Année universitaire 2006-2007

UNIVERSITÉ D'ORLÉANS

Olivier GARET

Intégration, Fourier et Probabilités

Table des matières

Chapitre 1

Un peu de théorie de la mesure

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

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

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

1.1 Tribus

1.1.1 Axiomes de base

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

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

1.1.2 Propriétés

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

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

– Pour toute suite $(A_i)_{1\leq i\leq n}$ d'éléments de $\mathcal{A}$, $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$.
– Pour toute suite $(A_i)_{1\leq i\leq n}$ d'éléments de $\mathcal{A}$, $\bigcap_{i=1}^{n} A_i \in \mathcal{A}$.

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

Exercice :

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

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

1.1.3 Sous-tribus

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

1.1.4 Opérations sur les tribus

Intersection de tribus

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

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

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

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

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

une tribu, $\underset{i\in I}{\cup} A_i \in \mathcal{A}$. Comme ceci est vrai pour tout $\mathcal{A} \in T$, on a $\underset{i\in I}{\cup} A_i \in \underset{A\in T}{\cap} \mathcal{A} = \mathcal{A}$.

□

Tribu engendrée par une famille de tribus

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

$$\underset{i\in I}{\wedge} \mathcal{A}_i.$$

Tribu engendrée par une famille d'ensembles

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

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

$$\underset{i\in I}{\wedge} \mathcal{A}_i.$$

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

1.1.5 Tribu borélienne, fonctions mesurables

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

Théorème 1 (Théorème fondamental de la mesurabilité). *Soit f une application quelconque d'un ensemble Ω dans un ensemble Ω'. Alors*

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

Démonstration. – Vérifions que $f^{-1}(\mathcal{T})$ vérifie les axiomes des tribus

- $\varnothing \in f^{-1}(\mathcal{T})$ car $\varnothing = f^{-1}(\varnothing)$ et $\varnothing \in \mathcal{T}$
- Soit $A \in f^{-1}(\mathcal{T})$: il existe $B \in \mathcal{T}$ avec $A = f^{-1}(B)$. $A^c = (f^{-1}(B))^c = f^{-1}(B^c)$; or $B^c \in \mathcal{T}$ donc $A^c \in f^{-1}(\mathcal{T})$

  - Soient $(A_i)_{i\geq 1} \in f^{-1}(\mathcal{T})$: pour tout i, il existe $B_i \in \mathcal{T}$ avec $A_i = f^{-1}(B_i)$. $\cup_i A_i = \cup_i (f^{-1}(B_i))^c = f^{-1}(\cup_i B_i)$; or $\cap_i B_i \in \mathcal{T}$ donc $\cup_i A_i \in f^{-1}(\mathcal{T})$
- $\mathcal{A} \subset \sigma(\mathcal{A})$, donc $f^{-1}(\mathcal{A}) \subset f^{-1}(\sigma(\mathcal{A}))$, puis

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

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

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

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

□

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

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

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

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

Théorème 2. *La tribu borélienne de $\mathbb{R}^d$ est également la tribu engendrée par les pavés ouverts de $\mathbb{R}^d$ dont les côtés ont des côtés rationnels ; les ensembles de la forme $\prod_{i=1}^d]a_i, b_i[$, avec $a_i < b_i$ et a_i, b_i dans $\mathbb{Q}$.*

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

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

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

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

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

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

Corollaire 2. *Soit $(A, \mathcal{A})$ un espace mesuré, f une application de A dans $\mathbb{R}$. Si pour tout $a \in \mathbb{R}$, l'ensemble $f^{-1}(]-\infty, a[)$ est dans $\mathcal{A}$, alors f est mesurable de $(A, \mathcal{A})$ dans $(R, \mathcal{B}(\mathbb{R}))$.*

Corollaire 3. *Soit A et B deux espaces topologiques. Toute application continue de $(A, \mathcal{B}(A))$ dans $(B, \mathcal{B}(B))$ est mesurable de $(A, \mathcal{B}(A))$ dans $(B, \mathcal{B}(B))$.*

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

Tribu produit

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

Commençons par une remarque simple : si π_1 est l'application de $\Omega \times \Omega'$ dans Ω qui à $(x, y) \in \Omega \times \Omega'$ associe $\pi_1(x, y) = x$, alors π_1 (la projection sur la première coordonnée) est une application $(\Omega \times \Omega', \mathcal{A} \otimes \mathcal{B}) - (\Omega, \mathcal{A})$ mesurable. En effet, si $A \in \mathcal{A}$, $\pi_1^{-1}(A) = A \times \Omega' \in \mathcal{A} \times \mathcal{B} \subset \mathcal{A} \otimes \mathcal{B}$. De même, si π_2 est l'application de $\Omega \times \Omega'$ dans Ω' qui à $(x, y) \in \Omega \times \Omega'$ associe $\pi_2(x, y) = y$, alors π_2 (la projection sur la deuxième coordonnée) est une application $(\Omega \times \Omega', \mathcal{A} \otimes \mathcal{B}) - (\Omega', \mathcal{B})$ mesurable.

Théorème 3. *On suppose que $\mathcal{A} = \sigma((A_i)_{i \in I})$ et $\mathcal{B} = \sigma((B_i)_{j \in J})$. Alors $\mathcal{A} \otimes \mathcal{B} = \sigma(A_i \times B_j)_{(i,j) \in I \times J}$.*

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

Théorème 4. *Soient f une application de C dans Ω, g une application de C dans Ω'.On définit une application F de C dans $\Omega \times \Omega'$ par $F(x) = (f(x), g(x))$. L'application F est $(C, \mathcal{C}) - (\Omega \times \Omega', \mathcal{A} \otimes \mathcal{B})$ mesurable si et seulement si f est $(C, \mathcal{C}) - (\Omega, \mathcal{A})$ mesurable et g $(C, \mathcal{C}) - (\Omega', \mathcal{B})$ mesurable.*

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

Théorème 5. *Soit $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2), (\Omega_3, \mathcal{F}_3)$ trois espaces mesurés. L'application $\Psi : ((\Omega_1 \times \Omega_2) \times \Omega_3) \to \Omega_1 \times (\Omega_2 \times \Omega_3)$ qui à $((x, y), z)$ associe $(x, (y, z)$ est bi-mesurable de $((\Omega_1 \times \Omega_2) \times \Omega_3), (\Omega_1 \otimes \Omega_2) \otimes \Omega_3)$ vers $(\Omega_1 \times (\Omega_2 \times \Omega_3), \Omega_1 \otimes (\Omega_2 \otimes \Omega_3))$. Ainsi les deux tribus $(\Omega_1 \otimes \Omega_2) \otimes \Omega_3$ et $\Omega_1 \otimes (\Omega_2 \otimes \Omega_3)$ peuvent s'identifier et on notera simplement $\Omega_1 \otimes \Omega_2 \otimes \Omega_3$ cette tribu sur $\Omega_1 \times \Omega_2 \times \Omega_3$.*

Démonstration. En utilisant le théorème 3, on voit que les ensembles $(A_1 \times A_2) \times A_3$ et $A_1 \times (A_2 \times A_3)$ engendrent respectivement les deux tribus considérées. Le corollaire 1 permet alors de conclure. □

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

Théorème 6. *Pour tout entier $d \geq 2$, on a*

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

Démonstration. Il suffit de montrer que pour tout $d \geq 1$, $\mathcal{B}(\mathbb{R}^{d+1}) = \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R})$, puis de conclure par récurrence. Or, d'après le théorème 2 la tribu $\mathcal{B}(\mathbb{R}^d)$ est la tribu engendrée par les ensembles A de la forme $\prod_{i=1}^d]a_i, b_i[$, avec $a_i < b_i$ et a_i, b_i dans $\mathbb{Q}$, tandis que $\mathcal{B}(\mathbb{R})$ est la tribu engendrée par les ensembles B de la forme $]a_{d+1}, b_{d+1}[$, avec $a_{d+1} < b_{d+1}$ et a_{d+1}, b_{d+1} dans $\mathbb{Q}$. D'après le théorème 3, les produits $A \times B$ engendrent la tribu $\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R})$; mais ces ensembles sont exactement les ensembles de la forme $\prod_{i=1}^{d+1}]a_i, b_i[$, avec $a_i < b_i$ et a_i, b_i dans $\mathbb{Q}$, qui, toujours d'après le théorème 2 engendrent la tribu $\mathcal{B}(\mathbb{R}^{d+1})$. □

Théorème 7. *Soit f, g deux applications mesurables de $(C, \mathcal{C})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et G une application mesurable de $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors H définie par $H(x) = G(f(x), g(x))$ est mesurable de $(C, \mathcal{C})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*
En particulier les choix $H(x, y) = x + y$ et $H(x, y) = xy$ nous disent que $f + g$ et fg sont mesurables de $(C, \mathcal{C})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Démonstration. Avec les notations du Théorème 4, $H = G \circ F$. Pour le cas particulier, notons que comme H est une application continue de $\mathbb{R}^2$ dans $\mathbb{R}$, c'est une application $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)) - (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable, ou de manière équivalente $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})) - (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable les applications continues sont mesurables par rapport aux tribus boréliennes associées aux topologies correspondantes. □

1.2 Mesures

1.2.1 Algèbres

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

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

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

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

Remarque : en anglais
– algèbre se dit "field", plus rarement algebra
– tribu (σ-algèbre) se dit "σ-field".

1.2.2 Espace mesuré

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

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

vérifiant les propriétés suivantes :

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

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

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

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

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

11. Si $(A_i)_{i\in\mathbb{N}}$ est une suite décroissante d'événements (c'est à dire que $\forall n \in \mathbb{N} \quad A_{n+1} \subset A_n$)) et que l'on pose $A = \overset{+\infty}{\underset{i=1}{\cap}} A_i$, alors la suite $(\mu(A_n))_{n\in\mathbb{N}}$ est monotone, décroissante, et converge vers $\mu(A)$.

Démonstration.

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

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

   car $A\backslash B$ $A \cap B$ sont disjoints, de réunion A, tandis que $B\backslash A$ et $A \cap B$ sont disjoints, de réunion B.
5. Il suffit d'appliquer la relation 5 en remarquant que $\mu(A \cap B) \geq 0$
6. Si $A \subset B$, on a B est la réunion disjointe de A et de $B\backslash A$. Donc $\mu(B) = \mu(A) + \mu(B\backslash A) \geq \mu(A)$.
7. $(A \cap B) \subset A$, donc d'après la propriété 6 $\mu(A \cap B) \leq \mu(A)$. De même $\mu(A \cap B) \leq \mu(B)$. Finalement $\mu(A \cap B) \leq \min(\mu(A), \mu(B))$.
8. $A \subset (A \cup B)$, donc d'après la propriété 6 $\mu(A) \leq \mu(A \cup B)$. De même $\mu(B) \leq \mu(A \cup B)$. Finalement $\max(\mu(A), \mu(B)) \leq \mu(A \cup B)$.
9. Posons $B_1 = A_1$ et, pour tout $n \geq 2$ $B_n = A_n\backslash(\overset{n-1}{\underset{i=1}{\cup}} B_i)$. Par construction, les $(B_k)_{k\geq1}$ sont deux à deux disjoints. De plus, on peut montrer par récurrence sur n que

$$\forall n \geq 1 \quad \overset{n}{\underset{i=1}{\cup}} A_i = \overset{n}{\underset{i=1}{\cup}} B_i.$$

   On en déduit

$$\overset{+\infty}{\underset{i=1}{\cup}} A_i = \overset{+\infty}{\underset{i=1}{\cup}} B_i.$$

Donc

$$\begin{aligned}\mu(\overset{+\infty}{\underset{i=1}{\cup}} A_i) &= \mu(\overset{+\infty}{\underset{i=1}{\cup}} B_i) \\ &= \sum_{i=1}^{+\infty} \mu(B_i) \\ &\leq \sum_{i=1}^{+\infty} \mu(A_i).\end{aligned}$$

10. Comme $A_n \subset A_{n+1}$, on a $\mu(A_n) \leq \mu(A_{n+1})$, donc la suite est croissante. Comme on a pour tout n : $A_n \subset A$, la suite $(\mu(A_n))_{n\geq 1}$ est majorée par $\mu(A)$. Posons $B_1 = A_1$ et, pour tout $n \geq 2$ $B_n = A_n \backslash A_{n-1}$. On a :

$$\forall n \geq 1 \quad A_n = \overset{n}{\underset{i=1}{\cup}} B_i$$

et

$$A = \overset{+\infty}{\underset{i=1}{\cup}} A_i = \overset{+\infty}{\underset{i=1}{\cup}} B_i$$

Ainsi,

$$\begin{aligned}\mu(A) &= \mu(\overset{+\infty}{\underset{i=1}{\cup}} B_i) \\ &= \sum_{i=1}^{+\infty} \mu(B_i) \\ &= \lim_{n\to+\infty} \sum_{i=1}^{n} \mu(B_i) \\ &= \lim_{n\to+\infty} \mu(\overset{n}{\underset{i=1}{\cup}} B_i) \\ &= \lim_{n\to+\infty} \mu(A_n)\end{aligned}$$

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

□

1.2.3 Extension d'une mesure

Théorème 8 (Caratheodory). *Une mesure sur une algèbre $\mathcal{A}$ admet une extension à la tribu $\sigma(\mathcal{A})$*

1.2.4 Mesure image

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

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

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

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

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

1.3 Exercices de théorie de la mesure

1. Soit a un réel et $\tau_a : \mathbb{R} \to \mathbb{R}$ la translation définie par $\tau_a(x) = x + a$. Montrer que la famille $\mathcal{A}_a = \{A \in \mathcal{P}(\mathbb{R}); \tau_a(A) = A\}$ des parties invariantes par τ_a est une tribu sur $\mathbb{R}$.
   Plus généralement, si f est une application de $\mathbb{R}$ dans $\mathbb{R}$, donner une condition suffisante sur f pour que la famille $\mathcal{A} = \{A \in \mathcal{P}(\mathbb{R}); f(A) = A\}$ soit une tribu.
2. Soient $\mathcal{A}$ et $\mathcal{B}$ deux tribus sur un ensemble Ω. Montrer que la tribu engendrée par $\mathcal{A}$ et $\mathcal{B}$ coïncide avec la tribu engendrée par les ensembles de la forme $A \cap B$, où (A, B) décrit $\mathcal{A} \times \mathcal{B}$.
3. On rappelle que la tribu borélienne de $\mathbb{R}$ est la tribu engendrée par les ensembles de la forme $]a, b[; (a, b)^2 \in \mathbb{Q}$. Montrer que la tribu borélienne est également la tribu engendrée par les familles
   - $\mathcal{C} = \{]-\infty, a[; a \in \mathbb{Q}\}$.
   - $\mathcal{D} = \{]-\infty, a]; a \in \mathbb{Q}\}$.
   - $\mathcal{E} = \{[a, b[; (a, b)^2 \in \mathbb{Q}\}$.
   - $\mathcal{F} = \{[a, b]; (a, b)^2 \in \mathbb{Q}\}$.
4. Pour n entier strictement positif, on note $A_n = n\mathbb{N}^*$. Notons $\mathcal{P}$ l'ensemble des nombres premiers positifs et $\mathcal{T}$ la sous-tribu de $(\mathbb{N}, \mathcal{B}(\mathbb{N}))$ engendrée par les $(A_p)_{p \in \mathcal{P}}$.
   (a) Montrer que l'ensemble C des entiers qui sont premiers avec 2000 est $\mathcal{T}$-mesurable.
   (b) Montrer que l'ensemble $B = \{2^k; k \in \mathbb{N}^*\}$ des puissances de deux est $\mathcal{T}$-mesurable.

5. Soit (E, d) un espace métrique. Montrer que la tribu borélienne $\mathcal{B}(E)$ engendrée par les ouverts de E est aussi la plus petite tribu rendant mesurables toutes les applications continues de (E, d) dans $\mathbb{R}$ (muni de la tribu borélienne et de la topologie usuelle).
6. *Lemme de Doob*
   Soit f et g deux applications mesurables de $(\Omega, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Montrer que g est $\sigma(f)$-mesurable si et seulement si il existe une application mesurable u de $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ dans lui-même telle que $g = u \circ f$.

1.4 Convergence et mesurabilité

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

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

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

1.4.2 Convergence et mesurabilité

Théorème 9. *Soit $(f_n)_{n\geq 1}$ une suite d'applications mesurables de $\Omega, \mathcal{F}$ dans $(\overline{\mathcal{R}}, \mathcal{B}(\overline{R})$. Alors les applications etsuivantes les événements suivants sont mesurables :*

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

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

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

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

où, en adoptant le formalisme probabiliste :

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

2. On peut simplement remarquer que $\inf\limits_{n\geq 1} f_n = -\sup\limits_{n\geq 1}(-f_n)$, et appliquer le point précédent, sachant que l'opposé d'une fonction mesurable est mesurable.

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

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

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

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

7. Notons $D = \{f_n$ converge vers $+\infty\} \cup \{f_n$ converge vers $-\infty\}$. D'après ce qui précède $D \in \mathcal{F}$. posons $F = \underline{\lim\limits_{n\to+\infty}}\ f_n$ et $G = \overline{\lim\limits_{n\to+\infty}}\ f_n$. Définissons alors des fonction F_1, G_1, H par :
   - $F_1 = F1\!\!1_{D^c}$
   - $G_1 = F1\!\!1_{D^c}$
   - $h = F_1 - G_1$

   h est bien définie par F_1 et G_1 sont à valeurs dans $\mathbb{R}$. $\{f_n$ converge dans $\mathbb{R}\} = h^{-1}(\{0\})\backslash D$ Ainsi, pour conclure, il suffit de montrer que h est $(\Omega, \mathcal{F})-(\overline{\mathcal{R}}, \mathcal{B}(\overline{R})$ mesurable. Comme h est à valeurs dans $\mathbb{R}$, il suffit de montrer que h est $(\Omega, \mathcal{F}) - (\mathbb{R}, \mathcal{B}(R)$ mesurable. Pour cela, il suffit de montrer que F_1 et G_1 sont $(\Omega, \mathcal{F}) - (\mathbb{R}, \mathcal{B}(R)$ mesurables. Montrons donc que F_1

et G_1 sont $(\Omega,\mathcal{F})-(\mathbb{R},\mathcal{B}(R)$ mesurable (l'autre preuve étant analogue). Soit A un borélien de $\mathbb{R}$: on a

$$F_1^{-1}(A) = F_1^{-1}(A\cap\{0\})\cup F_1^{-1}(A\cap\{0\}^c) = F_1^{-1}(A\cap\{0\})\cup F^{-1}(A\cap\{0\}^c).$$

$A\cap\{0\}$ ne peut prendre que deux valeurs : $\{0\}$ et $\varnothing$: bien sûr $F_1^{-1}(\varnothing) = \varnothing \in \mathcal{F}$; d'autre part $F_1^{-1}(\{0\}) = F^{-1}(\{0\}) \cup D$, qui est bien dans $\mathcal{F}$ car F est mesurable est $\{0\}$ un borélien. Enfin, $F^{-1}(A\cap\{0\}^c)$ est bien dans $\mathcal{F}$ car F est mesurable est $A\cap\{0\}^c$ un borélien.

$\square$

Chapitre 2

Espace probabilisé

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

2.1 Espace probabilisé

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

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

vérifiant les propriétés suivantes :

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

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

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

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

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

1. Pour toute suite $(A_i)_{1\leq i\leq n}$ d'éléments de $\mathcal{F}$ deux à deux disjoints, $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$.
2. $\forall A, B \in \mathcal{F} \quad (A \cap B = \varnothing) \Longrightarrow (P(A \cup B) = P(A) + P(B))$
3. $\forall A \in \mathcal{F} \quad P(A^c) = 1 - P(A)$
4. $P(\varnothing) = 0$
5. $\forall A, B \in \mathcal{F} \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$
6. $\forall A, B \in \mathcal{F} \quad P(A \cup B) \leq P(A) + P(B)$
7. $\forall A, B \in \mathcal{F} \quad (A \subset B) \Longrightarrow (P(A) \leq P(B))$
8. $\forall A, B \in \mathcal{F} \quad P(A \cap B) \leq \min(P(A), P(B))$
9. $\forall A, B \in \mathcal{F} \quad P(A \cup B) \geq \max(P(A), P(B))$
10. Pour toute suite $(A_i)_{i\geq 1}$ d'éléments de $\mathcal{F}$, $P(\bigcup_{i=1}^{+\infty} A_i) \leq \sum_{i=1}^{+\infty} P(A_i)$.
11. Si $(A_i)_{i\in\mathbb{N}}$ est une suite croissante d'événements (c'est à dire que $\forall n \in \mathbb{N} \quad A_n \subset A_{n+1}$)) et que l'on pose $A = \bigcup_{i=1}^{+\infty} A_i$, alors la suite $(P(A_n))_{n\in\mathbb{N}}$ est monotone, croissante, et converge vers $P(A)$.
12. Si $(A_i)_{i\in\mathbb{N}}$ est une suite décroissante d'événements (c'est à dire que $\forall n \in \mathbb{N} \quad A_{n+1} \subset A_n$)) et que l'on pose $A = \bigcap_{i=1}^{+\infty} A_i$, alors la suite $(P(A_n))_{n\in\mathbb{N}}$ est monotone, décroissante, et converge vers $P(A)$.

Démonstration.

1. Il suffit de poser $A_i = \varnothing$ pour $i \geq n+1$ et d'appliquer l'axiome 2.
2. Il suffit d'appliquer la propriété 1 avec $n = 2$, $A_1 = A$ et $A_2 = B$.
3. Il suffit d'appliquer la propriété 2 avec $B = A^c$: A et B sont disjoints donc $P(A) + P(A^c) = P(A \cup A^c) = P(\Omega) = 1$ d'après l'axiome 1.
4. On applique la propriété 3 avec $A = \Omega$: $P(\varnothing) = P(\Omega^c) = 1 - P(\Omega) = 1 - 1 = 0$.
5. Les ensembles $A \backslash B$, $B \backslash A$ et $A \cap B$ sont disjoints et leur réunion est $A \cup B$, donc l'après la propriété 1, on a

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

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

6. Il suffit d'appliquer la relation 5 en remarquant que $P(A\cap B)\geq 0$
7. Si $A\subset B$, on a B est la réunion disjointe de A et de $B\backslash A$. Donc $P(B)=P(A)+P(B\backslash A)\geq P(A)$.
8. $(A\cap B)\subset A$, donc d'après la propriété 6 $P(A\cap B)\leq P(A)$. De même $P(A\cap B)\leq P(B)$. Finalement $P(A\cap B)\leq\min(P(A),P(B))$.
9. $A\subset(A\cup B)$, donc d'après la propriété 6 $P(A)\leq P(A\cup B)$. De même $P(B)\leq P(A\cup B)$. Finalement $\max(P(A),P(B))\leq P(A\cup B)$.
10. Posons $B_1=A_1$ et, pour tout $n\geq 2$ $B_n=A_n\backslash(\bigcup_{i=1}^{n-1}B_i)$. Par construction, les $(B_k)_{k\geq 1}$ sont deux à deux disjoints. De plus, on peut montrer par récurrence sur n que

$$\forall n\geq 1\qquad \bigcup_{i=1}^{n}A_i=\bigcup_{i=1}^{n}B_i.$$

On en déduit

$$\bigcup_{i=1}^{+\infty}A_i=\bigcup_{i=1}^{+\infty}B_i.$$

Donc

$$\begin{aligned}P(\bigcup_{i=1}^{+\infty}A_i) &= P(\bigcup_{i=1}^{+\infty}B_i)\\ &= \sum_{i=1}^{+\infty}P(B_i)\\ &\leq \sum_{i=1}^{+\infty}P(A_i).\end{aligned}$$

11. Comme $A_n\subset A_{n+1}$, on a $P(A_n)\leq P(A_{n+1})$, donc la suite est croissante. Comme on a pour tout n : $A_n\subset A$, la suite $(P(A_n))_{n\geq 1}$ est majorée par $P(A)$. Posons $B_1=A_1$ et, pour tout $n\geq 2$ $B_n=A_n\backslash A_{n-1}$. On a :

$$\forall n\geq 1\quad A_n=\bigcup_{i=1}^{n}B_i$$

et

$$A=\bigcup_{i=1}^{+\infty}A_i=\bigcup_{i=1}^{+\infty}B_i$$

Ainsi,

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

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

□

2.2 Partitions et probabilités

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

Théorème 10. *Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé. Soit I un ensemble d'index fini ou dénombrable et $(\Omega_i)_{i\in I}$ une partition de Ω. Alors on a*

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

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

2.3 Probabilité conditionnelle

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

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

$P(A|B)$ se lit "Probabilité de A sachant B".

On a évidemment

$$P(A \cap B) = P(B)P(A|B). \tag{2.1}$$

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

2.3.1 Conditionnements en chaîne

Si A, B sont deux événements observables avec $A \subset B$ et $P(B) \neq 0$, la formule (2.1) devient

$$P(A) = P(B)P(A|B). \tag{2.2}$$

On a la généralisation suivante :

Théorème 11. *Soient $n \geq 2$ et $E_1, \ldots, E_n$ des événements observables vérifiant*

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

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

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

Démonstration. La formule se montre par récurrence sur n. Pour $n = 2$, c'est une conséquence immédiate de (2.2). Pour $n > 2$, on applique d'abord la formule pour $n = 2$ aux événements E_n et E_{n-1} :

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

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

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

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

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

Théorème 12. *Soient $n \geq 2$ et $A_1, \dots, A_n$ des événements observables avec $P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$. Alors*

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

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

2.3.2 Conditionnement par tous les cas possibles

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

Théorème 13. *Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé. Soit I un ensemble d'index fini ou dénombrable et $(\Omega_i)_{i \in I}$ une partition de Ω. Alors on a*

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

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

Démonstration. D'après le théorème 10, on a

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

□

2.3.3 Formule de Bayes

Théorème 14. *Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé. Soit I un ensemble d'index fini ou dénombrable et $(\Omega_i)_{i \in I}$ une partition de Ω telle que pour tout $i \in I$, $P(\Omega_i)$ soit non nul. Soit A un élément de probabilité non nulle.*

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

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

Démonstration.

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

et on applique le théorème précédent.

□

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

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

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

2.4 Indépendance

2.4.1 Événements indépendants

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

$$P(A\cap B) = P(A)P(B).$$

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

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

2.4.2 Tribus indépendantes

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

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

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

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

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

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

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

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

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

2.4.3 Indépendance et tribus engendrées

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

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

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

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

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

Alors $P = Q$.

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

Théorème 15. *Soit $\mathcal{C}$ et $\mathcal{D}$ deux familles de parties mesurables de $(\Omega, \mathcal{F})$. On suppose que $\mathcal{C}$ et $\mathcal{D}$ sont des π-systèmes et que pour tout $(A, B) \in \mathcal{C} \times \mathcal{D}$, on a $P(A \cap B) = P(A)P(B)$. Alors, les tribus $\mathcal{A} = \sigma(\mathcal{C})$ et $\mathcal{B} = \sigma(\mathcal{D})$ sont indépendantes.*

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

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

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

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

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

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

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

$$\forall (A, B) \in \mathcal{C} \times \mathcal{D} \quad P(A \cap B) = P(A)P(B) \Longrightarrow \forall (A, B) \in \sigma(\mathcal{C}) \times \sigma(\mathcal{D}) \quad P(A \cap B) = P(A)P(B),$$

ce qui était notre objectif

□

Théorème 16. *Les deux propriétés suivantes sont équivalentes :*

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

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

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

Il est facile de voir que $\mathcal{C}$ est un π-système qui engendre $\bigwedge_{i \in I \backslash \{j\}} \mathcal{A}_i$ et que $\forall (A, B) \in \mathcal{A}_j \times \mathcal{C} \quad P(A \cap B) = P(A)P(B)$. Le théorème 15 permet de conclure.

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

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

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

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

d'où

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

ce qui achève la preuve. □

2.5 Premiers exercices de probabilité

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

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

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

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

(d) Conclure.

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

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

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

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

   (On pose $d_0 = 1$.)

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

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

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

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

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

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

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

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

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

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

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

Chapitre 3

Integrales

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

On va commencer par rappeler la définition de l'intégrale dite “de Lebesgue” telle qu'elle a été vue au premier semestre et en rappeler les propriétés fondamentales.

3.1 Définition de l'intégrale et propriétés de base

3.1.1 Définition

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

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

où l'inf porte sur toutes les partitions finies de Ω.

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

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

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

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

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

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

3.1.2 Rappel des propriétés de bases

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

On rappelle sans démonstration les propriétés de base de l'intégrale :

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

3.1.3 Conséquences importantes

Théorème 17 (Lemme de Fatou). *Pour toute suite $(f_n)_{n\geq 1}$ de fonctions mesurables positives, on a*

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

Démonstration. Il suffit de poser $g_n = \inf_{k\geq n} f_k$. (g_n) est une suite croissante, dont la limite est, par définition, $\lim g_n = \varliminf f_n$. On a pour tout n

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

D'où

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

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

Théorème 18 (Convergence dominée). *Si $(f_n)_{n\geq 1}$ est une suite de fonctions convergeant presque partout vers f, et telle qu'il existe une fonction g intégrable vérifiant pour tout n, $|f_n| \leq g$ alors la suite $\int f_n \, d\mu$ converge vers $\int f \, d\mu$.*

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

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

et

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

soit

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

et

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

En simplifiant, on obtient

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

et

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

ce qui montre bien le résultat voulu. □

3.2 Intégration sur un ensemble, mesures à densité

3.2.1 Intégration sur un ensemble

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

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

Théorème 19. *Si f est intégrable et que $(A_n)_{n\geq 1}$ est une partition dénombrable de Ω, alors*

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

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

3.2.2 Fonctions simples (ou fonctions étagées)

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

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

Lemme 1. *Toute fonction mesurable positive f (éventuellement infinie) peut s'écrire comme limite simple d'une suite croissante de fonctions simples (f_n).*

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

3.2.3 Mesure à densité

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

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

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

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

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

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

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

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

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

Théorème 20. *On suppose que ν est une mesure qui admet une densité f par rapport à μ. Alors, pour toute fonction mesurable g*

$$\int |g|\ d\nu = \int |g| f\ d\mu. \tag{3.1}$$

Si cette quantité est finie, on a alors

$$\int g\ d\nu = \int g f\ d\mu. \tag{3.2}$$

Démonstration. Si $g = \mathbb{1}_A$ avec $A \in \mathcal{F}$ (3.2) est immédiat. Par linéarité, (3.2) est également vérifié lorsque g est une fonction simple positive. En utilisant le lemme (1) et le théorème de convergence monotone, il s'ensuit que (3.2) est vraie pour toute fonction mesurable positive, donc en particulier (3.1) est vraie pour toute fonction mesurable g. Supposons maintenant que $\int |g|\ d\nu = \int |g| f\ d\mu < +\infty$: on peut alors écrire $g = g_+ - g_-$ avec $\int g_+\ d\nu < +\infty$ et $\int g_-\ d\nu < +\infty$. Comme g_+ et g_- sont des fonctions mesurables positives, on a $\int g_+\ d\nu = \int g_+ f\ d\mu$ et $\int g_-\ d\nu = \int g_- f\ d\mu$. En faisant la différence, on obtient donc $\int (g_+ - g_-)\ d\nu = \int (g_+ - g_-) f\ d\mu$, soit (3.2). □

3.2.4 Intégration par rapport à une mesure image

Théorème 21. *Soit $(\omega, \mathcal{F}, \mu)$ un espace mesuré, T une application mesurable de $(\omega, \mathcal{F})$ dans $(\omega', \mathcal{F}')$. Soit f une application mesurable de $(\omega', \mathcal{F}')$ dans*

$(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. *Alors f est intégrable par rapport à μ_T si et seulement si $f \circ T$ est intégrable par rapport à μ. Dans ce cas, on a*

$$\int_{\Omega'} f(y)\ d\mu_T(y) = \int_{\Omega} (f \circ T)(x)\ d\mu(x) \tag{3.3}$$

Démonstration. Prenons d'abord le cas où f est l'indicatrice d'un ensemble $A \in \mathcal{F}'$: on a $\int_{\Omega'} f\ d\mu_T = \int_{\Omega'} 1\!\!1_A\ d\mu_T = \mu_T(A) = \mu(T^{-1}(A))$. D'un autre côté $1\!\!1_A \circ T = 1\!\!1_{T^{-1}(A)}$, donc $\int_\Omega f \circ T\ d\mu = \int_\Omega 1\!\!1_{T^{-1}(A)}\ d\mu = \mu(T^{-1}(A))$.L'égalité (3.3) est donc vérifiée lorsque lorsque f est l'indicatrice d'un ensemble $A \in \mathcal{F}'$. Par linéarité, elle est donc vérifiée pour toute fonction étagée mesurable.

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

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

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

3.3 Mesure produit

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

3.3.1 Construction de la mesure produit

Lemme 2. *Pour tout $A \in \mathcal{X} \otimes \mathcal{Y}$, $x \in X$ et $y \in Y$, on note*

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

et

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

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

$y \mapsto f(x,y)$ est mesurable par rapport à la tribu $\mathcal{Y}$, et de même pour chaque y fixé la fonction $x \mapsto f(x,y)$ est mesurable par rapport à la tribu $\mathcal{X}$.

Démonstration. Notons

$$\mathcal{T} = \{A \in \mathcal{X} \otimes \mathcal{Y} : \forall x \in X \times Y \quad A_y(x) \in \mathcal{Y}\}.$$

Pour tous $(x,y) \in X \times Y$ $\varnothing_y(x) = \varnothing \in \mathcal{Y}$ et $\varnothing_x(y) = \varnothing \in \mathcal{X}$, donc $\varnothing \in \mathcal{T}$. Supposons que $A \in \mathcal{T}$ et soit $x \in X$ $(A^c)_y(x) = \{y \in Y : (x,y) \in A^c\} = \{y \in Y : (x,y) \in A\}^c \in \mathcal{Y}$. Ainsi, $A^c \in \mathcal{T}$. Maintenant, si $(A_n)_{n\geq 1}$ est une suite d'éléments de $\mathcal{T}$ et que l'on pose $A = \cup_{n\geq 1} A_n$, on a

$$\begin{aligned} A_y(x) &= \{y \in Y : (x,y) \in A\} \\ &= \{y \in Y : \exists n \geq 1 \quad (x,y) \in A_n\} \\ &= \cup_{n\geq 1}\{y \in Y : \quad (x,y) \in A_n\} \\ &\in \mathcal{Y} \end{aligned}$$

$\mathcal{T}$ est une sous-tribu de $\mathcal{X} \otimes \mathcal{Y}$ qui contient $X \times Y$, donc $\mathcal{T} = \mathcal{X} \otimes \mathcal{Y}$, ce qui montre le résultat voulu pour $A_y(x)$. On procède de même pour $A_x(y)$. Passons au deuxième point : soit $A \in \mathcal{C}$, $x \in X$ et notons $g_x(y) = f(x,y)$

$$g_x^{-1}(A) = \{y : f(x,y) \in A\} = \{y : (x,y) \in f^{-1}(A)\}.$$

Mais comme f est mesurable $f^{-1}(A) \in \mathcal{X} \otimes \mathcal{Y}$, donc d'après le point précédent $g_x^{-1}(A) \in \mathcal{Y}$. □

Théorème 22. *Soit $(X, \mathcal{X}, \mu)$ et $(Y, \mathcal{Y}, \nu)$ deux espaces mesurés dont les mesures μ et ν sont σ-finies. Il existe une unique mesure m sur $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$ telles que pour tous $X \in \mathcal{X}$ et $Y \in \mathcal{Y}$, on ait*

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

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

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

Démonstration. Supposons d'abord que μ et ν sont finies. Soit E dans $X \times Y$; d'après le lemme précédent la fonction $x \mapsto \nu(E_y(x))$ est bien définie. Notons $\mathcal{T}$ la famille des ensembles E tels que cette fonction soit mesurable de $\mathcal{X}$ dans $\mathcal{B}(\mathbb{R})$. Il n'est pas difficile de voir que $\mathcal{T}$ est un λ-système. Mais $\mathcal{T}$ contient tous les pavés, donc d'après le théorème λ-π, $\mathcal{T}$ est $\mathcal{X} \otimes \mathcal{Y}$ tout entier. Ainsi, pour tout E dans $\mathcal{X} \otimes \mathcal{Y}$, on peut définir

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

et de même on pourrait définir

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

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

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

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

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

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

$\square$

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

3.3.2 Théorèmes de Fubini et Tonelli

Théorème 23 (Tonelli). *Soit $(X, \mathcal{X}, \mu)$ et $(Y, \mathcal{Y}, \nu)$ deux espaces mesurés dont les mesures μ et ν sont σ-finies et $f \in \mathcal{V}_+(X \times Y, \mathcal{X} \otimes \mathcal{Y})$.*

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

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

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

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

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

Enfin, on a les égalités

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

et

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

Démonstration. La mesurabilité de $y \mapsto f(x, y)$ est une conséquence immédiate du Théorème 2.

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

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

D'après la deuxième partie du Théorème 22, l'application $x \mapsto \int_Y f(x, y)\, d\nu(y)$ est donc mesurable et l'on a

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

Le résultat s'étend aisément à la classe des fonctions simples par linéarité, puis à $f \in \mathcal{V}_+(X \times Y, \mathcal{X} \otimes \mathcal{Y})$ en utilisant le lemme 1 et le théorème de convergence monotone. □

Théorème 24 (Fubini). *Soit* $(X, \mathcal{X}, \mu)$ *et* $(Y, \mathcal{Y}, \nu)$ *deux espaces mesurés dont les mesures* μ *et* ν *sont* σ*-finies et* $f \in \mathcal{V}(\mathcal{X} \times \mathcal{Y}, \mu \otimes \nu$. *On suppose que*

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

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

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

et

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

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

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

Il s'ensuit que si l'on pose

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

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

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

□

3.3.3 Associativité de la mesure produit

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

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

Démonstration.

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

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

Ainsi, les mesures coïncident. □

3.3.4 Convolution de mesures

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

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

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

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

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

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

3.4 Premiers exercices d'intégration

1. Le but de cet exercice est de montrer le *théorème du retour* de Poincaré. Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré et T une transformation, c'est à dire une application mesurable de $(\Omega, \mathcal{F}, \mu)$ dans lui-même. On suppose que μ est une mesure finie et qu'elle est invariante sous l'action de T, c'est à dire que la mesure image de μ par l'application T est μ elle-même. Alors, le théorème du retour dit que pour tout ensemble mesurable A de mesure non nulle, la suite des itérées $(T^n(x))_{n\geq 0}$ passe une infinité de fois dans A pour presque tout x appartenant à A.
   (a) On pose
   $$N(x) = \sum_{n=0}^{+\infty} \mathbb{1}_A(T^n(x))$$
   ainsi que $Y(x) = \exp(-N(x))$, avec la convention $\exp(-(+\infty)) = 0$. Montrer que Y est une application mesurable intégrable par rapport à μ.
   (b) On pose $Z(x) = Y(Tx)$. Montrer que $Y(x) = e^{-\mathbb{1}_A(x)} Z(x)$, puis que $\int Y(x)\ d\mu(x) = \int Z(x)\ d\mu(x)$.
   (c) Conclure.
2. *Intégration par rapport à une somme de mesures.*
   Soit $(\mu_i)_{i\in\mathbb{N}}$ une suite de mesure définies sur un espace $(\Omega, \mathcal{F})$.
   (a) Montrer que $\mu = \sum_{i\in\mathbb{N}} \mu_i$ est une mesure sur $(\Omega, \mathcal{F})$.
   (b) Montrer que pour f mesurable positive, puis pour f intégrable, on a
   $$\int_\Omega f\ d\mu = \sum_{i\in\mathbb{N}} \int_\Omega f\ d\mu_i.$$
3. Soit $(f_n)_{n\geq 0}$ une suite croissante d'applications μ-intégrables convergeant μ presque partout vers f. On suppose qu'il existe une constante K telle que
   $$\forall n \geq 0 \qquad \int f_n\ d\mu \leq K.$$
   Montrer que f est μ-intégrable et que $\int |f_n - f|\ d\mu$ tend vers 0 lorsque n tend vers l'infini.

4. Soit $(f_n)_{n\geq 0}$ une suite décroissante d'applications μ-mesurables positives. On suppose que f_1 est μ-intégrable. Montrer que (f_n) converge simplement vers une fonction mesurable f, que f est μ-intégrable et que $\int |f_n - f|\, d\mu$ tend vers 0 lorsque n tend vers l'infini.
5. (a) Soit φ une application de $\mathbb{N}$ dans $\mathbb{R}_+$ Montrer que l'on a
$$\int_{\mathbb{R}^d} \varphi(\mathrm{Ent}(\|x\|_\infty))\, d\lambda^d(x) = \sum_{n=0}^{+\infty} \left((2n+2)^d - (2n)^d\right) \varphi(n).$$
   (b) Soit $\alpha > 0$. À quelle condition la fonction $x \mapsto \frac{1}{\|x\|_2^\alpha}$ est elle intégrable sur le complémentaire de la boule unité ?
   (c) Montrer que l'application
$$f : x \mapsto \frac{x}{\|x\|^2}$$
réalise un C^1 difféomorphisme de $\{x \in \mathbb{R}^d; 0 < \|x\|_2 < 1\}$ sur $\{x \in \mathbb{R}^d; 1 < \|x\|_2\}$ et que sa différentielle est
$$h \mapsto Df_x.h = \frac{\|x\|_2^2 h - 2\langle x, h\rangle x}{\|x\|_2^4}.$$
   (d) Soit $\alpha > 0$. À quelle condition la fonction $x \mapsto \frac{1}{\|x\|_2^\alpha}$ est elle intégrable sur la boule unité ?
6. Soit f une fonction réelle intégrable sur $(\Omega, \mathcal{F}, \mu)$. Montrer que si $\int_A f\, d\mu = 0$ pour tout $A \in \mathcal{F}$, alors $f = 0$ presque partout.
7. Étudier la limite, lorsque n tend vers l'infini de
   (a) $\int_0^{+\infty} \frac{\log(x+n)}{n} e^{-x} \cos x\, dx$
   (b) $\int_0^1 \frac{1+nx}{(1+x)^n}\, dx$
   (c) $\int_1^{+\infty} \frac{1+nx}{(1+x)^n}\, dx$
8. (a) Calculer $I_n = \int_0^1 x^n \ln x\, dx$ pour $n \in \mathbb{N}$
   (b) En déduire la valeur de $\int_0^1 \frac{\ln x}{1-x}\, dx$, sachant que $\sum_{n=1}^{+\infty} n^{-2} = \pi^2/6$.
9. Soit $a > 0$. Montrer que la fonction, $f : (x,y) \mapsto e^{-xy}\sin x$, est intégrable sur $[0,a] \times [0,+\infty[$. On pose $I_a = \displaystyle\int_{[0,a]\times[0,+\infty[} f(x,y)\, dxdy$. Déterminer la limite de I_a quand a tend vers $+\infty$. En déduire la valeur de $\displaystyle\int_0^{+\infty} \frac{\sin x}{x}\, dx$.

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

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

Chapitre 4

Lois des variables et des vecteurs aléatoires

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

4.1 Définition

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

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

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

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

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

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

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

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

4.1.1 Fonction de répartition

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

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

Théorème 25. *Si deux variables (ou vecteurs) aléatoires ont la même fonction de répartition,alors elles ont même loi.*

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

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

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

Théorème 26. *La fonction de répartition F_X d'une variable aléatoire vérifie les propriétés suivantes*

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

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

- Si $s \le t$, on a $]-\infty, s] \subset]-\infty, s]$, d'où
$$F_X(s) = P(]-\infty, s]) \le P(]-\infty, t]) = F_X(t).$$
- Posons pour $n \ge 1$, $A_n =]-\infty, -n]$, on a $A_{n+1} \subset A_n$ et $\underset{n\ge 1}{\cap} A_n = \varnothing$, d'où $\lim\limits_{n\to+\infty} P_X(A_n) = P_X(\varnothing) = 0$. Soit $\varepsilon > 0$: d'après ce qui précède, il existe n tel que $P(A_n) < \varepsilon$. Comme F_X est croissante et positive, on a
$$t \le -n \quad \Longrightarrow 0 \le F_X(t) \le F_X(-n) \le \varepsilon,$$
ce qui prouve que $\lim\limits_{t\to-\infty} F_X(t) = 0$.
- Posons pour $n \ge 1$, $A_n =]-\infty, n]$, on a $A_n \subset A_{n+1}$ et $\underset{n\ge 1}{\cup} A_n = \mathbb{R}$, d'où $\lim\limits_{n\to+\infty} P_X(A_n) = P_X(\mathbb{R}) = 1$. Soit $\varepsilon > 0$: d'après ce qui précède, il existe n tel que $P(A_n) \ge< 1-\varepsilon$. Comme F_X est croissante et majorée par 1, on a
$$t \ge n \quad \Longrightarrow 1 \ge F_X(t) \ge F_X(n) \ge 1-\varepsilon,$$
ce qui prouve que $\lim\limits_{t\to+\infty} F_X(t) = 1$.
- Soit $t \in \mathbb{R}$, $(t_n)_{n\ge 1}$ une suite décroissante convergeant vers t. Posons pour $n \ge 1$, $A_n =]-\infty, t_n]$, on a $A_{n+1} \subset A_n$ et $\underset{n\ge 1}{\cap} A_n =]-\infty, t]$, d'où $\lim\limits_{n\to+\infty} F_X(t_n) = \lim_{n\to+\infty} P_X(A_n) = P_X(]-\infty, t]) = F_X(t)$. Comme cette égalité est obtenue pour toute suite décroissante convergeant vers t, ceci prouve que la limite à droite de F_X au point t est $F_X(t)$ (critère de

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

$\square$

4.1.2 Tribu engendrée par une ou plusieurs variables aléatoires

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

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

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

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

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

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

Démonstration. On a

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

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

$\square$

4.2 Indépendance des variables aléatoires

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

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

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

Théorème 27. *Soient $(X_i)_{i\in I}$ une collection de vecteurs aléatoires aléatoires indépendants. On suppose que X_i est à valeurs dans $\mathbb{R}^{n_i}$. Soient $(f_i)_{i\in I}$ une famille d'applications telles que pour tout i, f_i soit une application mesurable de $\mathbb{R}^{n_i}$ dans $\mathbb{R}^{p_i}$. Alors, si on pose $Y_i=f_i(X_i)$ les variables aléatoires $(Y_i)_{i\in I}$ sont indépendantes.*

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

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

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

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

Théorème 28. *Soient $(\mathcal{A}_i)_{i\in I}$ une famille de sous-tribus de $(\Omega,\mathcal{F})$ indépendantes sous P. Soient $J\subset I$ et $K\subset I$ disjoints.*

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

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

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

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

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

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

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

Comme $\mathcal{C}$ est un π-système qui engendre $\underset{j \in J}{\wedge} \mathcal{A}_j$ et $\mathcal{D}$ un π-système qui engendre $\underset{j \in K}{\wedge} \mathcal{A}_j$, le théorème 15 permet de conclure.

□

Théorème 29. *Soit* $(\Omega, \mathcal{F}, P)$ *un espace probabilisé et* $X_1, \dots, X_n$ *n vecteurs aléatoires. Les deux propositions suivantes sont équivalentes*

1. $X_1, \dots, X_n$ *sont indépendantes.*
2. $P_{X_1,\dots,X_n} = P_{X_1} \otimes \dots \otimes P_{X_n}$

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

$$\begin{aligned} P_{X_1,\dots,X_n}(A) &= P((X_1, \dots, X_n) \in A) \\ &= P(\overset{n}{\underset{i=1}{\cap}} \{X_i \in A_i\}) \\ &= \prod_{i=1}^{n} P(X_i \in A_i) \\ &= \prod_{i=1}^{n} P_{X_i}(A_i) \\ &= (P_{X_1} \otimes \dots \otimes P_{X_n})(A) \end{aligned}$$

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

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

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

ce qui prouve l'indépendance des tribus.

□

Corollaire 4. *Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé et $X_1, \dots, X_n$ n vecteurs aléatoires. On suppose qu'il existe des mesures de probabilités $\mu_1, \dots, \mu_n$ telles que $P_{X_1,\dots,X_n} = \mu_1 \otimes \cdots \otimes \mu_n$. Alors*

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

Démonstration. Soit B un borélien.

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

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

4.2.1 Application : loi $0 - 1$ de Kolmogorov

Théorème 30. *Soit S un ensemble infini et $(\mathcal{A}_i)_{i \in S}$ une famille de tribus indépendantes sous la loi P. On pose*

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

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

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

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

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

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

Comme $\bigcap_{i\in J\backslash\{\infty\}} A_i \in \bigwedge_{k\in J\backslash\{\infty\}} \mathcal{A}_k$, le théorème 28 implique que A_∞ et $\bigcap_{k\in J\backslash\{\infty\}} A_k$ sont indépendants. Il s'ensuit que

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

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

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

donc $P(A) \in \{0,1\}$. □

4.2.2 Variables aléatoires intépendantes et convolutions

Théorème 31. *Si deux variables aléatoires X et Y sont indépendantes sous $\mathbb{P}$, alors $P_X * P_Y = P_{X+Y}$.*

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

Théorème 32. *Soit μ, ν et χ trois mesures finies sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. On a*

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

et

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

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

$$P_X * PT = P_{X+T} = P_{S+Z} = P_S * P_Z.$$

Maintenant, il est facile de voir que $P_T = \frac{1}{\nu(\mathbb{R}^d)\chi(\mathbb{R}^d)} \nu * \chi$ et que $P_S = \frac{1}{\mu(\mathbb{R}^d)\nu(\mathbb{R}^d)} \mu * \nu$, d'où $(\mu * \nu) * \chi = \mu * (\nu * \chi)$. De même $P_X * P_Y = P_{X+Y} = P_{Y+X} = P_Y * P_X$ permet de montrer la première formule. □

Ainsi l'ensemble des mesures finies munies de (+,*,.) forme une $\mathbb{R}$-algèbre commutative.

4.3 Variables aléatoires discrètes

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

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

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

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

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

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

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

Théorème 33. *Soit X est une variable aléatoire discrète, et D un ensemble D fini ou dénombrable inclus dans $\mathbb{R}$ tel que $X(\Omega) = D$. Pour $i \in D$, on pose $p_i = P(X = i)$. Alors,*

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

$$P_X(A) = \sum_{i \in D \cap A} p_i.$$

2. $P_X = \sum_{i \in D} p_i \delta_i$.
3. *P_X admet comme densité par rapport à la mesure de comptage sur D la fonction $f(x)$ définie par*

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

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

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

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

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

On a d'une part

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

Comme A est quelconque, on en déduit que $\mu = P_X$. D'autre part

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

ce qui signifie que μ (c'est à dire P_X) admet f comme densité par rapport à la mesure de comptage.

□

On a la réciproque suivante :

Théorème 34. *Soit D un ensemble fini ou dénombrable, $(p_i)_{i\in D}$ une famille de réels positifs vérifiant*

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

Alors, on peut construire un espace probabilisé $(\Omega, \mathcal{F}, P)$ et une variable aléatoire X sur cet espace telle que

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

Démonstration. Comme on l'a déjà remarqué, le problème d'existence d'une variable aléatoire se ramène souvent à l'existence d'une loi. Ici,on peut prendre $\Omega = D$, $\mathcal{F} = \mathcal{P}(\Omega)$ et $X(\omega) = \omega$ avec

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

□

4.3.1 Fonction d'une variable aléatoire discrète

Théorème 35. *La loi image μ_f d'une loi discrète μ par une application mesurable f est une loi discrète.*

Démonstration. Soit D un ensemble fini ou dénombrable tel que $\mu(D) = 1$. $f(D)$ est un ensemble fini ou dénombrable. Il est donc mesurable par rapport à la mesure de Lebesgue. On a $\mu_f(f(D)) = \mu(f^{-1}(f(D)) \geq \mu(D) = 1$ car $f^{-1}(f(D) \supset D$. Il s'ensuit que $f(D)$ est un ensemble fini ou dénombrable dont la mesure sous μ_f est 1. □

Corollaire 5. *Soit X une variable aléatoire discrète définie sur un espace probabilisé $(\Omega, \mathcal{F}, P)$ et f une fonction quelconque définie sur $X(\Omega)$. Alors, la fonction Y définie par*

$$\forall \omega \in \Omega \quad Y = f(X(\omega))$$

est une variable aléatoire discrète sur $(\Omega, \mathcal{F}, P)$.

De manière plus concise, on écrit $Y = f(X)$.

Exemple : Soit X une variable aléatoire vérifiant $X(\Omega) = \{-1; 0; 1\}$, avec $P(X = -1) = P(X = 0) = P(X = 1) = \frac{1}{3}$.

On pose $Y = X^2$.

On a $Y(\Omega) = \{0; 1\}$, avec

$$\{Y = 0\} = \{X = 0\}$$

et

$$\{Y = 1\} = \{X = 1\} \cup \{X = -1\},$$

d'où

$$P(Y = 0) = P(X = 0) = \frac{1}{3}$$

et

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

4.4 Variables et vecteurs aléatoires à densité

On dit qu'une loi μ est à densité (sous-entendu par rapport à la mesure de Lebesgue) s'il existe une fonction mesurable f qui soit une densité de μ par rapport à la mesure de Lebesgue.

Ainsi, si f est une densité de la loi μ, on a pour tout borélien A

$$\mu(A) = \int_A f(\omega) d\lambda(\omega).$$

Exercice : Montrer que si f est la densité d'une loi, alors $\lambda(f < 0) = 0$.
Évidemment, si f est la densité d'une loi, on a

$$1 = \mu(\mathbb{R}^d) = \int_{\mathbb{R}^d} f(\omega) d\lambda(\omega).$$

Réciproquement, si f est une fonction mesurable, positive λ presque partout et d'intégrale 1, $\mu = f.\lambda$ est une mesure de probabilité admettant f pour densité.

Ainsi, on dit qu'une variable (ou un vecteur) aléatoire X est à densité si sa loi P_X est à densité.

4.4.1 Premières propriétés

Soit X une variable aléatoire de densité f. f a les propriétés suivantes :

- $\forall a, b \in \mathbb{R} \quad a \leq b \Longrightarrow \quad P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b) = \int_a^b f(x)\, dx.$
- $\forall a \in \mathbb{R} \quad P(a \leq X) = P(a < X) = \int_a^{+\infty} f(x)\, dx.$
- $\forall a \in \mathbb{R} \quad P(a \geq X) = P(a > X) = \int_{-\infty}^a f(x)\, dx.$
- $\int_{-\infty}^{+\infty} f(x)\, dx = 1.$

4.4.2 Densités et lois marginales

Théorème 36. *Soit $(\Omega, \mathcal{F}, \mu)$ et $(\Omega', \mathcal{F}', \mu')$ deux espaces mesurés. On suppose que la loi ν sur $(\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}')$ admet une densité h par rapport à $\mu \otimes \mu'$. Alors la loi image de ν par l'application*

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

admet comme densité par rapport à μ la fonction f définie par $f(x) = \int_{\Omega'} h(x, x') d\mu'(x')$.

Démonstration. Soit $B \in \mathcal{F}$

$$\begin{aligned}
\pi\nu(B) &= \nu(\pi^{-1}(B)) \\
&= \nu(B \times \Omega') \\
&= \int_{B\times\Omega'} d\nu(x,x') \\
&= \int_{B\times\Omega'} h(x,x')d(\mu\otimes\mu')(x,x') \\
&= \int_B (\int_{\Omega'} h(x,x')d\mu'(x'))d\mu(x) \\
&= \int_B f(x)d\mu(x),
\end{aligned}$$

ce qui prouve le résultat. $\square$

Théorème 37. *Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^n$ et Y un vecteur aléatoire à valeurs dans $\mathbb{R}^p$ définis sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$. Si $h(x,y)$ est une densité de (X,Y) par rapport à la mesure de Lebesgue sur $\mathbb{R}^{n+p}$, alors X admet la densité $f(x) = \int_{\mathbb{R}^p} h(x,y)d\lambda^p(y)$, tandis que*
Y admet la densité $g(x) = \int_{\mathbb{R}^n} f(x,y)d\lambda^n(x)$.

Démonstration. Le diagramme commutatif ci-dessous traduit que
$X = \pi \circ (X,Y)$.

$$\begin{array}{ccc}
(\Omega, \mathcal{F}, P) & \xrightarrow{(X,Y)} & (\mathbb{R}^n \times \mathbb{R}^p, \mathcal{B}(\mathbb{R}^n)\otimes\mathcal{B}(\mathbb{R}^p), P_{(X,Y)}) \\
 & X \searrow & \downarrow \pi \\
 & & (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P_X)
\end{array}$$

Il s'ensuit que P_X est la mesure image de $P_{(X,Y)}$ par π. Comme $h(x,y)$ est la densité de (X,Y) par rapport à $\lambda^n \otimes \lambda^p = \lambda^{n+p}$, la densité de X est bien $f(x) = \int_{\mathbb{R}^p} h(x,y)d\lambda^p(y)$, On procède de même pour calculer la densité de Y. $\square$

4.4.3 Indépendance et densités

Théorème 38. *Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^n$ et Y un vecteur aléatoire à valeurs dans $\mathbb{R}^p$ définis sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$. On suppose que X admet une densité f et Y une densité g. Alors, le vecteur aléatoire (X,Y) admet la fonction $h(x,y) = f(x)g(y)$ comme densité.*

Démonstration. Comme X et Y sont indépendantes, sous P, on a $P_{X,Y} = P_X \otimes P_Y$. Ainsi

$$\begin{aligned} P_{X,Y} &= P_X \otimes P_Y \\ &= f\lambda^n \otimes g\lambda^p \\ &= ((x,y) \mapsto f(x)g(y))\lambda^n \otimes \lambda^p \\ &= h\lambda^{n+p}, \end{aligned}$$

ce qui montre bien que (la loi de) (X,Y) admet h comme densité. □

Théorème 39. *Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^n$ et Y un vecteur aléatoire à valeurs dans $\mathbb{R}^p$ définis sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$. On suppose que (X,Y) admet une densité h_1 qui s'écrive sous la forme $h_1(x,y) = f_1(x)g_1(y)$, où f et g sont des fonctions positives. Alors, X et Y sont indépendantes ; X admet comme densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$ la fonction*

$$f(x) = \frac{f_1(x)}{\int_{\mathbb{R}^n} f_1(x')\ d\lambda^n(x')}$$

et Y admet comme densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$ la fonction

$$g(y) = \frac{g_1(y)}{\int_{\mathbb{R}^p} g_1(y')\ d\lambda^p(y')}$$

Démonstration. Posons $A = \int_{\mathbb{R}^n} f_1(x)d\lambda^n(x)$ et $B = \int_{\mathbb{R}^p} g_1(y)d(\lambda^p)(y)$. Comme f_1 et f_2 sont positives, on a $A \geq 0$ et $B \geq 0$. Comme h est une densité, on a

$$\begin{aligned} 1 &= \int_{\mathbb{R}^n \times \mathbb{R}^p} h(x,y)d(\lambda^n \otimes \lambda^p)(x,y) \\ &= \int_{\mathbb{R}^n \times \mathbb{R}^p} f_1(x)g_1(y)d(\lambda^n \otimes \lambda^p)(x,y) \\ &= (\int_{\mathbb{R}^n} f_1(x)d\lambda^n(x))(\int_{\mathbb{R}^p} g_1(y)d(\lambda^p)(y)) \\ &= AB \end{aligned}$$

On en peut donc dire que A et B sont tous deux strictement positifs. Ainsi les fonctions f et g sont bien définies. Elles sont positives, et, par construction, chacune d'entre elle admet 1 comme intégrale par rapport à la mesure de Lebesgue. Ainsi $\mu = f\lambda^n$ et $\nu = g\lambda^p$ sont des mesures de probabilité.

Comme dans le théorème précédent, la densité de $\mu \otimes \nu =$ est $h(x,y) = f(x)g(y)$. Donc $h(x,y) = f(x)g(y) = \frac{f_1(x)}{A}\frac{g_1(y)}{B} = \frac{f_1(x)g_1(y)}{AB} = h_1(x,y)$.

On en déduit $P_{(X,Y)} = \mu \otimes \nu$.

Il suffit d'appliquer le corollaire 4 pour conclure.

$\square$

4.5 Variables et lois discrètes classiques

4.5.1 Indicatrice d'un événement

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

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

1_A est une variable aléatoire à valeurs dans $\{0;1\}$.

4.5.2 Masse de Dirac

On appelle masse de Dirac en un point $x \in \Omega$ la mesure δ_x définie par

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

C'est bien une loi car elle est positive et $\delta_x(\Omega) = 1$.

Remarque : Si Ω est un groupe abélien $\delta_{x+y} = \delta_x * \delta_y = \delta_y * \delta_x$.

4.5.3 Loi de Bernoulli

On appelle loi de Bernoulli de paramètre p la loi $\mu = (1-p)\delta_0 + p\delta_1$.

Ainsi, on dit qu'une variable aléatoire X suit la loi de Bernoulli de paramètre p si on a $P(X=1) = p$ et $P(X=0) = 1-p$.

Remarques importantes :

- Pour tout évènement A, $\mathbb{1}_A$ suit la loi de Bernoulli de paramètre $P(A)$.
- Réciproquement, si une variable aléatoire X suit une loi de Bernoulli, elle vérifie $X = \mathbb{1}_{X=1}$.(Réfléchir un peu...)

Ainsi les variables aléatoires qui suivent des lois de Bernoulli sont exactement les indicatrices d'événements.

4.5.4 Loi uniforme sur un ensemble

Soit $E \subset \Omega$ un ensemble fini. On appelle loi uniforme sur E la loi définie sur $\mathcal{P}(\Omega)$ par

$$\begin{array}{rcl} \mathcal{P}(\Omega) & \to & [0,1] \\ A \mapsto \dfrac{|A \cap E|}{|E|} & & \end{array}$$

Ainsi, une variable aléatoire X suit la loi uniforme sur E si l'on a

$$\forall x \in E \quad P(X = x) = \frac{1}{|E|}.$$

Exemple : La variable aléatoire X représentant le résultat du lancer d'un dé non truqué suit la loi uniforme sur l'ensemble $\{1, 2, 3, 4, 5, 6\}$.

4.5.5 Loi binomiale

On appelle loi binomiale de paramètres n et p et on note $\mathcal{B}(n,p)$ la loi de la somme de n variables de Bernoulli indépendantes de même probabilité p.

Ainsi $\mathcal{B}(n,p) = (\mathrm{Ber}(p))^{*n}$.

Théorème 40. *$\mathcal{B}(n,p)$ charge les entiers $\{0, \ldots, n\}$. Plus précisément, on a*

$$\forall k \in \{0, \ldots, n\} \quad \mathcal{B}(n,p)(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Démonstration. Posons $\mu = \mathcal{B}(n,p)$. On a

$$\begin{aligned} \mu &= (\mathrm{Ber}(p))^{*n} \\ &= ((1-p)\delta_0 + (p\delta_1)^{*n} \end{aligned}$$

Comme l'ensemble des mesures positives munie de (+,*) est une algèbre commutative, la formule du binôme de Newton s'applique et l'on a

$$\begin{aligned}
\mu &= ((1-p)\delta_0 + p\delta_1)^{*n} \\
&= \sum_{k=0}^{n} \binom{n}{k} ((1-p)\delta_0)^{*(n-k)} * (p\delta_1)^{*k} \\
&= \sum_{k=0}^{n} \binom{n}{k} ((1-p)^{n-k}(\delta_0)^{*(n-k)}) * (p^k(\delta_1)^{*k}) \\
&= \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k}\delta_0^{*(n-k)} * \delta_1^{*k} \\
&= \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k}\delta_0 * \delta_k \\
&= \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k}\delta_k
\end{aligned}$$

□

Corollaire 6. *Soit $A_1, \dots, A_n$ n événements indépendants de même probabilité p. On pose*

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

Alors X suit la loi binomiale de paramètres n et p .

Remarque : X est le nombre de A_k qui sont réalisés.

Exemple : On lance n fois une pièce de monnaie équilibrée. Le nombre X de "pile" obtenus suit une loi binomiale de paramètres n et $\frac{1}{2}$.

4.5.6 Loi géométrique

On dit qu'une variable aléatoire X suit la loi géométrique de paramètre p si l'on a

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

Théorème 41. *Soit $(A_i)_{i\in\mathbb{N}^*}$ une suite d'événements indépendants de même probabilité $p > 0$. On pose*

$$X(\omega) = \inf\{k \in \mathbb{N}^* \quad \omega \in A_k\}.$$

Alors X suit la loi géométrique de paramètre p . De plus

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

Démonstration.

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

Comme les A_i sont dans $\mathcal{F}$, les A_i^c le sont aussi, et donc comme on peut l'écrire comme intersection finie d'éléments de $\mathcal{F}$,$\{X > k\}$ est dans $\mathcal{F}$, et donc, par passage au complémentaire

$$\{X \le k\} \in \mathcal{F}.$$

En utilisant l'indépendance des (A_i), on obtient

$$P(X > k) = (1-p)^k.$$

Comme

$$\{X = +\infty\} = \bigcap_{k \ge 1} \{X > k\},$$

on obtient, par continuité séquentielle décroissante

$$P(X = +\infty) = \lim_{k \to +\infty} P(\{X > k\}) = \lim_{k \to +\infty} (1-p)^k = 0.$$

Ainsi X est bien une variable aléatoire, et l'on a

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

De plus

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

□

Exemple : On lance une pièce de monnaie équilibrée jusqu'à obtention de "pile". Le nombre de lancers effectués suit une loi géométrique de paramètre $\frac{1}{2}$.

4.5.7 Loi de Poisson

On dit qu'une variable aléatoire X suit la loi de Poisson de paramètre $\lambda > 0$ si l'on a

$$\forall k \in \mathbb{N} \quad P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

La construction de telles variables est bien possible car $e^{-\lambda}\frac{\lambda^k}{k!} \ge 0$ et

$$\sum_{k=0}^{+\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

4.5.8 Loi hypergéométrique

La loi hypergéométrique $\mathcal{H}(N,n,k)$ modélise le phénomène suivant : on tire au hasard k individus dans une population de N individus, et l'on compte le nombre d'individus possédant une certaine particularité, sachant qu'il y a exactement n personnes dans la population totale qui possédaient cette particularité.

De manière théorique, la loi hypergéométrique est la loi image de la loi uniforme sur $\Omega = \mathcal{B}(N,k)$ par l'application

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

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

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

Démonstration. Notons P la loi uniforme sur Ω. On a

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

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

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

est une bijection, donc

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

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

4.6 Lois à densité usuelles

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

On dit qu'une variable aléatoire X suit la loi uniforme sur un compact K de $\mathbb{R}^d$ $[a,b]$ si elle admet la densité

$$x \mapsto \frac{1}{\lambda(K)} 1\!\!1_K(x)$$

4.6.2 Loi uniforme sur un intervalle

On dit qu'une variable aléatoire X suit la loi uniforme sur l'intervalle $[a, b]$ si elle admet la densité

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

4.6.3 Loi gaussienne de paramètres m et σ^2

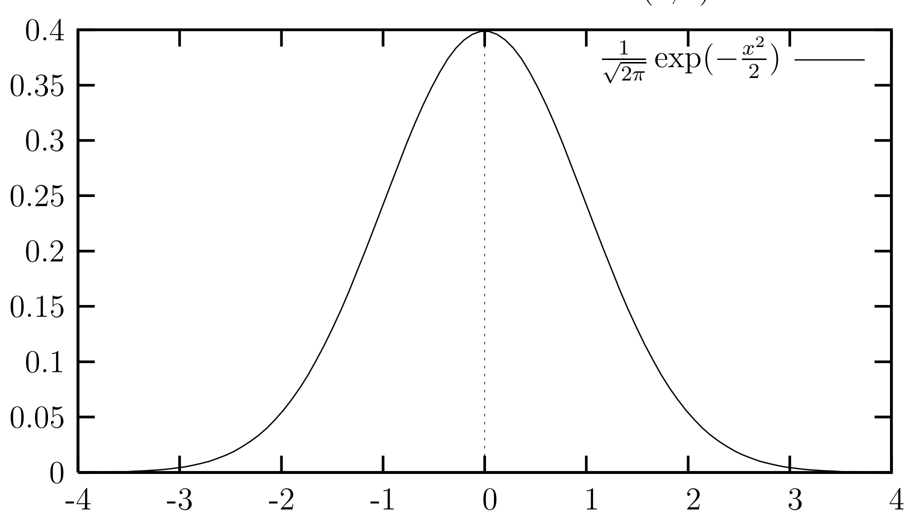
Soit $m \in \mathbb{R}$ et $\sigma^2 > 0$. On dit qu'une variable aléatoire X suit la loi gaussienne $\mathcal{N}(m, \sigma^2)$ de paramètres m et σ^2 si elle admet la densité

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

On emploie également parfois le mot "normale" à la place de "gaussienne" : ces deux mots signifient exactement la même chose. gamma On dit qu'une variable gaussienne est centrée lorsque $m = 0$.
On dit qu'une variable gaussienne est réduite lorsque $\sigma^2 = 1$.

Quelques résultats qui seront prouvés ultérieurement : si $X \sim \mathcal{N}(m, \sigma^2)$, alors $aX + b \sim \mathcal{N}(m + b, a^2\sigma^2)$. En particulier, si $X \sim \mathcal{N}(m, \sigma^2)$, alors $\frac{X-m}{\sigma} \sim \mathcal{N}(0, 1)$; et si $X \sim \mathcal{N}(0, 1)$, alors $\sigma X + m \sim \mathcal{N}(m, \sigma^2)$.

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

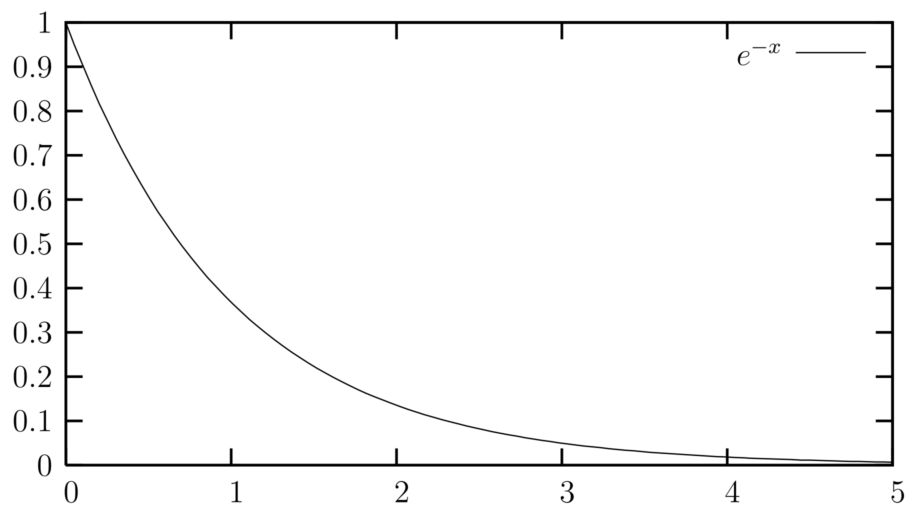


4.6.4 Loi exponentielle de paramètres a

Soit $a > 0$. On dit qu'une variable aléatoire X suit la loi exponentielle de paramètres a si elle admet la densité

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

Densité de la loi exponentielle de paramètre 1

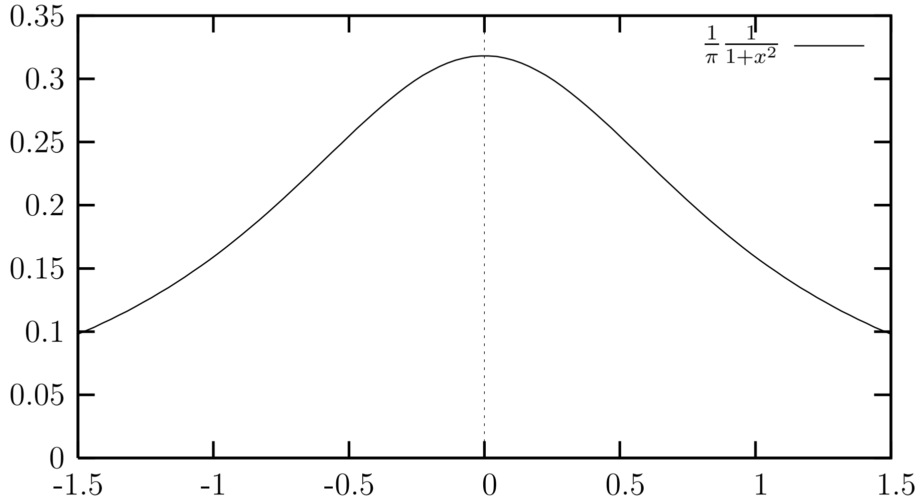


4.6.5 Lois de Cauchy

Soient $a \in \mathbb{R}, b > 0$. La loi de Cauchy $\mathcal{C}(a, b)$ admet comme densité par rapport à la mesure de Lebesgue :

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

Densité de la loi de Cauchy $\mathcal{C}(0,1)$



4.6.6 Lois Gamma

Soient a et λ des réels strictement positifs. On appelle loi Gamma $\Gamma(a,\lambda)$ la loi dont la densité par rapport à la mesure de Lebesgue est

$$x \mapsto \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \mathbb{1}_{]0,+\infty[}(x),$$

où $\Gamma(a)$ est la valeur au point a de la fonction Γ, définie par

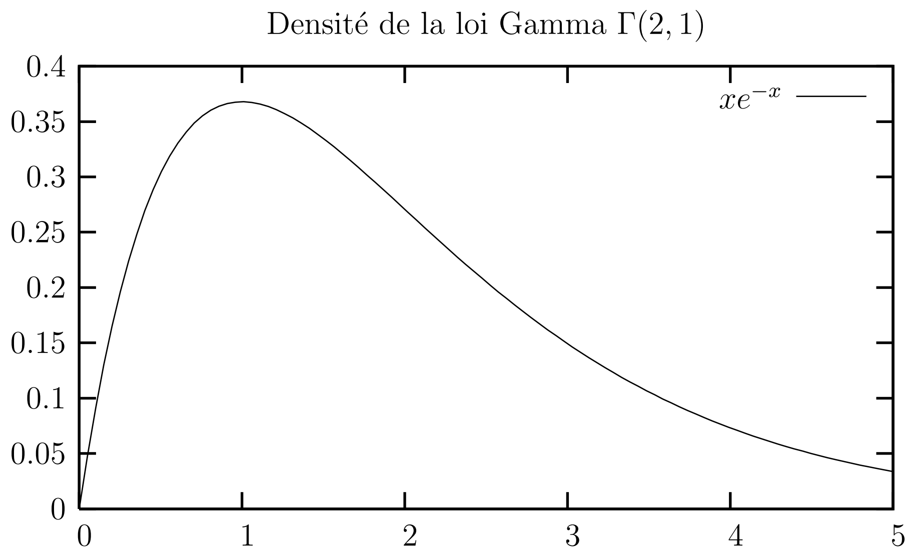
$$\Gamma(a) = \int_{\mathbb{R}+} x^{a-1} e^{-x}\ dx.$$

On dit parfois que a est le paramètre de forme et λ le paramètre d'échelle de la loi. En effet, on montrera plus loin que si $X \sim \Gamma(a,\lambda)$, alors pour tout $\mu > 0$, on a $\frac{1}{\mu}X \sim \Gamma(a, \lambda\mu)$.

On rappelle quelques propriétés classiques de la fonction Γ qui seront utiles dans la suite :

- $\forall a > 0 \quad \Gamma(a+1) = a\Gamma(a)$.
- $\forall n \in \mathbb{N} \quad \Gamma(n+1) = n!$

La preuve de ces deux propriétés sera vue en exercice.



4.7 Exercices sur les lois

1. Soit $s>1$. On dit que X suit une loi ζ de paramètre si l'on a

$$\forall n \in \mathbb{N}^* \quad P(X=n)=\frac{1}{\zeta(s)}\frac{1}{n^s},$$

où l'on a posé

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

Soit donc X suivant une loi ζ de paramètre s. On tire Y au hasard – c'est à dire avec équiprobabilité – entre 1 et X.

(a) Pour $n,k \in \mathbb{N}^*$, calculer $P(Y=k|X=n)$.

(b) On pose $Z=\frac{Y}{X}$. Montrer que la fonction de répartition F_Z est strictement croissante sur $[0,1]$.

(c) Soient p,q deux entiers positifs premiers entre eux, avec $p \le q$. Calculer $P(Z=\frac{p}{q})$.

(d) On rappelle que $\varphi(n)$ désigne le nombre d'entiers entre 1 et n qui sont premiers avec n. Déduire de ce qui précède une preuve probabiliste de l'identité

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

(e) Montrer

$$(\sum_{n=1}^{+\infty}\frac{1}{n^2})(\sum_{n=1}^{+\infty}\frac{\varphi(n)}{n^3})(\sum_{n=1}^{+\infty}\frac{\varphi(n)}{n^4}) = \int_2^3 x\ dx.$$

(On pourra admettre que $5\zeta(4) = 2\zeta(2)^2$.)

2. Donner un exemple de familles d'événements $\mathcal{C}$ et $\mathcal{D}$ telles que
   – $\forall A \in \mathcal{C} \quad \forall B \in \mathcal{D} \quad P(A \cap B) = P(A)P(B)$.
   – les tribus $\sigma(\mathcal{C})$ et $\sigma(\mathcal{D})$ ne sont pas indépendantes.
3. Donner un exemple de deux lois distinctes sur $(\Omega, \mathcal{F})$ coïncidant sur un système $\mathcal{C}$ engendrant $\mathcal{F}$.
4. On choisit de manière uniforme sur $[0,1]$ un réel Y. Quelle est la probabilité pour que le polynome

   $$p(x) = x^2 + x + Y$$

   ait des racines complexes ? des racines distinctes ?
5. Dans le segment $[AB]$ de longueur 1, on choisit au hasard un point M. Quelle est la probabilité pour que l'on ait $AM.MB \geq \frac{2}{9}$?
6. Soient $X_1, \dots, X_n, \dots$ des variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On pose $M_n = \max(X_1, \dots, X_n)$. Déterminer la fonction de répartition de M_n. Montrer que M_n admet une densité que l'on déterminera.
7. Soient $X_1, \dots, X_n, \dots$ des variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On pose $M_n = \max(X_1, \dots, X_n)$ et $m_n = \min(X_1, \dots, X_n)$. Montrer que M_n et $1 - m_n$ ont même loi.
8. Soit X une variable aléatoire suivant la loi $\mathcal{B}(2n, 1/2)$. On pose $Y = |X - n|$. Déterminer la loi de Y.
9. Soit (X,Y) un vecteur aléatoire suivant la loi uniforme sur le rectangle $[-1,2] \times [-1,1]$. Montrer que

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

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

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

(a) Montrer que $\mathcal{T} = \sigma(X)$.

(b) Déterminer le plus petit ensemble $\mathcal{T}$-mesurable contenant 1980.

(c) Montrer que A_n est $\mathcal{T}$-mesurable si et seulement si n n'est divisible par aucun carré.

(d) On munit $(\mathbb{N}^*, \mathcal{B}(\mathbb{N}^*))$ de la mesure de probabilité ζ de paramètre s, c'est à dire que

$$\forall n \in \mathbb{N}^* \quad P(\{n\}) = \frac{1}{\zeta(s)} \frac{1}{n^s},$$

où l'on a posé

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

Montrer que $P(A_n) = \frac{1}{n^s}$. À quelle condition les événements A_n et A_m sont-ils indépendants sous la loi P ?

(e) Soit $\mathcal{N} = \{\omega \in \mathbb{N}^*; \quad A_\omega \text{est } \mathcal{T} - \text{mesurable}\}$ Montrer que $\mathcal{N} = \underset{p \in \mathcal{P}}{\cap} A^c_{p^2}$, puis que $0 < P(\mathcal{N})$.

11. (a) Soit $(\mathcal{A}_n)_{n \geq 1}$ une suite de tribus indépendantes. Montrer que la tribu $\mathcal{Q} = \overset{+\infty}{\underset{k=1}{\cap}} \underset{i \geq k}{\wedge} \mathcal{A}_i$ est triviale.

(b) Soit $(A_n)_{n \geq 1}$ une suite d'événements indépendants. Soit A l'événement " une infinité de A_i se produisent. Montrer que $P(A)$ ne peut valoir que 0 ou 1

12. Démontrer les propriétés de la fonction Γ laissées en exercice :
– $\forall a > 0 \quad \Gamma(a+1) = a\Gamma(a)$.
– $\forall n \in \mathbb{N} \quad \Gamma(n+1) = n!$

13. *Queue de la gaussienne*
Soit X une variable aléatoire suivant la loi normale $\mathcal{N}(0,1)$. On pose $\psi(t) = P(X > t) = 1 - F_X(t)$. Montrer

$$\Psi(t) \sim \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-\frac{x^2}{2}}.$$

Chapitre 5

Espérances et calculs

5.1 Quelques rappels sur la construction de l'espérance

Définition Si X est une variable aléatoire intégrable définie sur $(\Omega, \mathcal{F}, P)$, on appelle espérance de X et on note $\mathbb{E}X$ le réel défini par

$$\mathbb{E}X = \int_\Omega X(\omega)\, dP(\omega).$$

Remarque : En toute rigueur, il faudrait écrire $\mathbb{E}_P X$.
Définition On note $L^1((\Omega, \mathcal{F}, P))$ l'ensemble des variables aléatoires intégrables sur $(\Omega, \mathcal{F}, P)$.
Définition Si $X = (X_1, \dots, X_n)$ est un vecteur aléatoire dont toutes les composantes sont intégrables, on note $\mathbb{E}X$ le vecteur $(\mathbb{E}X_1, \dots, \mathbb{E}X_n)$.

5.2 Quelques propriétés

- L^1 est un espace vectoriel.
- $\forall X, Y \in L^1 \quad \mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$.
- $\forall X \in L^1, \forall a \in \mathbb{R} \quad \mathbb{E}aX = a\mathbb{E}X$.
- $\forall A \in \mathcal{F} \quad \mathbb{E}(\mathbb{1}_A) = P(A)$.
- La variable aléatoire X est intégrable si et seulement si $|X|$ est intégrable.
- $\forall X \in L^1 \quad P(X \ge 0) = 1 \Longrightarrow \mathbb{E}X \ge 0$
- $\forall X \in L^1 \quad P(X \le a) = 1 \Longrightarrow \mathbb{E}X \le a$.
- $\forall X \in L^1 \quad P(X \ge b) = 1 \Longrightarrow \mathbb{E}X \ge b$.
- $\forall X \in L^1 \quad P(X = a) = 1 \Longrightarrow \mathbb{E}X = a$.
- $\forall X \in L^1 \quad P(|X| \le a) = 1 \Longrightarrow \mathbb{E}|X| \le a$.

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

Vocabulaire : on dit qu'une variable aléatoire X est centrée si $\mathbb{E}X = 0$. On définit de même ce qu'est un vecteur aléatoire centré.

5.3 Application : Formule de Poincaré et inégalités de Bonferroni

La formule de Poincaré est l'analogue de la formule du même nom du cours de dénombrement. On peut considérer que c'en est une généralisation.

Théorème 42 (Formule de Poincaré). *Pour tous événements $A_1, A_2, ..., A_n$ sous la probabilité P*

$$
\begin{aligned}
P\Big(\bigcup_{i=1}^{n} A_i\Big) &= \sum_{B\in\mathcal{P}(\{1,\dots,n\})\backslash\varnothing} (-1)^{1+|B|} P(\cap_{j\in B} A_j) && (5.1)\\
&= \sum_{i=1}^{n} P(A_i) - \sum_{1\leq i_1<i_2\leq n} P(A_{i_1}\cap A_{i_2}) + \cdots && (5.2)\\
&\quad \cdots + (-1)^{k+1} \sum_{1\leq i_1<i_2<\ldots<i_k\leq n} P(A_{i_1}\cap\cdots\cap A_{i_k}) + \cdots && (5.3)\\
&\quad \cdots + (-1)^{n+1} P(A_1\cap\cdots\cap A_n). && (5.4)
\end{aligned}
$$

Exemple : Pour $n = 3$, on a

$$P(A_1\cup A_2\cup A_3) = P(A_1)+P(A_2)+P(A_3)-P(A_1\cap A_2)-P(A_2\cap A_3)-P(A_1\cap A_3)+P(A_1\cap A_2\cap A_3).$$

Pour prouver la formule de Poincaré, on va utiliser un lemme qui va nous permettre d'obtenir des encadrements de la probabilité d'une réunion.

Lemme 3. *Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé ; $(A_x)_{x\in I}$ des événements.*

Pour $n \geq 1$, on pose

$$V_n = P_n\Big(\sum_{x\in I} \mathbb{1}_{A_x} - 1\Big)\mathbb{1}_{\cup_{x\in I} A_x},$$

où $(P_k)_{k\geq 0}$ est la suile de polynômes définie par

$$
\begin{cases}
P_0 = 1\\
P_1 = X\\
P_2 = \frac{X(X-1)}{2}\\
\dots\\
P_k = \frac{X(X-1)\dots(X-k+1)}{k!}
\end{cases}
$$

Ainsi pour $n \geq k$ $P_k(n) = \binom{n}{k}$ tandis que $P_k(n) = 0$ pour $0 \leq n < k$.
Alors

$$\forall n \in \mathbb{N}^* \quad P(\underset{x \in I}{\cup} A_x) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{J \in \mathcal{B}_k(I)} P(\underset{x \in J}{\cap} A_x) + (-1)^n \mathbb{E} V_n \quad (5.5)$$

Démonstration. Il suffit de montrer

$$\forall n \in \mathbb{N}^* \quad \mathbb{1}_{\underset{x \in I}{\cup} A_x} = \sum_{k=1}^{n} (-1)^{k+1} \sum_{J \in \mathcal{B}_k(I)} \prod_{x \in J} \mathbb{1}_{A_x} + (-1)^n V_n \qquad (5.6)$$

Soit $\omega \in \Omega$. Si $\omega \notin \cup_{x \in I} A_x$, les deux membres de l'égalité sont nuls. Sinon, posons $N = \sum_{x \in I} \mathbb{1}_{A_x}(\omega) = |\{x \in I; \omega \in A_x\}|$.

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

Ainsi $\sum_{J \in \mathcal{B}_k(I)} \prod_{x \in J} \mathbb{1}_{A_x} = P_k(N)$

On doit donc montrer

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

ce qui est équivalent à

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

On va montrer par récurrence sur n :

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

Pour $n = 0$, c'est vérifié. Pour passer de n à $n+1$, on a

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

□

Ensuite, il suffit d'intégrer (5.6) pour obtenir (5.5).

Démonstration. En prenant $I = J = \{1, \ldots, n\}$ dans le lemme précédent, on obtient

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

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

Théorème 43 (Inégalités de Bonferroni). *Soit* $(\Omega, \mathcal{F}, P)$ *un espace probabilisé ;* $(A_x)_{x\in I}$ *des événements. Soit* $n \in \mathbb{N}^*$.

Alors,

– *Si* n *est impair, on a*

$$P(\underset{x\in I}{\cup} A_x) \leq \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} P(\underset{x\in J}{\cap} A_x) \qquad (5.7)$$

– *Si* n *est pair, on a*

$$P(\underset{x\in I}{\cup} A_x) \geq \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} P(\underset{x\in J}{\cap} A_x) \qquad (5.8)$$

Démonstration. Il suffit d'appliquer le lemme 3 en remarquant que V_n est une variable aléatoire positive, et que donc son espérance l'est aussi. □

5.4 Inégalité de Jensen

Théorème 44. *Soit X une variable aléatoire intégrable à valeurs dans l'intervalle I. Soit f une fonction convexe de I dans $\mathbb{R}$. Alors*

$$\varphi(\mathbb{E}X) \le \mathbb{E}\varphi(X).$$

Démonstration. Soit Φ l'ensemble des fonctions affines φ telles que

$$\forall x \in I \quad \varphi(x) \le f(x).$$

Soit $\varphi \in \Phi$. On a presque sûrement

$$\varphi(X) \le f(X).$$

On a donc $\mathbb{E}\varphi(X) \le \mathbb{E}f(X)$. Mais comme φ est une fonction affine, $\mathbb{E}\varphi(X) = \varphi(\mathbb{E}X)$. Ainsi,

$$\forall \varphi \in \Phi \quad \varphi(\mathbb{E}X) \le \mathbb{E}f(X),$$

donc

$$\sup_{\varphi \in \Phi} \varphi(\mathbb{E}X) \le \mathbb{E}f(X).$$

Mais pour tout x $\sup\{\varphi(x); \varphi \in \Phi\} = f(x)$ (il suffit de considérer l'application affine tangente à droite (ou à gauche) au point x), et donc finalement en appliquant cette propriété au point $\mathbb{E}X$, on a

$$f(\mathbb{E}X) \le \mathbb{E}f(X).$$

□

Pour retenir quel est le sens de l'égalité, prendre la fonction convexe $\varphi(x) = |x|$.

5.5 Intégrale et queue de distribution

Théorème 45. *Soit X une variable aléatoire positive. On a*

$$\mathbb{E}X = \int_{\mathbb{R}_+} P(X > t) \, d\lambda(t).$$

Démonstration.

$$\begin{aligned}
\int_{\mathbb{R}_+} P(X > t)\ d\lambda(t) &= \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \mathbb{1}_{]t,+\infty[}(s)\ dP_X(s)\ d\lambda(t) \\
&= \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \mathbb{1}_{[0,s[}(t)\ dP_X(s)\ d\lambda(t) \\
&= \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \mathbb{1}_{[0,s[}(t)\ d\lambda(t)\ dP_X(s) \\
&= \int_{\mathbb{R}_+} s\ dP_X(s) \\
&= \mathbb{E}X
\end{aligned}$$

□

Corollaire 7. *Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. On a*

$$\mathbb{E}X = \sum_{k=0}^{+\infty} P(X > k).$$

Démonstration. D'après le théorème précédent, $\mathbb{E}X = \int_{\mathbb{R}_+} P(X > t)\ d\lambda(t)$. Comme $t \mapsto P(X > t)$ est une fonction positive, on a

$$\int_{\mathbb{R}_+} P(X > t)\ d\lambda(t) = \sum_{k=0}^{+\infty} \int_{[k,k+1[} P(X > t)\ d\lambda(t).$$

Mais comme X est à valeurs entières, on a

$$\forall t \in [k, k+1[\quad P(X > t) = P(X > k),$$

d'où le résultat. □

5.6 Théorèmes de transfert

Théorème 46. *Soit X est une variable aléatoire dont la loi P_X admet une densité f par rapport à la mesure m. Soit g une fonction mesurable.*

Alors, $g(X)$ est intégrable si et seulement si

$$\int_{\mathbb{R}} |g(x)|d(x)m(x) < +\infty.$$

Si cette intégrale est finie, on a alors

$$\mathbb{E}g(X) = \int_{\mathbb{R}} |g(x)|f(x)m(x) < +\infty.$$

Démonstration. D'après le théorème de transfert

$$\begin{aligned} \int_\Omega |g(X(\omega))| dP(\omega) &= \int_{\mathbb{R}} |g(x)| dP_X(x) \\ &= \int_{\mathbb{R}} |g(x)| f(x)\ dm(x) \end{aligned}$$

De même, si cette quantité est finie, le théorème de transfert nous dit encore que

$$\begin{aligned} \int_\Omega g(X(\omega)) dP(\omega) &= \int_{\mathbb{R}} g(x) dP_X(x) \\ &= \int_{\mathbb{R}} g(x) f(x)\ dm(x) \end{aligned}$$

□

5.6.1 Calcul de l'espérance d'une variable aléatoire discrète

Théorème 47. *Soit X est une variable aléatoire discrète, et D un ensemble fini ou dénombrable inclus dans $\mathbb{R}$ tel que $X(\Omega) = D$. Soit g une fonction quelconque de D dans $\mathbb{R}$. Alors, la variable aléatoire $Y = g(X)$ est intégrable si et seulement si*

$$\sum_{i \in D} |g(i)| p_i < +\infty,$$

où l'on a posé $p_i = P(X = i)$. Si cette somme est finie, on a alors

$$\mathbb{E}Y = \mathbb{E}g(X) = \sum_{i \in D} g(i) p_i.$$

Démonstration. D'après nos hypothèses, P_X admet une densité par rapport à la mesure de comptage de support D : c'est la fonction $i \mapsto p_i$. Il suffit donc d'appliquer le théorème 46 en prenant pour m la mesure de comptage sur D et pour f :

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

On a alors

$$\int_{\mathbb{R}} |g(x)| f(x)\ dm(x) = \sum_{x \in D} |g(x)| f(x),$$

et, si cette somme est finie :

$$\int_{\mathbb{R}} g(x)f(x)\ dm(x) = \sum_{x\in D} g(x)f(x).$$

□

Corollaire 8. *Soit X est une variable aléatoire discrète, et D un ensemble fini ou dénombrable inclus dans $\mathbb{R}$ tel que $X(\Omega) = D$. Alors, X est intégrable si et seulement si*

$$\sum_{i\in D} |i|p_i < +\infty,$$

où l'on a posé $p_i = P(X = i)$. Si cette somme est finie, on a alors

$$\mathbb{E}X = \sum_{i\in D} ip_i.$$

Démonstration. Il suffit d'appliquer le thèorème précédent avec $g(x) = x$. □

5.6.2 Calcul de l'espérance d'une variable aléatoire à densité

Voici maintenant le théorème de transfert pour les variables à densité

Théorème 48. *Soit X est une variable aléatoire admettant la fonction f comme densité, et g une fonction continue par morceaux définie sur $X(\Omega)$. Alors, la variable aléatoire $Y = g(X)$ est intégrable si et seulement si*

$$\int_{\mathbb{R}} |g(x)|f(x)\ d\lambda(x) < +\infty.$$

Si cette intégrale est convergente, on a alors

$$\mathbb{E}Y = \mathbb{E}g(X) = \int_{\mathbb{R}} g(x)f(x)\ d\lambda(x).$$

Démonstration. Il suffit d'appliquer le théorème 46 avec pour m la mesure de Lebesgue. □

Corollaire 9. *Soit X est une variable aléatoire admettant la fonction f comme densité. Alors, X est intégrable si et seulement si*

$$\int_{\mathbb{R}} |x|f(x)\ d\lambda(x) < +\infty.$$

Si cette intégrale est convergente, on a alors

$$\mathbb{E}X = \int_{\mathbb{R}} xf(x)\ dx.$$

Démonstration. Il suffit d'appliquer le thèorème précédent avec $g(x) = x$. □

Remarque importante : Si la densité de X est paire et que X est intégrable, alors $\mathbb{E}X = 0$.

Démonstration. On a

$$\begin{aligned} \mathbb{E}X &= \int_{-\infty}^{0} xf(x)\ dx + \int_{0}^{+\infty} xf(x)\ dx \\ &= \int_{0}^{+\infty} -xf(-x)\ dx + \int_{0}^{+\infty} xf(x)\ dx \\ &= \int_{0}^{+\infty} x(f(x) - f(-x))\ dx \\ &= 0 \end{aligned}$$

car $f(x) = f(-x)$. □

5.7 Moments d'ordre 2

On dit qu'une variable aléatoire X admet un moment d'ordre 2 si elle est de carré intégrable, c'est à dire si $X^2 \in L^1$.

On note $L^2(\Omega, \mathcal{F}, P)$ (ou encore L^2) l'ensemble des variables aléatoires de carré intégrable.

Lemme 4. *Soient $X, Y \in L^2$. Alors la variable aléatoire XY est intégrable.*

Démonstration. Pour tous les réels a, b, on a $|ab| \leq \frac{1}{2}(a^2 + b^2)$.
(En effet, $a^2+b^2+2ab = (a+b)^2 \geq 0$, d'où $(a^2+b^2)/2 \geq -ab$ et $a^2+b^2-2ab = (a-b)^2 \geq 0$, d'où $(a^2+b^2)/2 \geq ab$.)
On a donc

$$0 \leq |XY| \leq \frac{1}{2}(X^2 + Y^2),$$

Comme $X^2 + Y^2$ est intégrable, on en déduit que $|XY|$ est intégrable, ce qui est ce que l'on voulait montrer □

Corollaire 10. $L^2(\Omega, \mathcal{F}, P)$ *est un espace vectoriel.*

Démonstration. La stabilité par multiplication ne pose pas de problème. Pour la stabilité par addition, il faut remarquer que

$$(X+Y)^2 = X^2 + Y^2 + 2XY,$$

puis utiliser le lemme précédent et le fait que L^1 est un espace vectoriel. □

5.7.1 Covariance et variance

Soient X et Y deux variables aléatoires admettant chacune un moment d'ordre 2. On appelle covariance du couple (X,Y) le nombre

$$\mathrm{Covar}(X,Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)$$

On appelle variance de X le nombre

$$\mathrm{Var}\, X = \mathrm{Covar}(X,X) = \mathbb{E}(X - \mathbb{E}X)^2.$$

On appelle écart-type de X le nombre

$$\sigma(X) = (\mathrm{Var}\, X)^{1/2}.$$

Vocabulaire : On dit qu'une variable aléatoire est réduite si on a $\mathrm{Var}\, X = 1$ (ou de manière équivalente si $\sigma(X) = 1$).

On a les propriétés suivantes

1. $(X,Y) \mapsto \mathrm{Covar}(X,Y)$ est une forme bilinéaire symétrique positive.
2. $\forall a,b \in \mathbb{R} \quad \mathrm{Covar}(X-a, Y-b) = \mathrm{Covar}(X,Y)$.
3. $\mathrm{Var}(X+Y) = \mathrm{Var}\, X + \mathrm{Var}\, Y + 2\,\mathrm{Covar}(X,Y)$.
4. $\mathrm{Covar}(X,Y) = \mathbb{E}XY - (\mathbb{E}X)(\mathbb{E}Y)$.
5. $\mathrm{Var}\, X = \mathbb{E}X^2 - (\mathbb{E}X)^2$.
6. $|\mathbb{E}XY|^2 \le \mathbb{E}X^2\mathbb{E}Y^2$ (inégalité de Cauchy-Schwarz)
7. $|\mathrm{Covar}(X,Y)| \le \sigma(X)\sigma(Y)$.

Démonstration. 1. Notons $\langle X,Y\rangle = \mathbb{E}X,Y$. Il est facile de voir que $(X,Y) \mapsto \langle X,Y\rangle$ est une forme bilinéaire symétrique positive. Posons $L(X) = X - \mathbb{E}X$. $X \mapsto L(X)$ est une application linéaire de L^2 dans lui-même. On a $\mathrm{Covar}(X,Y) = \langle L(X), L(Y)\rangle$. Les deux observations faites ci-dessus permettent de dire que $(X,Y) \mapsto \mathrm{Covar}(X,Y)$ est une forme bilinéaire symétrique positive.

2.

$$\begin{aligned}\operatorname{Covar}(X-a,Y-b) &= \langle L(X-a), L(Y-b)\rangle \\ &= \langle L(X)-L(a), L(Y)-L(b)\rangle \\ &= \langle L(X)-0, L(Y)-0\rangle \\ &= \operatorname{Covar}(X,Y)\end{aligned}$$

3.

$$\begin{aligned}\operatorname{Var}(X+Y) &= \operatorname{Covar}(X+Y,X+Y) \\ &= \operatorname{Covar}(X,X)+2\operatorname{Covar}(X,Y)+\operatorname{Covar}(Y,Y) \\ &= \operatorname{Var} X + 2\operatorname{Covar}(X,Y)+\operatorname{Var} Y\end{aligned}$$

Pour passer de la ligne 1 à la ligne 2, on utilise le fait que la covariance est bilinéaire symétrique.

4. $(X-\mathbb{E}X)(Y-\mathbb{E}Y) = XY + \mathbb{E}X\mathbb{E}Y - (\mathbb{E}X)Y - (\mathbb{E}Y)X$.
D'où

$$\begin{aligned}\mathbb{E}(X-\mathbb{E}X)(Y-\mathbb{E}Y) &= \mathbb{E}XY + \mathbb{E}(EX\mathbb{E}Y) - \mathbb{E}X\mathbb{E}Y - \mathbb{E}Y\mathbb{E}X \\ &= \mathbb{E}XY + \mathbb{E}X\mathbb{E}Y - 2\mathbb{E}X\mathbb{E}Y \\ &= \mathbb{E}XY - (\mathbb{E}X)(\mathbb{E}Y).\end{aligned}$$

5. Il suffit d'appliquer la formule précédente avec $X = Y$.
6. Comme $\langle .,.\rangle$ est une forme bilinéaire symétrique positive, l'inégalité de Cauchy-Schwarz s'applique.
7. Il suffit d'appliquer l'inégalité de Cauchy-Schwarz à la forme bilinéaire symétrique positive Covar , puis de prendre la racine carrée.

□

Lorsque $\sigma(X)$ et $\sigma(Y)$ sont non nuls, on définit le coefficient de corrélation de X et Y par

$$\operatorname{Corr}(X,Y) = \frac{\operatorname{Covar}(X,Y)}{\sigma(X)\sigma(Y)}.$$

D'après ce qui précède, $\operatorname{Covar}(X,Y) \in [-1;1]$. Lorsque $\operatorname{Covar}(X,Y) = 0$ (ce qui implique $\operatorname{Corr}(X,Y) = 0$ si $\sigma(X)$ et $\sigma(Y)$ sont non nuls), on dit que X et Y ne sont pas corrélées.
Lorsque $\operatorname{Covar}(X,Y) \geq 0$ (ce qui implique $\operatorname{Corr}(X,Y) \geq 0$ si $\sigma(X)$ et $\sigma(Y)$ sont non nuls), on dit que X et Y sont positivement corrélées.
Lorsque $\operatorname{Covar}(X,Y) \leq 0$ (ce qui implique $\operatorname{Corr}(X,Y) \leq 0$ si $\sigma(X)$ et $\sigma(Y)$ sont non nuls), on dit que X et Y sont négativement corrélées.

5.7.2 Matrice de covariance

Si $X = (X_1, \dots, X_n)$ est un vecteur aléatoire dont toutes les composantes admettent un moment d'ordre deux, on convient de dire que le vecteur a un moment d'ordre deux et on appelle matrice de covariance de X la matrice $n \times n$ dont les coefficients sont $(\mathrm{Covar}(X_i, X_j))_{1 \le i,j \le n}$.

Théorème 49. *Si $X = (X_1, \dots, X_n)$ est un vecteur aléatoire admettant un moment d'ordre deux , la matrice de covariance de X est la matrice dans la base canonique de l'application bilinéaire positive*

$$\begin{array}{rcl} \mathbb{R}^n \times \mathbb{R}^n & \to & \mathbb{R} \\ (a,b) & \mapsto & \mathrm{Covar}(\langle X, a\rangle, \langle X, b\rangle) \end{array}$$

C'est une matrice symétrique positive.

Démonstration. À X fixé, l'application $X \mapsto \langle X, a\rangle$ est une application linéaire. Comme on a déjà montré que Covar était une forme bilinéaire symétrique positive, il s'ensuit que l'application considérée ici est une forme bilinéaire symétrique positive. Cette application envoie le couple (e_i, e_j) sur $\mathrm{Covar}(\langle X, e_i\rangle, \langle X, e_j\rangle) = \mathrm{Covar}(X_i, X_j)$. La matrice d'une forme bilinéaire symétrique positive est une matrice symétrique positive. □

Théorème 50. *Soit $X = (X_1, \dots, X_n)$ est un vecteur aléatoire admettant un moment d'ordre deux et de matrice de covariance C_X et d'espérance m_X. Soit A une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$, et b un vecteur de $\mathbb{R}^p$. Alors $Y = AX+b$ admet $C_Y = AC_XA^*$ comme matrice de covariance et l'espérance de Y vaut $Am_X + b$.*

Démonstration.

$$\begin{aligned} \mathbb{E}Y_i &= \mathbb{E}(\sum_{k=1}^n a_{i,k}X_k + b_i) \\ &= \sum_{k=1}^n a_{i,k}\mathbb{E}X_k + b_i \\ &= \sum_{k=1}^n a_{i,k}m_k + b_i \\ &= (Am+b)_i \end{aligned}$$

$$\begin{aligned}
\mathrm{Covar}(\langle Y,a\rangle,\langle Y,b\rangle) &= \mathbb{E}\,\mathrm{Covar}(\langle AX+c,a\rangle,\langle AX+c,b\rangle)\\
&= \mathrm{Covar}(\langle AX,a\rangle,\langle AX,b\rangle)\\
&= \mathrm{Covar}(\langle X,A^*a\rangle,\langle X,A^*b\rangle)\\
&= \langle C_X A^*a, A^*b\rangle\\
&= \langle AC_X A^*a, b\rangle
\end{aligned}$$

□

5.7.3 Espérance et indépendance

Le théorème suivant est très important :

Théorème 51. *Soient X, Y deux variables aléatoires intégrables indépendantes. Alors, leur produit XY est une variable aléatoire intégrable et l'on a*

$$\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y.$$

Démonstration. D'après le théorème de transfert, on a

$$\mathbb{E}|XY| = \int_{\mathbb{R}^2} |xy|\ dP_{(X,Y)}$$

Il vient

$$\begin{aligned}
\mathbb{E}|XY| &= \int_{\mathbb{R}^2} |xy|\ dP_{(X,Y)}\\
&= \int_{\mathbb{R}^2} |x|.|y|\ d(P_X\otimes P_Y)(x,y)\\
&= \int_{\mathbb{R}} |x| dP_X(x).\int_{\mathbb{R}} |y|\ dP_Y(y)\\
&= \mathbb{E}|X|.\mathbb{E}|Y|\\
&< +\infty
\end{aligned}$$

Ainsi, le théorème de Fubini nous a permis de montrer que XY était intégrable.

Maintenant, et on a

$$\begin{aligned}
\mathbb{E}XY &= \int_{\mathbb{R}^2} xy\ dP_{(X,Y)} \\
&= \int_{\mathbb{R}^2} xy\ d(P_X \otimes P_Y)(x,y) \\
&= \int_{\mathbb{R}} x dP_X(x). \int_{\mathbb{R}} y\ dP_Y(y) \\
&= \mathbb{E}X\mathbb{E}Y
\end{aligned}$$

□

Corollaire 11. *Soient X, Y deux variables aléatoires intégrables indépendantes. Alors X et Y ne sont pas corrélées.*

Démonstration. On a $\mathrm{Covar}(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = 0$. □

Remarque importante : Des variables aléatoires peuvent être non corrélées sans être indépendantes.
Exemple : soient deux variables aléatoires vérifiant

$P(\{X = 1\} \cap \{Y = 1\}) = P(\{X = 1\} \cap \{Y = -1\}) = P(\{X = -1\} \cap \{Y = 0\}) = 1/3.$

La matrice M associée à la loi du couple est

$$\begin{pmatrix} 0 & 1/3 & 0 \\ 1/3 & 0 & 1/3 \end{pmatrix}$$

La loi de Y s'obtient en faisant la somme des lignes : on obtient

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 \end{pmatrix}$$

On a donc $\mathbb{E}Y = \frac{1}{3} \times (-1) + \frac{1}{3} \times (0) + \frac{1}{3} \times (1) = 0$.
D'autre part $\mathbb{E}XY = \sum_{i \in \{-1;1\}} \sum_{j \in \{-1;0;1\}} ijP(\{X = i\} \cap \{Y = j\}) = 1/3 - 1/3 = 0$.
On a donc $\mathrm{Covar}(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = 0$.
Cependant

$$0 = P(\{X = 1\} \cap \{Y = 0\}) \neq P(X = 1)P(Y = 0) = \frac{2}{3} \times \frac{1}{3}.$$

Corollaire 12. *Soient X, Y deux variables aléatoires indépendantes de carré intégrable. Alors on a*

$$\mathrm{Var}(X + Y) = \mathrm{Var}\, X + \mathrm{Var}\, Y.$$

Démonstration. On a toujours $\mathrm{Var}\, X+Y = \mathrm{Var}\, X + \mathrm{Var}\, Y + 2\,\mathrm{Covar}(X,Y)$.Comme X et Y sont indépendantes, elles ne sont pas corrélées, d'où le résultat. □

5.8 Calculs de lois images

On rappelle un théorème de théorie de la mesure de Lebesgue très important :

Proposition 3. *Soit O_1 et O_2 deux ouverts de $\mathbb{R}^d, d \geq 1$. On suppose que T est un C^1-difféomorphisme de O_1 dans O_2. Soit maintenant μ_1 une mesure positive sur $\mathbb{R}^d$ telle que $\mu_1(\mathbb{R}^d \backslash O_1) = 0$ et admettant une densité f_1 par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. Alors, la mesure image de μ_1 par T admet comme densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ la fonction f_2 définie par*

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

5.8.1 Exemple fondamental

Théorème 52. *Soit $A \in Gl_d(\mathbb{R})$ et $b \in \mathbb{R}^d$. On suppose que le vecteur aléatoire X admet la densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. Alors, le vecteur aléatoire $Y = AX + b$ admet la densité*

$$g(y) = \frac{1}{|\det A|} f(A^{-1}(y-b)).$$

Démonstration. Ici $O_1 = O_2 = \mathbb{R}^d$ et $T^{-1}(y) = A^{-1}(y-b)$. La différentielle en un point d'une transformation affine se confond avec l'application linéaire associée, d'où le résultat. □

Applications :

1. Si X suit la loi uniforme sur un compact K de $\mathbb{R}^d$, alors $Y = AX + b$ suit la loi uniforme sur l'image de K par $x \mapsto Ax+b$. Cette application est particulièrement intéressante en dimension 1.
2. Si $X \sim \Gamma(a, \lambda)$, alors pour tout $\mu > 0$, on a $\frac{1}{\mu}X \sim \Gamma(a, \lambda\mu)$.
3. Si X suit la loi exponentielle de paramètre e et que $a > 0$ alors $\frac{1}{\mu}X$ suit la loi exponentielle de paramètre μa. (Remarquer que ceci constitue un cas particulier de la remarque prédente.)
4. Pour $a \in \mathbb{R}$ et $b > 0$, X suit la loi de Cauchy $\mathcal{C}(0,1)$ si et seulement si $Y = bX + a$ suit la loi de Cauchy $\mathcal{C}(a, b)$.
5. Soit $\sigma > 0$, $m \in \mathbb{R}$. On a $X \sim \mathcal{N}(m, \sigma^2) \iff \frac{X-m}{\sigma} \sim \mathcal{N}(0,1)$.

5.8.2 Application aux lois gaussiennes

Lemme 5. *Soient $X = (X_1, X_2)$ un vecteur aléatoire formé de deux variables aléatoires indépendantes suivant la loi normale $\mathcal{N}(0,1)$. On pose $Y_1 = \cos\theta X_1 + \sin\theta X_2$ et $Y_2 = -\sin\theta X_1 + \cos\theta X_2$. Alors Y_1 et Y_2 sont deux variables aléatoires indépendantes suivant la loi normale $\mathcal{N}(0,1)$.*

Démonstration. Si l'on note, pour $x \in \mathbb{R}^2$, $x = (x_1, x_2)$, la densité de X est

$$\frac{1}{\sqrt{2\pi}}\exp(-\frac{x_1^2}{2}) \times \frac{1}{\sqrt{2\pi}}\exp(-\frac{x_2^2}{2}) = \frac{1}{2\pi}\exp(-\frac{x_1^2+x_2^2}{2}) = \frac{1}{2\pi}\exp(-\frac{\|x\|_2^2}{2}).$$

On a donc $Y = MX$, avec

$$M = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Ainsi, le vecteur $Y = MX$ admet pour densité

$$y \mapsto \frac{1}{\det M}\frac{1}{2\pi}\exp(-\frac{\|M^{-1}y\|_2^2}{2}).$$

M est une matrice de rotation, donc son déterminant vaut 1 et c'est une isométrie pour la norme euclidienne, ce qui implique que pour tout $x \in \mathbb{R}^2$ on a $\|M^{-1}y\|_2 = \|y\|_2$: la densité de Y est donc

$$y \mapsto \frac{1}{2\pi}\exp(-\frac{\|y\|_2^2}{2}),$$

ce qui est précisément la densité de X : Y a donc même loi que X, donc ses composantes Y_1 et Y_2 sont indépendantes et suivent la loi normale $\mathcal{N}(0,1)$. □

Théorème 53. *Soient U_1 et U_2 deux variables aléatoires indépendantes, avec $U_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ et $U_2 \sim \mathcal{N}(m_2, \sigma_2^2)$. Alors $U_1 + U_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.*

Démonstration. Si $\sigma_1 = 0$ ou $\sigma_2 = 0$, la variable aléatoire associée est constante est donc le résultat provient de la remarque faite plus haut – l'application affine est une translation.

Supposons donc $\sigma_1 > 0$ et $\sigma_2 > 0$. On pose $X_1 = \frac{U_1 - m_1}{\sigma_1}$ et $X_2 = \frac{U_2 - m_2}{\sigma_2}$. On peut trouver θ tel que $\cos\theta = \frac{\sigma_1}{\sqrt{\sigma_1^2+\sigma_2^2}}$ et $\sin\theta = \frac{\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}$. Alors, si on pose $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$, on a

$$U_1 + U_2 = m_1 + m_2 + \sigma(\cos\theta X_1 + \sin\theta X_2).$$

D'après le lemme, $\cos\theta X_1 + \sin\theta X_2 \sim \mathcal{N}(0,1)$, donc $U_1 + U_2 \sim \mathcal{N}(m_1 + m_2, \sigma^2)$. □

5.8.3 Application : convolution de deux lois à densité

Théorème 54. *Soient X et Y deux variables aléatoires indépendantes sur $(\Omega, \mathcal{F}, P)$, de densité f et g. Alors $Z = X + Y$ admet comme densité la fonction*

$$x \mapsto \int_{-\infty}^{+\infty} f(x-t)g(t)\ dt.$$

Si, de plus, X et Y sont à valeurs positives, alors la densité est simplement

$$x \mapsto \mathbb{1}_{\mathbb{R}_+}(x) \int_0^x f(x-t)g(t)\ dt.$$

Démonstration. On pose

$$\begin{pmatrix} Z \\ T \end{pmatrix} = A \begin{pmatrix} X \\ Y \end{pmatrix}, \text{ avec } A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

La densité de (X, Y) est $h(x, y) = f(x)g(y)$. D'après l'exemple fondamental, la densité de (Z, T) est

$$g(z,t) = \frac{1}{|\det A|} f(A^{-1} \begin{pmatrix} z \\ t \end{pmatrix}).$$

On a

$$\det A = 1 \text{ et } A^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix},$$

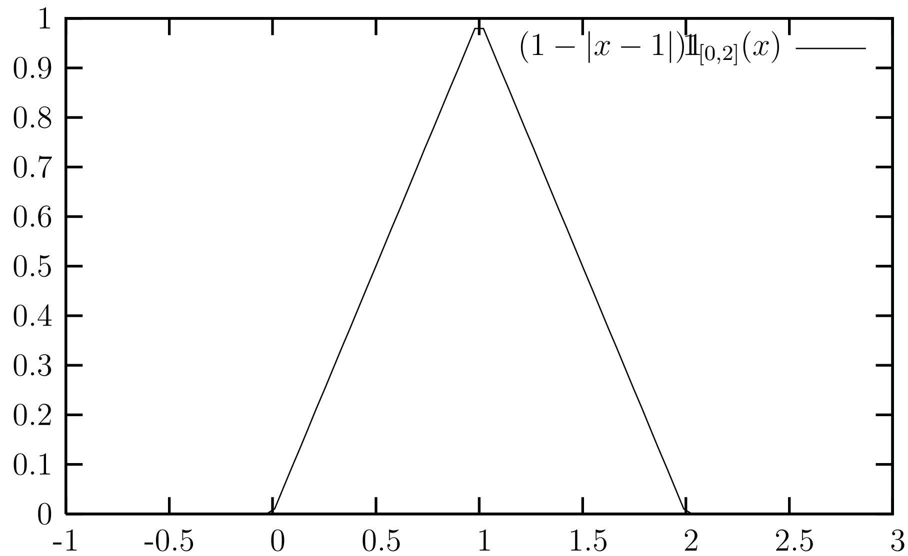
donc $g(z, t) = h(z - t, t) = f(z - t)g(t)$.

D'après le théorème 36, Z admet comme densité par rapport à la mesure de Lebesgue :

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

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

Exemple : ci-dessous, le graphe de la densité de Z lorsque X et Y suivent toutes les deux la loi uniforme sur $[0, 1]$.



Application : $\Gamma(a,\lambda) * \Gamma(b,\lambda) = \Gamma(a+b,\lambda)$

Théorème 55. *Soit a, b, λ strictement positifs, X et Y deux variables aléatoires indépendantes, X suivant la loi $\Gamma(a,\lambda)$ et Y la loi $\Gamma(b,\lambda)$. Alors $Z = X + Y$ suit la loi $\Gamma(a+b,\lambda)$.*

Démonstration. Pour tous a et λ strictement positifs, on note $f_{a,\lambda}$ la densité de la loi $\Gamma(a,\Lambda)$, soit (rappel)

$$f_{a,\lambda}(x) = 1\!\!1_{\mathbb{R}_+}(x)\frac{\lambda^a}{\Gamma(a)}x^{a-1}e^{-\lambda x}.$$

D'après le théorème précédent, Z admet une densité f_Z. Cette densité est nulle sur $\mathbb{R}_-$ tandis que pour x positif, on a

$$\begin{aligned} f_Z(x) &= \int_0^x f_{a,\lambda}(x-t)f_{b,\lambda}(t)\ dt \\ &= \int_0^x \frac{\lambda^a}{\Gamma(a)}t^{a-1}e^{-\lambda t}\frac{\lambda^b}{\Gamma(b)}(x-t)^{b-1}e^{-\lambda(x-t)}\ dt \\ &= \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)}\int_0^x t^{a-1}(x-t)^{b-1}\ dt \end{aligned}$$

On fait le changement de variable $t = \theta x$. On obtient

$$\begin{aligned} f_Z(x) &= \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} x^{a-1} x^{b-1} x \int_0^1 \theta^{a-1}(1-\theta)^{b-1}\, d\theta \\ &= \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} x^{a+b-1} \int_0^1 \theta^{a-1}(1-\theta)^{b-1}\, d\theta \\ &= K_{a,b} f_{a+b,\lambda}(x), \end{aligned}$$

où $K_{a,b} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \theta^{a-1}(1-\theta)^{b-1}\, d\theta$. Evidemment f_Z et $x \mapsto K_{a,b} f_{a+b,\lambda}(x)$ coincident également sur $\mathbb{R}_-$ où elles sont nulles. On a donc

$$\int_{\mathbb{R}} f_Z(x)\, d\lambda(x) = \int_{\mathbb{R}} K_{a,b} f_{a+b,\lambda}(x)\, d\lambda(x) = K_{a,b} \int_{\mathbb{R}} f_{a+b,\lambda}(x).$$

Mais f_Z et $f_{a+b,\lambda}$ sont des densités donc leur intégrale sur $\mathbb{R}$ vaut un. On en déduit $K_{a,b} = 1$, d'où

$$f_Z(x) = f_{a+b,\lambda}(x).$$

La densité de Z est la densité de $\Gamma(a+b, \Lambda)$, donc Z suit la loi $\Gamma(a+b, \Lambda)$.

Remarque : comme sous-produit de cette démonstration, on a obtenu le résultat non trivial suivant :

$$\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 \theta^{a-1}(1-\theta)^{b-1}\, d\theta.$$

□

5.8.4 Compléments méthodologiques

Cette sous-section peut être omise en première lecture.

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

- $\exists x \in O; D_x T$ non inversible.
- T non injectif

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

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

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

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

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

    qui est presque partout égale à

$$y \mapsto \frac{1}{\sqrt{2\pi}}\frac{1}{3}\frac{1}{|y|^{\frac{2}{3}}}e^{-\frac{|y|^{\frac{2}{3}}}{2}},$$

    dont l'écriture est tout de même plus simple.
  – Un exemple dans $\mathbb{R}^2$
    Soit X et Y indépendantes suivant toutes deux $\mathcal{N}(0,1)$. On veut calculer la loi du couple $(U,V) = (X^3, XY)$.
    On a ici $O = \mathbb{R}^2$ et $T(x) = (x^3, xy)$. On a $\{x \in O; \det D_xT = 0\} = \{(x,y) \in \mathbb{R}; 3x^3 = 0\} = \{0\}\times\mathbb{R}$. On prend donc $\mathbb{R}^2\backslash(\{0\}\times\mathbb{R})$ comme ouvert de départ : T est C^1 sur $\mathbb{R}^2\backslash(\{0\}\times\mathbb{R})$, sa différentielle y est partout inversible et T est injective sur $\mathbb{R}\backslash\{0\}$, de réciproque $(u,v) \mapsto T^{-1}(u,v) = (\sqrt[3]{u}, \frac{v}{\sqrt[3]{u}})$. T réalise donc un C^1 difféomorphisme de T sur son image, qui est $(\mathbb{R}\backslash\{0\})^2$.
    On trouve ainsi qu'une densité de (U,V) est

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

    qui est presque partout égale à

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

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

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

  où les O_i' sont des connexes et où I est dénombrable. (Je n'ai jamais rencontré d'exemple où $O' \backslash N$ a une infinité non dénombrable de composantes connexes, en fait en général $|I| = 1$ ou 2.)
  Je dis que si pour tout $i \in I$, T est injective sur O_i, alors on a gagné. Notons ν la mesure image de μ par T. On a pour tout borélien A

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

  où l'on a posé

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

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

  Ainsi, on obtient que la loi image de μ par T admet la densité

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

  où T_i^{-1} représente la réciproque de la restriction de T_i à O_i.
  Remarque : en dimension 1, l'injectivité de O_i sur les composantes connexes est gratuite, en vertu du théorème de Rolle
– Un exemple dans $\mathbb{R}$
  X suit la loi uniforme sur $[-1, 2]$ et $Y = X^2$. On a $O =]-1, 2[$ et $N = \{0\}$, de sorte que $O_1 =]-1, 0[$ et $O_2 =]0, 2[$ On trouve la densité

$$y \mapsto \frac{1}{2\sqrt{y}} \frac{1}{3} 1\!\!1_{]0,1[}(y) + \frac{1}{2\sqrt{y}} \frac{1}{3} 1\!\!1_{]0,4[}(y)$$

– Un exemple dans $\mathbb{R}^2$
Soient X et Y deux variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On pose $S = X+Y$ et $P = XY$. Déterminer la loi de (S,P).
Ici $N = \{(x,y) \in \mathbb{R}^2 ; x = y\}$, donc

$$O_1 = \{(x,y) \in \mathbb{R}^2 ; 0 < x < y < 1\} \text{ et } O_2 = \{(x,y) \in \mathbb{R}^2 0 < y < x < 1\}.$$

Ici la symétrie axiale $s(x,y) = (y,x)$ d'axe $x = y$ envoie laisse μ invariante – c'est à dire que μ est la mesure image de μ par s – , envoie O_2 sur O_2, et vérifie de plus $T \circ s = T$, on a

$$\begin{aligned}
\nu_2(A) &= \frac{\mu(T^{-1}A \cap O_2)}{\mu(O_2)} \\
&= \frac{\mu((T\circ s)^{-1}(A) \cap s^{-1}(O_1))}{\mu(O_2)} \\
&= \frac{\mu(s^{-1}(T^{-1}(A)) \cap s^{-1}(O_1))}{\mu(s^{-1}(O_1))} \\
&= \frac{\mu_s(T^{-1}(A) \cap O_1)}{\mu_s(O_1)} \\
&= \frac{\mu(T^{-1}(A) \cap O_1)}{\mu(O_1)} \\
&= \nu_1(A)
\end{aligned}$$

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

fermé de mesure nulle telle que T soit injective sur chaque composante de $O'' = O'\backslash F$ et remplacer O' par O''. Là, il n'y a pas vraiment de méthode générale. Exemple : $T(x,y) = (x^2+y^2, 2xy)$. Le seul point où la différentielle n'est pas inversible est l'origine, mais bien sûr T n'est pas injectif sur l'ouvert connexe $\mathbb{R}^2\backslash\{(0,0)\}$, puisque $T(-x,-y) = T(x,y)$. On constatera que $F = \{(x,0); x \in \mathbb{R}\}$ convient.

5.9 Calcul des premiers moments des lois discrètes usuelles

5.9.1 Indicatrice d'un événement

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

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

1_A est une variable aléatoire à valeurs dans $\{0;1\}$. Il est important de remarquer que, comme $\forall x \in \{0;1\} \quad x^2 = x$, on a $1\!\!1_A^2 = 1\!\!1_A$. Maintenant, on a

- $\mathbb{E}1\!\!1_A = P(A)$.
-

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

5.9.2 Loi binomiale

On a vu que la loi binomiale de paramètres n et p était la loi de

$$X = \sum_{k=1}^{n} 1\!\!1_{A_k},$$

où $A_1, \dots, A_n$ sont n événements indépendants de même probabilité p. On a donc

$$\mathbb{E}X = \sum_{k=1}^{n} \mathbb{E}1\!\!1_{A_k} = \sum_{k=1}^{n} P(A_k) = np,$$

et comme les variables aléatoires sont indépendantes

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

5.9.3 Loi géométrique

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

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

$$\begin{aligned}
\mathbb{E}X(X-1) &= \sum_{k=0}^{+\infty} k(k-1)P(X=k) \\
&= \sum_{k=2}^{+\infty} k(k-1)P(X=k) \\
&= \sum_{k=2}^{+\infty} k(k-1)p(1-p)^{k-1} \\
&= p(1-p)\sum_{k=2}^{+\infty} k(k-1)(1-p)^{k-2} \\
&= p(1-p)\frac{2}{(1-(1-p))^3} \\
&= \frac{2(1-p)}{2p^2}.
\end{aligned}$$

On a alors $\mathbb{E}X^2 = \mathbb{E}X(X-1) + \mathbb{E}X = \frac{1}{p} + \frac{2(1-p)}{p^2}$ et $\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{1}{p} + \frac{2(1-p)}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$.

5.9.4 Loi de Poisson

$$\begin{aligned}
\mathbb{E}X &= \sum_{k=0}^{+\infty} kP(X=k) \\
&= \sum_{k=1}^{+\infty} kP(X=k) \\
&= \sum_{k=1}^{+\infty} ke^{-\lambda}\frac{\lambda^k}{k!} \\
&= e^{-\lambda}\lambda\sum_{k=1}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!} \\
&= e^{-\lambda}\lambda e^{\lambda} \\
&= \lambda
\end{aligned}$$

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

On a alors $\mathbb{E}X^2 = \mathbb{E}X(X-1) + \mathbb{E}X = \lambda^2 + \lambda$ et $\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

5.9.5 Loi hypergéométrique

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

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

On va montrer que $\mathbb{E}X = k\frac{n}{N}$ et $\operatorname{Var} X = k\frac{n}{N}(1-\frac{n}{N})\frac{N-1}{N}$.

Démonstration. Notons P la loi uniforme sur Ω. Par souci de lisibilité, on définit l'ensemble aléatoire A par $A(\omega) = \omega$. Ainsi

$$\begin{aligned} X &= |\{1,\dots,n\}) \cap A| \\ &= \sum_{i=1}^n \mathbb{1}_{\{i\in A\}} \end{aligned}$$

Pour tout $k \in \{1,\dots,n\}$, on a

$$\mathbb{E}\mathbb{1}_{\{i\in A\}} = P(i\in A) = 1 - \frac{\binom{N-1}{k}}{\binom{N}{k}} = 1 - \frac{(N-1)!(N-k)!}{N!(N-k-1)!} = 1 - \frac{N-k}{N} = \frac{k}{N}.$$

Ainsi

$$\mathbb{E}X = \sum_{i=1}^n \mathbb{1}_{\{i\in A\}} = \frac{nk}{N} = k\frac{n}{N}.$$

Maintenant, on a

$$\operatorname{Var} X = \sum_{i=1}^n \sum_{j=1}^n \operatorname{Covar}(\mathbb{1}_{\{i\in A\}}, \mathbb{1}_{\{j\in A\}}).$$

Pour $i = j$, on a

$$\operatorname{Covar}(\mathbb{1}_{\{i\in A\}}, \mathbb{1}_{\{j\in A\}}) = \operatorname{Var} \mathbb{1}_{\{i\in A\}} = P(j\in A)(1-P(j\in A)) = \frac{k}{N}(1-\frac{k}{N}).$$

Pour $i \neq j$, on a

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

On en déduit

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

On remarque qu'une loi hypergéométrique a la même espérance qu'une loi binomiale $\mathcal{B}(k, \frac{n}{N})$ et que sa variance ne diffère de celle de cette binomiale que d'un facteur $\frac{N-1}{N}$.

□

5.10 Calcul des premiers moments des lois à densité usuelles

5.10.1 Loi uniforme sur un segment

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

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

On a donc

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

et

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

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

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

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

5.10.2 Loi gaussienne

Soit X une variable aléatoire suivant la loi $\mathcal{N}(0,1)$. On rappelle que la densité de X est
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$
Il est facile de vérifier que
$$\frac{\mathrm{d}}{\mathrm{dx}}(xf(x)) = (1-x^2)f(x).$$
On a donc
$$\forall a,b \in \mathbb{R} \quad bf(b) - af(a) = \int_a^b f(x)\ dx - \int_a^b x^2 f(x)\ dx$$

Comme $\lim\limits_{a\to-\infty} af(a) = \lim\limits_{b\to+\infty} bf(b) = 0$ et que $\int_{-\infty}^{+\infty} f(x)\ dx = 1$, on en déduit que
$$\int_{-\infty}^{+\infty} x^2 f(x)\ dx = 1.$$
Autrement dit, X admet un moment d'ordre 2 : $\mathbb{E}X^2 = 1$.

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

Passons au cas général. Si l'on a $Y = m + \sigma X$, on sait que Y suit la loi $\mathcal{N}(m,\sigma^2)$. On a alors $\mathbb{E}Y = m + \sigma\mathbb{E}X = m$ et $\operatorname{Var} Y = \sigma^2 \operatorname{Var} X = \sigma^2$.

5.10.3 Lois Gamma

Soit X une variable aléatoire suivant la loi $\Gamma(a,\lambda)$. Alors, X admet des moments de tout ordre, avec pour tout $\alpha \geq 0$, on a
$$\mathbb{E}X^\alpha$$
. En particulier $\mathbb{E}X = \frac{a}{\lambda}$ et $\operatorname{Var} X = \frac{a}{\lambda^2}$.

Démonstration. Pour tous a et λ strictement positifs, on note $f_{a,\lambda}$ la densité de la loi $\Gamma(a,\Lambda)$, soit (rappel)
$$f_{a,\lambda}(x) = \mathbb{1}_{\mathbb{R}_+}(x) \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}.$$

D'après le théorème de transfert,

$$\begin{aligned}\mathbb{E}X^\alpha &= \int_{\mathbb{R}_+} x^\alpha f_{a,\lambda}(x)\ dx\\ &= \int_{\mathbb{R}_+} \lambda^{-\alpha}\frac{\Gamma(a+\alpha)}{\Gamma(a)} f_{a+\alpha,\lambda}(x)\ dx\\ &= \lambda^{-\alpha}\frac{\Gamma(a+\alpha)}{\Gamma(a)}\end{aligned}$$

Ainsi $\mathbb{E}X = \lambda^{-1}\frac{\Gamma(a+1)}{\Gamma(a)} = \frac{a}{\lambda}$, $\mathbb{E}X^2 = \lambda^{-2}\frac{\Gamma(a+2)}{\Gamma(a)} = \frac{a(a+1)}{\lambda^2}$ $\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{a(a+1)}{\lambda^2} - \frac{a^2}{\lambda^2} = \frac{a}{\lambda^2}$. □

5.10.4 Lois exponentielles

Soit X suivant une loi exponentielle de paramètre $\lambda > 0$. La loi exponentielle est un cas particulier de la loi Gamma : on a $\mathcal{E}(\lambda) = \Gamma(1,\lambda)$. On déduit du calcul précédent que $\mathbb{E}X = \frac{1}{\lambda}$ et $\operatorname{Var} X = \frac{1}{\lambda^2}$.

5.10.5 Lois de Cauchy

Soient $a \in \mathbb{R}, b > 0$. La loi de Cauchy $\mathcal{C}(a,b)$ admet comme densité par rapport à la mesure de Lebesgue :

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

donc pour $k \geq 1$, on a

$$\mathbb{E}|X|^k = \frac{1}{\pi}\frac{b|x|^k}{(x-a)^2+b^2} = +\infty,$$

donc les lois de Cauchy n'admettent pas de moment d'ordre 1, ni , a fortiori, d'ordre supérieur.

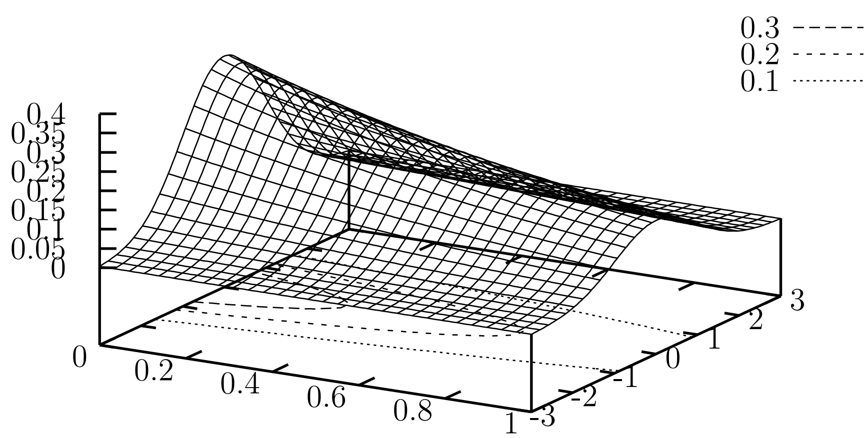
5.11 Exercice sur les espérances

1. Un jeu consiste à effectuer une mise en choisissant un nombre entre 1 et 6, puis à lancer simultanément trois dés. Si le numéro choisi sort une fois, le joueur récupère sa mise plus une somme égale à sa mise. Si le numéro choisi sort deux fois, le joueur récupère sa mise plus une somme égale à deux fois sa mise. Enfin, si le numéro choisi sort trois fois, le joueur récupère sa mise plus une somme égale à trois fois sa mise. Quelle est l'espérance de gain à ce jeu ?

2. Soient A, B deux éléments observables. On note
$$A \Delta B = \{x \in A; x \notin B\} \cup \{x \in B; x \notin A\}.$$
Ce sont donc les éléments qui sont dans A ou dans B, mais pas dans les deux. Montrer $\mathbb{1}_{A\Delta B} = (\mathbb{1}_A - \mathbb{1}_B)^2$. En déduire
$$P(A\Delta B) = P(A) + P(B) - 2P(A \cap B).$$
3. Soient A, B deux éléments observables. Montrer que
$$|P(A \cap B) - P(A)P(B)| \leq \sqrt{P(A)P(B)}.$$
4. Soit X une variable aléatoire à valeurs dans $\mathbb{N}^*$ telle la suite $(p_n)_{n\geq 1}$ définie par $p_n = P(X = n)$ soit décroissante. Montrer que pour toute injection σ de $\mathbb{N}^*$ dans lui-même, on a
$$\mathbb{E}\sigma(X) \geq \mathbb{E}X.$$
5. On suppose que $Y = \ln X$ vérifie $Y \sim \mathcal{N}(m, \sigma^2)$ (on dit alors que X est log-normale). Calculer $\mathbb{E}X$ et $\operatorname{Var} X$.
6. Calculer $\mathbb{E}\sin X$, où $P(X = \frac{\pi}{2}) = \frac{1}{6}, P(X = \frac{\pi}{3}) = \frac{1}{3}, P(X = \frac{\pi}{6}) = \frac{1}{2}$.
7. Soient X, Y deux variables aléatoires suivant chacune une loi uniforme sur $[a, b]$. Montrer que $\mathbb{E}|X - Y| \leq \frac{b-a}{2}$. Que vaut $\mathbb{E}|X - Y|$ lorsque X et Y sont indépendantes ?
8. Soit X une variable aléatoire suivant la loi uniforme sur $[0, 1]$. Déterminer la loi de $Z = -\ln(1 - X)$.
9. Soit X une variable aléatoire de densité f. Montrer que la variable aléatoire $|X|$ admet comme densité
$$x \mapsto \mathbb{1}_{R_+}(x)(f(x) + f(-x)).$$
10. Soit X une variable aléatoire positive de densité f. Montrer que la variable aléatoire $X^{1/2}$ admet comme densité
$$x \mapsto \mathbb{1}_{R_+}(x)2x(f(x^2)).$$
11. Soit X une variable aléatoire normale centrée réduite. Montrer que la variable aléatoire X^2 est à densité et la déterminer.

12. La figure ci-dessous représente la densité $f(x,y)$ d'un couple de variables aléatoires indépendantes X et Y. X suit une loi exponentielle de paramètre 1 et Y une loi normale centrée réduite.

    On a tracé quelques isoclines, c'est à dire des courbes reliant des points de même densité : $f(x,y) =$ constante. Quelle est la nature géométrique de ces isoclines ?



13. Soit X et Y deux variables aléatoires indépendantes. On suppose que X suit la loi normale $\mathcal{N}(0,1)$ et Y la loi gamma $\gamma(\frac{n}{2},\frac{1}{2})$. Calculer la loi de
$$Z = \frac{Y}{\sqrt{Y/n}}.$$
    La loi de Z est appelée loi de Student à n degrés de libertés.

14. Soient X et Y deux variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On pose $S = X + Y$ et $P = XY$.

    Déterminer la loi de (S,P).

15. Soit f une fonction réelle continue sur l'intervalle fermé $[0,1]$. Pour $n \in \mathbb{N}^*$, on note B_n le polynôme de Bernstein
$$B_n(x) = \sum_{k=0}^{n} f(\frac{k}{n})\binom{n}{k} x^k (1-x)^{n-k}\ .$$
    Pour tout $x \in]0,1[$ on se donne une suite (X_k) de variables de Bernoulli indépendantes de même paramètre x. On note $S_n = \sum_{k=1}^{n} X_k$.

    (a) Déterminer la moyenne $\mathbb{E}[f(\frac{S_n}{n})]$.

(b) Soit pour tout $\varepsilon > 0$, le réel $\delta(\varepsilon)$ défini par

$$\delta(\varepsilon) = \sup\{|f(x) - f(y)| : x, y \in [0,1] \text{ et } |x-y| \leq \varepsilon\} .$$

i. Démontrer que $\delta(\varepsilon)$ tend vers 0 avec ε.

ii. Démontrer que

$$\sup_{x\in[0,1]} |B_n(x) - f(x)| \leq \delta(\varepsilon) + \frac{2\|f\|_\infty}{n\varepsilon^2} .$$

En déduire que la suite des polynômes B_n converge vers f uniformément sur $[0,1]$.

16. On place 7 dames sur un échiquier torique 41×41 de telle manière qu'aucune dame ne puisse en prendre une autre. Montrer qu'il est possible de placer sur l'échiquier deux cavaliers pouvant se prendre mutuellement tels que chacun des cavaliers puisse être pris par au moins une des dames.

17. Soient n, r deux entiers tels que $1 \leq r \leq n$. On prend r nombres distincts au hasard dans $\{1, \dots, n\}$ et on note X le plus petit de ces r nombres.

(a) Quelles valeurs peut prendre X ? Montrer que pour $k \in \{0, n-r\}$, on a
$$P(X > k) = \frac{\binom{n-k}{r}}{\binom{n}{r}}.$$

(b) En déduire que
$$\mathbb{E}X = \frac{\binom{n+1}{r+1}}{\binom{n}{r}} = \frac{n+1}{r+1}.$$

18. Soient X et Y deux variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$.

(a) Soit r la rotation dans $\mathbb{R}^2$ de centre $(0,0)$ et d'angle $-\frac{\pi}{4}$. On pose $(U,V) = r(X,Y)$. Montrer que la loi du vecteur (U,V) est la loi uniforme sur un ensemble que l'on déterminera.

(b) Pour quelles valeurs de α la variable aléatoire $\frac{1}{|X-Y|^\alpha}$ est-elle intégrable ? Lorsqu'elle l'est, calculer sa valeur.

Chapitre 6

Espaces L^p

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

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

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

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

On convient également que 0 et l'infini sont des exposants conjugués.

6.1 De L^p à $\mathcal{L}^p$

6.1.1 Inégalité de Hölder

Théorème 56 (Inégalité de Hölder). *Soient p et q des exposants conjugués, $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, f et g deux éléments de $\overline{\mathcal{V}}(\Omega, \mathcal{A}, \mu)$. On a*

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

Démonstration. Si f est nulle μ presque partout, alors l'inégalité est évidente (c'est en fait une égalité). Idem pour g. Dans le cas inverse, on a $\left(\int_\Omega |f(x)|^p \ d\mu(x)\right)^{1/p} >$ 0et $\left(\int_\Omega |g(x)|^q \ d\mu(x)\right)^{1/q} > 0$. Ainsi, en remplaçant f par $f/\left(\int_\Omega |f(x)|^p \ d\mu(x)\right)^{1/p} >$ 0 et g par $g/\left(\int_\Omega |g(x)|^p \ d\mu(x)\right)^{1/q} > 0$, on peut se ramener au cas où $\left(\int_\Omega |f(x)|^p \ d\mu(x)\right)^{1/p} = \left(\int_\Omega |g(x)|^q \ d\mu(x)\right)^{1/q} = 1$.

Or pour tous x, y dans $\overline{\mathbb{R}}_+$, on a

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

Si x ou y est infini, c'est évident. Sinon, on peut écrire $x = e^{a/p}$, $y = e^{y/q}$ et appliquer la convexité de la fonction exponentielle.

Ainsi

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

d'où

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

□

6.1.2 Inégalité triangulaire

Théorème 57 (Inégalité triangulaire). *Soient p et q des exposants conjugués, $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, f et g deux éléments de $\overline{\mathcal{V}}(\Omega, \mathcal{A}, \mu)$. On a*

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

Démonstration. Comme précédemment, on peut supposer que f et g ne sont pas presque partout nulles. Par ailleurs, si $\int_\Omega |f(x)|^p\ d\mu(x) = +\infty$ ou que $\int_\Omega |g(x)|^p\ d\mu(x) = +\infty$, l'inégalité est évidente. On suppose donc que ces deux quantités sont finies. Comme $((f+g)/2)^p \le (f^p + g^p)/2$ par convexité de $x \mapsto x^p$, il sensuit qu'on a également $\int_\Omega |f(x)+g(x)|^p\ d\mu(x) < +\infty$.

On écrit

$$(f+g)^p = f(f+g)^{p-1} + g(f+g)^{p-1}.$$

L'inégalité de Hölder donne

$$\int_\Omega f(f+g)^{p-1}\ d\mu \le \left(\int_\Omega f^p\ d\mu\right)^{1/p} \left(\int_\Omega (f+g)^{(p-1)q}\ d\mu\right)^{1/q},$$

soit

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

De même,

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

En additionnant, on obtient

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

D'où

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

□

Ainsi, il est maintenant simple de constater que si l'on pose

$$\|f\|_p = \left(\int_\Omega f^p \, d\mu \right)^{1/p},$$

on définit ainsi une semi-norme sur l'espace vectoriel $L^p(\Omega, \mathcal{A}, \mu)$. Notons $V = \{v \in L^p; \|v\|_p = 0\}$. D'après l'inégalité triangulaire, V est un sous-espace vectoriel de L^p.

Notons $\mathcal{L}^p$ le quotient de l'espace vectoriel L^p par son sous-espace vectoriel V.

Soit f et g deux éléments de la même classe : $k = f - g \in V$. D'après l'inégalité triangulaire $\|f\|_p \leq \|g\|_p + \|k\|_p = \|g\|_p$. 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 $f \in \mathcal{L}^p$, on note $\|f\|_p = \|g\|_p$ où g est un quelconque représentant de la classe f. Evidemment, $\|f\|_p$ reste une sous norme.
Mais en réalité, $f \mapsto \|f\|_p$ est une norme sur $\mathcal{L}^p$. 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 $\mathcal{L}^p$.

6.2 Complétude de $\mathcal{L}^p$

Théorème 58. *$\mathcal{L}^p$ est complet*

Lemme 6. *Soit f_n une suite d'éléments de $\mathcal{L}^p$ avec*

$$\sum_{n\geq 1} \|f_n\|_< + \infty.$$

Alors la suite
$\sum_{k=1}^n f_k$ converge dans $\mathcal{L}^p$ 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 L^p telle que $\| \sum_{k=1}^n g_k - g\|_p$ tend vers 0, ce qui donnera la convergence de la suite $\sum_{k=1}^n f_k$ vers la classe de g.

Supposons d'abord que les (g_k) sont positives : dans ce cas la suite de fonctions $S_n = \sum_{k=1}^n g_k$ converge simplement vers une fonction g mesurable (éventuellement infinie en certains points) Cependant d'après l'inégalité triangulaire

$$\int S_n^p \leq (\sum_{k=1}^n \|g_k\|_p)^p,$$

et donc d'après le théorème de convergence dominée

$$\int g^p \leq (\sum_{k=1}^{+\infty} \|g_k\|_p)^p < +\infty.$$

Ainsi g est dans L^p Soient n et n' avec $n' \geq n$ On a $(S_n - S_{n'})^p = (\sum_{k=n+1}^{n'} g_k)^p$. Faisons tendre n' vers $+\infty$: d'après le théorème de convergence dominée, on a

$$\int_\Omega (S_n - g)^p \; d\mu = \lim_{n \to +\infty} \int_\Omega (S_n - S_{n'})^p \; d\mu.$$

Cependant, d'après l'inégalité triangulaire

$$\begin{aligned} \int_\Omega (S_n - S_{n'})^p \; d\mu &\leq (\sum_{k=n+1}^{n'} \|g_k\|_p)^p \\ &\leq (\sum_{k=n+1}^{+\infty} \|g_k\|_p)^p, \end{aligned}$$

donc

$$\int_\Omega (S_n - g)^p \; d\mu \leq (\sum_{k=n+1}^{+\infty} \|g_k\|_p)^p,$$

soit

$$\|S_n - g\|_p \leq \sum_{k=n+1}^{+\infty} \|g_k\|_p.$$

Mais on reconnait là le reste d'une série convergente, donc $\|S_n - g\|_p$ tend bien vers 0.

Dans le cas général, écrivons $g_k = g_k^+ - g_k^-$. On définit évidemment $g^+ = \sum g_k^+$ et $g^- = \sum g_k^-$, $S_n^+ = \sum_{k=1}^n g_k^+$, $S_n^- = \sum_{k=1}^n g_k^-$. La série de terme général $\|g_k^+\|_p$ est convergente car $\|g_k^+\|_p \leq \|g_k\|_p$. On montre ainsi que $\|S_n^+ - g^+\|_p$ tend bien vers 0, et de même que $\|S_n^- - g^-\|_p$ tend bien vers 0. Enfin, l'inégalité triangulaire permet de conclure que $\|S_n - g\|_p$ tend bien vers 0. □

Ainsi, on a montré que dans $\mathcal{L}^p$, toute série absolument convergente est convergente. Pour conclure, il suffit de s'appuyer sur le résultat d'analyse suivant.

Lemme 7. *Un espace vectoriel normé où toute série absolument convergente converge est complet.*

Démonstration. Remarquons d'abord que si une suite de Cauchy admet une sous-suite convergente, elle converge. En effet supposons (x_n) de Cauchy avec x_{n_k} qui converge vers l. Soit k_0 tel que $\|x_{n_k} - l\| \leq \varepsilon/2$ pour $k \geq k_0$ et b_0 tel que $\|x_k - x_{k'}\| \leq \varepsilon/2$ lorsque k et k' dépassent b_0. Alors $\|x_n - l\| \leq \varepsilon$ dès que n dépasse $\max(n_0, n_{b_0})$.

Soit maintenant x_n une suite de Cauchy dans un espace où toute série absolument convergente converge. On pose $n_0 = 1$, puis pour $k \geq 1$:

$$n_k = \inf\{n > n_{k-1} : k, k' \geq n \Longrightarrow \|x_k - x_{k'}\| \leq 2^{-k}\}.$$

Cette suite d'indices est strictement croissante et est bien définie car (x_k) est de Cauchy. Par construction $\|x_{n_k} - x_{n_{k+1}}\| \leq 2^{-k}$ pour $k \geq 1$, donc la série de terme général $x_{n_k} - x_{n_{k+1}}$ est absolument convergente. Mais on a fait l'hypothèse ici qu'une série absolument convergente est convergente, donc elle est convergente, ce qui veut dire que x_{n_k} est convergente. (x_n) est donc une suite de Cauchy qui admet une sous-suite convergente, elle est donc convergente. □

Théorème 59. *Soit $f, (f_n)_{n\geq 1}$ des fonctions dans L^p telles que $(f_n)_{n\geq 1}$ converge dans 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 $g_n = |f - f_n|^p$. (g_n) converge dans 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 $k \geq 1$:

$$n_k = \inf\{n > n_{k-1} : k, k' \geq n \Longrightarrow \|g_k - g_{k'}\|_1 \leq 2^{-k}\}.$$

Cette suite d'indices est strictement croissante et est bien définie car (g_k) est de Cauchy dans L^1. Par construction $\|g_{n_k} - g_{n_{k+1}}\| \leq 2^{-k}$ pour $k \geq 1$, donc la série de terme général $\|g_{n_k} - g_{n_{k+1}}\|_1$ est convergente. Mais

$$\begin{aligned}\sum_{n=1}^{+\infty} \|g_{n_k} - g_{n_{k+1}}\|_1 &= \sum_{n=1}^{+\infty} \int |g_{n_k} - g_{n_{k+1}}| \, d\mu \\ &= \int \sum_{n=1}^{+\infty} |g_{n_k} - g_{n_{k+1}}| \, d\mu\end{aligned}$$

La fonction positive

$$\sum_{n=1}^{+\infty} |g_{n_k} - g_{n_{k+1}}|$$

est intégrable, elle est donc en particulier finie presque partout. En un point x tel que

$$\sum_{n=1}^{+\infty} |g_{n_k}(x) - g_{n_{k+1}}(x)| < +\infty,$$

la suite $(g_{n_k}(x))_{k\geq 1}$ converge. $(g_{n_k})_{k\geq 1}$ converge presque partout vers une fonction g^*. Mais d'après le lemme de Fatou,

$$\int g^* d\mu = \int \underline{\lim}\ g_{n_k} d\mu \leq \underline{\lim} \int g_{n_k}\, d\mu = 0,$$

donc g^* est nulle presque partout, ce qui achève la preuve.

□

6.3 Théorèmes d'approximation

Théorème 60. *Soit $\mathcal{S}$ l'ensemble des fonctions simples s telles que*

$$\mu(\operatorname{supp} s) < +\infty.$$

Pour tout $p \in [1, +\infty[$, $\mathcal{S}$ est dense dans $L^p(\mu)$ (et donc les classes de ces fonctions sont denses dans $\mathcal{L}^p(\mu)$).

Démonstration. D'abord, il est facile de voir que $\mathcal{S}$ est dans $L^p(\mu)$. Soit $f \in L^p$. Supposons $f \geq 0$ et prenons f_n comme dans le lemme 1. On a

$$\mathbb{1}_{\{f_n>0\}} 2^{-np} \leq \mathbb{1}_{\{f_n>0\}} f_n^p \leq f^p,$$

d'où

$$\mu(\operatorname{supp} f_n) 2^{-np} \leq \int_\Omega f^p \, d\mu,$$

et donc $f_n \in \mathcal{S}$. On a $|f_n - f|^p \le f^p$, 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 L^p. Le cas général s'ensuit en séparant partie positive et partie néative, comme dans la preuve du théorème 58. □

Théorème 61. *Soit $p \in [1, +\infty[$. Les classes des fonctions continues à support compact forment une partie dense dans $\mathcal{L}^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$*

Ce théorème est admis.

6.4 Exercices sur les espaces L^p

1. Étudier l'appartenance à $L^1(\mathbb{R})$ et à $L^2(\mathbb{R})$ des fonction suivantes :
   (a) $f(t) = e^{-|t|}$.
   (b) $g(t) = \frac{\sin t}{t}$.
   (c) $g(t) = \frac{1}{\sqrt{|t|}(1+t^2)}$.
2. Étudier la convergence dans $L^1(\mathbb{R})$ et dans $L^2(\mathbb{R})$ des suites suivantes :
   (a) $f_n(t) = \sqrt{n} \exp(-n^2 t^2)$.
   (b) $g_n(t) = \frac{n^2 \sin nt}{2\pi} \mathbb{1}_{[-\pi/n, \pi/n]}(t)$.
   (c) $h_n(t) = \frac{2}{\pi n^2} \sqrt{n^2 - t^2} \mathbb{1}_{[-/n, n]}(t)$.
3. Soit f une fonction de $\mathbb{R}$ dans $\mathbb{R}$ intégrable et soit $\hat{f}$ la classe de f dans $L^1(\mathbb{R}, \lambda)$. Montrer que $\hat{f}$ contient au plus une fonction continue.
4. Soit $E = \{f \in L^1(\mathbb{R}, \lambda) \; ; \; |f| \le 1 \; \lambda - p.p\}$. Montrer que E est un sous-ensemble fermé de $L^1(\mathbb{R}, \lambda)$.
5. Montrer que la fonction $f \; : \; x \mapsto \dfrac{1}{\sqrt{x}(1 + |\ln(x)|)}$ est dans $\mathcal{L}^p(]0, +\infty[, \lambda)$ si et seulement si $p = 2$.
6. Montrer que si f et g appartiennent à $L^1(X, \mu)$ alors $\sqrt{|f^2 + g^2|}$ appartient aussi à $L^1(X, \mu)$.
7. Soient $\alpha \in \mathbb{R}$ et $p \in [1, \; +\infty[$, on note f_α l'application de $]0, \; +\infty[$ dans $]0, \; +\infty[$, définie par $f_\alpha(x) = x^\alpha$.
   (a) Pour quelle(s) valeur(s) de α, la fonction f_α appartient-elle à $L^p(]0, 1], \lambda)$? Calculer alors les normes de f_α dans chacun de ces espaces.
   (b) Même question avec les espaces $L^p([1, \infty[, \lambda)$
8. Donner un exemple de suite (f_n) dans $L^1(X, \mu)$ telle que

(a) (f_n) converge vers f presque partout mais (f_n) ne converge pas vers f au sens de la norme L^1 ;

(b) (f_n) converge vers f dans L^1 mais (f_n) ne converge pas vers f presque partout ;

(c) (f_n) converge vers f presque partout, $(\int f_n d\mu)$ converge vers $\int f d\mu$, mais (f_n) ne converge pas vers f au sens de la norme L^1.

9. Soient $(X,\ \mathcal{M},\ \mu)$ un espace mesuré avec $\mu(X) = 1$ et f, g des fonctions mesurables sur X à valeurs dans $[0,\ +\infty]$ telles que $fg \geq 1$. Montrer que l'on a $\left(\int_X f d\mu\right)\left(\int_X g d\mu\right) \ \geq\ 1$.

10. Soient $f \in L^p(X,\mu)$, $g \in L^q(X,\mu)$ et r tel que $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$. Montrer que $fg \in L^r(X,\mu)$ et que $||fg||_r \ \leq\ ||f||_p||g||_q$.

11. Soit $p \in]1,+\infty[$. Pour f dans $\mathcal{L}^p(]0,+\infty[)$ et pour $x > 0$, on pose

$$T(f)(x) \ =\ \frac{1}{x}\int_{]0,x[} f d\lambda\ .$$

(a) Montrer que $T(f)$ est bien définie sur $]0,+\infty[$.

(b) On suppose dans cette question que f est positive continue à support compact.

i. Montrer que $T(f)$ est dérivable sur $]0,+\infty[$ et calculer sa dérivée.

ii. Montrer que $T(f) \in \mathcal{L}^p(]0,+\infty[)$.

iii. Montrer que $\displaystyle\int_{]0,\infty[} T(f)^p d\lambda \ =\ \frac{p}{p-1}\int_{]0,\infty[} T(f)^{p-1} f d\lambda$.

iv. En déduire que $||T(f)||_p \ \leq\ \dfrac{p}{p-1}||f||_p$.

v. Montrer que cette inégalité reste vraie pour f de signe quelconque.

(c) Soit $f \in \mathcal{L}^p(]0,+\infty[)$.

i. Montrer que si (f_n) est une suite de fonctions continues à support compact qui converge vers f dans $\mathcal{L}^p(]0,+\infty[)$, alors $T(f_n)$ converge vers $T(f)$ $\lambda-$presque partout, la suite $(T(f_n))$ est de Cauchy dans $\mathcal{L}^p(]0,+\infty[)$ puis $(T(f_n))$ converge vers $T(f)$ dans $\mathcal{L}^p(]0,+\infty[)$.

ii. En déduire que $||T(f)||_p \ \leq\ \dfrac{p}{p-1}||f||_p$.

12. Soit $(X, \mathcal{T}, \mu)$ un espace mesuré tel que $\mu(X) < \infty$, $1 < p < \infty$ et $f : X \to \mathbb{R}$ une application borélienne. On suppose que pour toute fonction $g \in \mathcal{L}^p(X, \mu)$, la fonction fg est intégrable et il existe $C > 0$ telle que pour toute fonction $g \in \mathcal{L}^p(X, \mu)$ on ait $\left| \int fg \, d\mu \right| \leq C||g||_p$. Montrer que $f \in \mathcal{L}^q(X, \mu)$ où q est défini par $\frac{1}{p} + \frac{1}{q} = 1$

Chapitre 7

Convolution et transformation de Fourier

7.1 Produit de convolution

Remarques :

- Si f_1, f_2 sont deux fonctions de L^1 qui représentent le même élément de $\mathcal{L}^1$, alors $\int f_1 \, d\mu$ et $\int f_1 \, d\mu$ sont égales, donc on peut se permettre d'écrire $\int f \, d\mu$ pour $f \in \mathcal{L}^1$.
- L'application $T_t : f \mapsto (x \mapsto (x - t))$ passe au quotient dans $L^1(\mathbb{R}^d)$, car si $f_1 = f_2$ presque sûrement, alors $f_1(. - t) = f_2(. - t)$ presque sûrement.

Théorème 62. *Pour tout f dans $\mathcal{L}^p$, l'application*

$$t \mapsto T_t f$$

est continue sur $\mathbb{R}$.

Démonstration. $\|T_{t+h}f - T_t f\|_p = T_h(T_t f) - (T_t f)\|_p$, donc il suffit de montrer la continuité en 0. Traitons d'abord le cas où f est une fonction continue à support compact : comme f est continue $T_h h$ tend simplement vers f. En utilisant le théorème de convergence dominée, on optient alors la convergence dans L^p de $T_h f$ vers f. Passons au cas général. D'après le théorème 61, on peut trouver g et h, avec $f = g + h$, g continue à support compact et $\|h\|_p \leq \varepsilon$. On a

$$(T_t f - f) = (T_t g - g) + (T_t h - h),$$

d'où

$$\begin{aligned}\|T_t f - f\|_p &\leq \|T_t g - g\|_p + \|T_t h\|_p + \|h\|_p \\ &\leq \|T_t g - g\|_p + 2\|h\|_p,\end{aligned}$$

ce qui entraîne, en faisant tendre t vers 0

$$\overline{\lim_{t\to 0}}\ \|T_t f - f\|_p \leq 2\varepsilon.$$

Comme ε, on en déduit que $\overline{\lim_{t\to 0}}\ \|T_t f - f\|_p = 0$, ce qui est bien ce qu'on voulait montrer. □

7.1.1 convolution dans L^1

Soient f, g deux éléments de $L^1(\mu)$.

$$\begin{aligned}\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} |f(x-t)||g(t)|\ d\lambda^d(t)\right) d\lambda^d(x) &= \int_{\mathbb{R}^d\times\mathbb{R}^d} |f(x-t)||g(t)|\ d\lambda^d\otimes\lambda^d(t,x) \\ &= \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} |f(x-t)||g(t)|\ d\lambda^d(x)\ \right) d\lambda^d(t) \\ &= \int_{\mathbb{R}^d} |g(t)|\left(\int_{\mathbb{R}^d} |f(x-t)|\ d\lambda^d(x)\ \right) d\lambda^d(t) \\ &= \int_{\mathbb{R}^d} |g(t)|\left(\int_{\mathbb{R}^d} |f(x)|\ d\lambda^d(x)\ \right) d\lambda^d(t) \\ &= \left(\int_{\mathbb{R}^d} |f(x)|\ d\lambda^d(x)\ \right)\left(\int_{\mathbb{R}^d} |g(t)|\ d\lambda^d(t)\ \right) \\ &< +\infty.\end{aligned}$$

Ainsi, la fonction $f*g$ définie par

$$x \mapsto f*g(x) = \int_{\mathbb{R}^d} f(x-t)g(t)\ d\lambda^d(t)$$

est définie en presque tout point x et elle est dans L^1 : cette fonction est le produit de convolution de f par g

Les arguments évoqués plus haut fonctionnent encore : le produit de convolution "passe au quotient" et définit ainsi une application de $\mathcal{L}^1\times\mathcal{L}^1$ dans $\mathcal{L}^1$.

Au passage, notons qu'on a démontré

$$\|f * g\|_1 \leq \|f\|_1 \|g\|_1.$$

En reprenant le calcul précédant et en supposant que f et g sont dans L^1, le théorème de Fubini permet alors d'écrire

$$\begin{aligned}
\int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(x-t)g(t)\ d\lambda^d(t) \right) d\lambda^d(x) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x-t)g(t)\ d\lambda^d \otimes \lambda^d(t,x) \\
&= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(x-t)g(t)\ d\lambda^d(x) \right) d\lambda^d(t) \\
&= \int_{\mathbb{R}^d} g(t) \left(\int_{\mathbb{R}^d} f(x-t)\ d\lambda^d(x) \right) d\lambda^d(t) \\
&= \int_{\mathbb{R}^d} g(t) \left(\int_{\mathbb{R}^d} f(x)\ d\lambda^d(x) \right) d\lambda^d(t) \\
&= \left(\int_{\mathbb{R}^d} f(x)\ d\lambda^d(x) \right) \left(\int_{\mathbb{R}^d} g(t)\ d\lambda^d(t) \right),
\end{aligned}$$

soit

$$\int_{\mathbb{R}^d} (f * g)(x)\ d\lambda^d(x) = \left(\int_{\mathbb{R}^d} f(x)\ d\lambda^d(x) \right) \left(\int_{\mathbb{R}^d} g(t)\ d\lambda^d(t) \right). \tag{7.1}$$

7.1.2 autres produits

Supposons maintenant que $g \in L^1$ et que $f \in L^p$. On a

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

D'où

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

Cependant,

$$\begin{aligned}\int\left(\int |f(x-t)|^p|g(t)|\ d\lambda^d(t)\right) d\lambda^d(x) &= \int\left(\int |f(x-t)|^p|g(t)|\ d\lambda^d(x)\right) d\lambda^d(t)\\ &= \int\left(\int |f(x-t)|^p\ d\lambda^d(x)\right)|g(t)|d\lambda^d(t)\\ &= \int \|f\|_p^p|g(t)|d\lambda^d(t)\\ &= \|f\|_p^p\|g\|_1,\end{aligned}$$

donc finalement

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

Ainsi, l'intégrale

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

converge pour presque que tout x et l'application

$$x \mapsto f*g(t) = \int f(x-t)g(t)\ d\lambda^d(t)$$

représente un élément de L^p avec

$$\int |f*g(t)|^p\ d\lambda^d(t) \le \|f\|_p^p\|g\|_1^{1+p/q},$$

soit

$$\|f*g\|_p \le \|f\|_p\|g\|_1.$$

Remarque importante : quel que soit l'espace où on définit les choses, on a toujours

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

pour les x tels que

$$\int |f(x-t)g(t)|\ d\lambda^d(t) < +\infty,$$

de sorte que

$$f*g = g*f$$

toutes les fois où cela a un sens.

7.1.3 Approximations de l'unité

Théorème 63. *Soit φ une fonction positive avec*

$$\int_{\mathbb{R}^d} \varphi(x)\ d\lambda^d(x) = 1.$$

Pour tout $k \geq 1$ posons $\varphi_k(x) = k^d \varphi(kx)$.

Alors, pour tout f dans L^p la suite $f * \varphi_k$ converge vers f dans L^p.

Démonstration. Notons M_k l'application qui a f associe $f * \varphi_k$. C'est une application linéaire continue de L^p dans lui même. C'est même une contraction car

$$\forall f \in L^p \quad \|M_k f\|_p \leq \|f\|_p.$$

(C'est ce qu'on a montré dans la sous-section précédente.)

Soit $x \in \mathbb{R}^d$:

$$\begin{aligned} f * \varphi_k(x) - f(x) &= \int f(x-t)\varphi(kt)k^d\ d\lambda^d(t) - f(x) \\ &= \int f(x - t/k)\varphi(t)\ d\lambda^d(t) - f(x) \\ &= \int f(x - t/k)\varphi(t)\ d\lambda^d(t) - \int f(x)\varphi(t)\ d\lambda^d(t) \\ &= \int (f(x-t/k) - f(x))\varphi(t)\ d\lambda^d(t) \end{aligned}$$

Ainsi

$$\begin{aligned} |(M_k f - f)(x)| &= |\int (T_{t/k} f - f)(x)\varphi(t)\ d\lambda^d(t)| \\ &\leq \left(\int |(T_{t/k} f - f)(x)|^p \varphi(t)\ d\lambda^d(t) \right)^{1/p}, \end{aligned}$$

ce qui donne

$$\|M_k f - f\|_p \leq \left(\int \|T_{t/k} f - f\|_p^p \varphi(t)\ d\lambda^d(t) \right)^{1/p}$$

D'après de théorème 62, $\|T_{t/k} f - f\|_p^p$ tend vers 0 lorsque k tend vers l'infini. Comme

$$|\|T_{t/k} f - f\|_p^p \varphi(t)| \leq (2\|f\|_p)^p \varphi(t),$$

le théorème de convergence dominée permet de conclure. □

7.1.4 Régularisation

Théorème 64. *Soit $f \in L^p(\mathbb{R}^d)$, g C^1 à support compact. Alors $f * g$ est C^1 sur $\mathbb{R}^d$, avec*

$$D_x(f * g) = \int f(t) D_{x-t} g.$$

Démonstration. Soit M tel que $g(x) = 0$ pour $\|x\| \geq M$. Soit $R > 0$.Par définition

$$f * g(t) = \int f(x-t) g(t) \ d\lambda^d(t) = \int g(x-t) f(t) \ d\lambda^d(t).$$

Ici, c'est bien sûr la deuxième écriture qui va nous intéresser. Supposons $\|x\| \leq R$. La différentielle de $g(x-t)f(t)$, vue comme une fonction de x, est $f(t) D_{x-t} g$. Bien entendu

$$|f(t) D_{x-t} g| \leq |f(t)| \|Dg\|_\infty \mathbb{1}_{B(0,R+M)}(t).$$

$f \in L^p$ et $\|Dg\|_\infty \mathbb{1}_{B(0,R+M)} \in L^q$, donc $|f| \|Dg\|_\infty \mathbb{1}_{B(0,R+M)}$ est dans L^1. Le théorème de convergence dominée pour la différenciation sous le signe somme donne alors le résultat voulu. □

Corollaire 13. *Soit $f \in L^p(\mathbb{R}^d)$, g C^k à support compact. Alors $f * g$ est C^k sur $\mathbb{R}^d$, avec*

$$D_x^\alpha(f * g) = \int f(t) D_{x-t}^\alpha g,$$

où on a supposé que $|\alpha_1| + |\alpha_2| + \cdots + |\alpha_d| \leq k$.

Démonstration. Par récurrence sur k. □

Corollaire 14. *Les fonctions C^∞ à support compacts sont denses dans L^p.*

Démonstration. Cela provient immédiatement du Théorème 63 et du corollaire précédent. □

7.2 transformée de Fourier

Soit $f \in L^1(\mathbb{R}^d)$. On appelle transformée de Fourier de f, et l'on note $\hat{f}$ la fonction définie sur $\mathbb{R}^d$ par

$$\hat{f}(x) = \int e^{i\langle x,t\rangle} f(t) \ d\lambda^d(t).$$

Evidemment, $f \mapsto \hat{f}$ est linéraire, et comme $|e^{i\langle x,t\rangle} f(t)| \leq |f(t)|$, est on a

$$\forall f \in L^1 \quad \|\hat{f}\|_\infty \leq \|f\|_1.$$

7.2.1 propriétés élémentaires

Pour $f, g \in L^1$, on a
- $\widehat{f * g} = \hat{f}.\hat{g}$.
- $\widehat{T_t f}(x) = e^{i\langle x,t\rangle}\hat{f}(x)$
- Si $g(x) = f(x/\lambda)$, alors $\hat{g}(x) = \lambda^d \hat{f}(\lambda x)$.
- Si $g(x) = f(x)e^{i\langle x,\theta\rangle}$, alors $\hat{g}(x) = \lambda^d \hat{f}(x-\theta)$.
- $\int f(x)\, d\lambda^d(u) = \hat{f}(0)$

La première propriété mérite qu'on consacre quelques lignes à sa preuve :

$$\begin{aligned}
\widehat{f * g}(x) &= \int e^{i\langle x,t\rangle}\left(\int f(t-u)g(u)\ d\lambda^d(u)\right)\ d\lambda^d(t) \\
&= \int \left(\int e^{i\langle x,t\rangle} f(t-u)g(u)\ d\lambda^d(u)\right)\ d\lambda^d(t) \\
&= \int \left(\int f(t-u)e^{i\langle x,t-u\rangle} g(u)e^{i\langle x,u\rangle}\ d\lambda^d(u)\right)\ d\lambda^d(t) \\
&= \int (F * G)(t)\ d\lambda^d(t),
\end{aligned}$$

où $F(t) = f(t)e^{i\langle x,t\rangle}$ et $G(t) = g(t)e^{i\langle x,t\rangle}$. Mais d'après l'équation (7.1)

$$\int (F * G)(t)\ d\lambda^d(t) = \left(\int F(t) d\lambda^d(t)\right)\left(\int G(t) d\lambda^d(t)\right),$$

d'où le résultat voulu.

7.2.2 Théorème d'inversion

Théorème 65. *Soit $f \in L^1(\mathbb{R}^d)$ telle que $\hat{f} \in L^1(\mathbb{R}^d)$, alors on a*

$$f(x) = \frac{1}{(2\pi)^d}\int e^{-i\langle x,t\rangle}\ d\lambda^d(t)\ \lambda^d(t)\ p.p.$$

Démonstration. On aura besoin du lemme suivant, qui sera démontré (au moins une fois) en exercice.

Lemme 8. *Soit $G(x) = \frac{1}{(2\pi)^{d/2}}e^{-\|x\|_2^2}$. Alors $\hat{G}(x) = e^{-\|x\|_2^2} = (2\pi)^{d/2}G(x)$.*

Pour $k \geq 1$, posons $G_k(x) = k^n G(kx)$. On a

$$\hat{G_k}(x) = k^d k^{-d}\hat{G}(x/k) = (2\pi)^{d/2}G(x/k).$$

On recommence :

$$\widehat{\widehat{G_k}}(x) = (2\pi)^{d/2} k^d \hat{G}(kx) = (2\pi)^d k^d G(kx) = (2\pi)^d G_k(x),$$

et comme G_k est paire

$$\widehat{\widehat{G_k}}(-x) = (2\pi)^d G_k(x),$$

soit

$$\frac{1}{(2\pi)^d} \int e^{-i\langle x,t\rangle} \hat{G}_k(t)\ d\lambda^d(t) \quad = \quad G_k(x)$$

On a donc

$$\begin{aligned} f * G_k(x) &= \frac{1}{(2\pi)^d} \int \int e^{-i\langle x-y,t\rangle} \hat{G}_k(t) f(y)\ d\lambda^d(t)\ d\lambda^d(y) \\ &= \frac{1}{(2\pi)^d} \int e^{-i\langle x,t\rangle} \hat{G}_k(t) \hat{f}(t)\ d\lambda^d(t) \end{aligned}$$

En utilisant le théorème de convergence dominée, on voit que le terme de droite tend vers $\frac{1}{(2\pi)^d} \int e^{-i\langle x,t\rangle} \hat{f}(t)\ d\lambda^d(t)$ lorsque k tend vers l'infini. Mais d'après le théorème 63, le membre de gauche converge dans L^1 vers f. Comme la convergence dans L^1 entraîne la convergence d'une sous-suite presque partout, l'unicité de la limite donne l'égalité voulue. □

7.3 Exercices sur la convolution et la transformée de Fourier

1. Soient f et $g \in L^1(\mathbb{R}^n)$. Montrer que si f (resp. g) est nulle presque-partout en dehors d'un ensemble A (resp. B) alors $f * g$ est nulle presque-partout en dehors de $A + B = \{a + b; a \in A, b \in B\}$.
2. Calculer le produit de convolution $f * g$ des fonctions suivantes définies sur $\mathbb{R}$ $(a > 0, b > 0)$:
   (a) $f(x) = \exp(-\frac{x^2}{2a^2})$ et $g(x) = \exp(-\frac{x^2}{2b^2})$.
   (On admettra que $\int \exp(-\frac{x^2}{2})\ dx = \sqrt{2\pi}$.)
   (b) $f(x) = \mathbb{1}_{[-a,a]}(x)$ et $g(x) = \mathbb{1}_{[-b,b]}(x)$
3. Pour tout entier n on définit la fonction $g_n(x) = (1 - x^2)^n \mathbb{1}_{[-1,1]}(x)$. On pose $a_n = \displaystyle\int_{\mathbb{R}} g_n(x)\, dx$ et $k_n = a_n^{-1} g_n$.

   (a) Montrer que la suite (a_n) tend vers 0 et que $a_n \geq \frac{2}{n+1}$ pour tout entier n.
   (b) Soit f une fonction uniformément continue sur $\mathbb{R}$ et bornée. Montrer que $f * k_n$ converge uniformément vers f.
   (c) Soit f une fonction continue à support dans $[-\frac{1}{2}, \frac{1}{2}]$. Montrer que la restriction de $f * k_n$ à $[-\frac{1}{2}, \frac{1}{2}]$ est un polynôme de degré $\leq 2n$.
   (d) En déduire le théorème de Weierstrass : Toute fonction continue d'un intervalle $[a, b]$ dans $\mathbb{R}$ est limite uniforme sur $[a, b]$ d'une suite de polynômes.
4. Soit $f = 1\!\!1_{[-\frac{1}{2}, \frac{1}{2}]}$.
   (a) Déterminer $f * f$ et $f * f * f$.
   (b) On note $f^{(*)1} = f$ et pour $n \geq 2$, $f^{(*)n} = f^{(*)(n-1)} * f$. Vérifier que pour tout $n \geq 1$, $f^{(*)n} \in L^1(\mathbb{R})$ et que $\|f^{(*)n}\|_1 = 1$.
   (c) Montrer que pour tout $n \geq 2$, $f^{(*)n}$ est de classe $\mathcal{C}^{n-2}$.
5. Soit $E \in \mathcal{B}(\mathbb{R})$ tel que $0 < \lambda(E) < +\infty$.
   (a) Montrer que $1\!\!1_E * 1\!\!1_{-E}$ est continue sur $\mathbb{R}$.
   (b) En déduire que $E - E = \{x - y \,/\, x \in E,\, y \in E\}$ est un voisinage de 0.
6. Soit f la fonction de $\mathbb{R}$ dans lui-même définie par $f(x) = e^{-\frac{x^2}{2}}$ pour $x \in \mathbb{R}$.
   (a) Déterminer la transformée de Fourier de f en remarquant que $\widehat{f}$ est solution d'une équation différentielle linéaire.
   (b) Soit A une matrice carrée réelle symétrique d'ordre n définie positive. Déterminer la transformée de Fourier de la fonction de $\mathbb{R}^n$ dans $\mathbb{R}$ définie par $f(x) = e^{-<Ax,x>}$ pour $x \in \mathbb{R}^n$.
7. (a) Soit $f \in \mathcal{L}^1(\mathbb{R}^n)$ telle que $f * f = 0$. Montrer que $f = 0$.
   (b) Montrer que $L^1(\mathbb{R})$ n'a pas d'unité pour la convolution.
8. Déterminer la transformée de Fourier de la fonction indicatrice d'un intervalle $[a, b]$. Montrer que $1\!\!1_{[-1,1]} * 1\!\!1_{[-1,1]}$ est la transformée de Fourier d'une fonction de $\mathcal{L}^1(\mathbb{R})$ qu'on déterminera.
9. Calculer la transformée de Fourier de la fonction f de $\mathbb{R}$ dans lui-même définie par $f(x) = e^{-a|x|}$, pour $x \in \mathbb{R}$ (où $a > 0$). En déduire la transformée de Fourier de la fonction $g : x \mapsto \dfrac{1}{a^2 + x^2}$.

Chapitre 8

Fonctions génératrices et fonctions caractéristiques

8.1 Fonctions génératrices des variables aléatoires à valeurs dans $\mathbb{N}$

Définition : On appelle fonction génératrice d'une variable aléatoire X à valeurs dans $\mathbb{N}$ la fonction

$$z \mapsto G_X(z) = \mathbb{E}z^X = \sum_{k=0}^{+\infty} P(X = k)z^k.$$

Usuellement, on définit cette fonction sur l'intervalle réel $[0, 1]$, mais elle est en fait toujours définie sur la boule unité complexe fermée.

8.1.1 Fonction génératrice et indépendance

Théorème 66. *Si deux variables aléatoires X et Y sont indépendantes, on a $G_{X+Y} = G_X G_Y$.*

Démonstration. Soit $z \in B(0, 1)$. On a

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

□

8.1.2 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

$$\mathbb{E}z^X = P(X=0)z^0 + P(X=1)z^1 = (1-p)+pz.$$

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 loi binomiale de paramètres n et p. Ainsi, on déduit du théorème 66 que la fonction génératrice d'une loi binomiale de paramètres n et p est

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

Loi géométrique de paramètre $p \in]0,1[$

Soit $X\mathcal{G}(p)$ et $z \in B(0,1)$ On a

$$\begin{aligned}
\varphi_X(z) &= \mathbb{E}z^X \\
&= \sum_{n=1}^{+\infty} p(1-p)^{n-1}z^n \\
&= \sum_{n=0}^{+\infty} p(1-p)^n z^{n+1} \\
&= pz \sum_{n=0}^{+\infty} ((1-p)z)^n \\
&= \frac{pz}{1-(1-p)z}
\end{aligned}$$

Loi de Poisson

Soit X une variables aléatoire suivant une loi de Poisson de paramètre λ. On a

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

8.1.3 Fonction génératrice et loi

Théorème 67. *Soit X une variable aléatoire de loi ν sur $\mathbb{N}$. Sur $[0,1[$, la fonction $x \mapsto G_X(x)$ est infiniment dérivable et ces dérivées sont toutes strictement positives, avec*

$$G_X^{(n)}(s) = \mathbb{E}X(X-1)\dots(X-n+1)s^{X-n}$$

En particulier

$$P(X=n) = \frac{G_X^{(n)}(0)}{n!},$$

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 ouverte et y est infiniment dérivable, avec pour tout z dans le disque ouvert unité :

$$G_X^{(n)}(z) = \sum_{k=n}^{+\infty} k(k-1)\dots(k-n+1)P(X=k)z^{k-n}$$

Il suffit maintenant d'appliquer le théorème de transfert pour constater que le membre de droite est l'espérance de $X(X-1)\dots X(X-n+1)z^{X-n}$.

En prenant $z=0$, on obtient

$$\begin{aligned} G_X^{(n)}(0) &= \mathbb{E}X(X-1)\dots(X-n+1)1\!\!1_{\{X-n=0\}} \\ &= \mathbb{E}n(n-1)\dots(n-n+1)1\!\!1_{\{X-n=0\}} \\ &= n!P(X=n) \end{aligned}$$

La restriction à un intervalle de $\mathbb{R}$ d'une fonction holomorphe est évidemment une fonction infiniment dérivable et la notion de dérivée coincide. Lorsque $s \in [0,1[$, on a pour tout $\omega \in \Omega$:

$$X(\omega)(X(\omega)-1)\dots X(\omega)(X(\omega)-n+1)s^{X(\omega)-n} \geq 0.$$

Comme l'espérance d'une variable aléatoire positive est positive, le résultat s'ensuit.

□

8.1.4 Application : convolution des lois de Poisson

Théorème 68. *Soit 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 v.a.r. suivant une loi de Poisson de paramètre λ a comme fonction génératrice $s \mapsto e^{-\lambda(1-s)}$, ceci quelque soit $\lambda > 0$. En particulier, il s'ensuit que $G_X(s) = e^{-\lambda(1-s)}$ et $G_Y(s) = e^{-\mu(1-s)}$ Maintenant $G_{X+Y}(s) = G_X(s)G_Y(s) = e^{-\lambda(1-s)}e^{-\mu(1-s)} = e^{-(\lambda+\mu)(1-s)}$. Ainsi $X+Y$ a la même fonction génératrice qu'une loi de Poisson de paramètre $\lambda+\mu$. Mais d'après le théorème 67, la fonction génératrice détermine la loi, donc $X+Y$ suit la loi de Poisson de paramètre $\lambda+\mu$. □

8.1.5 Fonction génératrice et espérance

Théorème 69. *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 $x \in [0,1[$,

$$\frac{G_X(1)-G_X(x)}{1-x} = \sum_{n=0}^{+\infty} \frac{1-x^n}{1-x}\nu(n)$$

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 $\lim_{x\to 1^-} \frac{1-x^n}{1-x} = n$. D'après le théorème de convergence monotone (on intègre sur $\mathbb{N}$ par rapport à la mesure de comptage), on a donc on a

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

□

8.2 Fonctions caractéristiques

8.2.1 Motivations

La fonction caractéristique est un outil analogue à la fonction génératrice, qui permet de généraliser aux variables aléatoires à valeurs réelles et même aux vecteurs aléatoires les techniques des fonctions génératrices.
Définition: On appelle fonction caractéristique d'une mesure de probabilité sur $\mathbb{R}^d$ la fonction complexe définie en tout point de $\mathbb{R}^d$ par

$$\forall t \in \mathbb{R}^d \quad \varphi_\mu(t_1,\dots,t_n) = \varphi_{P_X}(t) = \int_{\mathbb{R}^d} \exp(i\langle t,x\rangle)\, d\mu(x) = \int_{\mathbb{R}^d} \exp(i(t_1x_1+t_2x_2+\dots+t_nx_n))\, d\mu(x)$$

Par extension, on appelle fonction génératrice d'un vecteur aléatoire X et on note φ_X la fonction génératrice de sa loi. Ainsi

$$\forall t \in \mathbb{R}^d \quad \varphi_X(t) = \int_{\mathbb{R}^d} \exp(i\langle t,x\rangle)\, dP_X(x) = \mathbb{E}e^{i\langle X,t\rangle}.$$

On va démontrer ici un important résultat d'analyse, qui justifie la dénomination de fonction caractéristique et rend cette outil pertinent pour nos besoins.

Théorème 70. *Soit μ et ν deux mesures de probabilité sur $\mathbb{R}^d$. On a*

$$\forall t \in \mathbb{R}^d \quad \varphi_\mu(t) = \varphi_\nu(t) \iff \mu = \nu.$$

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 :

$$\mu(]a,b[) + \frac{1}{2}(\mu(a)+\mu(b)) = \lim_{T\to+\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita}-e^{-itb}}{it}\varphi_\mu(t)\ dt.$$

Posons

$$I_T = \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita}-e^{-itb}}{it}\varphi_\mu(t)\ dt.$$

Remplaçons φ_μ par sa définition :

$$I_T = \frac{1}{2\pi}\int_{-T}^{T}\int \frac{e^{-ita}-e^{-itb}}{it}e^{itx}\ d\mu(x)\ dt.$$

On peut utiliser le théorème de Fubini, ce qui nous donne

$$\begin{aligned} I_T &= \frac{1}{2\pi}\int\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}e^{itx}\ dt\ d\mu(x) \\ &= \frac{1}{2\pi}\int\int_{-T}^{T}\frac{e^{it(x-a)}-e^{it(x-b)}}{it}\ dt\ d\mu(x) \end{aligned}$$

Or

$$\begin{aligned} \int_{-T}^{T}\frac{e^{i\lambda x}}{ix}\ dx &= \int_{-T}^{T}\frac{e^{i\lambda(-x)}}{i(-x)}\ dx \\ &= \int_{-T}^{T}\frac{\sin\lambda x}{x}\ dx \\ &= \operatorname{sign}(\lambda)\int_{-|\lambda|T}^{|\lambda|T}\frac{\sin x}{x}\ dx, \end{aligned}$$

d'où

$$I_T = \int\left(\frac{\operatorname{sign}(x-a)}{2\pi}\int_{-|x-a|T}^{|x-a|T}\frac{\sin x}{x}\ dx - \frac{\operatorname{sign}(x-b)}{2\pi}\int_{-|x-b|T}^{|x-b|T}\frac{\sin x}{x}\ dx\right)\ d\mu(x)$$

L'application

$$y \mapsto \int_{-y}^{y}\frac{\sin x}{x}\ dx$$

est une application continue qui admet comme limite π lorsque y tend vers l'infini. En particulier, sa norme est bornée par une 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 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 μ 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 $\mathbb{R}\backslash\{x \in \mathbb{R}; \mu(\{x\}) > 0$ ou $\nu(\{x\})$. Mais ces ensembles forment un π-système qui engendre la tribu, donc les deux mesures coïncident. □

Donnons une conséquence frappante de cette proposition qui nous sera utile dans l'étude des vecteurs gaussiens.

Théorème 71. *Soit 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 que X et Y ont même fonction caractéristique, ce qui assurera qu'ils ont même loi. Soit $a \in \mathbb{R}^d$ quelconque. On pose $Z = \langle X, a\rangle$ et $T = \langle Y, a\rangle$. Z et T ont même loi, donc $\mathbb{E}e^{iZ} = \mathbb{E}e^{iT}$. Mais $\mathbb{E}e^{iZ} = \mathbb{E}e^{i\langle X,a\rangle} = \varphi_X(a)$ et $\mathbb{E}e^{iT} = \mathbb{E}e^{i\langle Y,a\rangle} = \varphi_Y(a)$, donc $\varphi_X(a) = \varphi_Y(a)$. Ainsi $\varphi_X = \varphi_Y$, donc X et Y ont même loi. □

Le théorème très simple ci-après est d'usage courant.

Théorème 72. *Soit X une 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*

$$\forall t \in \mathbb{R}^n \quad \varphi_Y(t) = e^{i\langle b,t\rangle}\varphi_X(A^*t).$$

Démonstration.

$$\varphi_Y(t) = \mathbb{E}e^{i\langle Y,t\rangle} = \mathbb{E}e^{i\langle AX+b,t\rangle} = \mathbb{E}e^{i\langle AX,t\rangle}e^{i\langle b,t\rangle}\mathbb{E}e^{i\langle X,A^*t\rangle}e^{i\langle b,t\rangle} = e^{i\langle b,t\rangle}\varphi_X(A^*t)$$

□

8.2.2 Propriétés des fonctions caractéristiques

Théorème 73. *Soit μ une mesure de probabilité sur $\mathbb{R}^d$. On a les propriétés suivantes :*
- $\varphi_\mu(0) = 1$
- $|\varphi_\mu| \le 1$.
- *φ est uniformément continue sur $\mathbb{R}^d$.*
- *Si $e_1, \dots, e_p$ sont des éléments de $\mathbb{R}^n$, la matrice A de taille $p\times p$ définie par $a_{k,l} = \varphi_\mu(e_k - e_l)$ est hermitienne positive.*

Démonstration.
- $\varphi_\mu(0) = \int_{\mathbb{R}^d} e^{i\langle x,0\rangle}d\mu(x) = \int_{\mathbb{R}^d} 1\ d\mu = 1$.
- $|\varphi_\mu(t)| = |\int_{\mathbb{R}^d} e^{i\langle x,t\rangle}d\mu(x)| \le \int_{\mathbb{R}^d} |e^{i\langle x,t\rangle}|d\mu(x) = \int_{\mathbb{R}^d} 1\ d\mu = 1$.

– Soient $t, t' \in \mathbb{R}^d$

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

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.

– On remarque que

$$\begin{aligned}
\int_E |\sum_{k=1}^p x_k \exp(i\langle e_k, x\rangle)|^2\ d\mu(x) &= \int_E \sum_{k=1}^p \sum_{l=1}^p x_k x_l \exp(i\langle e_k - e_l, x\rangle)\ d\mu(x) \\
&= \sum_{k=1}^p \sum_{l=1}^p x_k x_l \int_E \exp(i\langle e_k - e_l, x\rangle)\ d\mu(x) \\
&= \sum_{k=1}^p \sum_{l=1}^p x_k x_l \varphi_\mu(e_k - e_l)
\end{aligned}$$

□

Exemple : pour $n = 1$, si f est une fonction positive 2π périodique intégrable sur $[-\pi, \pi[$, on peut prendre pour μ la mesure

$$d\mu(x) = \frac{f(x)\mathbb{1}_{[-\pi,\pi[}(x)}{2\pi}\ d\lambda(x)$$

$\varphi_{\mu(k)}$ 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 car

$$\begin{aligned}
\sum_{k=1}^p \sum_{l=1}^p x_k x_l \varphi_\mu(k-l) = 0 &\implies \int_E |\sum_{k=1}^p x_k \exp(ikx)|^2\ d\mu(x) = 0 \\
&\implies |\sum_{k=1}^p x_k \exp(ikx)|^2 f(x) = 0 \text{ p.p. sur}[-\pi, \pi],
\end{aligned}$$

Comme un polynôme trigonométrique non nul n'a qu'un nombre fini de zéro (ça se prolonge en un polynôme sur $\mathbb{C}$), on en déduit que f est presque partout nul, ce qu'on avait exclu.

En fait, un difficile théorème 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 (Bochner). *Soit $\varphi : \mathbb{R}^n \to \mathbb{C}$ une fonction continue en 0, vérifiant $\varphi(0) = 1$ 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 $a_{k,l} = \varphi(e_k - e_l)$ est hermitienne positive.*

Alors il existe une mesure de probabilité μ sur $\mathbb{R}^n$ telle que $\varphi = \varphi_\mu$.

8.2.3 Fonction caractéristique et indépendance

Théorème 74. *Soient X et Y deux vecteurs aléatoires indépendants à valeurs dans $\mathbb{R}^d$. Alors*

$$\forall t \in \mathbb{R}^d \quad \varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t).$$

Démonstration.

$$\varphi_{X+Y}(t) = \mathbb{E}e^{i\langle t,X+Y\rangle} = \mathbb{E}e^{i\langle t,X\rangle}e^{i\langle t,Y\rangle} = \mathbb{E}e^{i\langle t,X\rangle}\mathbb{E}e^{i\langle t,Y\rangle} = \varphi_X(t)\varphi_Y(t).$$

□

Théorème 75. *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*

$$\forall (s, t) \quad \varphi_{(X,Y)}(s, t) = \varphi_X(s)\varphi_Y(t).$$

Démonstration.

$$\varphi_{(X,Y)}(s,t) = \mathbb{E}e^{i(\langle s,X\rangle+\langle t,Y\rangle)} = \mathbb{E}e^{i\langle s,X\rangle}e^{i\langle t,Y\rangle} = \mathbb{E}e^{i\langle s,X\rangle}\mathbb{E}e^{i\langle t,Y\rangle} = \varphi_X(s)\varphi_Y(t).$$

□

Corollaire 15. *Si μ et ν sont des mesures de probabilité respectivement définies sur $\mathbb{R}^n$ et $\mathbb{R}^p$, alors*

$$\forall (s, t) \quad \varphi_{\mu\otimes\nu}(s, t) = \varphi_\mu(s)\varphi_\nu(t).$$

Démonstration. Il suffit de considérer un couple (X, Y) de variables aléatoires de loi $\mu \otimes \nu$. X et Y sont indépendantes, donc

$$\varphi_{\mu\otimes\nu}(s, t) = \varphi_{(X,Y)}(s, t) = \varphi_X(s)\varphi_Y(t) = \varphi_\mu(s)\varphi_\nu(t)$$

□

Le théorème 75 admet une réciproque

Théorème 76. *Soient X et Y deux vecteurs aléatoires, X étant à valeurs dans $\mathbb{R}^n$ et Y à valeurs dans $\mathbb{R}^p$. Si*

$$\forall (s,t) \quad \varphi_{(X,Y)}(s,t) = \varphi_X(s)\varphi_Y(t),$$

alors X et Y sont indépendants.

Démonstration. Ainsi $\varphi_{P_{(X,Y)}}(s,t) = \varphi_{P_X}(s)\varphi_{P_Y}(t)$. Mais d'après le corollaire précédent, $\varphi_{P_X}(s)\varphi_{P_Y}(t) = \varphi_{P_X \otimes P_Y}(s,t)$. Ainsi $\varphi_{P_{(X,Y)}}(s,t) = \varphi_{P_X \otimes P_Y}(s,t)$. Comme la fonction caractéristique caractérise la loi, on a $P_{(X,Y)} = P_X \otimes P_Y$, ce qui signifie que X et Y sont indépendants. □

En revanche, le théorème 74 n'admet pas de réciproque : on verra en exercice des variables aléatoires X et Y non indépendantes telles que $\varphi_{X+Y} = \varphi_X \varphi_Y$.

8.2.4 Fonction caractéristique et moments

Théorème 77. *Si X est une variable aléatoire sur $\mathbb{R}^d$ 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

$$\forall u \in \mathbb{R}^d \quad \frac{\partial^k \varphi_X}{\partial_1^{k_1} \dots \partial_d^{k_d}}(u) = i^k \int_{\mathbb{R}^d} e^{i\langle u,x\rangle} \prod_{j=1}^d x_j^{k_j} \; d\mu(x) \tag{8.1}$$

Ainsi

$$\frac{\partial^k \varphi_X}{\partial_1^{k_1} \dots \partial_d^{k_d}}(0) = i^k \mathbb{E} \prod_{j=1}^d X_j^{k_j}. \tag{8.2}$$

Démonstration. La première formule se prouve par récurrence sur k en utilisant le théorème de domination de Lebesgue pour la dérivation sous le signe somme. Pour la deuxième formule, il suffit de prendre $u = 0$. □

Dans le cas des variables aléatoires réelles, la forme du théorème est évidemment plus simple :

Corollaire 16. *Si X est une variable aléatoire réelle telle admettant un moment d'ordre N, alors la fonction caractéristique de X est de classe C^N et on a*

$$\forall k \in \{1, \dots, N\} \quad \varphi_X^{(k)}(0) = i^k \mathbb{E} X^k \tag{8.3}$$

En particulier, si X admet un moment d'ordre 2 et est centrée avec une variance σ^2 , on a le développement limité en 0 :

$$\varphi_X(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2)$$

Démonstration. La reformulation ne pose pas de mystère. Si X admet un moment d'ordre 2, φ_X est de classe C^2, donc

$$\varphi_X(t) = \varphi_X(0) + \varphi'_X(0)t + \frac{\varphi''_X(0)}{2}t^2 + o(t^2)$$

Maintenant $\varphi_X(0) = 1$, $\varphi'_X(0) = i\mathbb{E}X = 0$, $\varphi''_X(0) = -\mathbb{E}X^2 = -\operatorname{Var} X$ car $\mathbb{E}X = 0$. Il suffit de substituer pour conclure. □

8.2.5 Fonctions caractéristiques des variables aléatoires à valeurs dans $\mathbb{N}$

Pour une variable aléatoire à valeur dans $\mathbb{N}$, le calcul de la fonction caractéristique est équivalent à celui de la fonction génératrice.

En effet, on a la formule :

$$\forall t \in \mathbb{R} \varphi_X(t) = \mathbb{E}e^{itX} = \mathbb{E}(e^{it})^X = G_X(e^{it}).$$

On laisse ainsi au lecteur le soin de calculer la fonction caractéristique de la loi de Bernoulli, de la loi binomiale, de la loi de Poisson, de la loi géométrique.

8.2.6 Quelques fonctions caractéristiques de mesures à densité

Pour des mesures à densités, le calcul de la fonction caractéristique est en fait (au signe près) 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 peuvent souvent être calculées en utilisant des techniques issues de la théorie de la variable complexe, par exemple en utilisant une méthode des résidus ou encore en appliquant un théorème de prolongement analytique. C'est ce que nous allons voir dans les exemples qui vont suivre.

Il y a bien sûr d'autres méthodes pouvant être utiles : reconnaitre la transformée de Fourier d'une fonction connue et utiliser un théorème d'inversion ou expliciter une équation différentielle vérifiée par la fonction caractéristique, puis la résoudre. Nous ne traiterons pas cela ici et vous renvoyons à un cours d'analyse.

Loi uniforme sur $[a, b]$

On commence par calculer la fonction caractéristique de la loi uniforme sur $[-1, 1]$: on a

Pour $t \neq 0$, on a

$$\varphi_X(t) = \mathbb{E}e^{itx} = \int_{-1}^{1} \frac{1}{2} e^{itx}\ dx = [\frac{e^{itx}}{2it}]_{-1}^{1} = \frac{e^{it} - e^{-it}}{2it} = \frac{\sin t}{t}$$

La formule se prolonge par continuité pour $t = 0$.

Maintenant, si on pose $Y = \frac{a+b}{2} + (b-a)X$, Y suit la loi uniforme sur $[a, b]$ et on a

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

Loi exponentielle de paramètre λ

On commence par calculer la fonction caractéristique de la loi exponentielle de paramètre 1 : on a

Pour $t \neq 0$, on a

$$\varphi_X(t) = \mathbb{E}e^{itx} = \int_{0}^{+\infty} e^{-x} e^{itx}\ dx = [\frac{e^{(-1+it)x}}{-1+it}]_0^{+\infty} = \frac{0-1}{-1+it} = \frac{1}{1-it}$$

La formule se prolonge par continuité pour $t = 0$.

Maintenant, si on pose $Y = \frac{1}{\lambda}X$, Y suit la loi $\mathcal{E}(\lambda)$ et on a

$$\varphi_Y(t) = \varphi_X(\frac{1}{\lambda}t) = \frac{1}{1 - i\frac{t}{\lambda}} = \frac{\lambda}{\lambda - it}.$$

Variables aléatoires gaussiennes

Théorème 78. *La fonction caractéristique de la loi normale* $\mathcal{N}(m, \sigma^2)$ *est*

$$t \mapsto \exp(imt) \exp(-\frac{1}{2}\sigma^2 t^2).$$

Démonstration. On va d'abord calculer la fonction caratéristique de $X \sim \mathcal{N}(0, 1)$. Nous devons calculer

$$\varphi_X(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-x^2/2) \exp(itx)\ dx.$$

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 $x \mapsto \exp(-x^2/2)\exp(Rx)$, car

$$\begin{array}{rcl}\forall x \in \mathbb{R} \quad \forall z \in D(0,R) & & \\ |\exp(-x^2/2)\exp(xz)| & = & \exp(-x^2/2)\exp(x\mathrm{Re}\ z) \\ & \leq & \exp(-x^2/2)\exp(Rx)\end{array}$$

Il s'ensuit que $z \mapsto \int_{\mathbb{R}} f_z(x) d\lambda(x)$ est holomorphe.

Pour z réel, on a

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \exp(-x^2/2)\exp(zx)\ d\lambda(x) &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \exp(-((x-z)^2-z^2)/2)\ d\lambda(x) \\ &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \exp(-(x-z)^2/2)\ d\lambda(x)\exp(\frac{z^2}{2}) \\ &= \exp(\frac{z^2}{2}),\end{aligned}$$

car l'expression intégrée n'est autre que la densité de la loi $\mathcal{N}(z,1)$. Mais si deux fonctions holomorphes coincident sur $\mathbb{R}$, elles coïncident sur $\mathbb{C}$. On a donc

$$\forall z \in \mathbb{C} \quad \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \exp(-x^2/2)\exp(zx)\ d\lambda(x) = \exp(\frac{z^2}{2}).$$

On particularise alors z en it, t étant réel, et on obtient

$$\varphi_X(t) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \exp(-x^2/2)\exp(itx)\ dx = \exp(-\frac{t^2}{2}).$$

Pour passer au cas général, on pose $Y = \sigma X + m$; on a $Y \sim \mathcal{N}(m, \sigma^2)$, et alors $\varphi_Y(t) = \mathbb{E}e^{itY} = e^{it(\sigma X + m)} = e^{imt}\mathbb{E}e^{it\sigma X} = e^{imt}\varphi_X(\sigma t) = e^{imt}e^{-\frac{\sigma^2 t^2}{2}}$.

□

Lois de Cauchy

Théorème 79. *La fonction caractéristique de la loi de Cauchy $\mathcal{C}(a,b)$ est*

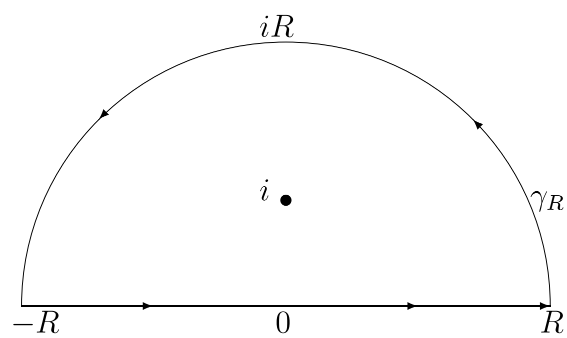
$$\varphi(t) = e^{iat}e^{-b|t|}.$$

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

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

Démonstration. On va d'abord calculer la fonction caractéristique de $\mathcal{C}(0,1)$ en un point $t>0$

Pour $R>1$, on intègre la forme différentielle $f(z)=\frac{e^{itz}}{1+z^2}$ sur le contour γ_R.



Le seul pôle de f à l'intérieur de la courbe est en 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

$$\forall R>1 \quad \frac{1}{\pi}\int_{\gamma_R} f(z)\ dz = e^{-t}.$$

Or

$$\int_{\gamma_R} f(z)\ dz = \int_{-R}^{R} f(t)\ dt + iR\int_0^{\pi} f(Re^{i\theta})e^{i\theta}\ d\theta.$$

Mais, lorsque $z=Re^{i\theta}$, on a

$$\begin{aligned}|f(Re^{i\theta})e^{i\theta}| &= \frac{e^{-Rt\sin\theta}}{|1+z^2|}\\ &\leq \frac{1}{R^2-1}\end{aligned}$$

(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

$$e^{-t} = \frac{1}{\pi} \int_{-\infty}^{+\infty} f(t)\ dt$$

Ainsi si $X \sim \mathcal{C}(0,1)$, on a

$$\forall t > 0 \quad \varphi_X(t) = e^{-t}.$$

Comme la loi de X est symétrique ($P_X = P_{-X}$) et que $\varphi(0) = 1$, on a

$$\forall t \in \mathbb{R} \quad \varphi_X(t) = e^{-|t|}.$$

Si on pose $Y = bX + a$, on a $Y \sim \mathcal{C}(a, b)$, et alors $\varphi_Y(t) = \mathbb{E}e^{itY} = e^{it(bX+a)} = e^{ia}\mathbb{E}e^{itbX} = e^{ia}\varphi_X(bt) = e^{ia}e^{-b|t|}$.

□

8.3 Exercices sur les fonctions génératrices et les fonctions caractéristiques

1. (a) On suppose que X et Y sont des variables aléatoires à valeurs entières, de fonctions génératrices f et g. Soit A un événement indépendant de X et de Y, avec $P(A) = p$. On note Z la variable aléatoire définie par

$$Z(\omega) = \begin{cases} X(\omega) \text{ si } \omega \in A \\ Y(\omega) \text{ si } \omega \notin A \end{cases}$$

   Montrer que la fonction génératrice de Z est $pf + (1-p)g$.

   (b) On lance 3 fois de suite un dé à 6 faces. À chaque série de 3 lancers est associé un score. Le score se calcule ainsi Si le troisième lancer est un "1", le score est le nombre de nombre pairs apparus dans les deux premiers lancers. Sinon, le score est le nombre de "6" apparus dans les deux premiers lancers.
   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.
   On demande de calculer la loi du score.

2. Soient (X_n) une suite de variables aléatoires indépendantes de même loi non dégénérée à valeurs dans $\mathbb{N}$ et T une variable aléatoire à valeurs dans $\mathbb{N}^*$ indépendante des précédentes. On définit pour tout $n \in \mathbb{N}^*$ la variable $S_n = \sum_{j=1}^{n} X_j$, puis $S(\omega) = S_{T(\omega)}(\omega)$ pour tout $\omega \in \Omega$.

(a) Si G_T et G_X désignent les fonctions génératrices de T et X_1, montrer que la fonction génératrice de S est donnée par $G_S = G_T \circ G_X$.

(b) *Formule de Wald*
Si X_1 et T admettent les moyennes (espérances) m et t, montrer que $\mathbb{E}[S] = mt$.

3. Soit N une variable aléatoire suivant une loi de Poisson de paramètre λ et $X_1, \dots, X_n \dots$ une suite de v.a.r.i.i.d de lois de Bernoulli de paramètre p, cette suite étant indépendante de N. Montrer que $S = X_1 + X_2 \cdots + X_N$ suit loi de Poisson de paramètre λp.

4. (a) 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 Bernoulli de paramètre p. On suppose que K et L sont indépendants. Déterminer la fonction génératrice de $\langle K, L \rangle$.

   (b) On suppose que pour tout $k \geq 0$, X_k suit une loi de Poisson de paramètre $n\lambda$. 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.

   (c) En déduire que $\langle K, L \rangle$ suit la même loi que X_T.

5. Un joueur joue à pile ou face avec $n \geq 2$ pièces équilibrées de la manière suivante : il lance simultanément ces n pièces ; 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 obtenu lors de cette deuxième série de jets. On note X_1 le nombre de piles obtenu à la première étape, et X_2 le gain du joueur.

   (a) Déterminer la fonction génératrice de X_2.

   (b) En déduire
$$\forall k \in \{0, \dots, n\} \quad P(X_2 = k) = \binom{n}{k} \frac{1}{3^k} (\frac{3}{4})^n$$

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

7. Donner un exemple de variable aléatoire telle que $(\varphi_X)^2 = \varphi_{2X}$. En déduire que la propriété
$$\forall t \in \mathbb{R} \quad \varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$$
n'implique PAS que X et Y sont indépendantes.

8. Soit $X \sim \mathcal{N}(0,1)$. Montrer que $\forall n \geq 0 \quad \mathbb{E}X^{2n} = \frac{(2n)!}{n!2^n}$.
9. Soit X et Y deux variables aléatoires admettant chacune un moment d'ordre 3. On pose $\varphi(x,y) = \mathbb{E}\exp(i(xX+yY))$. Montrer que X^2Y est intégrable. Régularité de φ ? Exprimer $\mathbb{E}X^2Y$ en fonction de φ.
10. Soit X une variable atéatoire suivant la loi exponentielle de paramètre 1 et ε une variable aléatoire prenant la valeur 1 avec probabilité 1/2 et la valeur -1 avec probabilité 1/2. On suppose que X et ε sont indépendantes. On appelle *Loi de Laplace* la loi de εX.
    (a) Montrer que la loi de Laplace est une loi à densité.
    (b) Calculer la fonction caractéristique de la loi de Laplace.
11. On considère quatre variables aléatoires indépendantes $X_{1,1}$, $X_{1,2}$, $X_{2,1}$, $X_{2,2}$ de même loi $\mathcal{N}(0,1)$.
    (a) On note $U = X_{1,1}X_{2,2}$ et $V = X_{1,2}X_{2,1}$. Déterminer la fontion caractéristique de U et de V.
    (b) 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.
12. (a) Montrer qu'il existe une constante K telle que

$$f(x) = \frac{K}{\|x\|_2^{1/2}} \mathbb{1}_{B(0,1)}(x)$$

    soit la densité d'une loi de probabilité sur $\mathbb{R}^3$.
    (b) Soit μ cette loi. Montrer qu'il existe une constante L non nulle telle que

$$\varphi_\mu(t) = L\frac{\cos\|t\|}{\|t\|^2} + O\left(\frac{1}{\|t\|^{2,5}}\right)$$

Chapitre 9

Lois des grands nombres

9.1 Inégalités classiques

9.1.1 Inégalité de Markov

Théorème 80. *Soit X une variable aléatoire positive, intégrable. Alors, on a*

$$\forall a > 0 \quad P(X \geq a) \leq \frac{\mathbb{E}X}{a}.$$

Démonstration. Comme X est positive, on a

$$\mathbb{E}X = \int_{\Omega} x \, dP(\omega) \geq \int_{\{X \geq a\}} x \, dP(\omega) \geq \int_{\{X \geq a\}} a \, dP(\omega) = aP(X \geq a).$$

□

9.1.2 Inégalité de Tchebytchef

Théorème 81. *Soit X une variable aléatoire admettant un moment d'ordre 2. Alors, on a*

$$\forall a > 0 \quad P(|X - \mathbb{E}X| \geq a) \leq \frac{\operatorname{Var} X}{a^2}.$$

Démonstration.

$$P(|X - \mathbb{E}X| \geq a) = P(|X - \mathbb{E}X|^2 \geq a^2)$$

Il suffit alors d'appliquer l'inégalité de Markov à la variable aléatoire $Y = |X - \mathbb{E}X|^2$. Comme $\mathbb{E}Y = \operatorname{Var} X$, l'inégalité s'ensuit. □

9.2 Convergence presque sûre

Définition : on dit qu'une suite de variables (ou de vecteurs) aléatoires $(X_n)_{n\geq 0}$ converge presque sûrement vers une variable (ou un vecteur) aléatoire X lorsqu'il existe un ensemble mesurable $\Omega' \subset \Omega$ tel que $P(\Omega') = 1$ et que

$$\forall \omega \in \Omega' \subset \Omega \quad X_n(\omega) \to X(\omega).$$

On écrit alors $X_n \xrightarrow[n\to+\infty]{\text{p.s.}} X$.

La convergence presque sûre n'est autre que la convergence presque partout relativement à une mesure de probabilité. On a alors les résultats classiques suivants : si $X_n \xrightarrow[n\to+\infty]{\text{p.s.}} X$ et $Y_n \xrightarrow[n\to+\infty]{\text{p.s.}} Y$, (avec X et Y dans $\mathbb{R}^d$, $d \geq 1$) alors

- $\forall a \in \mathbb{R} \quad aX_n \xrightarrow[n\to+\infty]{\text{p.s.}} aX$.
- $X_n + Y_n \xrightarrow[n\to+\infty]{\text{p.s.}} X + Y$.
- $\langle X_n, Y_n \rangle \xrightarrow[n\to+\infty]{\text{p.s.}} \langle X, Y \rangle$.

Plus généralement, si $X_1, \ldots, X_n, \ldots, X$ sont à valeurs dans un ouvert O et que $X_n \xrightarrow[n\to+\infty]{\text{p.s.}} X$, alors pour toute fonction f continue définie sur O, on a $f(X_n) \xrightarrow[n\to+\infty]{\text{p.s.}} f(X)$.

Il peut être intéressant de remarquer que la convergence presque sûre d'une suite de vecteurs aléatoires est équivalente à la convergence presque sûre de chacune des composantes.

9.2.1 Rappels d'analyse

En probabilités, le retour aux ε est très fréquent. Si l'on ne veut pas que cela devienne trop compliqué, il importe de bien connaître les outils d'analyse permettant de simplifier les choses.

Pour toute suite $(x_n)_{n\geq 0}$ de nombres réels, on peut définir

$$\overline{\lim_{n\to+\infty}} \; x_n = \lim_{n\to+\infty} \sup_{k\geq n} x_k$$

et

$$\underline{\lim_{n\to+\infty}} \; x_n = \lim_{n\to+\infty} \inf_{k\geq n} x_k.$$

Ces deux limites existent toujours dans $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty; +\infty\}$. La suite $(x_n)_n$ converge dans $\overline{\mathbb{R}}$ si et seulement si ces deux limites sont égales. Rappelons quelques propriétés des limites supérieures. Pour plus de détails, on pourra se reporter à un cours d'analyse.

– Pour $x \in \mathbb{R}$

$$\lim_{n\to+\infty} x_n = x \iff \overline{\lim_{n\to+\infty}} |x_n - x| = 0$$

–

$$\overline{\lim_{n\to+\infty}} x_n \leq M \iff \forall \varepsilon > 0 \quad \{n : x_n \geq M + \varepsilon\} \text{ est fini} \tag{9.1}$$

$$\overline{\lim_{n\to+\infty}} x_n \geq M \iff \forall \varepsilon > 0 \quad \{n : x_n \geq M - \varepsilon\} \text{ est infini} \tag{9.2}$$

– En prenant la contraposée de (9.1) et(9.2) , on a

$$\overline{\lim_{n\to+\infty}} x_n > M \iff \exists \varepsilon > 0 \quad \{n : x_n \geq M + \varepsilon\} \text{ est infini}$$

$$\overline{\lim_{n\to+\infty}} x_n < M \iff \exists \varepsilon > 0 \quad \{n : x_n \geq M - \varepsilon\} \text{ est fini}$$

–

$$\overline{\lim_{n\to+\infty}} (x_n + y_n) \leq \overline{\lim_{n\to+\infty}} x_n + \overline{\lim_{n\to+\infty}} y_n.$$

Nous verrons en exercice un exemple où l'inégalité est stricte

9.2.2 Limites supérieures, inférieures d'ensembles

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

$$\overline{\lim_{n\to+\infty}} A_n = \bigcap_{n\geq 1} \bigcup_{k\geq n} A_k$$

et

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

Ainsi la limite supérieure d'une suite d'ensembles est l'ensemble des points qui appartiennent à une infinité de ces ensembles, tandis que la limite inférieure d'une suite d'ensembles est l'ensemble des points qui appartient à tous ces ensembles à partir d'un certain rang.

Ainsi, dire que $\{n : X_n(\omega) \geq M - \varepsilon\}$ est infini , c'est dire que ω appartient à $\overline{\lim_{n\to+\infty}} \{\omega : X_n(\omega) \geq M - \varepsilon\}$.

On en déduit

$$\{\overline{\lim_{n\to+\infty}} X_n \geq M\} = \bigcap_{\varepsilon>0} \overline{\lim_{n\to+\infty}} \{X_n \geq M-\varepsilon\} \tag{9.3}$$

Par ailleurs, dire que

$$\{n : x_n \geq M+\varepsilon\} \text{ est fini },$$

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

$$\{\overline{\lim_{n\to+\infty}} X_n \leq M\} = \bigcap_{\varepsilon>0} \underline{\lim_{n\to+\infty}} \{X_n < M+\varepsilon\} \tag{9.4}$$

Si on remplace X_n par $-X_n$ et M par $-M$ dans (9.4), on obtient :

$$\{\overline{\lim_{n\to+\infty}} -X_n \leq -M\} = \bigcap_{\varepsilon>0} \underline{\lim_{n\to+\infty}} \{-X_n < -M+\varepsilon\}$$

Soit

$$\{\underline{\lim_{n\to+\infty}} X_n \geq M\} = \bigcap_{\varepsilon>0} \underline{\lim_{n\to+\infty}} \{X_n > M-\varepsilon\} \tag{9.5}$$

Et en passant aux complémentaires dans (9.4) et (9.5), on a

$$\{\overline{\lim_{n\to+\infty}} X_n > M\} = \bigcup_{\varepsilon>0} \overline{\lim_{n\to+\infty}} \{X_n \geq M+\varepsilon\} \tag{9.6}$$

et

$$\{\underline{\lim_{n\to+\infty}} X_n < M\} = \bigcup_{\varepsilon>0} \overline{\lim_{n\to+\infty}} \{X_n \leq M-\varepsilon\} \tag{9.7}$$

Si on fait subir à la formule (9.3) les mêmes transformations qu'à (9.4), on peut obtenir 3 autres formules.

Dans la pratique, comment fait-on si l'on veut montrer que $\overline{\lim}_{n\to+\infty} X_n = M$ presque sûrement ? Comme vous l'avez deviné, on montre $\overline{\lim}_{n\to+\infty} X_n \geq M$ presque sûrement, puis $\overline{\lim}_{n\to+\infty} X_n \leq M$ presque sûrement Comme la suite $\overline{\lim}_{n\to+\infty} \{X_n \geq M-\varepsilon\}$ est monotone en ε, on a

$$\{\overline{\lim_{n\to+\infty}} X_n \geq M\} = \cap_{\varepsilon>\in\mathbb{Q}_+^*} \overline{\lim_{n\to+\infty}} \{X_n \geq M-\varepsilon\} \tag{9.8}$$

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

Théorème 82. *L'intersection d'une famille dénombrable d'événements est de probabilité 1 si et seulement si chacun des événements est de probabilité 1.*

Démonstration. Soit D un ensemble d'indexs dénombrable. $(A_n)_{n\in D}$ une famille d'événements indexée par D. On pose $A = \bigcap_{n\in D} A_n$. Pour tout n, $A \subset A_n$, donc $P(A) \leq P(A_n)$. Ainsi si $P(A) = 1$, on a pour tout $n \in D$ $P(A_n) = 1$. Réciproquement, on a

$$\begin{aligned} P(A^c) &= P(\bigcup_{n\in D} A_n^c) \\ &\leq \sum_{n\in D} P(A_n^c) \\ &\leq \sum_{n\in D} 0 \end{aligned}$$

Donc $P(A) = 1 - P(A^c) = 1 - 0 = 1$. □

Pour prouver que $\overline{\lim_{n\to+\infty}} X_n \geq M$ presque sûrement, il suffit donc de prouver que $\forall a < M \quad P(\overline{\lim_{n\to+\infty}} \{X_n \geq a\}) = 1$

De la même manière, on voit que pour avoir $\overline{\lim_{n\to+\infty}} X_n \leq M$ presque sûrement, il suffit donc de prouver que $\forall a > M \quad P(\underline{\lim}_{n\to+\infty} \{X_n < a\}) = 1$, on de manière équivalente que $\forall a > M \quad P(\overline{\lim_{n\to+\infty}} \{X_n \geq a\}) = 0$.

On peut donc énoncer le théorème suivant

Théorème 83. *Soit X_n une suite de variables aléatoires et M un réel. On suppose que*

1. $\forall \quad a < M \quad P(\overline{\lim_{n\to+\infty}} \{X_n \geq a\}) = 1$
2. $\forall \quad a > M \quad P(\overline{\lim_{n\to+\infty}} \{X_n \geq a\}) = 0$

Alors

$$\overline{\lim_{n\to+\infty}} X_n = M \textit{ presque sûrement.}$$

Le théorème suivant très important en est une application directe

Théorème 84 (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*

$$\forall \varepsilon > 0 \quad P(\overline{\lim_{n\to+\infty}} \{|X_n - X| \geq \varepsilon\}) = 0.$$

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

9.3 Convergence en probabilité

Définition : On dit que (X_n) converge en probabilité vers X si

$$\forall \varepsilon > 0 \qquad \lim_{n\to+\infty} P(|X_n - X| \geq \varepsilon) = 0.$$

9.3.1 Comparaison avec les autres modes de convergence

Convergence dans L^p et convergence en probabilité

Théorème 85. *La convergence dans L^p $(p \geq 1)$ implique la convergence en probabilité*

Démonstration.

$$P(\|X_n - X\| \geq \varepsilon) = P(\|X_n - X\|^p \geq \varepsilon^p) \leq \frac{\mathbb{E}\|X_n - X\|^p}{\varepsilon^p}.$$

□

Convergence presque sûre et convergence en probabilité

Théorème 86. *La convergence presque sûre implique la convergence en probabilité.*

Démonstration. Soit $\varepsilon > 0$. D'après le théorème 84, on a

$$P(\overline{\lim_{n\to+\infty}} \{|X_n - X| \geq \varepsilon\}) = 0.$$

Or, d'après le théorème de continuité séquencielle décroissante, on a

$$P(\overline{\lim_{n\to+\infty}} \{|X_n - X| \geq \varepsilon\}) = \lim_{n\to+\infty} P(\bigcup_{k\geq n} \{|X_k - X| \geq \varepsilon\})$$

Comme

$$0 \leq P(|X_n - X| \geq \varepsilon) \leq P(\bigcup_{k\geq n} \{|X_k - X| \geq \varepsilon\}),$$

on en déduit que

$$\lim_{n\to+\infty} P(|X_n - X| \geq \varepsilon) = 0.$$

Comme ε est quelconque, on peut dire que X_n converge en probabilité vers X. □

9.3.2 Loi faible des grands nombres

Théorème 87. *Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires de même loi, admettant un moment d'ordre 2 et deux à deux non corrélées.*
On pose

$$S_n = \sum_{k=1}^{n} X_k \text{ et } M_n = \frac{1}{n} S_n.$$

Alors

1. $M_n \xrightarrow[n\to+\infty]{\mathrm{L}^2} \mathbb{E}X_0$. *On dit que M_n converge en moyenne quadratique vers $\mathbb{E}X_0$.*
2. *Et donc* $M_n \xrightarrow[n\to+\infty]{\mathbb{P}} \mathbb{E}X_0$.

Démonstration. $\mathbb{E}M_n = \frac{1}{n}\mathbb{E}S_n = \frac{1}{n}\sum_{k=1}^{n} EX_k = \frac{1}{n}n\mathbb{E}X_0 = \mathbb{E}X_0$. Par conséquent $\mathbb{E}|M_n - \mathbb{E}X_0|^2 = \operatorname{Var} M_n = \frac{1}{n^2}\operatorname{Var} S_n$. Comme les X_k sont 2 à 2 non corrélées, on a

$$\operatorname{Var} S_n = \sum_{k=1}^{n} \operatorname{Var} X_k = n \operatorname{Var} X_1.$$

On a donc

$$\mathbb{E}|M_n - \mathbb{E}X_0|^2 = \operatorname{Var} M_n = \frac{\operatorname{Var} X_1}{n}, \tag{9.9}$$

qui tend bien vers zéro.

□

9.4 Lemmes de Borel-Cantelli

9.4.1 Premier lemme de Borel-Canteli

Théorème 88. *Soit $(A_n)_{n\geq 1}$ une suite d'événements observables. Si la série de terme général $P(A_n)$ est convergente, alors $P(\overline{\lim_{n\to+\infty}} A_n) = 0$.*

Démonstration. On pose $B_n = \underset{k\geq n}{\cup} A_k$. la suite (B_n) est décroissante, et l'intersection des (B_n) est, par définition, $\overline{\lim_{n\to+\infty}} A_n$. D'après le théorème de continuité séquentielle décroissante, on a donc

$$0 \leq P(\overline{\lim_{n\to+\infty}} A_n) = \lim_{n\to+\infty} P(B_n).$$

Or

$$P(B_n) = P(\underset{k\geq n}{\cup} A_k) \leq \sum_{k\geq n} P(A_k) = r_n$$

Comme r_n est le reste d'ordre n d'une série convergente, r_n est de limite nulle, et donc, par comparaison $P(\overline{\lim_{n\to+\infty}} A_n) = 0$. □

9.4.2 Deuxième lemme de Borel-Cantelli

Théorème 89. *Soient $(A_n)_{n\geq 1}$ une suite d'évévements.*
On pose

$$N_n = \sum_{k=1}^{n} \mathbb{1}_{A_k} \text{ et } N = \sum_{k=1}^{+\infty} \mathbb{1}_{A_k}$$

$$m_n = \sum_{k=1}^{n} P(A_k) = \mathbb{E}N_n$$

$$\text{On a } \overline{\lim_{n\to+\infty}} A_n = \{N = +\infty\}.$$

Si

$$\lim_{n\to+\infty} m_n = +\infty \text{ et } \lim_{n\to+\infty} \frac{\operatorname{Var} N_n}{m_n^2} = 0,$$

alors

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

Démonstration. Pour $m_n > a$, on a

$$\begin{aligned} P(N \leq a) &\leq P(N_n \leq a) \\ &\leq P(|N_n - m_n| \geq m_n - a) \leq \frac{\operatorname{Var} N_n}{(m_n - a)^2} \end{aligned}$$

En faisant tendre n vers $+\infty$, on en déduit que

$$\forall a \in \mathbb{N} \quad P(N \leq a) = 0.$$

Ainsi

$$P(N < +\infty) = P(\lim_{a \to +\infty} \uparrow \{N \leq a\}) = \lim_{a \to +\infty} P(N \leq a) = \lim_{a \to +\infty} 0 = 0.$$

□

Théorème 90 (2ème lemme de Borel-Cantelli). *Soit $(A_n)_{n \geq 1}$ une suite d'événements indépendants. Si la série de terme général $P(A_n)$ est divergente, alors $P(\overline{\lim_{n \to +\infty}} A_n) = 1$.*

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

$$\operatorname{Var} N_n = \sum_{k=1}^{n} \operatorname{Var} \mathbb{1}_{A_k} = \sum_{k=1}^{n} P(A_k)(1 - P(A_k) \leq \sum_{k=1}^{n} P(A_k) = m_n$$

Ainsi $\frac{\operatorname{Var} N_n}{m_n^2} = \frac{1}{m_n}$. Comme $\lim_{n \to +\infty} m_n = \sum_{k=1}^{+\infty} P(A_k) = +\infty$, le résultat s'ensuit. □

Exercice : La conclusion du 2ème 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 ?

Théorème 91. *Soit $(X_n)_{n \geq 0}$ une suite convergeant en probabilité vers X. Alors, il existe une sous-suite X_{n_k} telle que $X_{n_k} \xrightarrow[k \to \infty]{p.s.} X$..*

Démonstration. On pose $n_0 = 0$, puis, pour $k \geq 1$:

$$n_k = \inf\{n > n_{k-1}; P(|X_n - X| \geq \frac{1}{k}) \leq \frac{1}{2^k}\}$$

À k fixé, $P(|X_n - X| \geq \frac{1}{k})$ tend vers 0 quand n tend vers l'infini, donc on a bien pour tout k : $n_k < +\infty$.

Maintenant, on a pour tout $k \geq 0$:

$$P(|X_{n_k} - X| \geq \frac{1}{k}) \leq \frac{1}{2^k}$$

Comme la série de terme général converge $\frac{1}{2^k}$, le premier lemme de Borel-Cantelli nous permet d'affirmer que

$$P(\overline{\lim_{k\to\infty}} \{|X_{n_k} - X| \geq \frac{1}{k})\}) = 0,$$

ce qui est équivalent

$$P(\underline{\lim_{k\to\infty}} \{|X_{n_k} - X| < \frac{1}{k})\}) = 1,$$

ce qui veut dire que pour presque tout ω, il existe un $k_0(\omega)$ tel que

$$k \geq k_0(\omega) \Longrightarrow |X_{n_k}(\omega) - X(\omega)| < \frac{1}{k},$$

ce qui implique bien sûr que $X_{n_k}(\omega)$ tend vers $X(\omega)$ pour P-presque tout ω. □

9.5 Loi forte des grands nombres

9.5.1 La loi forte des grands nombres

Théorème 92. *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires deux à deux indépendantes, de même loi μ. On suppose que μ admet un moment d'ordre 1. Alors*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to+\infty]{p.s.} \mathbb{E}X_1.$$

9.5.2 Probabilités et fréquences asymptotiques

Théorème 93. *Soit $(A_n)_{n\geq 0}$ une suite d'événements observables indépendants de même probabilité p. Pour ω dans l'univers Ω On note $N_n(\omega)$ le nombre d'événements qui sont réalisés parmi $A_1, \ldots, A_n$. Ainsi, on a*

$$N_n = \sum_{k=1}^{n} \mathbb{1}_{A_k} \text{ et } f_n = \frac{1}{n} N_n.$$

Alors il existe un événement observable $\tilde{\Omega} \subset \Omega$ avec $P(\tilde{\Omega} \subset \Omega) = 1$ et

$$\forall \omega \in \tilde{\Omega} \subset \Omega \quad f_n(\omega) \to p.$$

Démonstration. Il suffit de poser $X_k = \mathbb{1}_{A_k}$ et d'appliquer le théorème 92. X_k admet bien un moment d'ordre 1 car $0 \leq X_k \leq 1$ et l'on a $\mathbb{E}X_1 = P(A_1) = p$. □

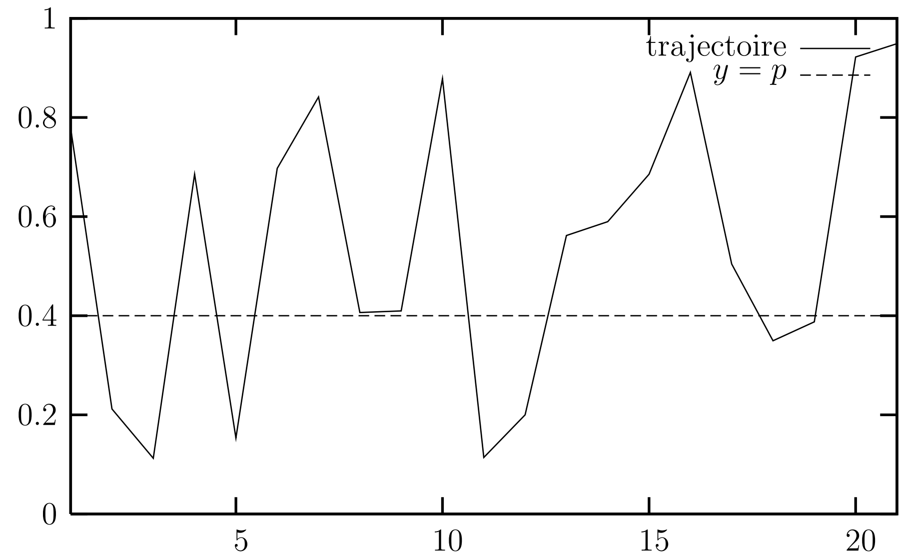
9.6 Exercices sur la convergence presque sûre

1. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi $\mathcal{N}(m, \sigma^2)$. Montrer que la suite $\frac{X_1^2+\dots+X_n^2}{n}$ convergence presque sûrement et déterminer sa limite.
2. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires identiquement distribuées telle qu'il existe $\alpha > 0$ avec $\mathbb{E}\exp(\alpha|X_1|) < +\infty$.
   Montrer que
   $$\lim_{n\to+\infty} \frac{X_n}{(\ln n)^{\frac{3}{2}}} = 0.$$
3. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant une loi exponentielle de paramètre 1.

   Calculer $\overline{\lim_{n\to+\infty}} \frac{X_n}{\ln n}$.
4. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi $\mathcal{N}(0, 1)$.

   (a) Calculer $\overline{\lim_{n\to+\infty}} \frac{X_n}{\sqrt{2\ln n}}$.

   (b) On se donne maintenant une deuxième suite $(Y_n)_{n\geq 1}$ de variables aléatoires indépendantes suivant la loi $\mathcal{N}(0, 1)$, cette deuxième suite étant indépendante de la première. Comparer
   $$\overline{\lim_{n\to+\infty}} \frac{X_n}{\sqrt{2\ln n}} + \overline{\lim_{n\to+\infty}} \frac{Y_n}{\sqrt{2\ln n}} \text{ et } \overline{\lim_{n\to+\infty}} \frac{(X_n+Y_n)}{\sqrt{2\ln n}}.$$
5. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes telles que pour tout n, X_n suive une loi binomiale $\mathcal{B}(n, \frac{1}{n^{1,01789}})$. Déterminer l'ensemble des valeurs d'adhérences de la suite (X_n)
6. Soit $p \in [0, 1]$ et $(U_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi uniforme sur $[0, 1]$. On note T_n le nombre de fois où le graphe associé à (U_n) coupe la droite d'équation $y = p$ avant le temps n. Dans notre exemple, $p = 0.4$ et $T_{20} = 8$.



Montrer que $\frac{T_n}{n}$ converge presque sûrement et déterminer la limite.

7. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires de Bernoulli indépendantes de paramètre $1/2$. On pose

$$M_n = \begin{pmatrix} 2+X_n & 1 \\ 1 & 2+X_n \end{pmatrix}$$

et

$$A_n = M_n \times M_{n-1} \times \dots M_2 \times M_1.$$

   (a) Montrer que la suite $(\det A_n)^{1/n}$ converge presque sûrement et déterminer la limite.

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

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

   Montrer que

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

8. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires telles que pour tout n, X_n suive une loi de Poisson de paramètre λ_n, où $(\lambda_n)_{n\geq 1}$ est une suite tendant vers 0 en l'infini. Montrer que la suite $(Y_n)_{n\geq 1}$ définie par $Y_n = X_1 X_2 \dots X_n$ est nulle à partir d'un certain rang.

9. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires telles que pour tout n, X_n suive une loi de Poisson de paramètre λ_n, avec

$$\sum_{n=1}^{+\infty} \lambda_n^2 < +\infty$$

Montrer que la suite $(X_n)_{n\geq 1}$ est presque sûrement bornée.

10. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes telles que pour tout n, X_n suive une loi de Poisson de paramètre λ_n, avec

$$\lambda_n = o(\ln n)$$

Montrer que la suite $(Y_n)_{n\geq 1}$ définie par $Y_n = X_1 X_2 \dots X_n$ est nulle à partir d'un certain rang.

11. Soit $(X_n)_{n\geq 2}$ une suite de variables aléatoires indépendantes telle que pour tout n, X_n suive une loi de Poisson de paramètre $2\ln n$. Montrer que

$$X_2 X_3 \dots X_n \xrightarrow[n\to+\infty]{\text{p.s.}} \begin{cases} 0 & \text{avec probabilité } p \\ +\infty & \text{avec probabilité } 1-p \end{cases},$$

où $p \in]0,1[$.

12. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes identiquement distribuées avec $\mathbb{E}|X_1| = +\infty$.

   (a) Soit $a > 0$. Montrer que $\overline{\lim\limits_{n\to+\infty}} \frac{|X_n|}{n} \geq a$ p.s.

   (b) On pose $S_n = \sum_{k=1}^n X_k$. Montrer que $\sup\limits_{n\geq 1} \frac{|S_n|}{n} = +\infty$ p.s.

Chapitre 10

Vecteurs gaussiens

Définition: On dit qu'un vecteur aléatoire $X \in \mathbb{R}^d$ est gaussien si pour tout $a \in \mathbb{R}^d$ la variable aléatoire $\langle X, a\rangle$ est gaussienne .

10.1 Image affine d'un vecteur gaussien

Théorème 94. *L'image d'un vecteur gaussien X d'espérance m_X et de matrice de covariance C_X par une application affine $x \mapsto Ax + b$ est un vecteur gaussien d'espérance $m_Y = Am_X + b$ et de matrice de covariance $C_Y = AC_XA^*$.*

Démonstration. On pose $Y = AX + b$. Soit $a \in \mathbb{R}^d$. $\langle Y, a\rangle = \langle AX, a\rangle + \langle b, a\rangle = \langle X, A^*a\rangle + \langle b, a\rangle$. Comme X est un vecteur gaussien $\langle X, A^*a$ est une variable aléatoire gaussienne. Quand on ajoute une constante à variable aléatoire gaussienne, on obtient une variable aléatoire gaussienne. Ainsi, pour tout A, $\langle Y, a\rangle$ est une variable aléatoire gaussienne, donc Y est un vecteur gaussien. L'expression de l'espérance et de la covariance est une conséquence du théorème 50. □

Corollaire 17. *Si $X = (X_1, \dots, X_d)$ est gaussien, alors pour tout $I \in \{1, \dots, d\}$, le vecteur $(X_i)_{i\in I}$ est gaussien*

Démonstration. $(X_i)_{i\in I}$ est l'image de $X = (X_1, \dots, X_d)$ par l'application linéaire

$$\begin{array}{ccc} \mathbb{R}^{\{1,\dots,d\}} & \to & \mathbb{R}^I \\ (x_i)_{i\in\{1,\dots,d\}} & \mapsto & (x_i)_{i\in D} \end{array}$$

□

10.2 Exemple fondamental

Théorème 95. *Soit $X_1, \dots, X_d$ d variables aléatoires gaussiennes indépendantes. Alors $X = (X_1, \dots, X_d)$ est un vecteur gaussien.*

Démonstration. Pour $k \in \{1, \dots d\}$, on pose $S_k = \sum_{i=1}^k a_X X_k$. On montre par récurrence sur k que S_k est une variable aléatoire gaussienne. Pour $k = 1$, $S_1 = a_1 X_1$: quand on multiplie une variable gaussienne par une constante, on a une variable aléatoire gaussienne. Supposons acquis que S_k est gaussienne : on a $S_{k+1} = S_k + a_{k+1} X_{k+1}$. $a_{k+1} X_{k+1}$ est une variable gaussienne indépendante de S_k, car S_k est $\sigma(X_1, \dots X_k)$-mesurable. Or d'après le théorème 53, la somme de deux variables aléatoires gaussiennes indépendantes est une variable aléatoire gaussienne, donc S_{k+1} est une variable aléatoire gaussienne.

Comme $\langle X, a \rangle = S_d$ quel que soit a , on en déduit que X est un vecteur gaussien. □

10.3 Lois gaussiennes

Théorème 96. *Soit C une matrice symétrique positive $d \times d$ et $m \in \mathbb{R}^d$. Alors, on peut construire un vecteur gaussien admettant m comme espérance et C comme matrice de covariance.*

Démonstration. Comme C est symétrique,on peut la diagonaliser avec une matrice de passage orthogonale. Comme elle est positive, les valeurs propres sont positives. Ainsi, on peut trouver une matrice O orthogonale et des réels positifs $\lambda_1, \dots, \lambda_d$ tels que

$$C = O \begin{pmatrix} \lambda_1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \lambda_2 & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \lambda_{d-1} & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \lambda_d \end{pmatrix} O^*$$

Posons

$$A = O \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \sqrt{\lambda_{d-1}} & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \sqrt{\lambda_d} \end{pmatrix} O^*$$

et soient $(X_1,\dots,X_d)$ d variables aléatoires indépendantes suivant la loi $\mathcal{N}(0,1)$: l'espérance de X est $m_X = 0$ sa matrice de covariance $C_X = \mathrm{Id}_{\mathbb{R}^d}$. D'après le théorème 95, X est gaussien donc, d'après le théorème 50, $Y = AX + m$ est un vecteur gaussien d'espérance $A.0 + m = m$ et de covariance $AC_XA^* = AA^* = C$. □

Théorème 97. *Soit X et Y deux vecteurs gaussiens ayant même espérance et même matrice de covariance. Alors X et Y ont même loi.*

Démonstration. Soit $a \in \mathbb{R}^d$. On pose $V = \langle X, A\rangle$ et $W = \langle Y, A\rangle$. L'espérance de V est $\langle m_X, a\rangle$ est la covariance de V est $\mathrm{Covar}(\langle X, a\rangle, \langle X, a\rangle) = \langle C_X a, a\rangle$. Comme X est gaussien, V est gaussienne, donc $V \sim \mathcal{N}(\langle m_X, a\rangle, \langle C_X a, a\rangle)$. De même $V \sim \mathcal{N}(\langle m_Y, a\rangle, \langle C_Y a, a\rangle)$. Comme $m_X = m_Y$ et $C_X = C_Y$, on en déduit que $V = \langle X, A\rangle$ et $W = \langle Y, A\rangle$ ont même loi. Comme cela est vrai quel que soit a, on déduit du théorème 71que X et Y ont même loi. □

Définition: Soit $m \in \mathbb{R}^d$ et C une matrice $d\times d$ symétrique positive. On note $\mathcal{N}(m, C)$ la loi commune à tous les vecteurs gaussiens admettant m comme espérance et C comme matrice de covariance.

La pertinence de cette définition est assurée par les théorèmes 96 et 97

10.4 Lois gaussiennes et indépendance

Théorème 98. *Soit $d_1,\dots,d_n$ n entiers positifs de somme d. Soit $C_1,\dots C_n$ n matrices symétriques positives, et $m_1,\dots m_n$ n vecteurs C_i étant de taille d_i. Alors*

$$\mathcal{N}(m_1, C_1) \otimes \mathcal{N}(m_2, C_2) \cdots \otimes \mathcal{N}(m_n, C_n) = \mathcal{N}(m, C),$$

avec

$$m = \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix} \text{ et } C = \begin{pmatrix} C_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & C_n \end{pmatrix}$$

Démonstration. Soit $Y_1, \dots, Y_n$ n vecteurs gaussiens indépendants, avec $Y_1 \sim \mathcal{N}(m_i, C_i)$. On va d'abord montrer que $X = (Y_1, \dots, Y_n)$ est gaussien. Soit $a \in \mathbb{R}^n$. On peut écrire $a = (a_1, \dots, a_n)$, avec a_i de taille d_i. On a

$$\langle X, a\rangle = \sum_{k=1}^{n} \langle Y_k, a_k\rangle.$$

Comme Y_k est gaussien, chaque variable aléatoire $\langle Y_k, a_k\rangle$ est une variable aléatoire gaussienne. Comme les Y_k sont indépendants, les variables aléatoires $\langle Y_k, a_k\langle$ sont indépendantes. Or on sait qu'une somme de variables aléatoires gaussiennes indépendantes est une variable aléatoire gaussienne, donc $\langle X, a\rangle$ est une variable aléatoire gaussienne. Il s'ensuit que X est un vecteur gaussien. L'expression de l'espérance et de la matrice de covariance ne pose pas de problème, puisque des variables aléatoires indépendantes ne sont pas corrélées. □

Théorème 99. *Soit $d_1, \dots, d_n$ n entiers positifs de somme d. On suppose que $X = (X_1, \dots, X_n)$ est un vecteur gaussien dont la matrice de covariance est diagonale par blocs :*

$$C_X = \begin{pmatrix} C_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & C_n \end{pmatrix}$$

Alors, si l'on pose $Y_1 = (X_1, \dots, X_{d_1})$, $Y_2 = (X_{d_1+1}, \dots, X_{d_1+d_2})$, $Y_n = (X_{d_1+d_2+\cdots+d_{n-1}+1}, \dots, X_{d_1+d_2+\cdots+d_n})$, les vecteurs $Y_1, \dots Y_n$ sont des vecteurs gaussiens indépendants.

Démonstration. On voit que X a même espérance et même matrice de covariance que le vecteur aléatoire considéré au théorème précédent. Comme X et le vecteur aléatoire considéré au théorème précédent sont tous deux gaussiens, ils ont tous deux la même loi, donc la loi de X est $\mathcal{N}(m_1, C_1) \otimes \mathcal{N}(m_2, C_2) \cdots \otimes \mathcal{N}(m_n, C_n)$, ce qui signifie que les Y_i sont indépendants et que pour tout i, on a $Y_i \sim \mathcal{N}(m_i, C_i)$. □

En particulier, on a le corollaire suivant :

Corollaire 18. *Si le vecteur gaussien $X = (X_1, \dots, X_d)$ a une matrice de covariance dont tous les termes non-diagonaux sont nuls, alors $X_1, \dots, X_d$ sont des variables aléatoires indépendantes.*

10.5 Lois gaussiennes à densité

Théorème 100. *Soit C une matrice symétrique définie positive et $m \in \mathbb{R}^d$. La loi sur $\mathbb{R}^d$ $\mathcal{N}(m, C)$ admet comme densité par rapport à la mesure de Lebesgue la fonction*

$$f_{m,C}(y) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det C}} \exp(-\frac{1}{2}\langle C^{-1}(y-m), y-m\rangle).$$

Démonstration. On reprend les notations de la preuve du théorème 96. On a $X \sim \mathcal{N}(0, \mathrm{Id})$ et $Y \sim \mathcal{N}(m, C)$. Par rapport au cas général, on gagne le fait que C est définie positive, ce qui implique que les λ_i sont strictement positifs, et donc que A est inversible. Comme X est composé de n variables aléatoires indépendantes à densité, la densité de X est le produit des densités, soit

$$f_X(x) = \prod_{k=1}^{d} \frac{1}{\sqrt{2\pi}} \exp(-\frac{x_k^2}{2}) = \frac{1}{(2\pi)^{d/2}} \exp(-\frac{1}{2}\langle x, x\rangle).$$

D'après le théorème 52, $Y = AX + m$ admet comme densité

$$\frac{1}{\det A} f_X(A^{-1}(y-m)) = \frac{1}{\det A} \frac{1}{(2\pi)^{d/2}} \exp(-\frac{1}{2}\langle A^{-1}(y-m), A^{-1}(y-m)\rangle).$$

Or

$$\begin{aligned}
\langle A^{-1}(y-m), A^{-1}(y-m)\rangle &= \langle (A^{-1})^* A^{-1}(y-m), y-m\rangle \\
&= \langle (A^{-1})^* A^{-1}(y-m), y-m\rangle \\
&= \langle (AA^*)^{-1}(y-m), y-m\rangle \\
&= \langle C^{-1}(y-m), y-m\rangle
\end{aligned}$$

D'autre part, $\det C = \det AA^* = (\det A)^2$, donc $\det A = \sqrt{\det C}$. On en déduit que la densité de Y (c'est à dire de la loi de Y) est

$$f_{m,C}(y) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det C}} \exp(-\frac{1}{2}\langle C^{-1}(y-m), y-m\rangle).$$

□

10.6 Fonction caractéristique des vecteurs gaussiens

Théorème 101. *En dimension quelconque, la fonction caractéristique de la loi normale $\mathcal{N}(m, C)$ est*

$$x \mapsto \exp(i\langle x, m\rangle \exp(-\frac{1}{2}\langle Cx, x\rangle)$$

Démonstration.

$$\mathbb{E}\exp(i\langle X,t\rangle) = \mathbb{E}\exp(iY) = \varphi_Y(1),$$

où $Y = \langle X,t\rangle$. comme X est gaussien de covariance C et d'espérance m, Y est gaussien de covariance $\langle Cx,x\rangle$ et d'espérance $\langle x,m\rangle$. On déduit donc le résultat de la formule précédente. □

10.7 Exercices sur les vecteurs gaussiens

1. Soient U et V deux variables aléatoires gaussiennes centrées et f une fonction croissante. Montrer que

$$\mathbb{E}\|U\|^2 < \mathbb{E}\|V\|^2 \Longrightarrow \mathbb{E}f(\|U\|) \leq \mathbb{E}f(\|V\|)$$

2. Soit X, Y deux variables aléatoires indépendantes suivant $\mathcal{N}(0,1)$. On pose

$$U = \begin{pmatrix} 2X \\ 2X \end{pmatrix} \text{ et } V = \begin{pmatrix} 2X \\ \sqrt{5}Y \end{pmatrix}.$$

Montrer que U et V sont gaussiens centrés, puis que l'on a

$$\mathbb{E}\|U\|_2^2 < \mathbb{E}\|V\|_2^2,$$

tandis que

$$\mathbb{E}\|U\|_2^4 > \mathbb{E}\|V\|_2^4.$$

Comparer avec le résultat de l'exercice précédent.

3. Soit X un vecteur gaussien de matrice de covariance C. Montrer que

$$\mathbb{E}\|X\|_2^2 = \mathrm{Tr}C.$$

4. (a) Montrer qu'il existe un vecteur gaussien dont la matrice de covariance est

$$C = \begin{pmatrix} 5 & 0 & 3 \\ 0 & 6 & 0 \\ 3 & 0 & 2 \end{pmatrix}$$

   (b) Déterminer

$$\sup_{a\in\mathbb{R}^3} \frac{\mathbb{E}|\langle X,a\rangle|}{\|a\|_2}.$$

5. Montrer qu'on peut trouver des variables aléatoires X_1, X_2, X_3 telles que pour tout $i \in \{1,2,3\}$ $X_i \sim \mathcal{N}(0,3)$ et que $\mathrm{Covar}(X_1, X_2) = \mathrm{Covar}(X_1, X_3) = \mathrm{Covar}(X_3, X_2) = 1$. Calculer la variance de $X_1 + 2X_2 + 3X_3$.
6. (a) Soit X et Y deux variables aléatoires indépendantes suivant la loi $\mathcal{N}(0,1)$. Calculer la matrice de covariance du couple $(X, \alpha X + \beta Y)$
   (b) Soit (Z,T) un vecteur gaussien centré, avec $\mathrm{Var}\, Z = 1$. Montrer qu'il existe α et β tels que la matrice de covariance de (Z,T) soit
   $$\begin{pmatrix} 1 & \alpha \\ \alpha & \alpha^2 + \beta^2 \end{pmatrix}.$$
   En déduire que $\mathbb{E}Z^2T^2 = 3\alpha^2 + \beta^2 = \mathrm{Var}\, Z + 2\,\mathrm{Covar}(Z,T)^2$,
   (c) Soit (Z,T) est un vecteur gaussien centré quelconque – on ne suppose plus que $\mathrm{Var}\, Z = 1$ –, montrer
   $$\mathbb{E}Z^2T^2 = \mathrm{Var}\, Z\, \mathrm{Var}\, T + 2\,\mathrm{Covar}(Z,T)^2.$$
   (d) Soit (Z,T) un vecteur gaussien de matrice de covariance
   $$\begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}.$$
   Calculer $\mathbb{E}Z^2T^2$.
   (e) Soit X un vecteur gaussien de matrice de covariance C. Montrer que
   $$\mathbb{E}\|X\|^4 = 2\mathrm{Tr} C^* C + (\mathrm{Tr} C)^2.$$
7. Soit X un vecteur gaussien dans $\mathbb{R}^n$ de loi $\mathcal{N}(0,C)$ et α est un réel vérifiant $\alpha < \rho(C)^{-1}$, où $\rho(C)$ est le rayon spectral de C. Montrer que
   $$\mathbb{E}\exp(\frac{\alpha}{2}\|X\|_2^2) = \prod_i (1 - \alpha\lambda_i)^{-1/2}.$$
   où les λ_i sont les valeurs propres de C.
8. (a) Soit X un vecteur gaussien dans $\mathbb{R}^n$ de loi $\mathcal{N}(0, I_n)$ et O une matrice orthogonale. Montrer que $Y = OX$ a la même loi que X.
   (b) On appelle loi du chi-deux à n degrés de liberté et on note $\chi^2(n)$ la loi de $\|X\|_2^2$. Montrer que $\chi^2(1)$ admet la densité
   $$x \mapsto \frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-\frac{x}{2}} \mathbb{1}_{\mathbb{R}_+}(x)$$
   En déduire que $\chi^2(n)$ est une loi à densité.
9. Soit A la matrice d'un projecteur orthogonal de rang r et X une variable aléatoire suivant la loi $\mathcal{N}(0,A)$. Montrer que $\|X\|_2^2$ suit la loi $\chi^2(r)$.

Chapitre 11

Convergence en loi

11.1 Convergence en loi

11.1.1 Définition

On dit qu'une suite $(\mu_n)_{n\geq 1}$ de mesures de probabilités 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

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

Par extension, on dit qu'une suite de variables aléatoires X_n converge en loi vers la variable aléatoire X (ou vers la loi μ) si la suite de mesures P_{X_n} converge faiblement vers 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 $\mathbb{E}f(X)$.

Rappel : si μ et ν sont deux mesures 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 102. *Soit g une fonction continue définie sur $\mathbb{R}^d$. Si $(X_n)_{n\geq 1}$ 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. $\mathbb{E}f(Y_n) = \mathbb{E}f(g(X_n)) = \mathbb{E}(f\circ g)(X_n)$. Comme f et g sont continues, $f\circ g$ est continue. Comme f est bornée, $f\circ g$ bornée. $f\circ g$ 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 19. *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*

- $X_n + Y_n$ *converge en loi vers* $X + Y$*.*
- $\langle, X_n Y_n\rangle$ *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 $(x, y) \mapsto xy$. □

11.1.2 Premiers exemples

Un critère de convergence en loi

Lemme 9 (Théorème de Scheffé). *Soit* $(\Omega, \mathcal{A}, \mu)$ *un espace mesuré ;* $f, (f_n)_{n\geq 1} : \Omega \to \mathbb{R}_+$ *des applications intégrables par rapport à* μ *telles que*

- $f_n \xrightarrow[n\to+\infty]{\mu\mathbf{p.p.}} f$*.*
- $\lim_{n\to\infty} \int_\Omega f_n dx = \int_\Omega f dx$*.*

Alors $f_n \xrightarrow[n\to+\infty]{L^1(\mu)} f$*.*

Démonstration. On a

$$\begin{cases} f + f_n = \max(f, f_n) + \min(f, f_n) \\ |f - f_n| = \max(f, f_n) - \min(f, f_n) \end{cases}$$

Alors

$$f + f_n - |f - f_n| = 2\min(f, f_n),$$

d'où

$$|f - f_n| = f + f_n - 2\min(f, f_n) = -f + f_n + 2(f - \min(f, f_n)).$$

Ainsi

$$\|f_n - f\|_{L^1(\mu)} = -\int_\Omega f(x)d\mu(x) + \int_\Omega f_n(x)d\mu(x) + 2\int_\Omega (f - \min(f, f_n))(x)d\mu(x)$$

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 $0 \leq f - \min(f, f_n) \leq f$ et $f - \min(f, f_n) \xrightarrow[n\to+\infty]{\mu\ \mathbf{p.s.}} 0$. D'après le théorème de convergence dominée, on a donc $\int_\Omega f - \min(f, f_n)\ d\mu \xrightarrow[n\to+\infty]{} 0$, d'où le résultat.

□

Corollaire 20. *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Soient $\nu, (\nu_n)_{n\geq 1}$ des mesures de probabilités sur $(\Omega, \mathcal{A}, \mu)$ admettant les densités $f, (f_n)_{n\geq 1}$ par rapport à μ. On suppose que $f_n \xrightarrow[n\to+\infty]{\mu\ p.p.} f$. Alors ν_n converge faiblement vers ν.*

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

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

qui tend vers 0 d'après le théorème de Scheffé. Comme cette convergence a lieu pour toute fonction continue bornée g, on peut dire que ν_n converge en loi vers ν. □

Corollaire 21. *Soit $X, (X_n)_{n\geq 1}$ des variables aléatoires discrètes à valeurs dans un ensemble dénombrable D. On suppose que*

$$\forall k \in D \qquad \lim_{n\to+\infty} P(X_n = k) = P(X = k)$$

Alors $(X_n)_{n\geq 1}$ converge en loi vers X.

Démonstration. Il suffit de remarquer que $P_X, P_{X_1}, \dots, P_{X_n} \dots$ 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 103. *Soit, pour $n \geq 1$ une variable aléatoire X_n suivant une loi binomiale de paramètres n et p_n. On suppose que l'on a*

$$\lim_{n\to\infty} np_n = \lambda > 0$$

Alors, la suite X_n converge en loi vers la loi de Poisson $\mathcal{P}(\lambda)$.

Démonstration. D'après le corollaire 21, il suffit de montrer que pour tout entier $k \geq 0$, $P(X_n = k)$ converge vers $e^{-\lambda}\frac{\lambda^k}{k!}$. On a

$$P(X_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} = \binom{n}{k} p_n^k (1-p_n)^n (1-p_n)^{-k}$$

On a $\binom{n}{k} = \frac{n(n-1)\dots(n-k+1)}{k!} \sim \frac{n^k}{k!}$, $p_n^k \sim (\lambda/n)^k = \lambda^k n^{-k}$. D'autre part, $\ln(1-p_n)^n = n\ln(1-p_n) = n(-p_n + o(p_n)) \sim -np_n \sim -\lambda$. Ainsi $\ln(1-p_n)^n$ converge vers $-\lambda$ donc $(1-p_n)^n$ converge vers $e^{-\lambda}$. En mettant bout à bout 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 etre interprété de la manière suivante : la loi de Poisson est une bonne modélisation pour le nombre de fois où un événement rare survient.

Convergence de la loi hypergéométrique vers la loi binomiale

Théorème 104. *Soit, pour $j \geq 1$ une variable aléatoire X_j suivant une loi hypergéométrique de paramètres $\mathcal{H}(N_j, n_j, k)$. 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, la suite $(X_j)_{j\geq 1}$ converge en loi vers la loi binomiale $\mathcal{B}(k,p)$.

Démonstration. D'après le corollaire 21, il suffit de montrer que pour tout entier $i \geq 0$, $P(X_j = i)$ converge vers $\binom{k}{i}p^i(1-p)^{k-i}$. On a

$$P(X_j = i) = \frac{\binom{n_j}{i}\binom{N_j-i}{k-i}}{\binom{N_j}{k}} \sim \frac{\frac{n_j^i}{i!}\frac{(N_j-n_j)^{k-i}}{(k-i)!}}{\frac{N_j^k}{k!}}.$$

Or

$$\frac{\binom{n_j}{i}\binom{N_j-i}{k-i}}{\binom{N_j}{k}} \sim \frac{\frac{n_j^i}{i!}\frac{(N_j-n_j)^{k-i}}{(k-i)!}}{\frac{N_j^k}{k!}} = \frac{k!}{i!(k-i)!}(\frac{n_j}{N_j})^i(1-\frac{n_j}{N_j})^{k-i},$$

qui converge vers $\binom{k}{i}p^i(1-p)^{k-i}$ lorsque j tend vers l'infini, ce qui achève la preuve.

□

11.1.3 Théorème du porte-manteau

Théorème 105. *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* $\lim_{n\to+\infty} \mu_n(A) = \mu(A)$.

Démonstration. On va prouver successivement (3) $\iff$ (4), (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (4) $\Longrightarrow$ (5) $\Longrightarrow$ (1).

– Pour (3) $\iff$ (4), il suffit de remarquer que
$$\begin{aligned}
\sup_{O\,\textbf{ouvert}} \big(\mu(O) - \underline{\lim_{n\to+\infty}}\ \mu_n(O)\big) &= \sup_{F\,\textbf{fermé}} \big(\mu(F^c) - \underline{\lim_{n\to+\infty}}\ \mu_n(F^c)\big) \\
&= \sup_{F\,\textbf{fermé}} \big((1-\mu(F)) - \underline{\lim_{n\to+\infty}}\ (1-\mu_n(F))\big) \\
&= \sup_{F\,\textbf{fermé}} \big(-\mu(F) - \underline{\lim_{n\to+\infty}}\ \mu_n(F)\big) \\
&= \sup_{F\,\textbf{fermé}} \big(-\mu(F) + \overline{\lim_{n\to+\infty}}\ \mu_n(F)\big)
\end{aligned}$$
Ainsi, si l'un des sup est négatif, l'autre aussi.

– Preuve de (2) $\Longrightarrow$ (3).
Soit F un fermé de $\mathbb{R}^d$. On pose $d_F(x) = d(x,F) = inf(|y-x|; y \in F)$ et, pour $\varepsilon > 0$, G_ε est la fonction continue par morceaux définie par $G_\varepsilon(x) = (1 - \frac{|x|}{\varepsilon})1\!\!1_{[-\varepsilon,+\varepsilon]}(x)$. On a $G_\varepsilon \circ d_F \geq 1\!\!1_F$, donc
$$\overline{\lim_{n\to+\infty}} \int G_\varepsilon \circ d_F\ d\mu_n \geq \overline{\lim_{n\to+\infty}} \int 1\!\!1_F\ d\mu_n$$
Comme $G_\varepsilon \circ d_F$ est uniformément continue (c'est la composée d'une application 1-lipschitzienne par 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

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

Or $\int G_\varepsilon \circ d_F \, d\mu = \int G_\varepsilon \, d\mu_{d_F}$ Lorsque ε tend vers 0, G_ε converge vers l'indicatrice de 0, donc par convergence dominée, $\int G_\varepsilon \circ d_F \, d\mu = \int G_\varepsilon \, d\mu_{d_F}$ converge vers

$$\int \mathbb{1}_{\{0\}} \, d\mu_{d_F} = \mu_{d_F}(\{0\}) = \mu(d_F = 0) = \mu(F).$$

– Preuve de $(3,4) \Longrightarrow 5$ On a $\overset{\circ}{A} \subset A \subset \overline{A}$ D'où

$$\varlimsup_{n\to+\infty} \mu_n(A) \leq \varlimsup_{n\to+\infty} \mu_n(\overline{A}) \leq \mu(\overline{A}).$$

D'autre part

$$\varliminf_{n\to+\infty} \mu_n(A) \geq \varliminf_{n\to+\infty} \mu_n(\overset{\circ}{A}) \geq \mu(\overset{\circ}{A}).$$

Ainsi, on a

$$\mu(\overset{\circ}{A}) \leq \varliminf_{n\to+\infty} \mu_n(A) \leq \varlimsup_{n\to+\infty} \mu_n(A) \leq \mu(\overline{A}).$$

Comme $\mu(\overline{A}) - \mu(\overset{\circ}{A}) = \mu(\partial A) = 0$, la suite $(\mu_n(A))_{n\geq 1}$ a une limite supérieure qui coïncide avec sa limite inférieure, donc converge vers $\mu(\overline{A}) = \mu(\overset{\circ}{A})$, c'est à dire vers $\mu(A)$, car $\mu(\overset{\circ}{A}) \leq \mu(A) \leq \mu(\overline{A})$.

– Preuve de $(5) \Longrightarrow (1)$ L'idée est d'approcher la fonction f par une somme d'indicatrice d'ensembles 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. 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 base. À k fixé, les ensembles $(\mathcal{P}_t^k)_{t\in[0,1[}$ sont disjoints. Comme μ est une probabilité, l'ensemble des $t \in [0,1[$ tels que $\mu(\mathcal{P}_t^k) > 0$ est au plus dénombrable. Comme T n'est pas dénombrable, il existe t_k tel que $\mu(\mathcal{P}_{t_k}^k) = 0$. Pour A entier, on pose $B_p = \{x \in \mathbb{R}^d; \|x - t\|_\infty \leq A.$

Soit $\varepsilon > 0$. On peut trouver A tel que $\mu(B_A) \geq 1 - \frac{\varepsilon}{2}$. D'après ce qui précède, $\mu(\partial B_A) = 0$. On peut donc trouver n_0 tel que

$$n \geq n_0 \Longrightarrow \mu_n(B_A) \geq 1 - \varepsilon.$$

B_A est compact. On note ω_f le module de continuité de la restriction de f à B_A :
$\omega_f(\eta) = \sup(|f(x) - f(y); \|x - y\|_\infty \leq \eta)$. Pour $p \geq 1$, on pose

$$f_p(x) = 1_{B_A}(x)\|f(t_1 + \frac{\mathbf{Ent}(p(x-t_1))}{p}, \dots, t_d + \frac{\mathbf{Ent}(p(x-t_d))}{p})$$

On a

$$|f_p(x) - f(x)| \leq \omega_f(1/p) + 2\|f\|_\infty \mathbb{1}_{B_A^c}(x)$$

Ainsi

$$\forall p \in \mathbb{N} \quad \int_{\mathbb{R}^d} |f_p(x) - f(x)| \leq \omega_f(1/p) + 2\|f\|_\infty \frac{\varepsilon}{2}$$

$$\forall p \in \mathbb{N} \quad \overline{\lim_{n\to\infty}} \int_{\mathbb{R}^d} |f_p(x) - f(x)| \leq \omega_f(1/p) + 2\|f\|_\infty \varepsilon$$

Soit p_0 un entier tel que $\omega_f(1/p_0) \leq \varepsilon$: on a

$$\int_{\mathbb{R}^d} |f_{p_0}(x) - f(x)| \ d\mu \leq \varepsilon + 2\|f\|_\infty \frac{\varepsilon}{2}$$

et

$$\overline{\lim_{n\to\infty}} \int_{\mathbb{R}^d} |f_{p_0}(x) - f(x)| \ d\mu_n \leq \varepsilon + 2\|f\|_\infty \varepsilon.$$

On a

$$f_{p_0} = \sum_{x \in (t + \frac{1}{p_0}\mathbb{Z}^d) \cap B_A} f(x) \mathbb{1}_{x + [0, \frac{1}{p_0}[^d}$$

Ainsi, f_p s'écrit comme une combinaison linéaire (finie !) d'indicatrices de parties de $\mathbb{R}^d$ dont la frontière est de μ-mesure nulle. On a donc

$$\lim_{n\to+\infty} \int f_p \ d\mu_n = \int f_p \ d\mu.$$

Comme

$$\int f \ d\mu - \int f \ d\mu_n = (\int f \ d\mu - \int f_p \ d\mu) + (\int f_p \ d\mu - \int f_p \ d\mu_n) + (\int f_p \ d\mu_n - \int f \ d\mu_n),$$

on a

$$\begin{aligned}\overline{\lim_{n\to\infty}}\ |\textstyle\int_{\mathbb{R}^d} f(x)\ d\mu(x) - \int_{\mathbb{R}^d} f(x)\ d\mu_n(x)| &\leq \textstyle\int_{\mathbb{R}^d} |f(x) - f_{p_0}(x)|\ d\mu(x) \\ &\quad + \overline{\lim_{n\to\infty}}\ |\textstyle\int_{\mathbb{R}^d} f_{p_0}(x)\ d\mu(x) - \int_{\mathbb{R}^d} f_{p_0}(x)\ d\mu_n(x)| \\ &\quad + \overline{\lim_{n\to\infty}}\ \textstyle\int_{\mathbb{R}^d} |f_{p_0}(x) - f(x)|\ d\mu_n(x) \\ &\leq (\varepsilon + \|f\|_\infty\varepsilon) + 0 + \varepsilon + (2\|f\|_\infty\varepsilon) \\ &\leq \varepsilon(2 + 3\|f\|_\infty).\end{aligned}$$

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 : (1) est donc vérifié.

□

Corollaire 22. *Soit X_n une suite de variables aléatoires réelles convergeant en loi vers une variable aléatoire X. Alors pour tout point x où F_X est continue, $F_{X_n}(x)$ tend vers $F(x)$ lorsque n tend vers l'infini.*

Cette dernière conséquence de la convergence en loi est très utile, par exemple en statistique.

Démonstration. Posons $A =]-\infty, x]$. A est un borélien dont la frontière $\{x\}$ est telle que $P_X(\{x\}) = 0$, car a est un point de continuité de F_X. Comme $P_{X_n}(F) = F_{X_n}(x)$ et $P_X(F) = F_X(x)$, il suffit donc d'appliquer le théorème du porte-manteau pour conclure. □

Le théorème suivant est également très utile :

Théorème 106. *Si une suite $(\mu_n)_{n\geq 1}$ de mesures de probabilités sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est telle 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 valant 1 sur $]-\infty, A/2]$, 0 sur $[A, +\infty[$, et affine sur $[A/2, A]$ 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

$$\int_{\mathbb{R}^d} G_A(\|x\|)\, d\mu \geq 1-\varepsilon.$$

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

$$n \geq n_0 \Longrightarrow \int_{\mathbb{R}^d} G_A(\|x\|)\, d\mu_n \geq 1-2\varepsilon,$$

d'où

$$n \geq n_0 \Longrightarrow \mu_n(B(0,A)) \geq 1-2\varepsilon.$$

Soit maintenant f une fonction continue positive. On a

$$\int f\ d\mu_n - \int f\ d\mu_n = (\int fH_{2A}\, d\mu_n - \int fH_{2A}\, d\mu_n) + (\int f(1-H_{2A})\, d\mu_n - \int f(1-H_{2A})\, d\mu_n)$$

Ainsi

$$|\int f\ d\mu_n - \int f\ d\mu_n| \leq (\int fH_{2A}\ d\mu_n - \int fH_{2A}\ d\mu_n) + \varepsilon\|f\|_\infty$$

Comme fH_{2A} est une fonction continue positive à support compact, $\int fH_{2A}\, d\mu_n$ converge vers $\int fH_{2A}\ d\mu$, d'où

$$\overline{\lim_{n\to+\infty}}\ |\int f\ d\mu_n - \int f\ d\mu_n| \leq \varepsilon\|f\|_\infty.$$

Comme ε est arbitraire, on en déduit que $\int f\ d\mu_n$ converge vers $\int f\ d\mu$.

Le passage à une fonction continue bornée de signe quelconque ne pose pas de problème, car $f = \max(f,0) - \max(-f,0)$ et le résultat s'ensuit par linéarité. □

11.1.4 Lien avec les autres modes de convergence

Théorème 107. *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires, 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 (c'est à dire vers une masse de Dirac δ_a, alors X_n converge en probabilité vers a.*

Démonstration. 1. Soit f une fonction continue bornée. On pose $x_n = \mathbb{E}f(X_n)$. La suite $(x_n)_{n\geq 1}$ est bornée. Soit a une valeur d'adhérence de $(x_n)_{n\geq 1}$, avec $\lim\limits_{k\to+\infty} x_{n_k} = a$. Comme X_{n_k} converge en probabilité vers X, on peut (d'après le théorème 91) en extraire une sous-suite $X_{n_{m_k}}$ telle que $\lim\limits_{k\to+\infty} X_{n_{m_k}} = X$ presque sûrement. Par continuité $\lim\limits_{k\to+\infty} f(X_{n_{m_k}}) = f(X)$ presque sûrement. D'après le théorème de convergence dominée, on a donc $\lim\limits_{k\to+\infty} \mathbb{E}f(X_{n_{m_k}}) = \mathbb{E}f(X)$, c'est à dire $\lim\limits_{k\to+\infty} x_{n_{m_k}} = \mathbb{E}f(X)$. Or $x_{n_{m_k}}$ est une sous-suite de (x_{n_k}) qui converge elle-même vers a, donc $a = \mathbb{E}f(X)$. $(x_n)_{n\geq 1}$ est une suite bornée dont $\mathbb{E}f(X)$ est l'unique valeur d'adhérence, donc elle converge vers $\mathbb{E}f(X)$. Comme pour toute fonction continue bornée f, $\mathbb{E}f(X_n)$ converge vers $\mathbb{E}f(X)$, on peut dire que $(X_n)_{n\geq 1}$ converge en loi vers X.

2. Soit $\varepsilon > 0$. On pose $F = \{x \in \mathbb{R}^d; \|a - x| \geq \varepsilon\}$. Ainsi $P(\|X_n - a\| \geq \varepsilon) = P_{X_n}(F)$. F est fermé et P_{X_n} converge faiblement vers δ_a, donc, d'après le théorème du porte-manteau,

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

Ainsi, pour tout $\varepsilon > 0$, $P(\|X_n - a\| \geq \varepsilon)$ tend vers 0 quand n tend vers l'infini : on peut dire que $(X_n)_{n\geq 1}$ converge en probabilité vers a.

□

11.2 Convergence en loi sur $\mathbb{R}^n$ grâce aux fonctions caractéristiques

11.2.1 Critère de convergence

Théorème 108 (Lévy). *Soit $(\mu_n)_{n\geq 1}$ une suite de mesures et μ une mesure 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*

$$\forall t \in \mathbb{R}^d \qquad \lim_{n\to+\infty} \varphi_{\mu_n}(t) = \varphi_\mu(t)$$

CE THÉROÈME EST ADMIS.

11.2.2 Théorème de continuité de Lévy

Théorème 109 (théorème de continuité de Lévy). *Soit $(\mu_n)_{n\geq 1}$ une suite de mesures, φ une fonction donnée. Si*

$$\forall t \in \mathbb{R}^d \quad \lim_{n\to+\infty} \varphi_{\mu_n}(t) = \varphi(t)$$

et que φ est continue en 0, alors, il existe une unique mesure μ telle que $\varphi = \varphi_\mu$ et la suite $(\mu_n)_{n\geq 1}$ converge faiblement vers μ.

Démonstration. Pour montrer l'existence d'une mesure μ telle que $\varphi = \varphi_\mu$, nous allons utiliser le théorème de Bochner. Comme φ_{μ_n} est une fonction caractéristique $\varphi_{\mu_n}(0) = 1$, d'où à la limite $\varphi(0) = 1$. Par hypothèse, φ est continue en 0. Il reste donc à montrer que φ est de type positif. Soient $e_1, \dots, e_p$ des éléments quelconques de $\mathbb{R}^d$; on va démontrer que la matrice A de taille $d\times d$ définie par $a_{k,l} = \varphi(e_k - e_l)$ est hermitienne positive. Pour $n \geq 1$, on note A_n la matrice de taille $d \times d$ définie par $a^n_{k,l} = \varphi_{\mu_n}(e_k - e_l)$. Soit $x \in \mathbb{C}^d$. On a $\langle A_n x, x\rangle \geq 0$, car A_n est hermitienne positive. Lorsque n tend vers l'infini A_n converge vers A car φ_{μ_n} converge vers φ. Ainsi $\langle Ax, x\rangle = \lim_{n\to\infty} \langle A_n x, x\rangle \geq 0$, donc A est hermitienne positive. Ainsi, d'après le théorème de Bochner, il existe une unique mesure μ telle que $\varphi = \varphi_\mu$ et le premier théorème de Lévy permet d'affirmer que μ_n converge faiblement vers μ. □

Ce dernier théorème peut être intéressant si la loi limite est une loi nouvelle, inconnue. L'appliquer lorsque la loi est une loi bien connue est assez maladroit.

11.2.3 Une application du théorème de Lévy

Le résultat qui suit peut être démontré sans l'aide du théorème de Lévy, mais ce dernier théorème en rend la preuve particulièrement simple.

Théorème 110. *Si μ_n tend faiblement vers μ et ν_n faiblement vers ν, alors la suite $(\mu_n \otimes \nu_n)_{n\geq 1}$ tend faiblement vers $\mu \otimes \nu$.*

Démonstration. Soit $(s,t) \in \mathbb{R}^d \times \mathbb{R}^d$. On a $\varphi_{\mu_n\otimes\nu_n}(s,t) = \varphi_{\mu_n}(s)\varphi_{\nu_n}(t)$. Comme μ_n converge faiblement vers μ, $\varphi_{\mu_n}(s)$ converge vers $\varphi_\mu(s)$. De même, $\varphi_{\nu_n}(t)$ converge vers $\varphi_\nu(t)$. Ainsi $\varphi_{\mu_n\otimes\nu_n}(s,t)$ converge vers $\varphi_\mu(s)\varphi_\nu(t) = \varphi_{\mu\otimes\nu}(s,t)$, donc d'après le théorème de Lévy, la suite $(\mu_n \otimes \nu_n)_{n\geq 1}$ tend faiblement vers $\mu \otimes \nu$. □

Théorème 111. *Si μ_n tend faiblement vers μ et ν_n 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 $\mu_n \otimes \nu_n$, (X, Y) de loi $\mu \otimes \nu$. D'après le théorème 110, (X_n, Y_n) converge en loi vers (X, Y), donc d'après le corollaire 19, $X_n + Y_n$ converge en loi vers $X + Y$. Mais la loi de $X_n + Y_n$ est $\mu_n * \nu_n$ et la loi de $X + Y$ est $\mu * \nu$, donc le résultat est démontré. □

11.3 Théorème de la limite centrale

11.3.1 Théorème de la limite centrale en dimension 1

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

Théorème 112. *Soit $X_1, \dots, X_n$ une suite de variables aléatoires réelles i.i.d. admettant un moment d'ordre 2. On note alors m l'espérance et σ^2 la variance communes à ces variables. Alors*

$$\frac{(X_1 + \dots X_n) - nm}{\sqrt{n}} \Longrightarrow \mathcal{N}(0, \sigma^2).$$

Démonstration. On pose $S_n = (X_1 + \dots X_n) - nm = \sum_{k=1}^n (X_k - m)$. Notons φ la fonction caractéristique de $X_1 - m$. Comme les variables aléatoires $X_1 - m, \dots, X_n - m$ sont indépendantes, la fonction caractéristique de $S_n/\sqrt{n}$ vaut

$$\varphi_{S_n/\sqrt{n}}(t) = \varphi_{S_n}(\frac{t}{\sqrt{n}}) = \prod_{k=1}^n \varphi_{X_k - m}(\frac{t}{\sqrt{n}}) = \varphi(\frac{t}{\sqrt{n}})^n.$$

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

$$\forall t \in \mathbb{R} \qquad \lim_{n \to +\infty} \varphi(\frac{t}{\sqrt{n}})^n = \exp(-\frac{\sigma^2}{2} t^2),$$

car $t \mapsto \exp(-\frac{\sigma^2}{2} t^2)$ est la fonction caractéristique de la loi $\mathcal{N}(0, \sigma^2)$.

Pour cela, on va utiliser le développement limité

$$\varphi(x) = 1 - \frac{\sigma^2}{2} x^2 + o(x^2) \tag{11.1}$$

établi au corollaire 16 (voir fonction caractéristique et moments).

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

$$\forall n \in \mathbb{N} \quad |z^n - u^n| = |(z - u)(\sum_{k=0}^{n-1} z^k u^{n-k})| \le n|z - u|.$$

Il s'ensuit que

$$\begin{aligned} |\varphi(\frac{t}{\sqrt{n}})^n - \exp(-\frac{\sigma^2}{2}t^2)| &= |\varphi(\frac{t}{\sqrt{n}})^n - \exp(-\frac{\sigma^2}{2n}t^2)^n| \\ &\leq n|\varphi(\frac{t}{\sqrt{n}}) - \exp(-\frac{\sigma^2}{2n}t^2)|, \end{aligned}$$

Or, on a d'une part $\exp(-\frac{\sigma^2}{2n}t^2) = 1 - \frac{\sigma^2}{2n}t^2 + o(1/n)$, et d'autre part d'après l'équation (11.1) $\varphi(\frac{t}{\sqrt{n}})) = 1 - \frac{\sigma^2}{2n}t^2 + o(1/n)$. Ainsi $n|\varphi(\frac{t}{\sqrt{n}}) - \exp(-\frac{\sigma^2}{2n}t^2)| = o(1)$, ce qui achève la preuve.

□

11.3.2 Théorème de la limite centrale en dimension d

Théorème 113. *Soit $X_1, \dots, X_n$ une suite i.i.d. de vecteurs aléatoires à valeurs dans $\mathbb{R}^p$. On suppose que $\mathbb{E}\|X_1\|^2 < +\infty$. On note alors m l'espérance et C la matrice des covariances. Alors*

$$\frac{(X_1 + \dots X_n) - nm}{\sqrt{n}} \Longrightarrow \mathcal{N}(0, C).$$

En fait, le théorème 113 peut être vu comme une conséquence du théorème 112. On va pour celà énoncer un lemme comparable dans son esprit au theorème 71.

Lemme 10. *Soit $(X_n)_{n\geq 1}$ une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^d$ et X un vecteur aléatoire à valeurs dans $\mathbb{R}^d$. Si pour tout $a \in \mathbb{R}^d$, $\langle X_n, a\rangle$ converge en loi vers $\langle X, a\rangle$, alors X_n converge en loi vers X.*

Démonstration. Soit $a \in \mathbb{R}^d$. On pose $Y_n = \langle X_n, a\rangle$. On a

$$\varphi_{X_n}(a) = \mathbb{E}e^{i\langle X_n, a\rangle} = \mathbb{E}e^{iY_n} = \varphi_{Y_n}(1)$$

Par hypothèse, Y_n converge en loi vers $\langle X, a\rangle$, donc $\varphi_{Y_n}(1)$ converge vers $\varphi_{\langle X,a\rangle}(1) = \mathbb{E}e^{i\langle X,a\rangle} = \varphi_X(a)$. Donc $\varphi_{X_n}(a)$ converge vers $\varphi_X(a)$. Comme c'est vrai pour tout $a \in \mathbb{R}^d$, le théorème de Lévy nous dit que $(X_n)_{n\geq 1}$ converge en loi vers X. □

On peut maintenant prouver le théorème 113.

Démonstration. Soit X un vecteur aléatoire suivant $\mathcal{N}(0, C)$. On pose

$$S_n = (X_1 + \dots X_n) - nm = \sum_{k=1}^{n} (X_k - m).$$

D'après le lemme précédent, il nous suffit de montrer que pour tout $a \in \mathbb{R}^d$ $\langle \frac{S_n}{\sqrt{n}}, a\rangle$ converge en loi vers $\langle X, a\rangle$.

Fixons $a \in \mathbb{R}^d$ et posons $Y_n = \langle X_n - m, a\rangle = a^*(X_n - m)$. Les Y_n sont des variables aléatoires indépendantes qui suivent toutes la même loi, la loi image de P_{X_1} par $x \mapsto \langle x-m, a\rangle = a^*(x-m)$. D'après le théorème 50, leur espérance est 0 et leur variance $a^*Ca = \langle Ca, a\rangle$. Ainsi, d'après le théorème 112, la suite $\frac{\sum_{k=1}^n Y_k}{\sqrt{n}}$ converge en loi vers $\mathcal{N}(0, \langle Ca, a\rangle)$.

Mais

$$\frac{\sum_{k=1}^n Y_k}{\sqrt{n}} = \langle \frac{S_n}{\sqrt{n}}, a\rangle,$$

et, encore d'après le théorème 50, la loi de $\langle X, a\rangle$ est précisément $\mathcal{N}(0, \langle Ca, a\rangle)$. $\langle \frac{S_n}{\sqrt{n}}, a\rangle$ converge donc en loi vers $\langle X, a\rangle$, ce qui achève la preuve.

□

11.4 Exercices sur la convergence en loi

1. Soit $(X_\lambda)_{\lambda>0}$ une famille de variables aléatoires telles que pour tout $\lambda > 0$, X_λ suive une loi de Poisson $\mathcal{P}(\lambda)$. Montrer que la suite

$$\frac{X_\lambda - \lambda}{\sqrt{\lambda}}$$

converge faiblement vers la loi $\mathcal{N}(0,1)$ lorsque λ tend vers l'infini.

2. (a) 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.

(b) 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 $+\infty$.

3. *Une suite convergente est tendue*
Soit $(\mu_n)_{n\geq 1}$ une suite de lois de probabilités convergeant faiblement vers la loi de probabilité μ. Montrer que la suite $(\mu_n)_{n\geq 1}$ est tendue,

c'est à dire que pour tout $\varepsilon > 0$, on peut trouver un compact K_ε tel que

$$\forall n \in \mathbb{N}^* \quad \mu_n(K) \geq 1 - \varepsilon.$$

4. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires convergeant en loi vers une variable aléatoire X, des suite de réels $(a_n)_{n\geq 1}$ et $(b_n)_{n\geq 1}$ convergeant respectivement vers les réels a et b. Montrer que la suite de variables aléatoires $(a_nX_n+b_n)_{n\geq 1}$ converge en loi vers la variable aléatoire $aX + b$.
5. Interpréter la quantité $e^{-n}\sum_{k=0}^n \frac{n^k}{k!}$ comme une probabilité. À l'aide du théorème de la limite centrale, démontrer la relation
$$\lim_{n\to+\infty} e^{-n}\sum_{k=0}^n \frac{n^k}{k!} = \frac{1}{2}.$$
6. 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 choisies au hasard (c'est à dire avec équiprobabilité), de manière indépendante. Ensuite, il tire au hasard un nombre X 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 alors 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
$$\lim_{n\to\infty} P(D_n \geq 0) = \frac{1}{2}.$$
7. *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
$$S_n = \sum_{k=0}^n X_k.$$
   (a) Montrer que $\frac{S_n - n}{\sqrt{n}}$ converge en loi vers la loi $\mathcal{N}(0,1)$.

(b) 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

$$h_n(x)=a_n 1\!\!1_{[-\sqrt{n},+\infty[}\frac{1}{\sqrt{2\pi}}e^{-\sqrt{n}x}(1+\frac{x}{\sqrt{n}})^n.$$

(c) Montrer que

$$\lim_{n\to+\infty}\int_0^1 g_n(x)\,dx=\int_0^1\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx$$

(d) Soit Ψ la fonction définie sur $[-1/2,1/2]$ par

$$\forall x\in[-1/2,+1/2]\quad \ln(1+x)=x-\frac{x^2}{2}+x^3\Psi(x).$$

Montrer que

$$x\geq-\sqrt{n}\Longrightarrow h_n(x)-\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}(e^{\frac{x^3}{n^{1/2}}\psi(\frac{x}{\sqrt{n}})}-1)$$

En déduire que $h_n(x)$ converge uniformément vers $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ sur tout compact.

(e) Montrer que

$$\lim_{n\to+\infty}\int_0^1 h_n(x)\,dx=\int_0^1\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx.$$

(f) En déduire la formule de Stirling :

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

Annexe A

Rappels de dénombrement

A.1 Rappels de vocabulaire ensembliste

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

On rappelle que l'ensemble vide (noté $\varnothing$) ne contient aucun élément et est inclus dans tous les ensembles

Pratiquement, si l'on veut montrer le résultat $A \subset \Omega$, la preuve ressemblera donc à "Soit $x \in A$...(raisonnement)...donc $x \in \Omega$. Comme on a choisi x quelconque dans A, on peut conclure que $A \subset \Omega$."

Si A est inclus dans Ω, on dit que A est un sous-ensemble, ou encore une partie de Ω.

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

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

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

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

A.2 Applications et cardinaux : définitions et notations

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

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

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

Une bijection d'un ensemble Ω dans lui-même est appelée *permutation* de Ω.

Un ensemble est Ω dit *fini* si

- ou bien c'est l'ensemble vide. ($\varnothing$)
- ou bien il existe un entier n tel qu'il existe une bijection entre Ω et $\{1, \dots, n\}$.

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 $\mathcal{B}_2(\{a, b, c\}) = \{\{a, b\}, \{b, c\}, \{a, c\}\}$. De même, on note $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω, quelque cardinal qu'elles aient. Par exemple $\mathcal{P}(\{a, b, c\}) = \{\varnothing, \{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{a, c\}, \{a, b, c\}\}$.

Soit A et D deux ensembles finis. On admettra les résultats suivants :

- Il existe (au moins) une bijection de D dans A si et seulement si $|A| = |D|$.
- Il existe (au moins) une injection de D dans A si et seulement si $|A| \geq |D|$.
- Il existe (au moins) une surjection de D dans A si et seulement si $|A| \leq |D|$.

Le premier des trois résultats énoncés est évidemment le plus utilisé lorsque l'on veut des dénombrements exacts, alors que les deux autres sont plutôt utilisés dans les cas trop complexes où l'on peut juste espérer des encadrements.

A.3 Principes de base du dénombrement

A.3.1 Principe de bijection

Dans la pratique, lorsque l'on veut compter les éléments d'un ensemble, on montre que cet ensemble est en bijection avec un ensemble dont on connaît (par coeur !) le nombre d'éléments. La section suivante énoncera un certain nombre de résultats qu'il faut connaître.

A.3.2 Principe d'indépendance

C'est juste la formule

$$|A \times B| = |A|.|B| \quad .$$

Considérée isolément, elle peut paraître sans intérêt mais elle est souvent utilisée en association avec le principe de bijection.

A.3.3 Principe de partition

On dit que les ensembles $(A_i)_{i\in I}$ forment une partition de A si l'on a $A = \underset{i\in I}{\cup} A_i$ et $i \neq j \Longrightarrow A_i \cap A_j = \varnothing$. On a alors

$$|A| = \sum_{i\in I} |A_i|.$$

Le résultat élémentaire suivant peut souvent être utile.

Théorème 114. *Soit Ω un ensemble quelconque, I un ensemble d'index fini ou dénombrable et $(\Omega_i)_{i\in I}$ une partition de Ω. Alors, si l'on pose $A_i = A\cap\Omega_i$, les ensembles $(A_i)_{i\in I}$ forment une partition de A.*

Démonstration. Comme $\Omega = \underset{i\in I}{\cup} \Omega_i$, on a $A = A\cap\Omega = \underset{i\in I}{\cup} A\cap\Omega_i = \underset{i\in I}{\cup} A_i$. D'autre part, pour $i \neq j$, on a $A_i \cap A_j \subset \Omega_i \cap \Omega_j = \varnothing$, d'où $A_i \cap A_j = \varnothing$. $\square$

A.3.4 Lemme des bergers

Le lemme suivant peut également être utile

Lemme 11 (Lemme des bergers). *Soit φ une application* surjective *de A dans D. On suppose qu'il existe un entier $a \geq 1$ tel que*

$$\forall y \in A \quad |\{x \in D ; \varphi(x) = y\}| = a$$

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

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

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

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

□

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

A.4 Quelques résultats incontournables

A.4.1 Nombre d'applications de D dans A

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

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

On pose $|A| = n$ et $|D| = p$. Un cas particulier important est celui où l'on a $D = \{1, \ldots, p\}$. Or, un p-uplet $(x_1, x_2, \ldots, x_p)$ dont les composantes sont des éléments de A peut être considéré comme la donnée d'une application de $\{1, \ldots, p\}$ dans A. Le nombre de p-uplets $(x_1, x_2, \ldots, x_p)$ dont les composantes sont des éléments de A est donc n^p.

Exemple : Un professeur note chaque étudiant d'une classe de 30 étudiants par une note entière de 0 à 20. Le nombre de résultats possibles est le nombre de fonctions de l'ensemble D des étudiants dans l'ensemble $A = \{0, \ldots, 20\}$ des notes possibles. Comme $|A| = 21$ et $|D| = 30$, il y a donc 21^{30} résultats possibles.

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

A.4.2 Nombre de permutations de Ω

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

$$n(n-1)\ldots 1.$$

On note $n! = n(n-1)\ldots 1$.

Remarque : $n!$ se lit "*factorielle* n" .

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

A.4.3 Nombre d'injections de D dans A

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

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

Démonstration. Soit n un entier. On pose $A = \{1, \ldots, n\}$ et on note $\mathcal{I}_p$ l'ensemble des injections de $\{1, \ldots, p\}$ dans A. On va montrer par récurrence sur $p \in \{1, \ldots, n\}$ 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 $\{1, \ldots, p+1\}$ dans A associe sa restriction à $\{1, \ldots, p\}$. Un peu de réflexion montre que

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

D'après le lemme des bergers, on a donc

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

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

Remarques :

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

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

A.4.4 Nombre de parties de Ω possédant p éléments

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

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

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

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

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

A.4.5 Nombre total de parties de Ω

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

Démonstration. Il suffit de remarquer que l'application

$$\begin{array}{rcl} \mathcal{P}(\Omega) & \to & \{0;1\}^{\Omega} \\ A & \mapsto & \mathbb{1}_A \end{array}$$

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

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

□

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

A.5 Équations et inéquations en entiers

Lemme 12. *Soit n, p entiers. Si $n \geq p$, 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 croissantes de p éléments pris dans $\{1, \dots, n\}$ revient à choisir une partie de $\{1, \dots, n\}$ possédant p éléments, puis à les ordonner dans l'ordre naturel. Or $\binom{n}{p}$ est, par définition, le nombre de parties de $\{1, \dots, n\}$ 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 115. *Pour n, p 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 à l'inéquation :*

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

Démonstration. Il suffit de remarquer que l'application

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

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

Théorème 116. *Pour n, p 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 à 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 117. *Pour n, p entiers positifs, il existe exactement $\binom{n+p-1}{p-1}$ p-uplets $(x_1, \dots, x_p) \in \mathbb{N}^p$ solutions à l'équation :*

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

Démonstration. Il suffit de remarquer que l'application

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

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

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

□

Exemple : 4 listes se présentent aux élections syndicales étudiantes où 9 sièges sont à pourvoir. Combien y a-t-il de répartitions des sièges possibles. Réponse : il s'agit de compter les solutions en entiers positifs ou nuls à l'équation

$$x_1 + x_2 + x_3 + x_4 = 9,$$

x_k représente le nombre d'élus de la liste k. Il y a donc $\binom{9+4-1}{4-1} = \binom{12}{3} = \frac{12 \times 11 \times 10}{1 \times 2 \times 3} = 220$ répartitions possibles.

Théorème 118. *Pour n, p entiers positifs, il existe exactement $\binom{n+p}{p}$ p-uplets $(x_1, \dots, x_p) \in \mathbb{N}^p$ solutions à l'inéquation :*

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

Démonstration. La preuve, analogue à celle du théorème précédent, est laissée en exercice. □

A.6 Formule de Poincaré (aussi appelée formule du crible)

Cette formule est très utile en combinatoire, son application la plus classique étant le calcul du nombre de permutations sans point fixe (nombre de dérangements).

Pour tous ensembles $A_1, A_2, ..., A_n$, on a

$$\begin{aligned} |\bigcup_{i=1}^{n} A_i| &= \sum_{B\in\mathcal{P}(\{1,\dots,n\})\backslash\varnothing} (-1)^{1+\mathrm{Card}(B)} |\cap_{j\in B} A_j| && \text{(A.7)} \\ &= \sum_{i=1}^{n} |A_i| - \sum_{1\le i_1<i_2\le n} |A_{i_1}\cap A_{i_2}| + \cdots && \text{(A.8)} \\ &\quad \cdots + (-1)^{k+1} \sum_{1\le i_1<i_2<\ldots<i_k\le n} |A_{i_1}\cap\cdots\cap A_{i_k}| + \cdots && \text{(A.9)} \\ &\quad \cdots + (-1)^{n+1}|A_1\cap\cdots\cap A_n|. && \text{(A.10)} \end{aligned}$$

Exemple : Pour $n = 3$, on a

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

Démonstration. On pourrait prouver la formule par récurrence sur n, mais c'est plutôt lourd. On préférera une preuve probabiliste (voir plus loin). □

A.7 Exercices

1. Combien existe-t-il de mots de n lettres construits avec l'alphabet $\{a; b\}$ et ne comportant pas deux "a" consécutifs ?
   Indication : montrer que si u_n est le nombre de tels mots se terminant par "a" et v_n est le nombre de tels mots se terminant par "b", on a la récurrence

$$\begin{pmatrix} u_{n+1} \\ v_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} u_n \\ v_n \end{pmatrix}$$

2. On donne un ensemble Ω de 15 entiers compris entre 1000 et 2000. Montrer que l'on peut en extraire deux sous-ensembles disjoints non-vides A et B tels que la somme des éléments de A soit égale à la somme des éléments de B.

Indication : montrer que l'application

$$\begin{aligned} \mathcal{P}(\Omega) &\to \mathbb{N} \\ A &\mapsto \sum_{x\in A} x \end{aligned}$$

n'est pas injective.
Raffiner le raisonnement pour montrer que si l'on remplace 15 par 14, le résultat est encore vrai.
Indication : remplacer $\mathcal{P}(\Omega)$ par $\bigcup\limits_{i=5}^{8} \mathcal{B}_i(\Omega)$.

3. Au tapis vert, il faut choisir 4 cartes (as,roi, dame, valet, dix, neuf, huit ou sept) par couleur (pique, coeur, carreau ou trèfle). On appelle grille l'un des résultats possibles.
   exemple de grille : as de pique, dame de cœur, 7 de carreau, 7 de trèfle.
   Dénombrer les grilles contenant :
   (a) la dame de cœur
   (b) une dame et une seulement
   (c) deux dames et deux seulement
   (d) aucune dame
   (e) au moins une dame
   (f) l'as de pique et une dame seulement
4. 5 prévenus sont amenés à choisir un avocat dans une liste de 10 avocats commis d'office.
   (a) Combien y a-t-il de choix possibles ?
   (b) Combien y a-t-il de choix tels que les 5 prévenus choisissent le même avocats ?
   (c) Combien y a-t-il de choix tels que 2 avocats soient appelés ?
5. De combien de manières peut-on disposer n personnes autour d'une table ronde ? Précision : seule la position relative à la maîtresse de maison importe, une rotation ne change donc rien.
   Si $n = 2k$, avec k hommes et k femmes, combien de dispositions permettent une alternance des sexes ?
6. Dans une assemblée de n personnes, on note la date d'anniversaire de chacun.
   (a) Donner le cardinal de toutes les répartitions possibles.
   (b) Calculer le cardinal des ensembles suivants :

i. tous les participants ont la meme date d'anniversaire.
ii. Pierre et Paul ont la meme date d'anniversaire.
iii. deux personnes ont la meme date d'anniversaire.
iv. toutes les personnes ont des dates d'anniversaires différentes.

7. On considère l'ensemble Ω des suites de n chiffres (les chiffres sont pris dans $0, 1, \ldots, 9$). Combien vaut $|\Omega|$? Combien y-a-il de chiffres comportant un nombre pair de zéros ?
8. Combien y a-t-il de nombres à trois chiffres (les chiffres sont pris dans $0, 1, \ldots, 9$) dont la somme des chiffres est égale à 22 ?
   Indication : noter
   $$\Omega_{22} = \{(x_1, x_2, x_3) \in \mathbb{N}^3; x_1 + x_2 + x_3 = 22\},$$
   $$B = \{(x_1, x_2, x_3) \in \{0, \ldots, 9\}^3; x_1 + x_2 + x_3 = 22\},$$
   $$M_1 = \{x \in \Omega_{22}; x_1 \geq 10\},$$
   $$M_2 = \{x \in \Omega_{22}; x_2 \geq 10\},$$
   $$M_3 = \{x \in \Omega_{22}; x_3 \geq 10\}.$$
   Remarquer que $\Omega \backslash B = M_1 \cup M_2 \cup M_3$. Ensuite, si l'on définit Ω_{12} et Ω_2 de manière analogue à Ω_{22}, on pourra exhiber une bijection entre Ω_{12} et M_1 et entre Ω_2 et $M_1 \cap M_2$.

Annexe B

Indications

B.1 Exercices de théorie de la mesure

1. Pour la deuxième question, on pourra d'abord observer que certaines des conditions des axiomes sont vérifiées sans hypothèse supplémentaire sur f.
2. On procèdera par double inclusion. On rappelle que si $\mathcal{C} \subset \mathcal{D}$, $\sigma(\mathcal{C}) \subset \sigma(\mathcal{D})$.
3. $\mathbb{Q}$ est dense dans $\mathbb{R}$, ainsi chaque réel est limite d'une suite croissante et d'une suite décroissante.
4. (a) Exprimer C en fonction de A_2 et A_5.
   (b) Exprimer B en fonction des $(A_p)_{p\in\mathcal{P}}$.
5. Pour F fermé de E, on pourra considérer la fonction $d_F(x) = \inf\{d(x,y); y \in F\}$.
6. Traiter le cas où g est une indicatrice, puis une fonction étagée, puis une fonction positive,...

B.2 Premiers exercices de probabilité

1. Remarquer qu'une union dénombrable d'événements est de probabilité nulle si et seulement si chacun est de probabilité nulle.
2. (a) Remarquer qu'aucun entier n'a une infinité de diviseurs premiers.
   (b) On pourra remarquer que
   $$(1 - \frac{1}{p_i})^{-1} \geq \sum_{k=0}^{N} p_i^{-k},$$
   où N est choisi tel que $p_i^{N+1} > p_n$.

   (c) Utiliser la continuité séquentielle décroissante.
   (d) Utiliser le résultat de l'exercice précédent.
3. Utiliser la formule de Bayes.
4. (a) On numérote les mathématiciens, et à chaque mathématicien, on asocie le numéro du propriétaire du chapeau qu'il prend
   (b) On pourra utiliser le principe de partition.
   (c) Quel est l'endomorphisme réciproque ?
   (d) Relire les résultats précédents
   (e) On pourra remarquer que la série est alternée.
5. (a) $A_d = d\mathbb{N} \cap \Omega_n$.
   (b) On déterminera un entier d tel que $\cap_{i=1}^r A_{d_i} = A_d$.
   (c) On remarquera que deux nombres sont premiers entre eux si et seulement si ils n'ont pas de diviseur premier commun.
   (d) Utiliser la formule de Poincaré.
6. Commencer par choisir clairement l'espace Ω.
7. On pourra conditionner par la valeur prise par l'*ensemble* des trois nombres tirés au sort.
8. (a) $\Omega = \{(x_1, \dots, x_{a+b}) \in \{\overrightarrow{i}\,;\,\overrightarrow{j}\}^{a+b}\ x_1 + \dots + x_{a+b} = a\overrightarrow{i} + b\overrightarrow{j}\}$ est un choix possible.
   (b) Si l'on note $I = \{$ le graphe coupe la diagonale.$\}$, on pourra montrer que

$$P(I \cap \{\text{A gagne le premier échange}\}) = P(I \cap \{\text{B gagne le premier échange}\}).$$

9. Faire en sorte que le résultat soit nul.

B.3 Premiers exercices d'intégration

1. (a) Pour tout t dans $\mathbb{R}$, on pourra montrer que l'événement $\{Y \leq t\}$ est mesurable.
   (b) Utiliser le fait que T préserve la mesure μ.
   (c) On pourra montra que $Y(x) = Z(x)$ pour μ presque tout x.
2. (a) Vérifier les axiomes.
   (b) Commencer par le cas où f est étagée.
3. Utiliser le théorème de convergence monotone.

4. Utiliser le théorème de convergence dominée.
5. (a) Utiliser une partition de l'espace $\mathbb{R}^d$.
   (b) On pourra montrer qu'il existe des constantes positives C et D telles que pour tout x avec $\|x\|_2 \geq 1$, on ait
   $$\frac{C}{\|x\|_2^\alpha} \leq \frac{1}{\mathrm{Ent}(\|x\|_\infty)^\alpha} \leq \frac{D}{\|x\|_2^\alpha}.$$
   (c) Le plus simple ici est de montrer que f est C^1 par des théorèmes généraux, puis de calculer la différentielle dans une direction donnée.
   (d) On pourra utiliser le théorème de C^1-difféomorphisme. Pour le calcul de la différentielle, il pourra être intéressant d'écrire la matrice de Df_x dans une base orthogonale bien choisie.
6. On pourra prendre $A = \{f > 0\}$.
7. (a) On pourra démontrer que pour $n \geq 1$, $0 \leq \frac{\log(x+n)}{n} \leq 1 + x$.
   (b) $(1+x)^n \geq \ldots$
   (c) Pour $n \geq 3$, $\frac{1+nx}{(1+x)^n} \leq \frac{3}{(1+x)^2}$.
8. (a) Procéder par récurrence.
   (b) Faire un développement en série.
9. Fubini est ton ami.
10. Tonelli aussi.

B.4 Exercices sur les lois

1. (a) Pensez à discuter suivant les positions relatives de n et k.
   (b) $\mathbb{Q}$ est dense dans $\mathbb{R}$.
   (c) Utiliser le principe de partition
   (d) Écrire $P(D) = 1$, avec D bien choisi.
   (e) Trivialiser la question !
2. Prendre X et Y deux variables indépendantes suivant chacune une loi de Bernoulli de paramètre $1/2$ et poser $Z = |X - Y|$.
3. On prend $\Omega = \{0,1\} \times \{0,1\}$, $\mathcal{F} = \mathcal{P}(\omega)$ $\mathcal{C} = \{(0,0),(0,1)\}, \{(0,0),(1,0)\}\}$. $P = \frac{1}{2}(\delta_0 + \delta_1) \otimes \frac{1}{2}(\delta_0 + \delta_1)$, $Q = \frac{1}{2}(\delta_{(0,0)} + \delta_{(1,1)})$
4. Traduire les événements considérés en fonction de Y
5. Poser $x = AM$ et résoudre l'inéquation.

6. Dire que le maximum de n nombres ne dépasse pas x, c'est dire que chacun ne dépasse pas x.
7. Remarquer que $1 - m_n = \max(1 - X_1, \dots, 1 - X_n)$.
8. À k fixé, il faut déterminer les valeurs de X qui sont telles que $Y = k$.
9. Représenter graphiquement les domaines considérés.
10. (a) On pourra montrer que X est $\mathcal{T}$-mesurable, puis que les A_p sont $\sigma(X)$-mesurables.
    (b) On montrera que l'ensemble cherché est $\{\omega \in \mathbb{N}; X(\omega) = 2 \times 3 \times 5 \times 11$.
    (c) Le sens direct est le plus simple. Pour la réciproque, remarquer que si A_n est $\mathcal{T}$-mesurable, il doit contenir $\{\omega \in \mathbb{N}^*; X(\omega) = X(n)\}$.
    (d) On montrera que $A_m \cap A_n = A_{m\wedge n}$.
    (e) Remarquer que pour p assez grand $1 - \frac{1}{p^{2s}} \geq e^{-\frac{2}{p^{2s}}}$.
11. (a) On pourra montrer que $\mathcal{Q}$ est la tribu de queue associée à la famille $(\mathcal{A}_n)_{n\geq 1}$.
    (b) Un ensemble d'entiers est infini si et seulement si il contient au moins un entier plus grand que n'importe quel entier fixé à l'avance.

    Ainsi, on pourra montrer que pour tout n_0, $A = \bigcap_{n\geq n_0} \bigcup_{k\geq n} A_k$.
12. Pour la première formule, faire une intégration par parties ; pour la deuxième, procéder par récurrence.
13. On pourra écrire $e^{-\frac{x^2}{2}} = \frac{1}{x}(xe^{-\frac{x^2}{2}})$ afin de procéder à une intégration par parties.

B.5 Exercice sur les espérances

1. Si X désigne le nombre de fois où l'on a obtenu le nombre choisi, le gain est $X - \mathbb{1}_{X=0}$.
2. Une probabilité est l'indicatrice d'une espérance.
3. Interpréter le membre de gauche comme la valeur absolue d'une covariance.
4. Effectuer une transformation d'Abel.
5. Appliquer le théorème de transfert, et penser à la forme canonique des polynômes du second degré.
6. Appliquer le théorème de transfert.

7. On pourra commencer par supposer que la loi est centrée (c'est à dire que $a+b=0$) et faire une majoration simple). On s'y ramènera dans le cas général.
8. Fonction de répartition, ou transformation : vous avez le choix !
9. On peut raisonner en termes de loi.
10. S'inspirer de l'exercice précédent et utiliser une transformation.
11. Appliquer l'exercice précédent.
12. Bien observer que X et Y sont indépendantes.
13. On pourra observer que l'application $(x,y) \mapsto (\frac{x}{\sqrt{y/n}}, y)$ réalise un C^1-difféomorphisme de $\mathbb{R}^* \times \mathbb{R}_+^*$ dans lui même.
14. Remarquer que tout se passe comme si (X,Y) suivait la loi uniforme sur $T = \{(x,y) \in \mathbb{R}^2; 0 < y < x < 1\}$.

    Si x et y sont solutions réelles de $x^2 - Sx + P = 0$, alors $|x-y| = \sqrt{S^2 - 4P}$
15. (a) Identifier la loi de S_n et appliquer le théorème de transfert.

    (b) i. Remarquer que $[0,1]$ est compact.

    ii. Remarquer que

    $$B_n(x) - f(x) = \int_{|\frac{S_n}{n} - x| \le \varepsilon} f(\frac{S_n}{n}) - f(x)\, dP + \int_{|\frac{S_n}{n} - x| > \varepsilon} f(\frac{S_n}{n}) - f(x)\, dP$$

    En déduire que la suite des polynômes B_n converge vers f uniformément sur $[0,1]$.
16. Si on note C_x l'ensemble des cases controlées par la dame x, et $C = \bigcup_{x=1}^{7} C_x$, on peut minorer $|C|$ grâce aux inégalités de Bonferroni.
17. (a) On prendra $\Omega = \mathcal{B}_r(\{1,\dots,n\})$.

    (b) Utiliser les relations entre l'espérance et la queue de distribution, puis utiliser la relation de récurrence du triangle de Pascal.
18. (a) On pourra remarquer qu'une rotation est une application linéaire isométrique.

    (b) Remarquer que $|X - Y| = \sqrt{2}U$.

B.6 Exercices sur les espaces L^p

1. (a) Cherchez un peu plus...

   (b) Découper $\mathbb{R}$ en intervalles de longeur 2π.
   (c) Utiliser des équivalents.
2. Étudier d'abord la convergence ponctuelle.
3. On pourra raisonner par l'absurde
4. On pourra utiliser des sous-suites.
5. Retrousser ses manches (ou équivalents).
6. Majorer $\sqrt{|f^2+g^2|}$ par une fonction intégrable.
7. Passer à l'intégrale de Riemann.
8. (a) Prendre $X = [0,1]$ et pour μ la mesure de Lebesgue sur $[0,1]$, choisir ensuite f_n telle que $f_n(x) \to 0$ pour tout $x \in [0,1[$ et que l'on ait $\int f_n = 1$.
   (b) Pour p entier s'écrivant $p = 2^n + k$, avec $0 \leq k < 2^n$, poser $u_p = \frac{k}{2^n}$. Ensuite, poser $\varphi_n(x) = \max(1 - \frac{n}{4}|x|, 0)$ et finalement $f_n(x) = \varphi_n(x - u_n)$.
   (c) Symétriser l'exemple trouvé à la première question.
9. Utiliser l'inégalité de Cauchy-Schwartz.
10. Utiliser l'inégalité de Hölder.
11. (a) Utiliser l'inégalité de Holder.
    (b) i. Considérer l'intégrale $\int_{]0,x[} f d\lambda$ comme intégrale de Riemann.
        ii. Remarquer que $T(f)(x)$ est bornée et décroit suffisamment vite à l'infini.
        iii. Remarquer que $f(x) = T(f)(x) + xT(f)'(x)$ et faire une intégration par parties.
        iv. Cherchez un peu plus. . .
        v. $f = f^+ - f^-$.
    (c) i. Pour le premier point, on pourra utiliser l'inégalité de Hölder.
        ii. Utiliser la densité des fonctions continues à support compact dans L^p.
12. 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 $\{f = 0\})$.

B.7 Exercices sur la convolution et la transformée de Fourier

1. Si $a = 0$ ou $b = 0$, alors $ab = 0$.
2. (a) Pleins de calculs en perspective. On conseille de commencer par exprimer $\int \exp(-p(x))\ dx$ en fonction de A, B, C, lorsque $p(x) = Ax^2 + Bx + C$ et $A > 0$.

   (b) Commencer par identifier les points où la convolée est nulle
3. (a) $1 - x^2 \geq 1 - x$ pour $0 \leq x \leq 1$.

   (b) On pourra montrer que pour tout $\delta \in]0, 1[$,
   $$f * k_n(x) - f(x) \leq \frac{4(1-\delta^2)^n}{a_n}\|f\|_\infty + \omega_f(\delta).$$

   (c) On pourra montrer que
   $$f * k_n(x) = \frac{1}{a_n}\int_{-1/2}^{1/2} \left(1 - (x-t)^2\right)^n f(t)\ dt.$$

   (d) On pourra commencer par le cas où $a = -1/4$ et $b = 1/4$.
4. (a) Pour éviter d'oublier des cas, se souvenir que le support de la convolée est inclus dans la "somme" des supports ; la parité peut également permettre de simplifier des choses...

   (b) Remarquer que $f^{(*)n}$ est positive.

   (c) Procéder par récurrence.
5. (a) Si l'on pose $g = 1\!\!1_E * 1\!\!1_{-E}$, on peut remarquer que $g(x) = \int_E T_x 1\!\!1_E d\mu$.

   (b) Remarquer que si $g(x) \neq 0$, $x \in E - E$.
6. (a) Ah bah non ! Vous l'avez déjà eue, l'indication.

   (b) Réduire A dans une base orthonormale.
7. (a) On pourra remarquer que la transformée de Fourier est injective dans $\mathcal{L}^1(\mathbb{R}^n))$

   (b) Utiliser la transformation de Fourier et une fonction bien choisie.
8. On pourra utiliser la formule d'inversion.
9. On pourra utiliser la formule d'inversion.

B.8 Exercices sur les fonctions génératrices et les fonctions caractéristiques

1. (a) Remarquer que $t^Z = 1\!\!1_A t^X + 1\!\!1_{A^c} t^Y$.

   (b) Remarquer que le score est une variable aléatoire fabriquée suivant le principe de la première question.

2. (a) Remarquer que

$$\forall s \in [0,1] \quad s^S = \sum_{n=1}^{+\infty} 1\!\!1_{T=n} s^{S_n}.$$

   (b) Utiliser les liens entre la fonction génératrice et l'espérance.

3. Utiliser la première question de l'exercice précédent.

4. (a) Commencer par déterminer la fonction génératrice de $K_1 L_1$. On pourra s'inspirer de l'exercice 1.

   (b) S'inspirer de l'exercice 2.

   (c) Relire le cours.

5. (a) On trouvera $f^n \circ f$, où $f(z) = (1+z)/2$.

   (b) Remarquer que $f^n \circ f = (f \circ f)^n$

6. Penser à la fonction caractéristique.

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

8. Utiliser les liens entre fonction caractéristique et moments.

9. Utiliser les liens entre fonction caractéristique et moments.

10. (a) Pour tout borélien A, on a $P_X(A) = P_X(A \cap [0,+\infty[) + P_X(A \cap]-\infty, 0[)$.

    (b) Appliquer le théorème de transfert.

11. (a) On pourra utiliser le théorème de Fubini et se servir de la valeur de la fonction caractéristique d'une loi gaussienne.

    (b) On pourra remarquer que (U,V) a même loi que $(U,-V)$.

12. (a) Passer en coordonnées sphériques.

    (b) Montrer que $\varphi_\mu(t)$ ne dépend que de $\|t\|$, puis écrire $\varphi_\mu(t)$ sous la forme $\varphi_\mu(t) = \frac{1}{\|t\|^{2,5}} \int_0^{\|t\|} f(x)\, dx$.

B.9 Exercices sur la convergence presque sûre

1. Appliquer la loi des grands nombres
2. Pour tout $\varepsilon > 0$, appliquer le lemme de Borel-Cantelli aux événements $\{\frac{X_n}{(\ln n)^{\frac{3}{2}}} > \varepsilon\}$.
3. Discuter en fonction de A la nature de la série de terme général $P(\frac{X_n}{\ln n} > A)$.
4. S'inspirer de l'exercice précédent en utilisant un équivalent pour la queue de la gaussienne.
5. Pour une suite à veleurs entières, les valeurs d'adhérences sont les valeurs qui sont prises une infinité de fois.
6. On a $T_n = \sum_{k=2}^n \mathbb{1}_{\{U_{n-1}<p\}\Delta\{U_n<p\}}$. On pourra découper T_n en deux sommes de variables aléatoires indépendantes.
7. (a) Passer au logarithme.
   (b) Écrire M_n sous la forme $M_n = PD_nP^{-1}$ et introduire la norme $\|x\|_* = \|P^{-1}x\|_\infty$.
8. On pourra remarquer que pour tout entier n, l'événement "$(Y_n)_{n\geq 1}$ est nulle à partir d'un certain rang" contient l'événement $\{X_n = 0\}$.
9. On pourra montrer que $X_n \in \{0,1\}$ à partir d'un certain rang.
10. On pourra montrer que $P(Y_n \neq 0) \leq \exp(-\sum_{k=1}^{n} \exp(-\lambda_k))$.
11. On pourra montrer que $(X_n)_{n\geq 2}$ ne prend la valeur 1 qu'un nombre fini de fois. Enfin, on montrera que
$$1 - p = \prod_{n=2}^{+\infty} (1 - \frac{1}{n^2}).$$
12. (a) Utiliser le lemme de Borel-Cantelli et le lien série-intégrale.
    (b) On pourra commencer par montrer que $\sup_{n\geq 1} \frac{|X_n|}{n} = +\infty$ p.s.

B.10 Exercices sur les vecteurs gaussiens

1. Remarquer que $\frac{\sigma(U)}{\sigma(V)}V$ a même loi que U.
2. On rappelle que si X suit $\mathcal{N}(0,1)$, $\mathbb{E}X^4 = 3$.
3. Trouver une matrice orthogonale O tel que la matrice de covariance de OX soit diagonale.

4. (a) Déterminer le spectre de C.
   (b) On pourra commencer par trouver une constante K telle que
   $$\forall a \in \mathbb{R}^3 \quad \mathbb{E}|\langle X, a\rangle| = (\mathbb{E}|\langle X, a\rangle|_2^2)^{1/2}.$$
5. On pourra montrer qu'il existe un vecteur gaussien dont la matrice de covariance est
   $$\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$$
6. (a) Relire le cours !
   (b) Se souvenir qu'une matrice de covariance est symétrique positive.
   (c) Si (Z,T) est gaussien $(\frac{Z}{\sigma(Z)}, T)$ aussi.
   (d) Utiliser la question précédente.
   (e)
   $$\mathbb{E}\|X\|^4 = \mathbb{E}\Big(\sum_{k=1}^{n} X_k^2\Big)^2.$$
7. Trouver une matrice orthogonale O tel que la matrice de covariance de OX soit diagonale.
8. (a) Revoir l'image d'un vecteur gaussien par une transformation affine.
   (b) On pourra utiliser un changement de variables.
9. On pourra diagonaliser A dans une base orthogonale.

B.11 Exercices sur la convergence en loi

1. Utiliser le premier théorème de Levy.
2. (a) Commencer par déterminer la loi de X_n.
   (b) Pour montrer la croissance, on pourra remarquer que $f(x) = \lim_{n\to+\infty} \mathbb{E}u \circ X_n^x$. Pour déterminer la limite, il est commode de se ramener au cas où $(u_n)_{n\geq 0}$ est à valeurs positives et remarquer qu'alors, pour tout n, il existe un polynôme P_n tel que
   $$\forall x > 0 \quad f(x) \geq (1 - \frac{P_n(x)}{e^x})u_n.$$
3. On pourra commencer par déterminer R tel que $\mu(B(0,R)) \geq 1 - \frac{\varepsilon}{2}$, puis utiliser le théorème du porte-manteau.

4. On pourra considérer la loi du triplet (a_n, b_n, X_n).
5. On pourra considérer une suite $(X_n)_{n\in\mathbb{N}^*}$ de variables aléatoires indépendantes de même loi de Poisson de paramètre 1, poser $S_n = \sum_{k=1}^{n} X_k$ et regarder la quantité $P(\frac{S_n - \mathbb{E}S_n}{\sqrt{n}} \le 0)$.
6. Notons $m_{x,y} = 1$ si la pièce en (x, y) est face, -1 sinon. Si on note $C'_n = \sum_{k\neq X} M_{k,X}$ et $L'_n = \sum_{k\neq X} M_{X,k}$, on a

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

7. (a) On pourra utiliser le théorème de la limite centrale et le résultat de l'exercice 4.
   (b) Se souvenir qu'une loi exponentielle est une loi Gamma.
   (c) Utiliser le théorème de la limite centrale.
   (d) Remarquer que Ψ est bornée.
   (e) Intégrer.
   (f) Cela revient à montrer que (a_n) tend vers 1.

Annexe C

Problèmes

C.1 Densité naturelle des couples premiers entre eux

Notations et résultats admis

- Si E est un ensemble fini, $\mathcal{P}(E)$ désigne l'ensemble des parties de E et $|E|$ le cardinal de E. Si E est non-vide, alors il existe une unique mesure P sur $(E, \mathcal{P}(E))$ vérifiant $\forall x \in E \quad P(\{x\}) = \frac{1}{|E|}$. Cette mesure est une mesure de probabilité, appelée (mesure de) probabilité uniforme sur E. Elle vérifie la propriété suivante :

$$\forall A \in \mathcal{P}(E) \quad P(A) = \frac{|A|}{|E|}.$$

- Pour $x \in \mathbb{R}$, on appelle partie entière de x et on note $\mathrm{Ent}(x)$ l'unique entier n vérifiant $n \le x < n+1$.
- $\displaystyle\sum_{n=1}^{+\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$.
- Lorsque n et k sont des entiers, l'écriture $k|n$ signifie "k divise n".
  Si les séries de terme général (u_n) et (v_n) sont absolument convergentes, alors on a l'identité

$$\Big(\sum_{n\ge 1} u_n\Big)\Big(\sum_{n\ge 1} v_n\Big) = \sum_{n\ge 1} (u * v)_n,$$

  où l'on a posé

$$(u * v)_n = \sum_{d|n} u_d v_{n/d}.$$

Par exemple

$$(u * v)_6 = u_1 v_6 + u_2 v_3 + u_3 v_2 + u_6 v_1.$$

– Formule de Poincaré (aussi appelée formule du crible) Pour tous événemements $A_1, A_2, ..., A_n$ sous la probabilité P

$$\begin{aligned} P\Big(\bigcup_{i=1}^{n} A_i\Big) &= \sum_{B \in \mathcal{P}(\{1,\dots,n\}) \setminus \varnothing} (-1)^{1+\mathrm{Card}(B)} P\big(\bigcap_{j \in B} A_j\big) \\ &= \sum_{i=1}^{n} P(A_i) - \sum_{1 \le i_1 < i_2 \le n} P(A_{i_1} \cap A_{i_2}) + \cdots \\ &\quad \cdots + (-1)^{k+1} \sum_{1 \le i_1 < i_2 < \ldots < i_k \le n} P(A_{i_1} \cap \cdots \cap A_{i_k}) + \cdots \\ &\quad \cdots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n). \end{aligned}$$

Question préliminaire

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

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

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

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

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

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

I

Soit $k \in \mathbb{N}^*$. On pose $E_k = \{0, 1\}^k$.

Soit $A_k = \{x \in E_k;\ \sum_{i=1}^{k} x_i \text{ est pair}\}$ et $B_k = \{x \in E_k;\ \sum_{i=1}^{k} x_i \text{ est impair}\}$.

1. Montrer que l'application
$$\begin{array}{rcl}\Psi : E_k & \to & E_k \\ (x_1, x_2, \dots, x_k) & \mapsto & (1 - x_1, x_2, \dots, x_k)\end{array}$$
réalise une bijection de A_k dans B_k.
2. Montrer que pour tout entier naturel k non nul
$$|A_k| = |B_k| = 2^{k-1}.$$
3. k étudiants assistent à un cours de mathématiques. Quelle est la probabilité pour que le nombre d'étudiants qui s'ennuyent soit pair ? Préciser les hypothèses simplificatrices que vous choisirez de prendre.
Indication : Repérer l'état de lassitude des étudiants en associant à chaque étudiant un élément de $\{0, 1\}$ suivant l'intérêt qu'il porte au cours.

II

Soit $n \in \mathbb{N}, n \geq 2$. On décompose n en produits de facteurs premiers :
$$n = \prod_{i=1}^{k} p_i^{\alpha_i},$$
où pour tout i dans $\{1, ..., k\}$, p_i est un nombre premier et α_i un entier strictement positif, les p_i étant deux à deux distincts. On pose
$$\mu(n) = \begin{cases} 0 & \text{si il existe } i \text{ tel que } \alpha_i > 1, \\ (-1)^k & \text{sinon.} \end{cases}$$
On pose également $\mu(1) = 1$.
On a ainsi défini une fonction $\mu : \mathbb{N}^* \to \{-1, 0, 1\}$.
Cette fonction est appelée fonction de Moëbius.
On pose
$$\Omega_n = \{0, \dots, \alpha_1\} \times \{0, \dots, \alpha_2\} \times \dots \times \{0, \dots, \alpha_k\}$$
et on note Ω'_n l'ensemble des diviseurs positifs de n. On note Q la mesure de probabilité uniforme sur Ω_n et Q' la mesure de probabilité uniforme sur Ω'_n.

1. Montrer que $|\Omega_n| = \prod_{i=1}^{k} (\alpha_i + 1)$.

2. Soit f l'application :
$$\begin{array}{rcl} \Omega_n & \to & \mathbb{N} \\ (\nu_1, \ldots, \nu_k) & \mapsto & \prod_{i=1}^{k} p_i^{\nu_i} \end{array}$$
Montrer que Q' est la mesure image de Q par f.
3. On pose $Z_k = \Omega_n \backslash E_k$. Montrer que pour toute fonction g de Ω_n dans $\mathbb{R}$, on a
$$\int_{\Omega_n} g(\omega)\ dQ(\omega) = \int_{A_k} g(\omega)\ dQ(\omega) + \int_{B_k} g(\omega)\ dQ(\omega) \int_{Z_k} g(\omega)\ dQ(\omega)$$
4. Montrer $\forall \nu = (\nu_1, \ldots, \nu_k) \in \Omega_n$
$$(\mu \circ f)(\nu) = \begin{cases} (-1)^{\left(\sum_{i=1}^{k} \nu_i\right)} & \text{si } \nu \in E_k, \\ 0 & \text{sinon.} \end{cases}$$
5. En utilisant le théorème d'intégration par rapport à une mesure image – appelé aussi théorème de transfert – , montrer que
$$\int_{\Omega'_n} \mu(x)\ dQ'(x) = Q(A_k) - Q(B_k).$$
6. Montrer $\forall n \in \mathbb{N}^*$
$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{si } n = 1, \\ 0 & \text{sinon.} \end{cases}$$
7. On choisit au hasard - avec équiprobabilité - un diviseur de 2000. Quelle est la probabilité pour que ce nombre soit divisible par un carré autre que 1 ?

III

Pour n fixé, soit Ω l'ensemble des couples d'entiers $(u, v) \in \{1, ..., n\}^2$. On suppose que tous les couples sont équiprobables. On note donc P_n la probabilité définie sur $\mathcal{P}(\Omega)$ par
$$P_n(A) = \frac{1}{n^2}|A|.$$
On note p_n la probabilité pour que deux entiers pris “au hasard” entre 1 et n soient premiers entre eux.
Pour $k \in \mathbb{N}^*$, on pose $F_k = \{(u, v) \in \Omega;\ k|u \text{ et } k|v\}$.

1. Montrer $P_n(F_k) = (\text{Ent}(n/k))^2/n^2$.
2. Montrer que si k et l sont premiers entre eux, alors
$$P_n(F_k \cap F_l) = P_n(F_{kl}).$$
Généraliser au cas de m entiers deux à deux premiers entre eux.
3. Soient $\beta_1, ..., \beta_k$ les nombres premiers compris entre 2 et n.
On pose $F = \bigcup\limits_{i=1}^{k} F_{\beta_i}$.
Montrer $p_n = 1 - P_n(F)$.
4. En utilisant la formule de Poincaré, montrer
$$p_n = \sum_{d=1}^{n} \mu(d) \frac{1}{n^2} (\text{Ent}(\frac{n}{d}))^2.$$
Indication : il pourra être utile de remarquer que pour tout entier d compris entre 1 et n, $\mu(d)$ est non-nul si et seulement si d peut s'écrire comme produit d'éléments deux à deux distincts pris dans l'ensemble $\{\beta_1, \ldots, \beta_k\}$.
5. (a) On pose
$$a(k,n) = \begin{cases} \mu(k) \frac{1}{n^2} (\text{Ent}(\frac{n}{k}))^2 & \text{si } 1 \leq k \leq n, \\ 0 & \text{sinon.} \end{cases}$$
Montrer
$$\forall (n,k) \in (\mathbb{N}^*)^2 \quad |a(k,n)| \leq \frac{1}{k^2}$$
ainsi que
$$\forall k \in \mathbb{N}^* \qquad \lim_{n \to \infty} a(k,n) = \frac{\mu(k)}{k^2}.$$
   (b) Établir que la suite $(p_n)_{n \geq 2}$ converge vers
$$p = \sum_{d=1}^{+\infty} \frac{\mu(d)}{d^2}.$$
6. Calculer p.
Indication : on pourra appliquer le résultat sur les séries absolument convergentes admis en introduction avec $u_n = \frac{\mu(n)}{n^2}$ et $v_n = \frac{1}{n^2}$, puis utiliser II.6.
7. Que pensez vous de l'affirmation suivante :
« La probabilité pour que deux entiers pris au hasard soient premiers entre eux est $6/\pi^2$. » ?

C.2 Preuve de la loi forte des grands nombres

Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires positives de même loi deux à deux indépendantes, admettant un moment d'ordre un. On pose

$$S_n = \sum_{i=1}^{n} X_i \text{ et } Q_n = \frac{1}{n} S_n.$$

On considère également les variables aléatoires tronquées : $X_i^* = X_i 1\!\!1_{X_i \leq i}$ et les sommes et quotients associés : $S_n^* = \sum_{i=1}^n X_i$ et $Q_n^* = S_n^*/n$.

1. Montrer
$$\operatorname{Var} S_n^* \leq \sum_{i=1}^{n} \mathbb{E}(X_i^*)^2.$$
En déduire
$$\operatorname{Var} S_n^* \leq n \mathbb{E} X_1^2 1\!\!1_{X_1 \leq n}.$$

2. Soit $\beta > 1$. On note u_n l'entier le plus proche de β^n. Montrer $u_n \sim \beta^n$. En déduire
$$\sum_{n=N}^{+\infty} \frac{1}{u_n} = O(\frac{1}{u_N}).$$

3. Montrer qu'il existe une constante $C > 0$ telle que
$$\forall N \geq 1 \quad \sum_{n=N}^{+\infty} \frac{1}{u_n} \leq C \frac{1}{u_N}.$$

4. Montrer
$$\sum_{n=1}^{+\infty} \operatorname{Var} Q_{u_n}^* \leq \mathbb{E}\left(X_1^2 \sum_{n=1}^{+\infty} \frac{1}{u_n} 1\!\!1_{X_1 \leq u_n} \right),$$
puis
$$\sum_{n=1}^{+\infty} \operatorname{Var} Q_{u_n}^* \leq \mathbb{E}\left(X_1^2 \sum_{n : u_n \geq X_1} \frac{1}{u_n} \right).$$

5. En déduire que
$$\sum_{n=1}^{+\infty} \operatorname{Var} Q_{u_n}^* < +\infty.$$

6. Montrer que
$$Q_{u_n}^* - \mathbb{E} Q_{u_n}^* \to 0 \text{ p.s.}$$

7. Montrer que
$$\lim_{n\to+\infty} \mathbb{E}Y_n = \mathbb{E}X_1.$$
8. Montrer que $\lim \mathbb{E}Q_n^* = \mathbb{E}X_1$. (Indication : on pourra utiliser le théorème de Césaro)
9. En déduire que
$$Q_{u_n}^* \to \mathbb{E}X_1 \text{ p.s.}$$
10. Montrer que la série de terme général $P(X_n \neq X_n^*)$ converge.
11. A l'aide du lemme de Borel-Cantelli, montrer que pour presque tout ω, il existe un $n_0(\omega)$ tel que les suites $(X_n(\omega))$ et $(X_n^*(\omega))$ coincident à partir du rang $n_0(\omega)$.
12. Montrer que
$$Q_{u_n} \to \mathbb{E}X_1 \text{ p.s.}$$
13. Si $u_n \leq k \leq u_{n+1}$, montrer que
$$\frac{u_n}{u_{n+1}} Q_{u_n} \leq Q_k \leq \frac{u_{n+1}}{u_n} Q_{u_{n+1}}.$$
14. En déduire que
$$\frac{1}{\beta}\mathbb{E}X_1 \leq \underline{\lim}_{k\to+\infty} Q_k \leq \overline{\lim}_{k\to+\infty} Q_k \leq \beta\mathbb{E}X_1 \text{ p.s.}$$
15. On note
$$\Omega_\beta = \{\omega \in \Omega; \frac{1}{\beta}\mathbb{E}X_1 \leq \underline{\lim}_{k\to+\infty} Q_k(\omega) \leq \overline{\lim}_{k\to+\infty} Q_k(\omega) \leq \beta\mathbb{E}X_1\}.$$
Montrer que
$$P(\bigcap_{n=1}^{+\infty} \Omega_{1+1/n}) = 1.$$
16. Montrer que
$$\lim_{k\to+\infty} Q_k = \mathbb{E}X_1 \text{ p.s..}$$
17. Montrer que le résultat de la dernière question demeure vrai si l'on ne suppose plus que les (X_n) sont des variables positives.