mathbb{R}} \frac{1}{v}\mathbb{1}_O(u, v)\,d\lambda(v)$. Sur $]0, 1[$, elle vaut $\int_0^1 \frac{1}{v}\mathbb{1}_{\{u<v\}}\,dv = \int_u^1 \frac{dv}{v} = -\log u$.

1.5 Les lois uniformes

Les lois uniformes ont une place particulière dans la littérature des exercices de probabilités. D'abord elles sont souvent mal nommées, ou plutôt pas nommées : quand un énoncé dire qu'une variable aléatoire est tirée "au hasard", il veut souvent dire qu'elle suit la loi uniforme sur un ensemble que le contexte permet de préciser. Cependant, elles ont des propriétés très intéressantes.

1.5.1 Loi uniforme sur un ensemble fini

La loi uniforme U_C sur l'ensemble fini C étant définie par

$$\forall A \in \mathcal{P}(C) \quad U_C(A) = \frac{|A|}{|C|},$$

lorsque une variable suit une loi uniforme, le calcul des probabilités est ramené à un dénombrement.

Par ailleurs, on a quelques propriétés très simples, mais bien utiles :

Produit

Si C et D sont finis, alors $U_{C\times D} = U_C \otimes U_D$.

Démonstration. Il suffit de vérifier l'égalité des mesures sur des ensembles produits

$$U_{C\times D}(A\times B) = \frac{|A\times B|}{|C\times D|} = \frac{|A|.|B|}{|C|.|D|}$$

$$= \frac{|A|}{|C|}\frac{|B|}{|D|} = U_C(A)U_D(B) = (U_C \otimes U_D)(A\times B)$$

□

Par exemple, si $X_1, \dots, X_n$ suivent la loi uniforme sur C, le vecteur $(X_1, \dots, X_n)$ suit la loi uniforme sur C^n.

Conditionnement

Rappelons que si $(\Omega, \mathcal{F}, \mu)$ est un espace probabilisé et $D \in \mathcal{F}$ avec $\mu(D) > 0$, l'application

$$\mu(\cdot|D) : A \mapsto \mu(A|D) = \frac{\mu(A\cap D)}{\mu(D)}$$

est une mesure de probabilité sur $(\Omega, \mathcal{F})$. Il est aisé de constater que la loi uniforme sur C conditionnée par $D \subset C$ (ou de manière équivalente $D \in \mathcal{P}(C)$) est la loi uniforme sur D.

Démonstration.

$$U_C(A|D) = \frac{U_C(A\cap D)}{U_C(D)} = \frac{|A\cap D|/|C|}{|D|/|C|} = \frac{|A\cap D|}{|D|} = U_D(A).$$

□

Par exemple, si $(X_1, \dots, X_n)$ sont des variables indépendantes suivant la loi de Bernoulli de paramètre $1/2$, la loi de $(X_1, \dots, X_n)$ sachant $X_1 + \dots + X_n = k$ est la loi uniforme sur $\{(x_1, \dots, x_n) \in \{0,1\}^n : x_1 + \dots + x_n = k\}$: c'est la loi de la suite des tirages sans remise dans une urne.

1.5.2 Loi uniforme sur un compact de $\mathbb{R}^d$

Rappelons qu'une variable aléatoire X suit la loi uniforme sur un compact K de $\mathbb{R}^d$ si elle admet la densité

$$x \mapsto \frac{1}{\lambda(K)}\,\mathbb{1}_K(x).$$

Par voie de conséquence, la loi uniforme $\mathcal{U}_K$ sur le compact K de $\mathbb{R}^d$ vérifie pour A borélien de $\mathbb{R}^d$:

$$\mathcal{U}_K(A) = \frac{\lambda(K\cap A)}{\lambda(K)}.$$

Ainsi, lorsque une variable suit une loi uniforme, le calcul des probabilités est ramené à un calcul d'aire.

Cette remarque est très importante car de nombreuses aires classiques se calculent à l'aide de formules de géométrie mémorisées par tous, alors que les calculs intégraux correspondants sont plus délicats. Par Exemple si μ est la loi uniforme sur le carré $[0,1]$, la mesure du disque unité est l'aire d'un quart de disque, classiquement égale à $\pi/4$, alors que le théorème de Fubini amène à calculer $\int_0^1 \sqrt{1-x^2}\,dx$, ce qui peut être un peu plus délicat...

Produit

Si C et D sont finis, alors $\mathcal{U}_{C\times D} = \mathcal{U}_C \otimes \mathcal{U}_D$.

Démonstration. Là encore, il suffit de vérifier l'égalité des mesures sur des ensembles produits. La preuve est laissée au lecteur. □

Par exemple, si X et Y sont des variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$, le vecteur aléatoire (X,Y) suit la loi uniforme sur le carré $[0,1]^2$.

Conditionnement

La loi uniforme sur C conditionnée par le borélien D (ou de manièle équivalente $D\in\mathcal{P}(C)$) est la loi uniforme sur $C\cap D$.

Par exemple, si X et Y sont des variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$, le vecteur aléatoire (X,Y) conditionné par $\{X+Y\le 1\}$ suit la loi uniforme sur le triangle $T=\{(x,y)\in\mathbb{R}_+^2; x+y\le 1\}$.

1.5.3 Application

Théorème 1.11. *Soit F une fonction de $\mathbb{R}$ dans $\mathbb{R}$, croissante, continue à droite, dont la limite est nulle en $-\infty$ et vaut 1 en $+\infty$. On suppose que sur $(\Omega,\mathcal{F},\mathbb{P})$, U est une variable aléatoire suivant la loi uniforme sur $[0,1]$. On pose*

$$\forall u \in]0,1[\quad Q^*(u) = \min\{x\in\mathbb{R}:\ 1-F(x)\le u\}.$$

et

$$\forall u \in]0,1[\quad F^*(u) = \inf\{s\in\mathbb{R}:\ F(s)>u\}.$$

Alors $F^*(U)$ et $Q^*(u)$ sont des variables aléatoires réelles dont la fonction de répartition est F.

Démonstration. On fait seulement la preuve pour F^* ; celle pour Q^* est analogue. Notons que pour tout $u \in]0,1[$, $F^*(u) < +\infty$ car F a une limite 1 en l'infini et $F^*(u) > -\infty$ car F a une limite 0 en l'infini. Comme U prend presque sûrement ses valeurs dans $]0,1[$, la variable $F^*(U)$ est bien définie. On va calculer la fonction de répartition de $F^*(U)$, et on doit donc travailler sur l'événement $\{F^*(U) \le t\}$. Si $F(t) > U$, alors $F^*(U) \le t$ par définition de l'$\inf$. D'autre part, si $F^*(U) \le t$, alors, comme F est croissante, pour tout $t' > t$, $F(t') > U$. Donc pour tout $t' > t$, on a

$$\{F(t) > U\} \subset \{F^*(U) \le t\} \subset \{F(t') > U\}$$

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

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 1.12. *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 1.13.

— *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 1.11 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 $i \ge 1$, il suffit de poser $X = x_{f(U)}$, où $f(x) = \inf\{n : s_n > x\}$, avec $s_0 = 0$ et $s_n = p_1 + \dots + p_n$.*
En effet, on a alors $\{X = x_n\} = \{f(U) = i\} = \{s_{n-1} \le U < s_n\}$ 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.

1.5.4 Exercices de la série 1

Exercice 1. On rappelle que pour x réel, $\{x\}$ désigne la partie fractionnaire de x : $x = \lfloor x \rfloor + \{x\}$.
Soit $\alpha \in \mathbb{R}$ et X une variable aléatoire sur $(\Omega, \mathcal{F}, \mathbb{P})$ suivant la loi uniforme sur $[0,1]$. Montrer que $Y = \{X + \alpha\}$ a la même loi que X.

Exercice 2. 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 $X_3 = X_1 X_2$. Montrer que les variables X_2 et X_3 sont indépendantes, ainsi que X_1 et X_3. En revanche, montrer que X_1, X_2, X_3 ne le sont pas.

Exercice 3. Les points X et Y sont répartis uniformément et indépendamment respectivement sur les côtés AB et BC d'un triangle ABC. Soient S_{ABC} l'aire du triangle ABC et S_{XBY} celle du triangle XBY.
Trouver $\mathbb{P}(S_{ABC} > 2S_{XBY})$.

Exercice 4. Soit X une variable aléatoire de densité $f(x) = \frac{\mathbb{1}_{[0,1]}(x)}{(1+x)\log 2}$. Montrer que $\{\frac{1}{X}\} = \frac{1}{X} - \lfloor \frac{1}{X} \rfloor$ a même loi que X.

Exercice 5.
1. On suppose que $X_1, \dots X_n$ sont des variables aléatoires indépendantes suivant la loi uniforme sur $[0, 1]$.
   Calculer $\mathbb{E}\left(\max\{X_1, \dots, X_n\}\right)$.
2. On suppose que $Y_1, \dots Y_n$ sont des variables aléatoires indépendantes suivant la loi exponentielle de paramètre 1.
   Calculer $\mathbb{E}\left(\min\{Y_1, \dots, Y_n\}\right)$.

Exercice 6. *volume de la boule unité* de $(\mathbb{R}^n, \|.\|_p)$ On rappelle que pour $x \in \mathbb{R}^n$, on note $\|x\|_p = (\sum_{k=1}^n |x_k|^p)^{1/p}$.

1. Soient $Y_1, \dots, Y_n$ des variables aléatoires indépendantes suivant la loi $\Gamma(\frac{1}{p}, 1)$. On pose $X_k = Y_k^{1/p}$ et $S = X_1^p + \dots + X_n^p$. Montrer que X_1 admet la densité $x \mapsto \frac{e^{-x^p}\mathbb{1}_{\mathbb{R}_+}(x)}{\Gamma(\frac{1}{p}+1)}$ et que $S \sim \Gamma(n/p, 1)$.
2. Soit $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ une fonction mesurable positive. Montrer que
   $$\int_{\mathbb{R}_+^n} \phi(\|x\|_p^p)\, d\lambda^{\otimes n}(x) = \mathbb{E}[\psi(S)], \text{ où } \psi(x) = \Gamma\left(\frac{1}{p}+1\right)^n \phi(x)e^x.$$
   En déduire que
   $$\int_{\mathbb{R}^n} \phi(\|x\|_p^p)\, d\lambda^{\otimes n}(x) = \frac{2^n\Gamma(\frac{1}{p}+1)^n}{\Gamma(n/p)} \int_{\mathbb{R}_+} u^{\frac{n}{p}-1}\phi(u)\, du.$$
3. Montrer que le volume de la boule unité de $(\mathbb{R}^n, \|.\|_p)$ est $\frac{2^n\Gamma(\frac{1}{p}+1)^n}{\Gamma(\frac{n}{p}+1)}$.
4. Soit $n \geq 2$. On pose $T_n = \{(x_1, \dots, x_n) \in \mathbb{R}_+^n ; x_1 + \dots + x_n \leq 1\}$.
   Calculer $\displaystyle\int_{T_n} \frac{d\lambda^{\otimes n}(x)}{x_1 + \dots + x_n}$.

Exercice 7.
1. Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes et identiquement distribuées suivant la loi exponentielle $\mathcal{E}(1)$. Calculer la loi du vecteur $(U_1, \dots, U_{n-1}, S_n)$ 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 $U = (U_1, \dots, U_{n-1})$ puis celle du vecteur $V = (U_1, \dots, U_n)$ 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.

Exercice 8.
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 $\zeta(s) = \sum\limits_{n=1}^{+\infty} \frac{1}{n^s}$. 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 $\mathbb{P}(X \wedge X' = 1) = \frac{1}{\zeta(s+t)}$.

1.5.5 Exercices de la série 2

Exercice 9. *Application de Hénon* Soient a et b des réels. On suppose que le vecteur (X, Y) suit la loi uniforme sur le compact K. On pose $U = 1 + Y - aX^2$ et $V = bX$. Montrer que (U, V) suit la loi uniforme sur un compact.

Exercice 10. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes identiquement distribuées telles que $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = 1/2$. Pour tout $n \geq 1$, on pose $Z_n = \prod_{i=1}^n X_i$. Montrer que les $(Z_n)_{n \geq 1}$ sont indépendantes.

Exercice 11.
1. Soient X, Y des variables aléatoires suivant la loi uniforme sur $\{0, 1\}$. On pose $Z = \{X + Y\}$, où $\{\cdot\}$ désigne la partie fractionnaire. Montrer que Z suit la loi uniforme sur $\{0, 1\}$, qu'elle est indépendante de X, et également de Y. Les variables X, Y, Z sont elles globalement indépendantes ?
2. Soit $(U_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes suivant la loi uniforme sur $[0, 1]$. Pour $n \geq 1$, on pose $S_n = U_1 + \ldots U_n$ et $W_n = \{S_n\}$. Montrer que les $(W_n)_{n \geq 1}$ forment une suite de variables indépendantes. En déduire que
$$P(S_n < 1) = \mathbb{P}(W_1 \leq W_2 \leq \cdots \leq W_n) = \mathbb{P}(W_1 < W_2 < \cdots < W_n) = \frac{1}{n!}.$$
3. Calculer $\mathbb{E}[T]$, où $T = \inf\{n \geq 0 \quad S_n \geq 1\}$.

Exercice 12. 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 $\zeta(s) = \sum\limits_{n=1}^{+\infty} \frac{1}{n^s}$. 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 : $\mathbb{P}_{Y|X=x} = \mathcal{U}(\{1, \ldots, x\})$ la loi uniforme sur $\{1, \ldots, x\}$.

1. Pour $n, k \in \mathbb{N}^*$, calculer $\mathbb{P}(Y = k | X = n)$.
2. On pose $Z = \frac{Y}{X}$. Montrer que la fonction de répartition F_Z est strictement croissante sur $[0, 1]$.
3. Soient p, q deux entiers positifs premiers entre eux, avec $p \leq q$. Calculer $\mathbb{P}(Z = \frac{p}{q})$.
4. On rappelle que $\phi(n)$ désigne le nombre d'entiers compris entre 1 et n qui sont premiers avec n. Déduire de ce qui précède une preuve probabiliste de l'identité
$$\zeta(s + 1) \sum_{n=1}^{+\infty} \frac{\phi(n)}{n^{s+1}} = \zeta(s).$$

Exercice 13. Soit $\mathcal{P}$ l'ensemble des nombres premiers. Pour $p \in \mathcal{P}$, on note $\nu_p(n)$ le plus grand entier k tel que p^k divise n. Soit X une variable aléatoire suivant la loi Zêta de paramètre s (voir exercice précédent). Montrer que les variables aléatoires $(1+\nu_p(X))_{p\in\mathcal{P}}$ sont des variables indépendantes, avec $1+\nu_p(X) \sim \mathcal{G}(1-\frac{1}{p^s})$.

Exercice 14. Dans le segment $[AB]$ de longueur 1, on choisit au hasard un point M. Quelle est la probabilité pour que l'on ait $AM.MB \geq \frac{2}{9}$?

Exercice 15. Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$.
On pose $M_n = \max(X_1, \dots, X_n)$. Déterminer la fonction de répartition de M_n. Montrer que M_n admet une densité que l'on déterminera.

Exercice 16. Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On pose

$$M_n = \max(X_1, \dots, X_n) \text{ et } m_n = \min(X_1, \dots, X_n).$$

Montrer que M_n et $1-m_n$ ont même loi.

Exercice 17. La tradition veut que l'Épiphanie soit l'occasion de « tirer les rois » : une fève est cachée dans une galette, découpée entre les convives et la personne qui obtient cette fève devient le roi de la journée. Lorsque le premier coup de couteau est porté sur la fève, c'est la consternation ! 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, de rayon r. Enfin, on suppose que

— la position du centre de la fève suit la loi uniforme sur le disque de rayon $1-r$ ayant le même centre que la galette
— le coup de couteau est un rayon du disque représentant la galette

Application numérique avec une fève de 2,7 centimètres de diamètre dans une galette de 23 centimètres de diamètre achetée ce matin.

Exercice 18. Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes telles que pour tout $1 \leq i \leq n$, X_i suit la loi exponentielle $\mathcal{E}(\lambda_i)$.
On note $T = \inf(X_1, \dots, X_n)$ et $N = \inf\{i \geq 1; X_i = T\}$.

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. Pour i entre 1 et N, on pose $Y_i = \inf(X_j; j \in \{1, \dots, n\} \backslash \{i\})$. Montrer que Y_i est indépendant de X_i, puis déterminer sa loi.
3. Soit $t > 0$. Montrer que pour tout i compris entre 1 et n,
$$\mathbb{P}(T > t, N = i) = \mathbb{P}\left(Y_i > X_i > t\right)$$
4. On pose $\lambda = \sum\limits_{j=1}^{n} \lambda_j$. Montrer que $\mathbb{P}(T > t, N = i) = \frac{\lambda_i}{\lambda} \exp(-\lambda t)$.
5. Montrer que T et N sont indépendantes et préciser leurs lois.

Exercice 19. Soit n un entier naturel. On considère X une variable aléatoire exponentielle de paramètre 1 et Y une binomiale $\mathcal{B}(n, \frac{1}{2})$. On suppose que X et Y sont indépendantes.

Montrer que $Z = \frac{X}{Y+1}$ est une variable à densité et déterminer sa densité.

Chapitre 2

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.

2.1 De $\mathcal{L}^p$ à L^p

2.1.1 Inégalité de Hölder

Théorème 2.1. *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, $\left|\int_\Omega fg\ d\mu\right| \leq \int_\Omega |f|.|g|\ d\mu$.
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

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

$\square$

Remarque 2.2. *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

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

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.

2.1.2 Inégalité triangulaire (ou inégalité de Minkowski)

Théorème 2.3. *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*

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

1. Parfois appelée inégalité de Young.

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

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

□

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 2.4. *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 $1 \leq p \leq +\infty$ 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 $1 \leq p \leq +\infty$.*

Lemme 2.5. *Soit μ une mesure finie. Si $1 \leq p \leq r \leq +\infty$, 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 $|f|^p \leq 1 + |f|^r$ (pour voir cela, il suffit de séparer les cas $|f| \leq 1$ et $|f| > 1$). On a alors

$$\int_\Omega |f|^p \, d\mu \leq \int_\Omega (1 + |f|^r) \, d\mu = \mu(\Omega) + \int_\Omega |f|^r \, d\mu < +\infty$$

car f est dans L^r et donc f est dans L^p.
Si $f \in L^\infty$, on a $|f(x)|^p \leq \|f\|_\infty^p$ et donc $\int_\Omega |f|^p \, d\mu \leq \|f\|_\infty^p \mu(\Omega) < +\infty$. □

Remarque 2.6. *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 $1 \leq p < +\infty$.*

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 2.7. *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.

2.2 Complétude de L^p

Théorème 2.8. *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 2.9. *Soit (f_n) une suite d'éléments de L^p avec*

$$\sum_{n=1}^{+\infty} \|f_n\|_p < +\infty.$$

Alors la suite $\left(\sum_{k=1}^{n} f_k \right)$ *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 $\left\| \sum_{k=1}^{n} g_k - g \right\|_p$ tend vers 0, ce qui donnera la convergence de la suite $\left(\sum_{k=1}^{n} f_k \right)$ 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

$$\|S_{n'} - S_n\|_p \leq \sum_{k=n+1}^{n'} \|g_k\|_p \leq \sum_{k=n+1}^{+\infty} \|g_k\|_p,$$

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 2.10. *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 2.11. *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 2.12. *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 $\max\limits_{1\leq i\leq N_n} \lambda(A_{n,i}) < \varepsilon$ 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 $\{(n, i); n \in \{1, \dots, N_0\} \text{ et } i \in \{1, \dots, N_n\}\}$. Par suite, l'ensemble des n tels que $\lambda(A_{\phi(n)}) \geq \varepsilon$ est également fini, ce qui montre que $\lim\limits_{n\to+\infty} \lambda(A_{\phi(n)}) = 0$.

On conclut que $f_n = \mathbb{1}_{A_{\phi(n)}}$ converge dans L^p vers 0, tandis que

$$\forall x \in \mathbb{R}, \qquad \overline{\lim}_{n\to+\infty} f_n(x) = 1.$$

2.3 Théorèmes d'approximation

Théorème 2.13. *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 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
$$\phi_n(x) = 2^{-n} \lfloor 2^n x \rfloor \mathbb{1}_{[0,n]}(x) \text{ pour } x < +\infty \text{ et } \phi_n(+\infty) = n.$$
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 2.8. $\square$

Théorème 2.14. *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.8 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 2.13. $\square$

2.4 Exercices sur les espaces L^p

2.4.1 Exercices de la série 1

Exercice 20. 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 $\overline{\lim\limits_{p\to+\infty}}\, N(p) \leq \|X\|_{\infty,\text{ess}}$.
   (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 $\underline{\lim\limits_{p\to+\infty}}\, N(p) \geq M$.
   (d) Montrer que $\lim\limits_{p\to+\infty} N(p) = \|X\|_{\infty,\text{ess}}$.
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.

Exercice 21. 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 $(\{s_n\})_{n\geq 1}$?

Exercice 22. *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 $B_{k,n} = \bigcap_{i\geq n} \{x \in E : |f_i(x) - f(x)| > \frac{1}{k}\}$, 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.

Exercice 23. *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).

2.4.2 Exercices de la série 2

Exercice 24. 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.

Exercice 25. 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. $h(t) = \frac{1}{\sqrt{|t|}(1+t^2)}$.

Exercice 26. 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. $h_n(t) = \frac{2}{\pi n^2}\sqrt{n^2-t^2}\mathbb{1}_{[-/n,n]}(t)$.

Exercice 27. 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 $\mathcal{L}^1(\mathbb{R},\lambda)$.

Exercice 28. 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 $p = 2$.

Exercice 29. 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 $\ell^1(\mathbb{N}^*)$.

Exercice 30. Montrer que si f et g appartiennent à $\mathcal{L}^1(X,\mu)$ alors $\sqrt{|f^2+g^2|}$ appartient aussi à $\mathcal{L}^1(X,\mu)$.

Exercice 31. 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 $\mathcal{L}^p([1,\infty[,\lambda)$.

Exercice 32. 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 au sens de la norme $\mathcal{L}^1$;
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 au sens de la norme $\mathcal{L}^1$.

Exercice 33. 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 $\displaystyle (\int_X f d\mu)(\int_X g d\mu) \geq 1$.

Exercice 34. 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 $||fg||_r \leq ||f||_p ||g||_q$.

Exercice 35. *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 $\displaystyle ||T(f)||_p \leq \frac{p}{p-1}||f||_p$.
   (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 $\displaystyle ||T(f)||_p \leq \frac{p}{p-1}||f||_p$.

Exercice 36. Soit $(X,\mathcal{T},\mu)$ un espace mesuré tel que $\mu(X) < +\infty$. On se donne $p \in]1,+\infty[$ et $f : X \to \mathbb{R}$ 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 $\frac{1}{p} + \frac{1}{q} = 1$

Chapitre 3

Convolution et transformation de Fourier

Dans ce chapitre, p est un réel appartenant à $[1, +\infty[$.

3.1 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 3.1. *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 2.14, 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

$\overline{\lim\limits_{t \to 0}} \ \|T_t f - f\|_p = 0$, ce qui est le résultat voulu. $\square$

3.1.1 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

$$\int_{\mathbb{R}^d} (f * g)(x) \, d\lambda^d(x) = \left(\int_{\mathbb{R}^d} f(x) \, d\lambda^d(x) \right) \left(\int_{\mathbb{R}^d} g(t) \, d\lambda^d(t) \right). \tag{3.1}$$

3.1.2 Autres produits

Supposons maintenant que $g \in \mathcal{L}^1$ et que $f \in \mathcal{L}^p$. On a par l'inégalité de Hölder

$$\begin{aligned} & \int |f(x-t)||g(t)| \; d\lambda^d(t) \\ = & \int |f(x-t)||g(t)|^{1/p}|g(t)|^{1/q} \; d\lambda^d(t) \\ \leq & \left(\int |f(x-t)|^p|g(t)| \; d\lambda^d(t)\right)^{1/p} \left(\int |g(t)| \; d\lambda^d(t)\right)^{1/q}. \end{aligned}$$

D'où

$$\int \left(\int |f(x-t)||g(t)| \; d\lambda^d(t)\right)^p d\lambda^d(x) \leq \iint |f(x-t)|^p|g(t)| \; d\lambda^d(t)d\lambda^d(x)\|g\|_1^{p/q}.$$

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,

$$\int \left(\int |f(x-t)||g(t)| \; d\lambda^d(t)\right)^p \; d\lambda^d(x) \quad \leq \|f\|_p^p\|g\|_1^{1+p/q}.$$

Ainsi, l'intégrale

$$\int f(x-t)g(t) \; d\lambda^d(t)$$

converge pour presque 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 3.2 (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.

3.1.3 Approximations de l'unité

Théorème 3.3. *Soit $(\phi_k)_{k\geq 1}$ une suite de fonctions mesurables positives telles que*

$$\forall k \geq 1 \qquad \int_{\mathbb{R}^d} \phi_k(x)\, d\lambda^d(x) = 1$$

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

$$|f * \phi_k(x) - f(x)| \leq \int_{B(0,\delta)} \omega_f(\delta)\phi_k(t)\, d\lambda^d(t) + \int_{\mathbb{R}^d\setminus B(0,\delta)} 2\|f\|_\infty \phi_k(t)\, d\lambda^d(t),$$

où ω_f désigne le module de continuité de f. En passant au supremum en x, on obtient

$$\|f * \phi_k - f\|_\infty \leq \omega_f(\delta) + 2\|f\|_\infty \int_{\mathbb{R}^d\setminus B(0,\delta)} \phi_k(t)\, d\lambda^d(t).$$

Puis, faisant tendre k vers l'infini, on trouve

$$\overline{\lim_{k\to+\infty}}\ \|f * \phi_k - f\|_\infty \leq \omega_f(\delta).$$

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

$$\begin{aligned} &|f * \phi_k(x) - f(x)| \\ &= |\int f(x-t) - f(x)\phi_k(t)^{\frac{1}{p}}\phi_k^{1-\frac{1}{p}}(t)d\, d\lambda^d(t)| \\ &\leq \left(\int \phi_k(t)d\, d\lambda^d(t)|\right)^{1-1/p} \left(\int |f(x-t) - f(x)|^p\phi_k(t)\, d\lambda^d(t)\right)^{1/p} \\ &\leq \left(|\int T_t f(x) - f(x)|^p \phi_k(t)\, d\lambda^d(t)\right)^{1/p}. \end{aligned}$$

En élevant à la puissance p et en intégrant, le théorème de Tonelli nous donne

$$\|f * \phi_k - f\|_p^p \leq \int_{\mathbb{R}^d} \|T_t f - f\|_p^p \phi_k(t)\, d\lambda^d(t).$$

Soit $\varepsilon > 0$. D'après le théorème 3.1, $\|T_t f - f\|_p^p$ tend vers 0 lorsque t tend vers 0. On peut donc se donner δ tel que $\|T_t f - f\|_p \leq \varepsilon$ sur $B(0,\delta)$. Comme pour tout t, $|\|T_t f - f\|_p^p \leq (2\|f\|_p)^p$, en découpant comme dans la première partie, on a

$$\|f * \phi_k - f\|_p^p \leq \varepsilon^p + (2\|f\|_p)^p \int_{\mathbb{R}^d \setminus B(0,\delta)} \phi_k(t)\, d\lambda^d(t).$$

En passant à la limite supérieure, on a $\overline{\lim_{k \to +\infty}} \|f * \phi_k - f\|_p^p \leq \varepsilon^p$, soit $\overline{\lim_{k \to +\infty}} \|f * \phi_k - f\|_p \leq \varepsilon$. On conclut en faisant tendre ε vers 0. $\square$

Théorème 3.4. *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. $\square$

3.1.4 Régularisation

Théorème 3.5. *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. $\square$

Corollaire 3.6. *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 3.7. *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 3.4 et du corollaire précédent. □

3.2 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 3.8. *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.

3.2.1 Propriétés élémentaires

Proposition 3.9. *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 quelquues 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 (3.1), 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 3.10 (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 3.7). 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. □

3.2.2 Théorème d'inversion

Théorème 3.11. *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 3.12. *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. 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 3.11. 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 3.4, 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. □

3.3 Exercices sur la transformation de Fourier

3.3.1 Exercices de la série 1

Exercice 37. 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 $g(x) = \mathbb{1}_{[-b,b]}(x)$.

Exercice 38. 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.

Exercice 39. 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 $x \in E - E$.

3.3.2 Exercices de la série 2

Exercice 40. 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 $\mathcal{C}^{n-2}$.

Exercice 41. 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 $A + B = \{a + b; a \in A, b \in B\}$.

Exercice 42. 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 $x \in \mathbb{R}^n$.

Exercice 43. 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.

Exercice 44. 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.

Exercice 45. 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 $g : x \mapsto \dfrac{1}{a^2 + x^2}$.

Exercice 46. *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

$$\int_{\mathbb{R}^d} \widehat{\mu}(x)\, d\nu(x) = \int_{\mathbb{R}^d} \widehat{\nu}(x)\, d\mu(x).$$

Chapitre 4

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.

4.1 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 4.1. *Si la loi de X est à support fini, alors G_X est un polynôme.*

4.1.1 Fonction génératrice et indépendance

Théorème 4.2. *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

$$G_{X+Y}(z) = \mathbb{E}[z^{X+Y}] = \mathbb{E}[z^X z^Y] = \mathbb{E}[z^X]\mathbb{E}[z^Y] = G_X(z)G_Y(z).$$

□

4.1.2 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 $X_1, \dots, X_n$ sont des variables aléatoires indépendantes suivant la loi de Bernoulli de paramètre p, alors $S_n = X_1 + \dots + X_n$ suit la loi binomiale de paramètres n et p. Ainsi, on déduit du théorème 4.2 que la fonction génératrice d'une loi binomiale de paramètres n et p est

$$G_{S_n}(z) = G_{X_1}(z) \times \dots \times G_{X_n}(z) = ((1-p)+pz)^n.$$

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

$$\begin{aligned} G_X(s) &= \sum_{k=0}^{+\infty} \mathbb{P}(X=k)s^k = \sum_{k=0}^{+\infty} e^{-\lambda}\frac{\lambda^k}{k!}s^k = e^{-\lambda}\sum_{k=0}^{+\infty}\frac{(\lambda s)^k}{k!} \\ &= e^{-\lambda}e^{\lambda s} = e^{-\lambda(1-s)}. \end{aligned}$$

4.1.3 Fonction génératrice et loi

Théorème 4.3. *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. □

4.1.4 Application : convolution de lois de Poisson

Théorème 4.4. *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 4.3, la fonction génératrice détermine la loi, donc $X+Y$ suit la loi de Poisson de paramètre $\lambda+\mu$. □

4.1.5 Fonction génératrice et espérance

Théorème 4.5. *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

$$\lim_{x\to 1^-} \frac{G_X(1)-G_X(x)}{1-x} = \sum_{n=0}^{+\infty} n\nu(n) = \int x\ d\nu(x) = \mathbb{E}X.$$

□

4.2 Fonctions caractéristiques

4.2.1 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 3. 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*

$$\begin{aligned}&\forall t = (t_1, \dots, t_d) \in \mathbb{R}^d,\\ &\qquad \phi_\mu(t_1, \dots, t_d) \quad = \quad \int_{\mathbb{R}^d} \exp(i(t_1x_1 + t_2x_2 + \dots + t_dx_d))\, d\mu(x).\end{aligned}$$

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 4.6. *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 4.7. *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

$$\begin{aligned} &I_T(a,b) \\ &= \int\left(\frac{\operatorname{sign}(x-a)}{2\pi}\int_{-|x-a|T}^{|x-a|T}\frac{\sin t}{t}\, dt - \frac{\operatorname{sign}(x-b)}{2\pi}\int_{-|x-b|T}^{|x-b|T}\frac{\sin t}{t}\, dt\right) d\mu(x). \end{aligned}$$

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 4.8. *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. □

4.2.2 Propriétés des fonctions caractéristiques

Théorème 4.9. *Soit μ une mesure de probabilité sur $\mathbb{R}^d$. On a les propriétés suivantes :*

i) $\phi_\mu(0) = 1$.

ii) $|\phi_\mu| \leq 1$.

iii) *ϕ est uniformément continue sur $\mathbb{R}^d$.*

iv) *Si $e_1, \ldots, e_p$ 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) $|\phi_\mu(t)| = |\int_{\mathbb{R}^d} e^{i\langle x,t\rangle} d\mu(x)| \leq \int_{\mathbb{R}^d} |e^{i\langle x,t\rangle}| d\mu(x) = \int_{\mathbb{R}^d} 1 \; d\mu = 1$.

iii) Soient $t, t' \in \mathbb{R}^d$. On a

$$\begin{aligned}
|\phi_\mu(t) - \phi_\mu(t')| &= \left| \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} - e^{i\langle t',x\rangle} \; d\mu(x) \right| \\
&= \left| \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} (1 - e^{i\langle (t'-t),x\rangle}) \; d\mu(x) \right| \\
&\leq \int_{\mathbb{R}^d} |1 - e^{i\langle (t'-t),x\rangle}| \; d\mu(x).
\end{aligned}$$

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 $|1 - e^{iay}| \leq \min(|ay|, 2)$.

iv) On remarque que, pour tout $p-$uplet de nombres réels $x_1, \ldots, x_p$, 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

$$\begin{aligned}\sum_{k=1}^{p}\sum_{l=1}^{p} x_k x_l \phi_\mu(k-l) = 0 &\Longrightarrow \int_{[-\pi,\pi[}\left|\sum_{k=1}^{p} x_k \exp(ikx)\right|^2 \, d\mu(x) = 0\\ &\Longrightarrow \frac{1}{2\pi}\int_{[-\pi,\pi[}\left|\sum_{k=1}^{p} x_k \exp(ikx)\right|^2 f(x)\, d\lambda(x) = 0\\ &\Longrightarrow \left|\sum_{k=1}^{p} x_k \exp(ikx)\right|^2 f(x) = 0 \text{ p.p. sur } [-\pi,\pi].\end{aligned}$$

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 4.10 (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 4.11. *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

$$\begin{aligned}\phi_Y(t) &= \mathbb{E}e^{i\langle Y,t\rangle} = \mathbb{E}e^{i\langle AX+b,t\rangle} = \mathbb{E}e^{i\langle AX,t\rangle}e^{i\langle b,t\rangle}\\ &= \mathbb{E}e^{i\langle X,A^*t\rangle}e^{i\langle b,t\rangle} = e^{i\langle b,t\rangle}\phi_X(A^*t).\end{aligned}$$

□

4.2.3 Fonction caractéristique et indépendance

Théorème 4.12. *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 4.13. *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 4.14. *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 4.13 admet une réciproque.

Théorème 4.15. *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 4.12 n'admet pas de réciproque : on verra en exercice des variables aléatoires X et Y non-indépendantes telles que $\phi_{X+Y}=\phi_X\phi_Y$.

4.2.4 Fonction caractéristique et moments

Théorème 4.16. *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 $k_1,\dots,k_d$ sont des entiers naturels dont la somme $k=k_1+\dots+k_d$ ne dépasse pas N, on a alors

$$\forall u\in\mathbb{R}^d\quad \frac{\partial^k\phi_X}{\partial_1^{k_1}\dots\partial_d^{k_d}}(u)=i^k\int_{\mathbb{R}^d}e^{i\langle u,x\rangle}\prod_{j=1}^d x_j^{k_j}\,d\mu(x).\tag{4.1}$$

Ainsi, on a en particulier

$$\frac{\partial^k\phi_X}{\partial_1^{k_1}\dots\partial_d^{k_d}}(0)=i^k\mathbb{E}\left[\prod_{j=1}^d X_j^{k_j}\right].\tag{4.2}$$

Démonstration. On montre d'abord le lemme suivant : si $P \in \mathbb{R}[X_1, \dots, X_d]$ est le monôme de degré k : $P(X) = X_1^{\alpha_1} \dots X_d^{\alpha_d}$ 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 $j \in \{1, \dots, d\}$:

$$\forall u \in \mathbb{R}^d \quad \frac{\partial \Psi}{\partial_j}(u) = \mathbb{E}(ie^{i\langle u, X\rangle} X_j P(X)).$$

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 $u = 0$. $\square$

Dans le cas des variables aléatoires réelles, le théorème s'énonce évidemment plus simplement.

Corollaire 4.17. *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*

$$\forall k \in \{1, \dots, N\}, \quad \phi_X^{(k)}(0) = i^k \mathbb{E}[X^k]. \tag{4.3}$$

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. $\square$

4.2.5 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.

4.2.6 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

$$\phi_Y(t) = e^{i\frac{a+b}{2}} \phi_X((b-a)t) = e^{i\frac{a+b}{2}} \frac{\sin \frac{(b-a)t}{2}}{\frac{(b-a)t}{2}}.$$

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 4.18. *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 3.12.
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 s'appuie sur le lemme suivant :

Lemme 4.19. *Soit g une fonction dérivable sur $\mathbb{R}$ telle qu'il existe $A > 0$ et $c > 0$ tels que*

$$\forall x \in \mathbb{R} \quad |g(x)| + |g'(x)| \le A\exp(+c|x|).$$

Si $X \sim \mathcal{N}(0,1)$, alors $g'(X)$ et $Xg(X)$ sont intégrables et on a

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

Démonstration. Il est facile de vérifier que

$$(g(x)f(x))' = (g'(x) - xg(x))f(x).$$

On a donc pour tout $a, b \in \mathbb{R}$,

$$g(b)f(b) - g(a)f(a) = \int_a^b g'(x)f(x)\, dx - \int_a^b xg(x)f(x)\, dx.$$

Les hypothèses faites sur g et g' assurent l'intégrabilité sur $\mathbb{R}$ de $g'f$ et gf. Comme, de plus $\lim\limits_{a\to-\infty} g(a)f(a) = \lim\limits_{b\to+\infty} g(b)f(b) = 0$, on en déduit que

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

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

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

$$\phi'_X(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-x^2/2) ix \exp(itx)\, dx = i\mathbb{E}[X g_t(X)].$$

Mais d'après la formule d'intégration par parties de la variable gaussienne (vue au lemme 4.19), 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 $\phi_Y(t) = \mathbb{E}e^{itY} = e^{it(\sigma X+m)} = e^{imt}\mathbb{E}e^{it\sigma X} = e^{imt}\phi_X(\sigma t) = e^{imt}e^{-\frac{\sigma^2 t^2}{2}}$. □

Loi de Cauchy

Théorème 4.20. *La fonction caractéristique de la loi de Cauchy $\mathcal{C}(a, b)$ est*

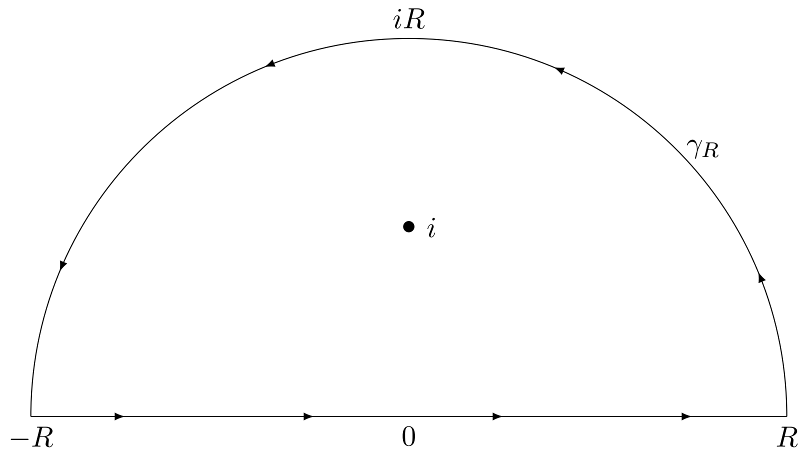
$$t \mapsto \phi(t) = e^{iat}e^{-b|t|}.$$

Démonstration. Rappelons que la loi de Cauchy $\mathcal{C}(a, b)$ admet comme densité par rapport à la mesure de Lebesgue :

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

On va d'abord calculer la fonction caractéristique de $\mathcal{C}(0, 1)$.

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 $\mathrm{Res}_i f(z) = \frac{e^{-t}}{2i}$. 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

$$\phi_Y(t) = \mathbb{E}e^{itY} = e^{it(bX+a)} = e^{iat}\mathbb{E}e^{itbX} = e^{iat}\phi_X(bt) = e^{iat}e^{-b|t|}.$$

□

4.3 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 4.21. *$\mathcal{L}_\mu$ 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 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 B.3. 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.

Théorème 4.22 (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 renvoie à Garet–Kurtzmann pour la preuve. □

4.4 Exercices sur les fonctions génératrices et les fonctions caractéristiques

4.4.1 Exercices de la série 1

Exercice 47. 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 $X_1+\cdots+X_n$.
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 $X_1^2+\cdots+X_n^2$ (appelée loi du χ^2).

Exercice 48. 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 $S_n=\sum\limits_{j=1}^{n} X_j$, 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 $\mathbb{E}[S]=mt$.

Exercice 49. *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.

Exercice 50. 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.

Exercice 51. *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 $\mathbb{P}(S_{2n}=0)\sim\dfrac{1}{\sqrt{\pi n}}$.

Exercice 52. 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.

4.4.2 Exercices de la série 2

Exercice 53.

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.

Exercice 54. 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.

Exercice 55.

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.

Exercice 56. 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

$$\forall k \in \{0, \ldots, n\}, \quad \mathbb{P}(X_2 = k) = \binom{n}{k} \frac{1}{3^k} \left(\frac{3}{4}\right)^n.$$

Exercice 57. Montrer que la convolée de deux lois de Cauchy est une loi de Cauchy.

Exercice 58. 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.

Exercice 59. Soit $X \sim \mathcal{N}(0,1)$. Montrer que $\forall n \geq 0, \quad \mathbb{E}[X^{2n}] = \frac{(2n)!}{n!2^n}$.

Exercice 60. 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.

Exercice 61. 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.

Exercice 62. 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.

Exercice 63. 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.

Exercice 64. 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

$$\forall t \in \mathbb{R} \backslash 2\pi\mathbb{Z} \quad \phi_Y(t) = \frac{p^2 e^{it}}{(1 - (1-p)e^{it})^2}.$$

5. Montrer que dans ce cas, $Z + 1$ suit la loi géométrique de paramètre p.

Chapitre 5

Convergences, lois des grands nombres

5.1 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 5.1. *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 2 les liens entre convergence dans L^p et convergence presque partout. Ainsi, d'après le théorème 2.12, 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 21.

5.1.1 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

$$\overline{\lim_{n\to+\infty}} x_n \le M \iff \forall \varepsilon > 0 \quad \{n : x_n \ge M + \varepsilon\} \text{ est fini} \tag{5.1}$$

et

$$\overline{\lim_{n\to+\infty}} x_n \ge M \iff \forall \varepsilon > 0 \quad \{n : x_n \ge M - \varepsilon\} \text{ est infini.} \tag{5.2}$$

5.1.2 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 5.2 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

$$\left\{ \overline{\lim_{n\to+\infty}} X_n \ge M \right\} = \bigcap_{\varepsilon\in\mathbb{Q}_+^*} \overline{\lim_{n\to+\infty}} \{X_n \ge M - \varepsilon\}. \tag{5.3}$$

L'avantage est que l'intersection est maintenant dénombrable. Or, on a le résultat classique très utile suivant.

Théorème 5.2. *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 $\varlimsup_{n\to+\infty} X_n \leq M$ presque sûrement, il suffit donc de prouver que $\forall a > M, \quad \mathbb{P}\left(\varliminf_{n\to+\infty} \{X_n < a\}\right) = 1$ ou, de manière équivalente, que $\forall a > M, \quad \mathbb{P}\left(\varlimsup_{n\to+\infty} \{X_n \geq a\}\right) = 0$.

On peut donc énoncer le théorème suivant.

Théorème 5.3. *Soient X_n une suite de variables aléatoires et M un réel. On suppose que*

1. $\forall a < M, \quad \mathbb{P}\left(\varlimsup_{n\to+\infty} \{X_n \geq a\}\right) = 1,$
2. $\forall a > M, \quad \mathbb{P}\left(\varlimsup_{n\to+\infty} \{X_n \geq a\}\right) = 0.$

Alors, $\varlimsup_{n\to+\infty} X_n = M$ *presque sûrement.*

Le théorème suivant très important en est une application directe

Théorème 5.4 (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*

$$\forall \varepsilon > 0 \quad \mathbb{P}\left(\varlimsup_{n\to+\infty} \{|X_n - X| \geq \varepsilon\}\right) = 0.$$

Démonstration. Il suffit d'appliquer le lemme précédent à la suite de variables aléatoires $(|X_n - X|)_{n\geq 0}$, avec $M = 0$ et $a = \varepsilon$. □

5.2 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 $\mathbb{P}(|(X_n + Y_n) - (X + Y)| \geq \varepsilon) \leq \mathbb{P}(|X_n - X| \geq \varepsilon/2) + \mathbb{P}(|Y_n - Y| \geq \varepsilon/2)$.

5.2.1 Comparaison avec les autres modes de convergence

Convergence dans L^p et convergence en probabilité

Théorème 5.5. *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 66).

Convergence presque sûre et convergence en probabilité

Théorème 5.6. *La convergence presque sûre implique la convergence en probabilité.*

Démonstration. Soit $\varepsilon > 0$. D'après le théorème 5.4, 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 5.7. *Il y a une différence fondamentale entre la convergence presque sûre et les autres modes de convergence évoqués ici : pour une suite $(X_n)_{n\geq1}$, 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.*

5.2.2 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 5.8. *Soit $(X_n)_{n\geq0}$ 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

$$\mathbb{E}\left(|M_n - \mathbb{E}X_0|^2\right) = \frac{\operatorname{Var} X_1}{n}, \tag{5.4}$$

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.

5.3 Lemmes de Borel-Cantelli

5.3.1 Premier lemme de Borel–Cantelli

Théorème 5.9. *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 5.10. — *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.*

5.3.2 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 5.11 (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

$$\mathbb{P}\left(\overline{\lim_{n\to\infty}}\ A_n\right) = 1.$$

Remarque 5.12. *On a* $\overline{\lim\limits_{n\to\infty}}\ A_n = \{N = +\infty\}$.

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 5.13 (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* $\mathbb{P}\left(\overline{\lim\limits_{n\to+\infty}}\ A_n\right) = 1$.

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 5.14. *La propriété "*$\mathbb{P}\left(\overline{\lim}_{n\to+\infty} A_n\right) = 1$*" 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 5.15 (loi 0–1 de Borel)**.** *Soit $(A_n)_{n\geq 1}$ une suite d'événements indépendants. La probabilité de l'événement $\mathbb{P}\left(\overline{\lim}_{n\to+\infty} A_n\right)$ 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 5.16. *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

$$\mathbb{P}\left(\underline{\lim}_{k\to+\infty} \{|X_{n_k} - X| < \frac{1}{k}\}\right) = 1.$$

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 5.17. *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 $f(X)$.*

1. C'est-à-dire que $\mathbb{P}(A_i \cap A_j) \leq \mathbb{P}(A_i)\mathbb{P}(A_j)$ pour $i \neq j$.

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 $\lim_{k\to+\infty} u_{n_k} = \overline{\lim}_{n\to+\infty} u_n = 0.$ □

5.4 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 ».

5.4.1 Deux lois fortes des grands nombres

Théorème 5.18. *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 5.19. *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*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{\text{p.s.}} \mathbb{E}X_1.$$

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 5.20 (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*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{\text{p.s.}} \mathbb{E}X_1.$$

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 5.21. *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.*

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.

5.4.2 Probabilités et fréquences asymptotiques

Théorème 5.22. *Soit $(A_n)_{n\ge 0}$ 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 $A_1, \dots, A_n$. 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 5.20. X_k admet bien un moment d'ordre 1 car $0 \le X_k \le 1$ et l'on a $\mathbb{E}X_1 = \mathbb{P}(A_1) = p$. □

5.4.3 Exercice : une preuve de la loi forte des grands nombres

La preuve, désormais classique, présentée ici sous forme d'exercice, est due à Etemadi [2]. 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 [1] comment elle peut être utilisée pour généraliser le théorème 5.18 à certaines familles de variables aléatoires dépendantes.

Soit $(X_n)_{n\ge 1}$ 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 : $X_i^* = X_i \mathbb{1}_{\{X_i \le i\}}$ et les sommes et quotients associés : $S_n^* = \sum_{i=1}^{n} X_i^*$ et $Q_n^* = S_n^*/n$.

1. Montrer que
$$\operatorname{Var} S_n^* \le \sum_{i=1}^{n} \mathbb{E}\left((X_i^*)^2\right).$$
En déduire que
$$\operatorname{Var} S_n^* \le n\mathbb{E}[X_1^2 \mathbb{1}_{X_1 \le n}].$$
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
$$\forall N \ge 1 \quad \sum_{n=N}^{+\infty} \frac{1}{u_n} \le C \frac{1}{u_N}.$$

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 $u_n \leq k \leq u_{n+1}$, montrer que
$$\frac{u_n}{u_{n+1}} Q_{u_n} \leq Q_k \leq \frac{u_{n+1}}{u_n} Q_{u_{n+1}}.$$
14. En déduire que
$$\frac{1}{\beta}\mathbb{E}[X_1] \leq \varliminf_{k\to+\infty} Q_k \leq \varlimsup_{k\to+\infty} Q_k \leq \beta\mathbb{E}[X_1] \text{ p.s.}$$
15. On note
$$\Omega_\beta = \left\{ \omega \in \Omega ; \frac{1}{\beta}\mathbb{E}[X_1] \leq \varliminf_{k\to+\infty} Q_k(\omega) \leq \varlimsup_{k\to+\infty} Q_k(\omega) \leq \beta\mathbb{E}[X_1] \right\}.$$
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 $u_N \sum\limits_{n=N}^{+\infty} \frac{1}{u_n}$ 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. $\mathbb{E}X_n^* = \mathbb{E}X_n\mathbb{1}_{\{X_n\le n\}} = \mathbb{E}X_1\mathbb{1}_{\{X_1\le n\}}$, car X_1 et X_n ont même loi. Comme la suite $(X_1\mathbb{1}_{\{X_1\le n\}})_{n\ge 1}$ 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. $\mathbb{P}(X_n \ne X_n^*) = \mathbb{P}(X_n > n) = \mathbb{P}(X_1 > n)$. Pour tout $n \ge 0$, on a $\mathbb{P}(X_1 > n) \le \int_{]n-1,n]} \mathbb{P}(X_1 > t)\, d\lambda(t)$. Ainsi
$$\sum_{n=1}^{+\infty} \mathbb{P}(X_n \ne X_n^*) = \sum_{n=1}^{+\infty} \mathbb{P}(X_1 > n) \le \int_0^{,+\infty} \mathbb{P}(X_1 > t)\, dt \le \mathbb{E}X_1 < +\infty.$$
11. Ainsi, avec le lemme de Borel-Cantelli, $\mathbb{P}\left(\overline{\lim}_{n\to+\infty} \{X_n \ne X_n^*)\}\right) = 0$, 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 $n \ge n_0(\omega)$, 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 $u_n \ge n_0(\omega)$
$$Q_{u_n} - \mathbb{E}X_1 = \frac{S_{u_{n_0-1}} - S^*_{u_{n_0-1}}}{u_n} + \left(Q_{u_n}^* - \mathbb{E}X_1^*\right).$$
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 $S_{u_n} \le S_k \le S_{u_{n+1}}$. On en déduit que
$$\frac{u_n}{u_{n+1}} Q_{u_n} \le \frac{u_n}{k} Q_{u_n} \le Q_k \le \frac{u_{n+1}}{k} Q_{u_{n+1}} \le \frac{u_{n+1}}{u_n} Q_{u_{n+1}}.$$
14. Si n_k désigne la partie entière du logarithme de n en base β, on a
$$\frac{u_{n_k}}{u_{n_k+1}} Q_{u_{n_k}} \le Q_k \le \frac{u_{n_k+1}}{u_{n_k}} Q_{u_{n_k+1}}.$$
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
$$\frac{1}{\beta}\mathbb{E}X_1 \le \underline{\lim}_{k\to+\infty} Q_k \le \overline{\lim}_{k\to+\infty} Q_k \le \beta\mathbb{E}X_1 \text{ p.s.}$$

15. On 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 $\mathbb{E}X_1^+ - \mathbb{E}X_1^- = \mathbb{E}X_1$.

5.5 Exercices sur la convergence presque sûre

5.5.1 Exercices de la série 1

Exercice 65. 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 5.18, montrer que $S_n/n^{3/2}$ converge presque sûrement vers 0.

Exercice 66. 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 ?

Exercice 67. 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 ?

Exercice 68. 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.

Exercice 69. 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” ?

Exercice 70. 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 $\overline{\lim_{n\to+\infty}} \frac{|X_n|}{n} \geq a$ 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.

Exercice 71. *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.

Exercice 72. *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

$$\omega = \sum_{i=1}^{\infty} a_i g^{-i} \quad \text{où} \quad a_i \in \{0,1,\ldots,g-1\}$$

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 $\{0,1,\ldots,g-1\}$. Plus précisément, on dit que ω est g-normal si pour tout $\ell \geq 1$ et quels que soient $b_1,\ldots,b_\ell$ compris entre 0 et $g-1$, on a

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{a_i=b_1, a_{i+1}=b_2, \ldots, a_{i+\ell-1}=b_\ell\}} = \frac{1}{g^\ell} \text{ p.s.}$$

(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

$$\omega = X_0(\omega) = \sum_{i=0}^{+\infty} \frac{A_i^g(\omega)}{g^{i+1}} \text{ avec } A_i^g \in \{0,1,\ldots,g-1\}$$

   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 $\{0,\ldots,g-1\}\times[0,1[$ qui à x associe $G(x) = (\lfloor gx \rfloor, \{gx\})$. Montrer que la loi image de $\mathbb{P}$ par G est $\mathcal{U}(\{0,\ldots,g-1\}) \otimes \mathbb{P}$.
   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 71, 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 $g \geq 2$.

Exercice 73. *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 $U_i = \sum\limits_{k=1}^{+\infty} \frac{X_{2^k(2i+1)}}{2^k}$?
4. Soit (μ_i) une suite de mesures de probabilités sur $\mathbb{R}$. En utilisant le théorème 1.11 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.

Exercice 74. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant une loi exponentielle de paramètre 1. Calculer $\overline{\lim\limits_{n\to+\infty}} \frac{X_n}{\log n}$.

Exercice 75. 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.

Exercice 76. *Lemme de Kochen–Stone [5].*

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 $\sum\limits_{n=1}^{+\infty} \mathbb{P}(B_n) = +\infty$.
   On pose $N_n = \sum\limits_{k=1}^{n} \mathbb{1}_{B_k}$ et on rappelle que
   $\{\overline{\lim\limits_{n\to+\infty}} B_n\} = \{\lim\limits_{n\to+\infty} N_n = +\infty\}$. 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 ?

Exercice 77. 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 $(a_n X_n)_{n\geq 1}$ tend en probabilité vers zéro. Montrer que $(a_n)_{n\geq 1}$ tend vers zéro. Montrer que la réciproque est aussi vraie.

Exercice 78. 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) $\lambda_n = 2\log(n+1)$.

Exercice 79. *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 $\sum_{n=1}^{+\infty} \lambda_n^2 < +\infty$. Montrer que

$$\overline{\lim_{n\to+\infty}} X_n \leq 1 \text{ p.s.}$$

   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 $\sum_{n=1}^{+\infty} \lambda_n^k < +\infty$, 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 $(X_n)_{n\geq 1}$.

5.5.2 Exercices de la série 2

Exercice 80. 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 $\lim_{n\to+\infty} m_n = \ell$.

Exercice 81. 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.

Exercice 82. 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 $N_n \geq n/3$.

Exercice 83. 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.

Exercice 84. 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

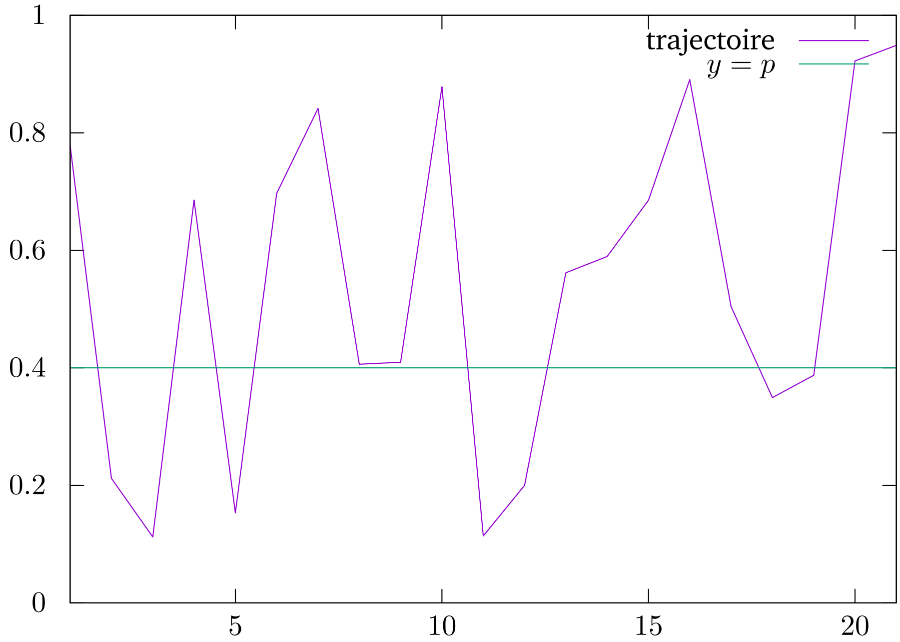
$$\lim_{n\to+\infty} \frac{X_n}{(\log n)^{\frac{3}{2}}} = 0.$$

Exercice 85. 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

$$\overline{\lim_{n\to+\infty}} \frac{X_n}{\sqrt{2\log n}} + \overline{\lim_{n\to+\infty}} \frac{Y_n}{\sqrt{2\log n}} \quad \text{et} \quad \overline{\lim_{n\to+\infty}} \frac{(X_n + Y_n)}{\sqrt{2\log n}}.$$

Exercice 86. 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 $(0, U_0), (1, U_1), (2, U_2), \ldots$ 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.

Exercice 87. 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 $(x, y) \in \mathbb{R}^2 \backslash \{(0,0)\}$. On pose

$$Y_n = A_n \begin{pmatrix} x \\ y \end{pmatrix}.$$

Montrer que

$$\|Y_n\|^{1/n} \xrightarrow{\text{p.s.}} \begin{cases} \sqrt{2} & \text{si } x + y = 0 \\ \sqrt{12} & \text{si } x + y \neq 0. \end{cases}$$

Chapitre 6

Convergence en loi

6.1 Convergence en loi

6.1.1 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]

$$\lim_{n\to+\infty} \int_{\mathbb{R}^d} f\, d\mu_n = \int_{\mathbb{R}^d} f\, d\mu. \tag{6.1}$$

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.

Théorème 6.1. *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 6.2. *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. 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 (6.1) est vérifiée pour les fonctions continues bornées, et de convergence vague lorsque (6.1) 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.

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 $(x, y) \mapsto \langle x, y\rangle$. □

6.1.2 Premiers exemples

Un critère de convergence en loi

Théorème 6.3 (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) $\lim_{n\to\infty} \int_\Omega f_n \, d\mu(x) = \int_\Omega f \, d\mu(x)$.

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 $0 \leq f - \min(f, f_n) \leq f$ 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 6.4. *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Soient ν et $(\nu_n)_{n\geq 1}$ des mesures de probabilité sur $(\Omega, \mathcal{A}, \mu)$ admettant les densités f et $(f_n)_{n\geq 1}$ par rapport à μ. On suppose que $f_n \to f$ μ–p.p. Alors (ν_n) converge faiblement vers ν.*

Démonstration. Soit g une fonction continue bornée sur Ω. On a

$$\begin{aligned} \left| \int g(x) \, d\nu_n(x) - \int g(x) \, d\nu(x) \right| &= \left| \int g(x) f(x) \, d\mu(x) - \int g(x) f_n(x) \, d\mu(x) \right| \\ &\leq \int |g(x)(f(x) - f_n(x))| \, d\mu(x) \\ &\leq \|g\|_\infty \int |f(x) - f_n(x)| \, d\mu(x), \end{aligned}$$

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 6.5. *Soient X et $(X_n)_{n\geq 1}$ 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 $(X_n)_{n\geq 1}$ converge en loi vers X.

Démonstration. Il suffit de remarquer que $\mathbb{P}_X$ et $\mathbb{P}_{X_n}$ pour $n \geq 1$ 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 6.6. *Soit, pour $n \geq 1$, 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 6.5, 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 $\binom{n}{k} = \frac{n(n-1)\dots(n-k+1)}{k!} \sim \frac{n^k}{k!}$ 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 6.7. *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, on peut montrer 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 6.8. *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 6.5, 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. □

6.1.3 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 6.9. *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 [6].

— Pour voir que $(3) \iff (4)$, il suffit de remarquer que

$$\begin{aligned}
&\sup_{O \text{ ouvert}} \left(\mu(O) - \underline{\lim}_{n\to+\infty} \mu_n(O)\right) \\
= &\sup_{F \text{ fermé}} \left(\mu(F^c) - \underline{\lim}_{n\to+\infty} \mu_n(F^c)\right) \\
= &\sup_{F \text{ fermé}} \left(1 - \mu(F) - \underline{\lim}_{n\to+\infty} (1 - \mu_n(F))\right) \\
= &\sup_{F \text{ fermé}} \left(-\mu(F) - \underline{\lim}_{n\to+\infty} - \mu_n(F)\right) \\
= &\sup_{F \text{ fermé}} \left(-\mu(F) + \overline{\lim}_{n\to+\infty} \mu_n(F)\right).
\end{aligned}$$

Ainsi, si l'un des supremum est négatif, alors l'autre l'est aussi.

— Preuve de $(2) \implies (3)$. 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

$$\overline{\lim}_{n\to+\infty} \int G_\varepsilon \circ d_F \, d\mu_n \geq \overline{\lim}_{n\to+\infty} \int \mathbb{1}_F \, d\mu_n.$$

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

$$\overline{\lim}_{n\to+\infty} \int G_\varepsilon \circ d_F \, d\mu_n = \int G_\varepsilon \circ d_F \, d\mu,$$

d'où

$$\forall \varepsilon > 0, \quad \int G_\varepsilon \circ d_F \, d\mu \geq \overline{\lim}_{n\to+\infty} \mu_n(F).$$

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 $(3, 4) \implies 5$. On a $\subset A \subset \overline{A}$, d'où

$$\begin{aligned}
\overline{\lim}_{n\to+\infty} \mu_n(A) &\leq \overline{\lim}_{n\to+\infty} \mu_n(\overline{A}) \leq \mu(\overline{A}), \\
\text{et} \quad \underline{\lim}_{n\to+\infty} \mu_n(A) &\geq \underline{\lim}_{n\to+\infty} \mu_n() \geq \mu().
\end{aligned}$$

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

$$\begin{aligned} \forall N \in \mathbb{N}, \quad \int_{\mathbb{R}^d} |f_N(x) - f(x)| d\mu(x) &\leq \omega_f(1/N) + 2\|f\|_\infty \frac{\varepsilon}{2}, \\ \forall N \in \mathbb{N}, \quad \overline{\lim_{n \to \infty}} \int_{\mathbb{R}^d} |f_N(x) - f(x)| d\mu_n(x) &\leq \omega_f(1/N) + 2\|f\|_\infty \varepsilon. \end{aligned}$$

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,

$$f_{n_0}(x) = \sum_{y \in \left(t + \frac{1}{n_0}\mathbb{Z}^d\right) \cap \bar{B}_p} f(y) \ \mathbb{1}_{\left\{y + \left]-\frac{1}{n_0}, 0\right]^d\right\}}(x).$$

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.

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 6.10. *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 $F_X(a) = \lim_{n\to+\infty} F_{X_n}(a)$ et

$F_X(b) = \lim_{n\to+\infty} F_{X_n}(b)$, 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 6.11. *Cette dernière conséquence de la convergence en loi est très utile, par exemple en statistique (voir le chapitre 7).*

Le théorème suivant est également très utile.

Théorème 6.12. *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

$$\begin{aligned} \int f\, d\mu_n - \int f\, d\mu &= \left(\int f h_{2A}\, d\mu_n - \int f h_{2A}\, d\mu\right) \\ &\quad + \left(\int f(1 - h_{2A})\, d\mu_n - \int f(1 - h_{2A})\, d\mu\right). \end{aligned}$$

Ainsi pour $n \geq n_0$, on a

$$\left|\int f\, d\mu_n - \int f\, d\mu\right| \leq \left|\int f h_{2A}\, d\mu_n - \int f h_{2A}\, d\mu\right| + 3\varepsilon \|f\|_\infty.$$

Comme $f h_{2A}$ est une fonction continue positive à support compact, la suite des intégrales $\int f h_{2A}\, d\mu_n$ converge vers $\int f h_{2A}\, d\mu$, 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. □

6.1.4 Lien avec les autres modes de convergence

Théorème 6.13. *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 5.16) 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,

$$\overline{\lim_{n\to\infty}}\ \mathbb{P}_{X_n}(F) \leq \delta_a(F) = 0.$$

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 6.14. *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 6.1 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 $\varlimsup_{n \to +\infty} |\mathbb{E}f(V_n) - \mathbb{E}f(U)| \leq \varepsilon$. Comme ε est quelconque, on en déduit le résultat cherché. □

6.2 Convergence en loi sur $\mathbb{R}^d$ grâce aux fonctions caractéristiques

6.2.1 Critère de convergence

Théorème 6.15 (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 dépasse le cadre de ce cours. On pourra la trouver dans Garet–Kurtzmann.

6.2.2 Théorème de continuité de Lévy

Théorème 6.16 (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.

6.2.3 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 6.17. *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 6.18. *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 6.17, (X_n, Y_n) converge en loi vers (X, Y), et donc d'après le corollaire 6.2, $X_n + Y_n$ 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. □

6.3 Théorème central limite en dimension 1

En dimension 1, le théorème s'énonce comme suit.

Théorème 6.19. *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 $X_1 - m, \ldots, X_n - m$ 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 4.17 :

$$\phi(x) = 1 - \frac{\sigma^2}{2}x^2 + o(x^2). \tag{6.2}$$

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 (6.2), $\phi(\frac{t}{\sqrt{n}}) = 1 - \frac{\sigma^2}{2n}t^2 + o(1/n)$. 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 6.20. *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*

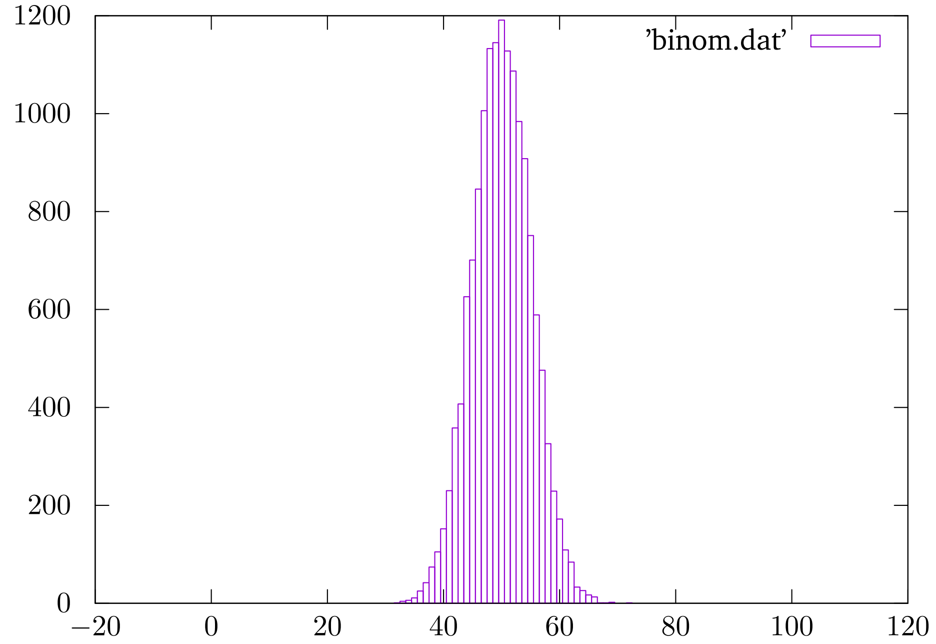
$$N_n = \sum_{k=1}^{n} \mathbb{1}_{A_k}$$

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 $\mathcal{B}(100, \frac{1}{2})$.



6.4 Exercices sur la convergence en loi

6.4.1 Exercices de la série 1

Exercice 88. Soit $(U_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes de même loi $\mathcal{N}(0,1)$. Pour tout $\theta \in \mathbb{R}$, on définit la suite $(X_n)_{n\geq 0}$ par la relation de récurrence $X_n = \theta X_{n-1} + U_n$ pour tout $n \geq 1$ 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 $(X_n)_{n\geq 0}$.

Exercice 89. Soit X_n suivant la loi uniforme sur $\{0, \ldots, n-1\}$. Montrer que X_n/n converge en loi vers la loi uniforme sur $[0,1]$.

Exercice 90. 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 $\int_0^{+\infty} |\sin t| e^{-nt}\, dt$.

Exercice 91. 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 $M_n = \max\limits_{1\leq j\leq n} X_j$.
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.

Exercice 92. *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 :
$$n!\sim\sqrt{2\pi}n^{n+1/2}e^{-n}.$$

Exercice 93. *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 $|f(x)|\leq\min\left(\frac{2}{x^{1+\alpha}},\frac{1}{x^\alpha},\frac{1}{2}\frac{1}{x^{\alpha-1}}\right)$.)
3. Soit α tel que $0<\alpha\leq 2$. Montrer qu'il existe une mesure m_α dont la fonction caractéristique est la fonction $\theta\mapsto\exp(-|\theta|^\alpha)$.
4. Soient $X_1,\dots,X_n$ 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.

Exercice 94. Soit (X_n) une suite de variables aléatoires convergeant en loi vers une variable aléatoire X.

1. Montrer que la famille $(X_n)_{n\geq 1}$ est tendue.
2. On pose $\phi_n = \phi_{X_n}$. Montrer que la famille $(\phi_n)_{n\geq 1}$ est uniformément équicontinue, c'est-à-dire que

$$\forall\varepsilon > 0\,\exists\eta > 0\,\forall n \geq 1,\ \forall(x,y)\in\mathbb{R}^2,\ |x-y| \leq \eta \Longrightarrow |\phi_n(x)-\phi_n(y)| \leq \varepsilon.$$

Exercice 95. *Renouvellement.*
Soit $(X_n)_{n\geq 1}$ 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 $N_t = \inf\{n \geq 1; S_n \geq t\}$.
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 $+\infty$).

Exercice 96.

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.

Exercice 97. 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.

Exercice 98. 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 $\mathbb{E}(|S_n|) \sim A\sqrt{n}$.

Exercice 99. *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 $N \geq 1$, $\mathbb{P}(N|X_n) \to \mathbb{P}(N|X)$.
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 $(\nu_{p_1}(X_n), \dots, \nu_{p_N}(X_n))_{n\geq 1}$ converge en loi vers le vecteur $(\nu_{p_1}(X), \dots, \nu_{p_N}(X))$. 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 $\{1, \dots, n\}$. 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 [4] une extension des résultats de cet exercice à l'anneau $\mathbb{Z}[i]$.

6.4.2 Exercices de la série 2

Exercice 100. 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.

Exercice 101. 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 $+\infty$.

Exercice 102. 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

$$\lim_{n\to+\infty} P(D_n \geq 0) = \frac{1}{2}.$$

Exercice 103. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires positives convergeant en loi vers X. Montrer que

$$\mathbb{E}[X] \leq \underline{\lim}_{n\to+\infty} \mathbb{E}[X_n].$$

Indication : on pourra utiliser le fait que $\mathbb{E}[X_n] = \int_{\mathbb{R}^+} \mathbb{P}(X_n > t)\, d\lambda(t)$.

Exercice 104. 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.

Chapitre 7

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 $(X_1, \dots, X_n)$ de n variables aléatoires sur $(\Omega, \mathcal{F})$, telles que pour tout $\theta \in \Theta$, les variables aléatoires $X_1, \dots, X_n$ forment une famille de variables aléatoires indépendantes identiquement distribuées.*

Remarque 7.1. *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 $\omega = (i_1, \dots, i_n) \in \Omega$ et $X(\omega) = S_n(\omega) = X_1(\omega) + \dots + X_n(\omega) = i_1 + \dots + i_n$, on a $\mathbb{P}_\theta(\omega) = \theta^{X(\omega)}(1 - \theta)^{n - X(\omega)}$ pour $\theta \in [0, 1]$.*

7.1 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 $f(\theta)$.

7.1.1 Lois empiriques

Si $X_1, \dots, X_n$ 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

$$F_n(x) = \mu_n(]-\infty, x]) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}.$$

   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 7.2. *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 7.2, 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

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n).$$

7.1.2 Théorème de Glivenko–Cantelli

Théorème 7.3 (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 $X^{(1)} \le X^{(2)} \le \cdots \le X^{(n)}$ (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

$$\{X^{(k)} \le t\} = \left\{ \sum_{i=1}^{n} \mathbb{1}_{[0,t]}(X_i) \ge k \right\}.$$

Pour $0 \le k \le n$, 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

$$\|F_n - F\|_\infty = \max\left(\max_{1\le k\le n} \left| \frac{k}{n} - F(X^{(k)}) \right|, \max_{1\le k\le n} \left| \frac{k-1}{n} - F(X^{(k)}-0) \right| \right),$$

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 1.11, 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 B.1 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 7.4. *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 7.5. *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).

7.1.3 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 7.6. *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 $\Omega = \{0,\dots,n\}$ 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 7.7. *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 $X_1, X_2, \ldots, X_n$.

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,

$$\mathbb{E}\left[\frac{1}{n}\sum_{k=1}^n X_k^2\right] = \mathbb{E}[X_1^2] = \mathrm{Var}\,X_1 + (\mathbb{E}X_1)^2 = \sigma^2 + m^2,$$

et d'autre part,

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{k=1}^n X_k\right)^2\right] = \frac{1}{n^2}\mathbb{E}\left[\left(\sum_{k=1}^n X_k\right)^2\right] = \frac{1}{n^2}\mathbb{E}\left(\sum_{k=1}^n X_k^2 + 2\sum_{1\le i<j\le n} X_iX_j\right).$$

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

$$\mathbb{E}\left[\left(\sum_{k=1}^{n} X_k\right)^2\right] = \sum_{k=1}^{n} \mathbb{E}\left[X_k^2\right] + 2\sum_{1\leq i<j\leq n} \mathbb{E}X_iX_j = n(\sigma^2+m^2) + n(n-1)m^2.$$

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 7.8. *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 $(X_1, \ldots, X_n)$ 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 7.9. *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 $\mu = \frac{1}{n} \sum\limits_{i=1}^{n} \delta_{a_i}$. Appliquons maintenant l'inégalité de Cauchy-Schwarz à la fonction identité. On obtient alors, car $\sum\limits_{i=1}^{n} a_i = 1$

$$\frac{1}{n^2} = \left(\frac{1}{n}\sum_{i=1}^{n} a_i\right)^2 = \left(\int_{\mathbb{R}} x \, d\mu(x)\right)^2 \leq \int_{\mathbb{R}} x^2 \, d\mu(x) = \frac{1}{n}\sum_{i=1}^{n} a_i^2.$$

On a donc $\sum\limits_{i=1}^{n} a_i^2 \geq \frac{1}{n}$ avec égalité si et seulement si $a_i = 1/n$ et dans ce cas, $T = \bar{X}_n$. □

7.2 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

$$\begin{aligned}\mathbb{P}_\theta\left(\left|\frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}}\right| \geq a\right) &= \mathbb{P}_\theta(\theta \in [T_n - \frac{a\sqrt{\theta(1-\theta)}}{\sqrt{n}}, T_n - \frac{a\sqrt{\theta(1-\theta)}}{\sqrt{n}}]) \\ &\geq \mathbb{P}_\theta(\theta \in [T_n - \frac{a}{2\sqrt{n}}\sqrt{n}, T_n - \frac{a}{2\sqrt{n}}])\end{aligned}$$

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

$$[T - 0,98/\sqrt{n}, T + 0,98/\sqrt{n}], \tag{7.1}$$

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

$$[T - \frac{1}{\sqrt{n}}, T + \frac{1}{\sqrt{n}}]. \tag{7.2}$$

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 $\left[\bar{X}_n - 1,96\frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1,96\frac{\sigma}{\sqrt{n}}\right]$.

7.3 Tests d'hypothèses

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* $\mathbb{P}_.(D) : \Theta \to [0,1]$, $\theta \mapsto \mathbb{P}_\theta(D)$ *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.

7.4 Exercices de statistiques

7.4.1 Exercices de la série 1

Exercice 105. 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 ?

Exercice 106. *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

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i) - \int_{[0,1]^d} g(x)\,dx\right| > \varepsilon\right) \leq 2\exp\left(-\frac{n\varepsilon^2}{4\|g\|_2^2}\right).$$

7.4.2 Exercices de la série 2

Exercice 107. 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 avec 5 comme écart-type. Pour cela, on prélève un échantillon dans chacune des deux populations, ce qui donne les résultats suivants :

	Effectif	Moyenne
Bons sauteurs	$n_1 = 55$	$\bar{X}_1 = 164$
Mauvais sauteurs	$n_2 = 50$	$\bar{X}_2 = 161,5$

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 $n_2 = 10$.

Exercice 108. 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.)

Annexe A

Rappels de dénombrement

A.1 Rappels de vocabulaire ensembliste

Un ensemble Ω est constitué de points, tous distincts. On dit qu'un ensemble A est inclus dans Ω, et l'on écrit$A \subset \Omega$, lorsque tous les éléments de A appartiennent à Ω.

On rappelle que l'ensemble vide (noté $\varnothing$) ne contient aucun élément et est inclus dans tous les ensembles. Pratiquement, si l'on veut montrer le résultat $A \subset \Omega$, la preuve ressemblera donc à « Soit $x \in A$ ... (raisonnement) ... donc $x \in \Omega$. Comme on a choisi x quelconque dans A, on conclut que $A \subset \Omega$. » Si A est inclus dans Ω, on dit que A est un sous-ensemble, ou une partie de Ω.

Si A et B sont des parties de Ω, l'ensemble $A \cup B$ est constitué des éléments de Ω qui sont dans A ou dans B, éventuellement dans les deux. Plus généralement, si I est un ensemble quelconque et $(A_i)_{i\in I}$ une famille de parties de Ω indexée par I, $\underset{i\in I}{\cup} A_i$ est constitué des points de Ω qui sont dans au moins un des A_i.

Pratiquement, si l'on veut montrer le résultat $x \in \underset{i\in I}{\cup} A_i$, la preuve ressemblera donc à « ... (raisonnement) ... Il existe donc $i_0 \in I$ tel que $x \in A_{i_0}$. Donc $x \in \underset{i\in I}{\cup} A_i$. »

Si A et B sont des parties de Ω, l'ensemble $A \cap B$ est constitué des éléments de Ω qui sont dans A et dans B. Plus généralement, si I est un ensemble quelconque et $(A_i)_{i\in I}$ une famille de parties de Ω indexée par I, $\underset{i\in I}{\cap} A_i$ est constitué des points de Ω qui sont dans tous les A_i.

Pratiquement, si l'on veut montrer le résultat $x \in \underset{i\in I}{\cap} A_i$, la preuve ressemblera donc à « Soit $i \in I$ ... (raisonnement) ... Donc $x \in A_i$. Comme i est quelconque, on a donc $x \in \underset{i\in I}{\cap} A_i$. »

A.2 Applications et cardinaux : définitions et notations

Pour A, D deux ensembles non vides quelconques, on note A^D ou $\mathcal{F}(D, A)$ l'ensemble des fonctions de D (ensemble de départ) vers A (ensemble d'arrivée). Soit f une application de D dans A. On dit que f est

— *injective* si $\forall x, y \in D, \quad x \neq y \Longrightarrow f(x) \neq f(y)$
— *surjective* si $\forall z \in A, \quad \exists x \in D : \quad f(x) = z$.
— *bijective* si elle est à fois injective et surjective.

Une application injective (resp. surjective, bijective) est une injection (resp. surjection, bijection).

Une bijection d'un ensemble Ω dans lui-même est appelée *permutation* de Ω. On note $\mathfrak{S}(\Omega)$l'ensemble des permutations de Ω, et simplement $\mathfrak{S}_n$ pour l'ensemble des permutations de $\{1, \dots, n\}$.

Un ensemble Ω est dit *fini* si
— ou bien c'est l'ensemble vide $\varnothing$,
— ou bien il existe un entier n tel qu'il existe une bijection entre Ω et $\{1, \dots, n\}$.

Cet entier n est unique : on l'appelle le *cardinal* de l'ensemble Ω. On le note $|\Omega|$. De manière intuitive, c'est le nombre d'éléments de Ω.
Le cardinal de l'ensemble vide est zéro.
Pour Ω fini de cardinal n, et $p \in \{0, \dots, n\}$, on note $\mathcal{B}_p(\Omega)$l'ensemble des parties de Ω de cardinal p. Par exemple $\mathcal{B}_2(\{a, b, c\}) = \{\{a, b\}, \{b, c\}, \{a, c\}\}$.
On note de plus $\mathcal{P}(\Omega)$l'ensemble des parties de Ω, quel que soit leur cardinal.
Par exemple $\mathcal{P}(\{a, b, c\}) = \{\varnothing, \{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{a, c\}, \{a, b, c\}\}$.

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 $A = \underset{i\in I}{\cup} A_i$ 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 $\Omega = \underset{i\in I}{\cup} \Omega_i$, on a

$$A = A \cap \Omega = \underset{i\in I}{\cup} (A\cap\Omega_i) = \underset{i\in I}{\cup} A_i.$$

D'autre part, pour $i \neq j$, on a $A_i \cap A_j \subset \Omega_i \cap \Omega_j = \varnothing$, d'où $A_i \cap A_j = \varnothing$. □

Lemme A.2. *Soit $\phi : D \to A$ une application surjective. Alors les ensembles $(\phi^{-1}(\{a\}))_{a\in A}$ forment une partition de D.*

La preuve de ce résultat est laissée en exercice au lecteur.

A.3.4 Lemme des bergers

Le lemme suivant peut également être utile

Lemme A.3 (des bergers)**.** *Soit ϕ une application* surjective *de D dans A. On suppose qu'il existe un entier $a \geq 1$ tel que*

$$\forall y \in A \quad |\phi^{-1}(\{y\})| = |\{x \in D; \phi(x) = y\}| = a$$

(autrement dit si tout élément de A admet exactement a antécédents), on a

$$|A| = \frac{|D|}{a}.$$

Démonstration. On applique le principe de partition avec $I = A$. Si l'on pose, pour $y \in A$, $D_y = \{x \in D; \phi(x) = y\}$, les D_y forment clairement une partition de D, d'où

$$|D| = \sum_{y\in A} |D_y| = \sum_{y\in A} a = |A|a.$$

□

Le nom du lemme est dû à la procédure prétendument employée par les bergers chaldéens pour compter le nombre de leurs moutons : il s'agit de compter le nombre de pattes et de diviser par 4. Dans cet exemple, A est l'ensemble des moutons, D l'ensemble des pattes de mouton, et ϕ l'application qui à une patte associe le mouton auquel elle appartient.

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.4 Quelques résultats incontournables

A.4.1 Nombre d'applications de D dans A

Il existe exactement $|A|^{|D|}$ applications de D dans A, ce qui peut s'écrire

$$|A^D| = |A|^{|D|}.$$

On pose $|A| = n$ et $|D| = p$. Un cas particulier important est celui où l'on a $D = \{1, \dots, p\}$. Or, un p-uplet $(x_1, x_2, \dots, x_p)$ dont les composantes sont des éléments de A peut être considéré comme la donnée d'une application de $\{1, \dots, p\}$ dans A. Le nombre de p-uplets $(x_1, x_2, \dots, x_p)$ dont les composantes sont des éléments de A est donc n^p.
Exemple: Un professeur note chaque étudiant d'une classe de 30 étudiants par une note entière de 0 à 20. Le nombre de résultats possibles est le nombre de fonctions de l'ensemble D des étudiants dans l'ensemble $A = \{0, \dots, 20\}$ des notes possibles. Comme $|A| = 21$ et $|D| = 30$, il y a donc 21^{30} résultats possibles.

Remarque A.4. *Au lycée, vous avez vu ce résultat sous la dénomination "choix indépendant (avec remise) de p objets dans un ensemble de cardinal $|A| = n$."*

A.4.2 Nombre de permutations de Ω

On pose $|\Omega| = n$. Le nombre de permutations de Ω est

$$n! = n \times (n-1) \times \dots \times 1.$$

Remarque A.5. $n!$ *se lit* "factorielle n" *ou* "n factorielle".

Exemple: Un professeur doit faire passer dans la journée cinq étudiants à l'oral de contrôle. Il a 5 !=120 manières de choisir l'ordre dans lequel il va les interroger.

A.4.3 Nombre d'injections de D dans A

Proposition A.6. *On pose $|A| = n$ et $|D| = p$. En vertu de la remarque faite en A.2, il existe une injection de D dans A si et seulement si $p \leq n$. Alors, le nombre d'injections de D dans A est*

$$n(n-1)\dots(n-p+1).$$

Démonstration. Soit n un entier. On pose $A = \{1, \dots, n\}$ et on note $\mathcal{I}_p$ l'ensemble des injections de $\{1, \dots, p\}$ dans A. On va montrer par récurrence sur $p \in \{1, \dots, n\}$ que $|\mathcal{I}_p| = \frac{n!}{(n-p)!}$. Il est évident que $|\mathcal{I}_1| = 1 = \frac{n!}{(n-1)!}$. Considérons l'application

$$R_p : \mathcal{I}_{p+1} \to \mathcal{I}_p$$

qui à chaque injection de $\{1, \dots, p+1\}$ dans A associe sa restriction à $\{1, \dots, p\}$. Avec un peu de réflexion, on montre que

$$\forall f \in \mathcal{I}_p \quad |\{g \in \mathcal{I}_{p+1}; R_p(g) = f\}| = n - p.$$

D'après le lemme des bergers, on a donc

$$|\mathcal{I}_{p+1}| = (n-p)|\mathcal{I}_p|.$$

Cette identité permet d'achever la preuve par récurrence. □

Remarque A.7. — *Comme on l'a vu dans la preuve, ce nombre peut s'écrire aussi* $\frac{n!}{(n-p)!}$.
— *Lorsque* $n = p$*, on trouve* $n!$*. En fait, une injection entre deux ensembles de même cardinal est une bijection.*

Exemple: 3500 personnes se présentent au concours de l'Agrégation de Mathématiques. 300 places sont mises au concours. Combien y a-t-il de palmarès possibles, en supposant qu'il n'y ait pas d'ex-æquos ?
Réponse : $3500\times 3499\times\cdots\times 3202\times 3201$. Ici D est l'ensemble des rangs, on a donc $D = \{1,\dots,300\}$ et A l'ensemble des candidats (donc $|A| = 3500$). On compte bien le nombre d'applications injectives puisqu'une même personne ne peut avoir deux rangs différents.

A.4.4 Nombre de parties de Ω possédant p éléments

Proposition A.8. *On pose* $|\Omega| = n$*. Par définition, on note* $\binom{n}{p}$ *le nombre de parties à* p *éléments d'un ensemble de* n *éléments. Il s'agit donc de calculer* $|\mathcal{B}_p(\Omega)|$*. On va montrer que*

$$\binom{n}{p} = \frac{n(n-1)\dots(n-p+1)}{p(p-1)\dots 1} = \frac{n!}{p!(n-p)!}.$$

Démonstration. Il suffit d'appliquer le lemme des bergers à
— D : ensemble des injections de $\{1,\dots,p\}$ dans Ω,
— $A = \mathcal{B}_p(\Omega)$,
— ϕ définie par $\phi(f) = \text{Image}(f) = \{f(k); k \in \{1,\dots,p\}\}$.

On a vu précédemment que $|A| = n(n-1)\dots(n-p+1)$. Il n'est pas difficile de voir que ϕ est surjective. Une partie $\{e_1,\dots,e_p\}$ de Ω étant donnée, combien existe-t-il d'injections (en fait de bijections) de $\{1,\dots,p\}$ dans Ω telles que $\{f(1),\dots,f(p)\} = \{e_1,\dots,e_p\}$? C'est évidemment le nombre d'injections de $\{1,\dots,n\}$ dans $\{e_1,\dots,e_p\}$, c'est-à-dire $p!$. Le lemme des bergers s'applique donc avec $a = p!$, d'où le résultat. □

Exemple: 3500 personnes se présentent au concours de l'Agrégation de Mathématiques. 300 places sont mises au concours. Combien y a-t-il de listes alphabétiques des reçus possibles ? Réponse : $\binom{3500}{300}$. Ici, Ω est l'ensemble des candidats et $p = 300$ le nombre de reçus.

A.4.5 Nombre total de parties de Ω

Proposition A.9. *Le nombre total de parties de* Ω *est* $|\mathcal{P}(\Omega)| = 2^{|\Omega|}$.

Démonstration. Il suffit de remarquer que l'application

$$\mathcal{P}(\Omega) \to \{0;1\}^{\Omega}, \quad A \mapsto \mathbb{1}_A$$

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

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

□

Exemple: 200 étudiants se présentent à un examen. Combien y a-t-il de listes alphabétiques des reçus possibles ? Réponse : 2^{200}. Ici Ω est l'ensemble des candidats. La grande différence avec l'exemple précédent est qu'ici, le nombre de reçus n'est pas fixé à l'avance.

A.5 Équations et inéquations en entiers

Lemme A.10. *Soient n et p des entiers. Si $n \geq p$, alors il existe exactement $\binom{n}{p}$ applications strictement croissantes de $\{1, \dots, p\}$ dans $\{1, \dots, n\}$. Sinon, il n'en existe aucune.*

Démonstration. Une application strictement croissante étant injective, il est nécessaire que $n \geq p$. Mais se donner une suite strictement croissante de p éléments pris dans $\{1, \dots, n\}$ revient à choisir une partie de $\{1, \dots, n\}$ possédant p éléments, puis à les ordonner avec l'ordre naturel. Or $\binom{n}{p}$ est, par définition, le nombre de parties de $\{1, \dots, n\}$ possédant p éléments, d'où le résultat. □

Exemple: Un enseignant devrait faire un cours de 70 pages en 7 séances. Combien y a-t-il de progressions possibles, en admettant qu'à chaque séance, l'enseignant progresse d'un nombre entier strictement positif de pages, mais sans être astreint à terminer le programme ?
Réponse : une progression correspond donc à une fonction strictement croissante de $\{1, \dots, 7\}$ dans $\{1, \dots, 70\}$ qui au numéro de chaque cours associe le numéro de la dernière page étudiée à ce cours. Il y a donc $\binom{70}{7}$ progressions possibles.

Théorème A.11. *Pour n et p des entiers vérifiant $n \geq p$, il existe exactement $\binom{n}{p}$ p-uplets $(x_1, \dots, x_p) \in (\mathbb{N} \backslash \{0\})^p$ solutions de l'inéquation :*

$$x_1 + x_2 + \cdots + x_p \leq n. \tag{A.1}$$

Démonstration. Il suffit de remarquer que l'application

$$(x_1, \dots, x_p) \quad \mapsto \quad (x_1, x_1 + x_2, x_1 + x_2 + x_3, \dots, x_1 + x_2 + \cdots + x_p)$$

réalise une bijection entre l'ensemble des solutions recherchées de l'inéquation et l'ensemble des suites strictement croissantes de p éléments à valeurs dans $\{1, \dots, n\}$. □

Théorème A.12. *Pour n, p des entiers vérifiant $n \geq p$, il existe exactement $\binom{n-1}{p-1}$ p-uplets $(x_1, \dots, x_p) \in (\mathbb{N} \backslash \{0\})^p$ solutions de l'équation*

$$x_1 + x_2 + \cdots + x_p = n. \tag{A.2}$$

Démonstration. Il suffit de remarquer que les solutions $(x_1, \dots, x_p) \in (\mathbb{N} \backslash \{0\})^p$ de l'équation (A.2) sont exactement les solutions de l'inéquation (A.2) qui ne sont pas solutions de l'inéquation

$$x_1 + x_2 + \cdots + x_p \leq n - 1. \tag{A.3}$$

Il y en a donc $\binom{n}{p} - \binom{n-1}{p} = \binom{n-1}{p-1}$. □

Théorème A.13. *Pour n, p des entiers positifs tels que $p \geq 1$, il existe exactement $\binom{n+p-1}{p-1}$ p-uplets $(x_1, \dots, x_p) \in \mathbb{N}^p$ solutions de l'équation*

$$x_1 + x_2 + \cdots + x_p = n. \tag{A.4}$$

Démonstration. Il suffit de remarquer que l'application

$$(x_1, \dots, x_p) \quad \mapsto (x_1 + 1, \dots, x_p + 1)$$

réalise une bijection entre les solutions $(x_1, \dots, x_p) \in \mathbb{N}^p$ de l'équation (A.4) et les solutions $(x_1, \dots, x_p) \in (\mathbb{N}\backslash\{0\})^p$ de l'équation

$$x_1 + x_2 + \dots + x_p = n + p. \tag{A.5}$$

□

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 $\binom{9+4-1}{4-1} = \binom{12}{3} = \frac{12\times 11\times 10}{1\times 2\times 3} = 220$ répartitions possibles.

Théorème A.14. *Soient n, p des entiers positifs.*
Il existe exactement $\binom{n+p}{p}$ p-uplets $(x_1, \dots, x_p) \in \mathbb{N}^p$ solutions de l'inéquation :

$$x_1 + x_2 + \dots + x_p \leq n. \tag{A.6}$$

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 $A_1, A_2, \dots, A_n$, on a

$$\begin{aligned}\left|\bigcup_{i=1}^{n} A_i\right| &= \sum_{B\in\mathcal{P}(\{1,\dots,n\})\backslash\varnothing} (-1)^{1+\text{Card}(B)} \left|\bigcap_{j\in B} A_j\right| \\ &= \sum_{i=1}^{n} |A_i| - \sum_{1\leq i_1<i_2\leq n} |A_{i_1}\cap A_{i_2}| + \dots + \\ &\quad + \dots + (-1)^{k+1} \sum_{1\leq i_1<i_2<\dots<i_k\leq n} |A_{i_1}\cap \ldots \cap A_{i_k}| + \dots + \\ &\quad + \dots + (-1)^{n+1} |A_1 \cap \ldots \cap A_n|.\end{aligned} \tag{A.7}$$

Exemple: Pour $n = 3$, on a

$$|A_1\cup A_2\cup A_3| = |A_1|+|A_2|+|A_3|-|A_1\cap A_2|-|A_2\cap A_3|-|A_1\cap A_3|+|A_1\cap A_2\cap A_3|.$$

On pourrait prouver la formule par récurrence sur n, mais c'est plutôt lourd. On préférera une preuve probabiliste (voir par exemple Garet–Kurtzmann).

A.7 Développement d'un produit de sommes

A.7.1 Développement d'un produit dans un anneau

Dans un anneau quelconque, on a l'identité très utile

$$\prod_{i=1}^{n} \left(\sum_{j=1}^{m} X_{i,j}\right) = \sum_{\phi\in\mathcal{F}(\{1,\dots,n\},\{1,\dots,m\})} \prod_{i=1}^{n} X_{i,\phi(i)}.$$

A.7.2 Formule du multinôme

En particulier si l'anneau est commutatif et si X_{ij} ne dépend pas de i, on a

$$\left(\sum_{j=1}^{m} X_j\right)^n = \sum_{\phi\in\mathcal{F}(\{1,\dots,n\},\{1,\dots,m\})} \prod_{i=1}^{n} X_{\phi(i)}.$$

Ainsi, si l'on note $\Psi^n(a_1,\dots,a_m)$, l'ensemble des applications ϕ de $\{1,\dots,n\}$ dans $\{1,\dots,m\}$ telles que pour tout $i\in\{1,\dots,m\}$, on a $|\phi^{-1}(i)| = a_i$ et si l'on note $\binom{n}{a_1,a_2,\dots,a_m} = |\Psi^n(a_1,\dots,a_m)|$, on obtient en regroupant les termes la formule du multinôme :

$$\left(\sum_{j=1}^{m} X_j\right)^n = \sum_{(a_1,\dots,a_m)} \binom{n}{a_1,a_2,\dots,a_m} \prod_{k=1}^{m} X_k^{a_k},$$

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 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

Rappels et compléments d'analyse

B.1 Analyse réelle

B.1.1 Le théorème de Dini-Polyà

Théorème B.1. *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 $|f_k(x) - f(x)| \leq \varepsilon$. □

B.1.2 Théorème de Helly

Théorème B.2 (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

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

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

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

On obtient finalement que $F(x) \leq \underline{\lim}_{k \to +\infty} F_{n_k}(x)$. De la même manière, on montre que $\overline{\lim}_{k \to +\infty} F_{n_k}(x) \leq F(x)$. Finalement, on a

$$F(x) \leq \underline{\lim}_{k \to +\infty} F_{n_k}(x) \leq \overline{\lim}_{k \to +\infty} F_{n_k}(x) \leq F(x),$$

ce qui prouve bien que $\lim_{k \to +\infty} F_{n_k}(x) = F(x)$. □

B.2 Intégration

B.2.1 Holomorphie d'une intégrale dépendant d'un paramètre

Théorème B.3. *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 $|\frac{\partial^n}{\partial^n z} f(x,z)| \leq \frac{n!}{r^n} g_{K+\overline{B}(0,r)}(x)$, 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 $|z - z_0| \leq r$, on a

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

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 $\mathbb{C}$. □

B.2.2 Intégration des fonctions radiales

Théorème B.4. *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*

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

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 $\mathbb{R}_+ = \overset{+\infty}{\underset{i=0}{\cup}} [0,i]$, le théorème 1.5 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 B.5. 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 B.6. *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 B.7. *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 $\|x\|_p^p = \sum\limits_{i=1}^n |x_i|^p$, en prenant $\phi(x) = \exp(-x^p)$ (exercice laissé au lecteur ; on trouvera comme volume $\frac{2^n \Gamma(\frac{1}{p}+1)^n}{\Gamma(\frac{n}{p}+1)}$).*

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.8. *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

$$\mathcal{C} = \left\{ A \in \mathcal{B}(\mathbb{R}^d) \middle| {}_{\text{tels que } F \subset A \subset O \text{ et } m(O \backslash F) < \eta}^{\forall \eta > 0, \exists O \text{ ouvert et } F \text{ fermé}} \right\}.$$

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 $\bigcap\limits_{n=1}^{+\infty} O_{1/n} = F$, on a donc $\bigcap\limits_{n \geq 1} \left(O_{1/n} \backslash F\right) = \varnothing$. 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 $m'_n(O_n) \leq m'_n(A \cap U_n) + \frac{\varepsilon}{3^d 2^{\|n\|_1}}$, soit encore $m(O_n) \leq m(A \cap U_n) + \frac{\varepsilon}{3^d 2^{\|n\|_1}}$. L'ensemble $O = \bigcup_{n \in \mathbb{Z}^d} O_n$ est ouvert, contient A et on a

$$m(O) \leq \sum_{n \in \mathbb{Z}^d} m(O_n) \leq \sum_{n \in \mathbb{Z}^d} m(A \cap U_n) + \frac{\varepsilon}{3^d 2^{\|n\|_1}} = m(A) + \varepsilon.$$

$\square$

Annexe C

Indications des exercices

C.1 Exercices sur les calculs de loi

Indication 1 On peut par exemple calculer la fonction de répartition.

Indication 2 Pour la première question, il suffit de considérer les probabilités des différentes issues. Pour la suite considérer le produit $X_1X_2X_3$.

Indication 3 On pourra poser $a = AB$, $b = BC$ et noter α l'angle formé par les droites (AB) et (BC). Alors X et Y suivent respectivement les lois uniformes sur $[0, a]$ et $[0, B]$. Un dessin est nécessaire.

Indication 4 On peut par exemple montrer que pour toute fonction H mesurable bornée, on a $\mathbb{E}[H(\{\frac{1}{X}\})] = \mathbb{E}[H(X)]$.

Indication 5 Dans les deux cas, on peut, au choix, d'abord déterminer la loi, ou utiliser le lien entre espérance et probabilité de queue.

Indication 6

1. Utiliser le théorème de calcul d'une mesure image par un C^1-difféomorphisme, ainsi que le théorème de transfert.
2. Choisir une fonction ϕ très simple de manière à retrouver le volume de la boule unité dans le membre de gauche.
3. Prendre encore une fonction ϕ appropriée, avec $p = 1$, et utiliser les symétries de la boule associée à la norme $\|\cdot\|_1$.

Indication 7

1. Appliquer le théorème de C^1-difféomorphisme, puis intégrer la densité trouvée par rapport à la variable à supprimer.
2. Se ramener au cas où T est le triangle de sommets $a = (1, 0, 0)$, $b = (0, 1, 0)$, $c = (0, 0, 1)$. L'aire du triangle est proportionnelle au volume du prisme de sommet $(0, 0, 0)$ et de base le triangle (a, b, V) ; ce volume s'exprime à l'aide d'un déterminant.

Indication 8

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 9 Utiliser le théorème de C^1-difféomorphisme.

Indication 10 Il y a au moins deux méthodes possibles :

— montrer par récurrence que pour tout n, $(Z_1, \dots Z_n)$ suit la loi $(\frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1})^{\otimes n}$.

— Remarquer que la loi de $(Z_1, \dots Z_n)$ est la loi image d'une loi uniforme par une bijection.

Indication 11 1. On notera que pour a et b dans $[0,1]$, $\mathbb{P}(X \le a, Z \le b) = \mathbb{P}(X \le a, X+Y \le b) + \mathbb{P}(X \le a, 1 \le\le X+Y \le 1+b)$. Pour la dernière question, on pourra calculer $\mathbb{P}(X \le 1/4, Y \le 1/4, Z \le 1/2)$.

2. On pourra montrer par récurrence que pour tout n, $(W_1, \dots W_n)$ suit la loi $U[0,1]^{\otimes n}$. Pour la suite on peut noter (démontrer) que $\{(i \neq j) \Longrightarrow W_i \neq W_j\}$ est de probabilité 1, puis utiliser les symétries.
3. Utiliser le lien entre espérances et probabilité de queue.

Indication 12 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 13 Calculer $\mathbb{P}(1 + \nu_{p_1}(X) > k_1, \dots, 1 + \nu_{p_n}(X) > k_n)$.

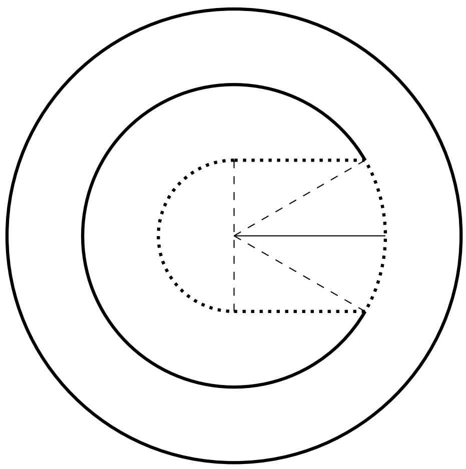
Indication 14 Poser $x = AM$ et résoudre l'inéquation.

Indication 15 Dire que le maximum de n nombres ne dépasse pas x revient à dire que chaque nombre ne dépasse pas x.

Indication 16 Remarquer que $1 - m_n = \max(1 - X_1, \dots, 1 - X_n)$.

Indication 17 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 18 1. On pourra remarquer que pour $i \neq j$, $X_i - X_j$ est une variable à densité.

2. Utiliser le théorème d'associativité des indépendances (ou lemme des coalitions), puis calculer la fonction de queue (ou la fonction de répartition).
3. Utiliser la première question.
4. Appliquer le théorème de Fubini.
5. On pourra par exemple calculer $\mathbb{P}_{(T,N)}$ sur des ensembles de type $]a, +\infty[\times]b, +\infty[$.

Indication 19 Commencer par calculer la fonction de répartition.

C.2 Exercices sur les espaces L^p

Indication 20

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 21 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 $\{s_n\} \geq x \geq \{s_{n+1}\}$. 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 22

1. Suivre l'indication...
2. Idem.
3. On applique la question précédente à $|f|\mathbb{1}_{\{|f|\leq n\}}$ (pourquoi est-elle intégrable ?)

Indication 23

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 24 On pourra raisonner par l'absurde.

Indication 25

1. Cherchez un peu plus...
2. Découper $\mathbb{R}$ en intervalles de longeur 2π.
3. Utiliser des équivalents.

Indication 26 Étudier d'abord la convergence ponctuelle.

Indication 27 On pourra utiliser des sous-suites.

Indication 28 Retrousser ses manches (ou équivalents).

Indication 29 Qu'est-ce que la mesure de comptage ? Qu'intégrer par rapport à la mesure de comptage ?

Indication 30 Majorer $\sqrt{|f^2 + g^2|}$ par une fonction intégrable.

Indication 31 Passer à l'intégrale de Riemann.

Indication 32

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 33 Utiliser l'inégalité de Cauchy-Schwartz.

Indication 34 Utiliser l'inégalité de Hölder.

Indication 35

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 36 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.3 Exercices sur la convolution et Fourier

Indication 37

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 38

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 39

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 40
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 41 Si $a = 0$ ou $b = 0$, alors $ab = 0$.

Indication 42
1. Ah bah non ! Vous l'avez déjà eue, l'indication.
2. Réduire A dans une base orthonormale.

Indication 43
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 44 On pourra utiliser la formule d'inversion.

Indication 45 On pourra à nouveau utiliser la formule d'inversion.

Indication 46 On pourra utiliser le théorème de Fubini.

C.4 Exercices sur les fonctions caractéristiques

Indication 47
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 48
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 49
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 50
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 51
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 52
1. Appliquer le théorème de Fubini.
2. Majorer le sinus.
3. Couper l'intégrale en deux.

Indication 53 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 54 Utiliser la première question de l'exercice précédent.

Indication 55 1. S'inspirer de l'exercice 48.

2. Commencer par déterminer la fonction génératrice de $K_1 L_1$. On pourra s'inspirer de l'exercice 53.

3. Relire le cours.

Indication 56 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 57 Penser à la fonction caractéristique.

Indication 58 Regarder la liste des fonctions caractéristiques connues, ou/et chercher l'équation fonctionnelle que doit vérifier ϕ_X.

Indication 59 Utiliser encore les liens entre fonction caractéristique et moments.

Indication 60 1. Utiliser les hypothèses.

2. Utiliser les liens entre fonction caractéristique et moments.

3. Idem.

Indication 61 1. Remarquer que Re $(1 - \exp(itX - i\theta)) \geq 0$.

2. Prendre t et t' rationnels différents.

Indication 62 On pourra donner un équivalent de la fonction caractéristique en l'origine.

Indication 63 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 64 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.5 Exercices sur la convergence presque sûre

Indication 65 1. Penser à l'inégalité de Chebychev.

2. Elle est dans l'énoncé.
3. Idem.

Indication 66 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 67 La série de terme général $\mathbb{P}(X_n \neq 0)$ converge-t-elle ?

Indication 68 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 69 Utiliser le second lemme de Borel–Cantelli pour une sous-suite d'événements indépendants.

Indication 70 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 71 1. Elles sont données dans l'énoncé.

2. Utiliser la question précédente.
3. Idem.

Indication 72 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 73 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 74 Penser à utiliser les deux lemmes de Borel-Cantelli pour calculer $\mathbb{P}\left(\overline{\lim_{n\to+\infty}}\ \{\frac{X_n}{\log n} \geq a\}\right)$ selon que $a > 1$ ou $a \leq 1$.

Indication 75 On remarque d'abord que pour tout entier a, on a :
$\{a$ est valeur d'adhérence de $X_n\} = \overline{\lim_{n\to+\infty}}\ \{X_n = a\}$. Puis on pense à la loi du $0-1$ de Borel.

Indication 76 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 77 On raisonne par l'absurde.

Indication 78 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 79 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 80 Considérer $S_n = X_1 + \dots + X_n$ où (X_n) est une suite de variables de Bernoulli de paramètre s, et observer la suite (x_{S_n}).

Indication 81 Appliquer la loi forte des grands nombres.

Indication 82 On pourra montrer que N_n/n tend presque sûrement vers $1/2$.

Indication 83 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 84 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 85 1. S'inspirer de l'exercice sur le calcul de $\overline{\lim_{n\to+\infty}}\ \frac{X_n}{\log n}$ 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 86 On a $T_n = \sum_{k=2}^{n} \mathbb{1}_{\{U_{k-1}<p\}\Delta\{U_k<p\}}$. 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 87 1. Passer au logarithme.

2. On peut écrire M_n sous la forme $M_n = PD_nP^{-1}$ et introduire la norme $\|x\|_* = \|P^{-1}x\|_\infty$.

C.6 Exercices sur la convergence en loi

Indication 88 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 89 Appliquer le théorème de Lévy.

Indication 90 1. On pourra commencer par étudier la convergence en probabilité.

2. Évaluer $\mathbb{E}(\phi(X_n))$ pour ϕ bien choisie.

Indication 91 1. Calculer la fonction de répartition.

2. Penser au lemme de Borel–Cantelli.
3. Calculer la fonction de répartition.

Indication 92 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 93 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 94 1. On pourra penser au théorème de Portmanteau.

2. Noter que $|e^{ixs} - e^{ixt}| \leq \min(|s-t||x|, 2)$.

Indication 95 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 96 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 $S_n = \sum\limits_{k=1}^{n} X_k$ 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 $\{1,\ldots,n\}$. L'espérance étudiée peut se calculer à l'aide de la fonction de queue.

Indication 97 On peut regarder de deux manières la suite $(Z_{2n}-Z_n)_{n\geq 1}$

Indication 98 Écrire $\mathbb{E}(\frac{|S_n|}{\sqrt{n}}$ à l'aide de la fonction de queue de $\frac{|S_n|}{\sqrt{n}}$.

Indication 99 1. (a) On peut noter que

$$\begin{aligned}\mathbb{1}_{\{Y^1=a_1,\ldots,Y^N=a_N\}} &= \prod_{k=1}^{N}(\mathbb{1}_{\{Y^k\geq a_k\}} - \mathbb{1}_{\{Y^k\geq a_k+1\}})\\ &= \prod_{k=1}^{N}\sum_{j=0}^{1}(-1)^j\mathbb{1}_{\{Y^k\geq a_k+j\}}\\ &= \sum_{(b_1,\ldots,b_N)\in\{0,1\}^N}\prod_{k=1}^{N}(-1)^{b_1+\cdots+b_N}\mathbb{1}_{\{\forall k\in\{1,\ldots,n\},Y^k\geq a_k+b_k\}}\end{aligned}$$

(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 100 Utiliser le premier théorème de Lévy.

Indication 101 1. Commencer par déterminer la loi de X_n^x.

2. Pour montrer la croissance, noter que $f(x) = \lim\limits_{n\to+\infty} \mathbb{E}[u\circ X_n^x]$. 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 102 Notons $M_{x,y} = 1$ si la pièce en (x,y) est face, -1 sinon. Si on note $C'_n = \sum\limits_{k\neq X} M_{k,X}$ et $L'_n = \sum\limits_{k\neq X} M_{X,k}$, on a

$$D_n = \sum_{(k,l)\in\{1,\ldots,n\}\backslash\{X\}} M_{k,l} + M_{X,X} + |C'_n - L'_n|.$$

Indication 103 Utiliser le lemme de Fatou.

Indication 104 On pourra déterminer ν_n telle que $\mu_n = \nu_n * \nu_n$.

C.7 Exercices sur les statistiques

Indication 105

1. On utilise la formule (7.2).
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 106 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 107 Utiliser l'hypothèse gaussienne.

Indication 108 Chercher encore un peu…

Annexe D

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

Bibliographie

[1] 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.

[2] N. Etemadi. An elementary proof of the strong law of large numbers. *Z. Wahrsch. Verw. Gebiete*, 55(1) :119–122, 1981.

[3] O. Garet and A. Kurtzmann. *De l'intégration aux probabilités*. Ellipses, Paris, 2nde edition, 2019.

[4] Olivier Garet. Les lois zêta pour l'arithmétique. *Quadrature*, (96) :10–18, 2015.

[5] Simon Kochen and Charles Stone. A note on the Borel-Cantelli lemma. *Illinois J. Math.*, 8 :248–251, 1964.

[6] J.-P. Portmanteau. Un espoir pour l'ensemble vide. *Annales de l'Université de Felletin*, 1915.

Index