Notes du cours “Intégration et Probabilités”

Licence de Mathématiques, ENS de Cachan

Alain Trouvé

11 janvier 2010

Chapitre 1

Espaces mesurés

1.1 Riemann intégrabilité, intégrale de Riemann

- On se place sur un intervalle compact $I = [a, b]$ et on note E l'ensemble des fonctions en escaliers sur I. On rappelle que $g \in E$ si il existe une suite finie croissante $x_0 = a < \cdots < x_n = b$ appelée subdivision[1] de I telle g prend une valeur constante (notée ici g_i) sur chaque intervalle $]x_i, x_{i+1}[$.
- Pour tout $g \in E$, on note

$$I(g) \doteq \sum_{i=0}^{n-1} g_i(x_{i+1} - x_i)$$

(qui ne dépend pas de la subdivision choisie...).

Pour toute fonction f sur I, on définit

$$R^*(f) = \inf_{g \geq f,\ g \in E} I(g),\ R_*(f) = \sup_{g \leq f,\ g \in E} I(g)$$

respectivement l'intégrale par valeur supérieure et celle par valeur inférieure.

Definition 1. Une fonction f sur I est dite Riemann intégrable si f est bornée et $R^*(f) = R_*(f)$. On note alors $R(f)$ la valeur commune où selon la notation plus classique $\int_I f$.

L'intégrale de Riemann est l'intégrale que vous connaissez en entrant à l'ENS bien que la théorie de l'intégration vu en classes préparatoires soit construite sur la classe plus restreinte des fonctions continues par morceaux sur I ou disons sur les fonctions réglées (limite uniforme de fonctions en escaliers). La définition générale de fonction Riemann intégrable n'est elle pas au programme mais c'est en gros la classe la plus grande sur laquelle on peut essayer d'étendre le concept d'intégrale au sens de Riemann. L'intérêt du concept de fonction Riemann intégrable est la possibilité de calculer les intégrales comme limite de ... sommes de Riemann c'est à dire en approximant l'intégrale par la sommation des aires algébriques de petites tranches rectangulaires *verticales* :

Théorème 1. *Si f est une fonction Riemann intégrable sur I, alors pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que pour toute subdivision $(x_i)_{0 \leq i \leq n}$ de I de pas inférieur à δ on a*

$$|\int_I f - \sum_{i=0}^{n-1} f(\xi_i)(x_{i+1} - x_i)| \leq \epsilon$$

lorsque $\xi_i \in [x_i, x_{i+1}]$ pour tout $0 \leq i < n$.

[1]On appellera pas de la subdivision la valeur maximale des $x_{i+1} - x_i$

Démonstration. Pour tout $\epsilon > 0$, il existe deux fonctions en escaliers $g \leq f \leq h$ et une subdivision $(y_j)_{0\leq j\leq q}$ subordonnée à g et h telle que $|I(h) - I(f)| \leq \epsilon$. Notons $\delta_0 = \min_{0\leq j<q}(y_{j+1} - y_j)$ et $M = \sup_I |f|$ et considérons $0 < \delta < \delta_0$. On vérifie facilement qu'il existe un entier N (on peut prendre par exemple $N = 2q$) tel que pour toute subdivision $(x_i)_{0\leq i<n}$ de pas au plus δ, il existe au plus N intervalles $[x_i, x_{i+1}]$ qui contiennent l'un des points y_j. Par suite comme pour $[x_i, x_{i+1}] \subset]y_j, y_{j+1}[$ on a $g_j \leq f(\xi_i) \leq h_j$ et que si $[x_i, x_{i+1}]$ est "à cheval" sur l'un des y_j on peut majorer $f(\xi_i)(x_{i+1} - x_i)$ par $M\delta$ on a :

$$|\int_I f - \sum_{i=0}^{n-1} f(\xi_i)(x_{i+1} - x_i)| \leq |I(h) - I(g)| + N\delta M\,.$$

En choisissant $\delta \leq \epsilon/NM$ on obtient une majoration par 2ϵ. □

A priori, par mal de fonctions sur I sont Riemann intégrables avec cependant des exceptions notables :

Exercice 1. *1. Vérifier que si* f *est continue ou monotone sur* I *alors* f *est Riemann intégrable sur* I.

2. Vérifier que $f : I \to \mathbb{R}$ *définie par* $f(x) = 1$ *si* $x \in \mathbb{Q}$ *et* 0 *sinon n'est pas Riemann intégrable*

Quel est donc l'ensemble des fonctions Riemann intégrables ? Pour les caractériser il nous faut parler d'ensemble négligeable (dont nous verrons une défintion plus générale plus loin) :

Definition 2. Soit $A \subset \mathbb{R}$. On dit que A est négligeable si pour tout $\epsilon > 0$, il existe une suite d'intervalle $(I_i)_{i\geq 0}$ telle $A \subset \cup_{i\geq 0} I_i$ et $\sum_{i\geq 0} \ell(I_i) \leq \epsilon$ où $\ell(I_i)$ est la longueur de l'intervalle I_i.

Une caractérisation frappante des fonctions Riemann intégrables est donnée par le théorème suivant dont la preuve sera proposée en exercice :

Théorème 2. *Une fonction* f *est Riemann intégrable sur* I *ssi l'ensemble de ses points de discontinuité est un ensemble négligeable.*

Le principal problème de la notion de Riemann intégrabilité est qu'elle *n'est pas stable par limite simple* même en se restreignant aux fonctions à valeurs dans un intervalle compact fixé (afin de gérer les problèmes de bornitudes) comme on le vérifie facilement (cf ex1). De plus, certains théorèmes qui ne concernent que des fonctions continues sont très compliqués à montrer. Il suffit pour s'en persuader d'essayer de montrer à *mains nues* que si (f_n) est une famille de fonctions continues sur I, uniformement bornées par M et qui tendent simplement vers 0 alors $\int_I f_n \to 0$. Pour régler ce problème, il faut pouvoir calculer l'intégrale de fonctions plus irrégulières et changer de procédé d'intégration en intégrant par *tranches rectangulaires horizontales* dont la base peut être alors des ensembles plus complexes. Bref il faut changer de théorie et introduire la notion de mesure et de fonctions mesurables.

1.2 Tribus, σ-algèbres

Definition 3 (Tribu)**.** Soit E un ensemble. On appelle tribu sur E tout ensemble $\mathcal{A}$ de parties de E tel que

1. $E \in \mathcal{A}$
2. si $A \in \mathcal{A}$ alors $A^c \in \mathcal{A}$
3. Si $(A_n)_{n\in\mathbb{N}}$ est une famille d'éléments de $\mathcal{A}$, alors $\cup_{n\geq 0} A_n \in \mathcal{A}$ (stabilité par union dénombrable)

Le couple $(E, \mathcal{A})$ est alors appelé espace mesurable.

Remarque 1. *1. On dit aussi de façon équivalente que* $\mathcal{A}$ *est une* σ*-algèbre.*

2. Comme $\emptyset \in \mathcal{A}$*, (3) contient la stabilité par union finie.*

3. *Si au lieu de 3, on a seulement la stabilité par union finie, on dit que $\mathcal{A}$ est une* algèbre.

Exemple 1. – $\mathcal{P}(E)$ *est une tribu.*
– $\{A \subset E \mid A$ *ou* A^c *est au plus den.* $\}$ *est une tribu.*

Proposition 1. *Si* $(\mathcal{A}_i)_{i\in I}$ *est une famille de tribus de* E *alors* $\cap_{i\in I}\mathcal{A}_i$ *est une tribu.*

Démonstration. en exercice. □

Corollaire 1. *Soit* $\mathcal{C} \subset \mathcal{P}(E)$. *Il existe une unique tribu contenant* $\mathcal{C}$ *et qui est la plus petite tribu contenant* $\mathcal{C}$. *On la note* $\sigma(\mathcal{C})$ *et on l'appelle la* tribu engendrée *par* $\mathcal{C}$.

Démonstration. On construit $\sigma(\mathcal{C})$ comme l'intersection de toutes les tribus contenant $\mathcal{C}$. On note qu'il en existe au moins une qui est $\mathcal{P}(E)$. □

Definition 4 (Tribu borélienne)**.** Soit E un espace topologique et $\mathcal{B}(E)$ la tribu engendrée par les ouverts. $\mathcal{B}(E)$ est appelé tribu borélienne et les éléments de $\mathcal{B}(E)$ sont appelés boréliens.

Remarque 2. $\mathcal{B}(E)$ *est gros mais généralement on a* $\mathcal{B}(E) \neq \mathcal{P}(E)$. *On montrera qu'en utilisant l'axiome du choix, on peut construire un ensemble non mesurable de* $\mathbb{R}$.

Exercice 2.
1. *Vérifier que* $\mathcal{B}(\mathbb{R}) = \sigma(\{]a,b[\mid a < b \in \mathbb{R}\}) = \sigma(\{]-\infty,a] \mid a \in \mathbb{R}\}) = \sigma(\{]-\infty,a] \mid a \in \mathbb{Q}\})$*, etc, etc.*
2. *Si* $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty,+\infty\}$ *est muni de la topologie engendrée par la famille d'intervalles* $[-\infty,a[$, $]a,b[$, $]b,+\infty]$ *pour tous* $a,b \in \mathbb{R}$, *montrer que* $\mathcal{B}(\overline{\mathbb{R}}) = \sigma(\{[-\infty,a[\mid a \in \mathbb{R}\}) = \sigma(\{[-\infty,a[\mid a \in \mathbb{Q}\}) = \sigma(\{[-\infty,a] \mid a \in \mathbb{Q}\})$ *(etc, etc)..*

Definition 5 (Tribu trace)**.** Soit $(E,\mathcal{A})$ un espace mesurable et $B \subset E$. On vérifie facilement que $\mathcal{A}_B \doteq \{A \cap B \mid A \in \mathcal{A}\}$ est une tribu sur B appelée *tribu trace* de $\mathcal{A}$ sur B ou *tribu induite* par $\mathcal{A}$ sur B.

Exercice 3.
1. *Montrer que si* $\mathcal{A} = \sigma(\mathcal{C})$ *et* $B \subset E$, *alors la tribu trace* $\mathcal{A}_B$ *de* $\mathcal{A}$ *sur* B *est la tribu engendrée dans* B *par la trace* $\mathcal{C}_B \doteq \{C \cap B \mid C \in \mathcal{C}\}$ *de* $\mathcal{C}$ *sur* B *que l'on peut noter* $\sigma_B(\mathcal{C}_B)$. *On a donc* $\sigma(\mathcal{C})_B = \sigma_B(\mathcal{C}_B)$.
2. *Déduire que* $\mathcal{B}(\mathbb{R})$ *est la tribu induite sur* $\mathbb{R}$ *par* $\mathcal{B}(\overline{\mathbb{R}})$.

Definition 6 (Tribu produit)**.** Soient $(E_1,\mathcal{A}_1)$ et $(E_2,\mathcal{A}_2)$ 2 espaces mesurables. On appelle tribu produit la tribu sur $E_1 \times E_2$, notée $\mathcal{A}_1 \otimes \mathcal{A}_2$ et définie par

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma\{A_1 \times A_2 \mid A_1 \in \mathcal{A}_1 \text{ et } A_2 \in \mathcal{A}_2\}$$

1.3 Mesures positives

Soit $(E,\mathcal{A})$ un espace mesurable.

Definition 7 (Mesure)**.** On appelle mesure positive sur $(E,\mathcal{A})$ toute fonction $\mu : \mathcal{A} \to [0,+\infty]$ telle que

1. $\mu(\emptyset) = 0$
2. Si $(A_n)_{n\in\mathbb{N}}$ est une famille de mesurables *disjoints* alors

$$\mu(\bigcup_{n\geq 0} A_n) = \sum_{n\geq 0} \mu(A_n)\,.\ (\sigma\text{-additivité})$$

Remarque 3. *Le point (2) contient l'additivité sur les unions finies.*

Proposition 2. *Soit* μ *une mesure positive sur* $(E, \mathcal{A})$. *Alors*

1. *Si* $A \subset B$ *sont deux mesurables, on a* $\mu(A) \leq \mu(B)$.
2. *Si* $(A_n)_{n\in\mathbb{N}}$ *est une suite croissante de mesurable, alors*

$$\mu(\bigcup_{n\geq 0} A_n) = \lim_{n\to\infty} \uparrow \mu(A_n)\,.$$

On note évidemment que les $\mu(A_n)$ *sont croissants.*

3. *Si* $(A_n)_{n\in\mathbb{N}}$ *est une suite décroissante de mesurables telle que* $\mu(A_0) < \infty$, *alors*

$$\mu(\bigcap_{n\geq 0} A_n) = \lim_{n\to\infty} \downarrow \mu(A_n)\,.$$

4. *Si* $(A_n)_{n\in\mathbb{N}}$ *est une suite quelconque de mesurables, alors*

$$\mu(\bigcup_{n\geq 0} A_n) \leq \sum_{n\geq 0} \mu(A_n) \text{ (σ-sous-additivité)}$$

Remarque 4. *(3) n'est pas vraie si on enlève la condition* $\mu(A_0) < \infty$ *: il suffit de considérer la mesure de Lebesgue et la famille* $([n, +\infty[)_{n\geq 0}$

Démonstration.

1. Si $A \subset B$, en écrivant $B = A \cup (B \setminus A)$, on a par additivité $\mu(B) = \mu(A) + \mu(B \setminus A)$.
2. Posons $B_0 = A_0$ et $B_n = A_n \setminus A_{n-1}$ pour $n \geq 1$. On a alors $\cup_{n\geq 0} B_n = \cup_{n\geq 0} A_n$ et comme la famille des (B_n) est une famille disjointe, on a par σ-aditivité

$$\mu(\cup_{n\geq 0} A_n) = \sum_{n\geq 0} \mu(B_n) = \lim_{n\to\infty} \sum_{k=0}^{n} \mu(B_k) = \lim_{n\to\infty} \mu(A_n)$$

3. Posons $\tilde{A}_n = A_0 \backslash A_n$. On a alors que $(\tilde{A}_n)_{n\geq 0}$ est une famille croissante pour laquelle $\cup_{n\geq 0} \tilde{A}_n = A_0 \setminus \cap_{n\geq 0} A_n$. D'après (2), on a $\mu(A_0 \setminus \cap_{n\geq 0} A_n) = \lim \mu(A_0 \setminus A_n)$. Or $\mu(A_0 \setminus \cap_{n\geq 0} A_n) = \mu(A_0) - \mu(\cap_{n\geq 0} A_n)$ et $\lim \mu(A_0 \setminus A_n) = \mu(A_0) - \lim \mu(A_n)$. Comme $\mu(A_0) < \infty$, on ne travaille que sur des quantités finies, et l'on déduit le résultat.
4. Le dernier point est laissé en exercice.

□

Exemple 2.

- *(Masse de Dirac) Soit* $x \in E$ *et* $\delta_x : \mathcal{A} \to [0, +\infty]$ *tel que* $\delta_x(A) = 1$ *si* $x \in A$ *et* 0 *sinon.* δ_x *est appelée* masse de Dirac *en* x.
- *(Mesure de comptage)* $\mu(A) = |A|$ *(cardinal de* A*) definie sur* $\mathcal{P}(E)$.
- *(Mesure de Lebesgue) En anticipant sur le cours à venir, on appelle mesure de Lebesgue, l'unique mesure positive définie sur les boréliens de* $\mathbb{R}$ *telle que* $\lambda(]a, b[) = b - a$ *pour tous* $a < b \in \mathbb{R}$ *(la partie difficile est l'*existence *de cette mesure).*

Definition 8. Soit μ une mesure sur $(E, \mathcal{A})$. On dit que

- μ est *finie* si $\mu(E) < \infty$
- μ est une *mesure de probabilité* si $\mu(E) = 1$.
- μ est σ*-finie* si il existe une suite croissante $(A_n)_{n\in\mathbb{N}}$ de mesurables telle que $E = \cup_{n\geq 0} A_n$ et $\mu(A_n) < \infty$ pour tout $n \geq 0$.

1.4 Fonctions mesurables

Definition 9 (Fonctions mesurables)**.** Soient $(E, \mathcal{A})$ et $(F, \mathcal{B})$ deux espaces mesurables. On dit que $f : E \to F$ est *mesurable* (sous-entendu pour les tribus respectives $\mathcal{A}$ et $\mathcal{B}$ sur les ensembles de départ et d'arrivée) si pour tout $B \in \mathcal{B}$ on a $f^{-1}(B) \in \mathcal{A}$ oú $f^{-1}(B) = \{ x \in E \,|\, f(x) \in B \}$.

Si E et F sont topologiques munis de leur tribu borélienne i.e. $\mathcal{A} = \mathcal{B}(E)$ et $\mathcal{B} = \mathcal{B}(F)$, alors on dit que f est *borelienne*.

Proposition 3. *1. Si $f : E \to F$ et $g : F \to G$ sont mesurables alors $g \circ f$ l'est.*

2. Si $\mathcal{B} = \sigma(\mathcal{C})$ et $f : E \to F$, alors f est mesurable de $(E, \mathcal{A})$ dans $(F, \mathcal{B})$ si $f^{-1}(C) \in \mathcal{A}$ pour tout $C \in \mathcal{C}$.

Démonstration. 1. exo.

2. *La façon de procéder ici est typique de beaucoup de démonstrations en intégration.* Si $\mathcal{M} = \{ B \in \mathcal{B} \mid f^{-1}(B) \in \mathcal{A}\}$, on vérifie que $\mathcal{M}$ est une tribu qui contient $\mathcal{C}$ (le faire). Elle contient donc la plus petite tribu contenant $\mathcal{C}$ c'est-à-dire $\sigma(\mathcal{C}) = \mathcal{B}$.

$\square$

Corollaire 2. *Soient E et F deux espaces topologiques. Si $f : E \to F$ est continue, alors f est borelienne.*

Démonstration. soit $\mathcal{T}$ l'ensemble des ouverts de F. Pour tout $U \in \mathcal{T}$, $f^{-1}(U)$ est un ouvert, donc un borélien de E. Par suite, comme $\mathcal{B}(F) = \sigma(\mathcal{T})$ la proposition donne le résultat. $\square$

Definition 10. Soient E un ensemble, $(F_i, \mathcal{B}_i)_{i \in I}$ une famille d'espaces mesurables et pour tout $i \in I$, une fonction $f_i : E \to F_i$. On note $\sigma(f_i, i \in I)$ la plus petite tribu sur E qui rend toutes les fonctions f_i mesurables.

Exercice 4. *Soit I un ensemble fini. Si $E = \prod_{i \in I} F_i$ et $p_i : E \to F_i$ est la projection canonique sur F_i (i.e. $p_j((x_i)_{i \in I}) = x_j$) alors*

$$\sigma(p_i, i \in I) = \otimes_{i \in I} \mathcal{B}_i \,.$$

Dans le cas où I est quelconque, on peut prendre l'egalité précédente comme la définition d'un produit tensoriel quelconque de tribus.

Exercice 5. *Montrer que $\mathcal{B}(\mathbb{R}^2)$ n'est rien d'autre que $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$*

Proposition 4. *1. Soient $f_1 : (E, \mathcal{A}) \to (F_1, \mathcal{B}_1)$ et $f_2 : (E, \mathcal{A}) \to (F_1, \mathcal{B}_2)$ deux fonctions mesurables. Alors $h : E \to F_1 \times F_2$ définie par $h(x) = (f_1(x), f_2(x))$ est mesurable pour la tribu produit $\mathcal{B}_1 \otimes \mathcal{B}_2$ sur l'espace d'arrivée.*

2. Si $f_1, f_2 : (E, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ sont mesurables alors $f_1 + f_2$, $f_1 f_2$, $f_1 \vee f_2$, $f_1 \wedge f_2$, f_1^+, f_1^- sont mesurables.

3. Si pour tout $n \geq 0$, $f_n : (E, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})$ sont mesurables, alors $\inf_{n \geq 0} f_n$, $\sup_{n \geq 0} f_n$, $\underline{\lim} f_n$, $\overline{\lim} f_n$ sont mesurables dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$.

Démonstration. 1. Comme $\mathcal{B}_1 \otimes \mathcal{B}_2 = \sigma(\mathcal{B}_1 \times \mathcal{B}_2)$, il suffit de vérifier d'après la proposition 3 (2) que $h^{-1}(B_1 \times B_2) \in \mathcal{A}$ pour tous $(B_1, B_2) \in \mathcal{B}_1 \times \mathcal{B}_2$. Or

$$h^{-1}(B_1 \times B_2) = f_1^{-1}(B_1) \cap f_2^{-1}(B_2)$$

et $\mathcal{A}$ est stable par intersections dénombrable.

2. On considérant h comme précédemment, puis $g : \mathbb{R}^2 \to \mathbb{R}$ définie par $g(x, y) = x + y$ alors $f_1 + f_2 = g \circ h$. Or g est continue, donc borélienne i.e. mesurable pour la tribu $\mathcal{B}(\mathbb{R}^2)$ sur $\mathbb{R}^2$ et $\mathcal{B}(\mathbb{R})$ sur $\mathbb{R}$. Comme $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ (cf exo), on déduit de la proposition 3 (1) que $f_1 + f_2$ est mesurable. On continue de la même façon pour le produit $f_1 f_2$, etc.

3. Soit $g = \inf f_n$. On a pour tout $a \in \mathbb{R}$, $g^{-1}([a,\infty]) = \cap_{n\geq 0} f_n^{-1}([a,\infty])$. Comme $\mathcal{B}(\overline{\mathbb{R}}) = \sigma(\{ [a,+\infty] \mid a \in \mathbb{R} \}$, on déduit de la proposition 3 (2) que g est mesurable. De même $\sup_{n\geq 0} f_n$ est mesurable. Comme $\liminf_{n\to\infty} f_n = \sup_{n\geq 0} \inf_{k\geq n} f_k$ on déduit que $\liminf f_n$ est mesurable.

$\square$

Remarque 5. *Si* $(f_n)_{n\geq 0}$ *converge simplement vers* $f : (E,\mathcal{A}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$*, alors* f *est mesurable.*

Exercice 6. *Si* $(f_n)_{n\in\mathbb{N}}$ *est une suite de fonctions mesurables de* $(E,\mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *alors* $\{ x \in E \mid \lim_{n\geq 0} f_n(x)$ *existe* $\}$ *est un ensemble mesurable.*

Definition 11 (Mesure image)**.** Soient $(E,\mathcal{A})$ et $(F,\mathcal{B})$ deux ensembles mesurables. Soit $f : E \to F$ mesurable et μ une mesure sur $(E,\mathcal{A})$. On appelle *mesure image* de μ par f notée $f(\mu)$ ou encore $\mu \circ f^{-1}$, la mesure ν sur $(F,\mathcal{B})$ définie par $\nu(B) = \mu(f^{-1}(B))$ pour tout $B \in \mathcal{B}$.

1.5 Théorème de Dynkin

Definition 12 (λ-système, π-système)**.** Soit E un ensemble.

1. On appelle π-*système* sur E toute famille $\mathcal{C}$ de partie de E qui est stable par intersection finie.
2. On appelle λ-*système* sur E toute famille Λ de sous-ensembles de E telle que
   (a) $E \in \Lambda$
   (b) Si $A \in \Lambda$ et $B \in \Lambda$ avec $A \subset B$ alors $B \setminus A \in \Lambda$.
   (c) Si $(A_n)_{n\in\mathbb{N}}$ est une suite croissante d'éléments de Λ (i.e. $A_n \subset A_{n+1}$ pour tout $n \geq 0$) alors $\cup_{n\geq 0} A_n \in \Lambda$

Remarque 6. – *Si $\mathcal{A}$ est une tribu, alors $\mathcal{A}$ est un λ-système.*
- *Par contre la réciproque est fausse car on a seulement la stabilité par union ou intersection denombrable* croissante. *Pourtant, la σ-algèbre n'est pas si loin car il suffit de montrer la stabilité par intersection finie (ou union finie).*
- *Une intersection quelconque de λ-systèmes sur* E *est un λ-système sur* E*. Par suite on peut parler de* λ-système engendré *par une famille de parties.*

Théorème 3 (Dynkin)**.** *Si* $\mathcal{C}$ *est un π-système sur* E *alors* $\Lambda(\mathcal{C}) = \sigma(\mathcal{C})$.

Démonstration. Comme $\sigma(\mathcal{C})$ est un λ-système, on a $\Lambda(\mathcal{C}) \subset \sigma(\mathcal{C})$.

Montrons $\sigma(\mathcal{C}) \subset \Lambda(\mathcal{C})$. Pour ce, il suffit de montrer que $\Lambda(\mathcal{C})$ est une tribu et donc (cf remarque) que $\Lambda(\mathcal{C})$ est stable par intersection finie.

Soit $C \in \Lambda(\mathcal{C})$ et $\Lambda_C = \{ A \in \Lambda(\mathcal{C}) \mid A \cap C \in \Lambda(\mathcal{C}) \}$. Vérifions que Λ_C est un λ-système :

1. $E \in \Lambda_C$ car $E \cap C = C \in \Lambda(\mathcal{C})$
2. Si $A \subset B \in \Lambda_C$, alors $(B \setminus A) \cap C = (B \cap C) \setminus (A \cap C)$. Comme par hypothèse, $B \cap C$ et $A \cap C$ sont dans $\Lambda(\mathcal{C})$, $(B \cap C) \setminus (A \cap C) \in \Lambda(\mathcal{C})$ d'après le point (2) de la définition d'un λ-système.
3. Si $(A_n)_{n\in\mathbb{N}}$ est une suite croissante d'éléments de Λ_C, alors $(\cup_{n\geq 0} A_n) \cap C = \cup_{n\geq 0}(A_n \cap C)$ et $(A_n \cap C)_{n\geq 0}$ est une suite croissante d'éléments de $\Lambda(\mathcal{C})$. C'est donc (point 3) un élément de $\Lambda(\mathcal{C})$.

Par suite si $C \in \mathcal{C}$, on deduit que $\mathcal{C} \subset \Lambda_C$ et donc $\Lambda(\mathcal{C}) \subset \Lambda_C \subset \Lambda(\mathcal{C})$ i.e. $\Lambda(\mathcal{C}) = \Lambda_C$. On déduit en particulier que pour tout $B \in \Lambda(\mathcal{C})$ et $C \in \mathcal{C}$, $B \cap C \in \Lambda(\mathcal{C})$. C'est-à-dire que $\mathcal{C} \subset \Lambda_B$ pour tout $B \in \Lambda(\mathcal{C})$. Comme Λ_B est un λ-système, on déduit $\Lambda_B = \Lambda(\mathcal{C})$. $\square$

Corollaire 3 (Unicité de mesures)**.** *Soient μ et ν deux mesures sur $(E,\mathcal{A})$ qui coïncident sur un π-système $\mathcal{C}$ tel que $\mathcal{A} = \sigma(\mathcal{C})$. Alors*

1. *si* $\mu(E) = \nu(E) < \infty$ *alors* $\mu = \nu$
2. *Si il existe une suite* croissante $(A_n)_{n\in\mathbb{N}}$ *d'éléments de* $\mathcal{C}$ *telle que* $\cup_{n\geq 0} A_n = E$ *et* $\mu(A_n) = \nu(A_n) < \infty$ *alors* $\mu = \nu$.

Démonstration. 1. Si $\Lambda = \{ A \in \mathcal{A} \mid \mu(A) = \nu(A) \}$ alors on montre que Λ est un λ-système qui contient $\mathcal{C}$ et donc par le théorème de Dynkin, $\Lambda = \sigma(\mathcal{C}) = \mathcal{A}$. En effet,
   - $E \in \Lambda$ car $\mu(E) = \nu(E) < \infty$.
   - De plus si $A \subset B \in \Lambda$, alors $\mu(B \setminus A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B \setminus A)$ (noter que la première et la dernière égalité n'a de sens que parce que $\mu(A) < \infty$ et $\nu(A) < \infty$ puisque les deux mesures sont finies).
   - Si $(A_n)_{n\in\mathbb{N}}$ est une suite croissante d'éléments de Λ, alors $\mu(\cup_{n\geq 0} A_n) = \lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \nu(A_n) = \nu(\cup_{n\geq 0} A_n)$

2. En considérant pour tout $n \geq 0$ la tribu trace $\mathcal{A}_n$ de $\mathcal{A}$ sur A_n, on a que μ et ν coïncident sur $\mathcal{A}_n$: il suffit de remarquer que si $\mathcal{C}_n = \{ C \cap A_n \mid C \in \mathcal{C}\}$ est la trace de $\mathcal{C}$ sur A_n, alors $\mathcal{C}_n$ est un π-système tel que $\mathcal{A}_n = \sigma(\mathcal{C}_n)$ et appliquer le point (1). Par suite, pour tout $A \in \mathcal{A}$, on a $\mu(A) = \lim_{n\to\infty} \mu(A \cap A_n) = \lim_{n\to\infty} \nu(A \cap A_n) = \nu(A)$.

□

Chapitre 2

Intégration par rapport à une mesure

2.1 Intégration par rapport à une mesure

Henri Léon Lebesgue

11. *La définition de l'intégrale.* — Il est donc pratiquement suffisant de se poser le problème de la mesure pour les seuls ensembles mesurables et le problème d'intégration pour les seules *fonctions mesurables*, c'est-à-dire celles pour lesquelles, quels que soient a et b, l'ensemble des points en lesquels on a $a < f(P) < b$ est mesurable. Par l'emploi des opérations I et II on reconnait d'ailleurs de suite que, pour ces fonctions, on peut remplacer si l'on veut un signe $<$ par le signe $=$ ou $\leq$ sans que l'ensemble défini cesse d'être mesurable ; on voit de même que l'ensemble formé par les points en lesquels une fonction mesurable est définie est nécessairement mesurable.

Ceci posé, soit f une fonction mesurable bornée définie dans un ensemble mesurable borné E. Soient m et M les bornes inférieure et supérieure de f ; soient $a_1, a_2, \ldots, a_{n-1}$ des nombres tels qu'on ait

$$m = a_0 < a_1 < a_2 < \ldots < a_{n-1} < a_n = M = a_{n+1}$$

et que $(a_i - a_{i-1})$ ne surpasse jamais ε arbitrairement choisi ou positif. Soit $\mathcal{E}_i$ l'ensemble des points P en lesquels on a

$$a_i \leq f(P) < a_{i+1} \qquad (i = 0, 1, \ldots, n-1);$$

$\mathcal{E}_i$ est mesurable ; $\mathcal{E}_n$ sera l'ensemble mesurable des points ou $f(P) = a_n$.

Des conditions du problème d'intégration il résulte qu'on a

$$\sum_{i=0}^{i=n} a_i m(\mathcal{E}_i) \leq \int_E f(P)\, dP \leq \sum_{i=0}^{i=n} a_{i+1} m(\mathcal{E}_i),$$

"Sur l'intégration des fonctions discontinues", Annales Scientifiques de l'É.N.S., 1910.

L'idée centrale de l'intégration de Lebesgue est d'intégrer les fonctions par *tranches*. Soit $(E, \mathcal{A})$ un espace mesuré.

Definition 13 (Fonctions étagées)**.**

1. On appelle *fonction étagée* toute fonction mesurable $f : (E, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ne prenant qu'un nombre fini de valeurs. On note $\mathcal{E}$ l'ensemble des fonctions étagées et $\mathcal{E}^+$ les fonctions étagées positives.
2. Pour tout $f \in \mathcal{E}$, f peut s'écrire comme une somme finie $f = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}$ où $(\alpha_i, A_i) \in \mathbb{R} \times \mathcal{A}$ pour tout $1 \leq i \leq n$ et les ensembles $(A_i)_{1 \leq i \leq n}$ sont disjoints. Si $f \geq 0$, on appelle *intégrale de* f par rapport à μ, notée $\mu(f)$ ou $\int f d\mu$:

$$\int f d\mu = \sum_{i=1}^n \alpha_i \mu(A_i)$$

avec la convention $0 \times \infty = \infty \times 0 = 0$.

Remarque 7. *On dit parfois que $\mathcal{E}$ est l'espace des* fonctions simples.

Remarque 8. *Si* $f = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i} = \sum_{j=1}^m \beta_j \mathbb{1}_{B_j}$ *correspond à deux décompositions de* f*, alors en notant* $A_0 = (\cup_{i=1}^n A_i)^c$, $B_0 = (\cup_{j=1}^m B_j)^c$, $\alpha_0 = 0$ *et* $\beta_0 = 0$*, on a*

$$\begin{aligned} \sum_{i=1}^n \alpha_i \mu(A_i) &= \sum_{i=0}^n \sum_{j=0}^m \alpha_i \mu(A_i \cap B_j) && (2.1)\\ &= \sum_{i=0}^n \sum_{j=0}^m \beta_j \mu(A_i \cap B_j) && (2.2)\\ &= \sum_{j=1}^n \beta_j \mu(B_j) && (2.3) \end{aligned}$$

car $\{B_0, B_1, \cdots, B_m\}$ *est une partition de* E*. Comme* $\alpha_i = \beta_j$ *si* $A_i \cap B_j \neq \emptyset$*, on déduit que l'intégrale est bien définie.*

Definition 14. Soit $f : (E, \mathcal{A}) \to [0, +\infty]$ mesurable. Alors on définit

$$\int f d\mu = \sup_{h \in \mathcal{E}^+, h \leq f} \int h d\mu\,.$$

Remarque 9.
- *On introduira parfois une variable muette d'intégration et on écrira* $\int f(x) d\mu(x)$ *ou encore* $\int f(x) \mu(dx)$.
- *Si* $f \in \mathcal{E}^+$*, la nouvelle définition coïncide avec la précédente.*
- *Si* $0 \leq f \leq g$ *alors* $\int f d\mu \leq \int g d\mu$

Théorème 4 (Convergence monotone)**.** *Soit* $(f_n)_{n \geq 0}$ *une suite de fonctions mesurables de* E *dans* $[0, +\infty]$ *telle* $f_n \leq f_{n+1}$ *pour tout* $n \geq 0$ *(monotonie). Alors si* $f = \lim_{n\to\infty} \uparrow f_n$ *est la limite simple (éventuellement infinie) de la suite, on a*

$$\int f d\mu = \lim \uparrow \int f_n d\mu$$

Remarque 10. *Il s'agit de notre premier théorème d'intervertion limite-intégrale :*

$$\int \lim\uparrow f_n d\mu = \lim \uparrow \int f_n d\mu$$

Démonstration. Comme $f_n \leq f$, on a $\int f_n d\mu \leq \int f d\mu$ et donc $\lim \int f_n d\mu \leq \int f d\mu$.

Montrons l'autre inégalité. Il nous sufit de montrer que pour tout $h \in \mathcal{E}^+$ tel que $h \leq f$ alors $\int h d\mu \leq \lim \int f_n d\mu$. Décomposons h sur une famille finie de partie disjointes

$$h = \sum_{i=1}^m \alpha_i \mathbb{1}_{A_i}$$

et prenons $\epsilon > 0$ tel que $\alpha_i - \epsilon > 0$ pour tout i. Si $A_{i,n} = \{ x \in A_i \mid f_n(x) \geq \alpha_i - \epsilon\}$, alors $f_n \geq \sum_{i=1}^m (\alpha_i - \epsilon) \mathbb{1}_{A_{i,n}}$ et donc pour tout $k \geq 0$

$$\lim \int f_n d\mu \geq \int f_k d\mu \geq \sum_{i=1}^m (\alpha_i - \epsilon) \mu(A_{i,k}).$$

Or $(A_{i,k})_{k \geq 0}$ est une suite croisante de mesurables telle que $\cup_{k \geq 0} A_{i,k} = A_i$. Par suite, $\mu(A_i) = \lim \uparrow \mu(A_{i,k})$ et $\lim \int f_n d\mu \geq \sum_{i=1}^m (\alpha_i - \epsilon) \mu(A_i)$. Comme $\epsilon > 0$ est arbitraire, on obtient le résultat. □

Exercice 7. *Soit f mesurable positive, montrer qu'il existe une suite croissante de fonction étagées $(f_n)_{n\in\mathbb{N}}$ convergeant simplement vers f et telle que $\int f d\mu = \lim \int f_n d\mu$.*

Proposition 5. *1. Soient f, g mesurables, $\alpha, \beta \geq 0$. Alors*

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

2. Soit $(f_n)_{n\in\mathbb{N}}$ mesurables positives

$$\int (\sum_{n\geq 0} f_n) d\mu = \sum_{i=1}^{n} \int f_n d\mu .$$

Démonstration. 1. On considère deux suites $(f_n)_{n\in\mathbb{N}}$ et $(g_n)_{n\in\mathbb{N}}$ croissantes de fonctions étagées tendant resp. vers f et g. Alors $\alpha f_n + \beta g_n \in \mathcal{E}^+$ pour tout n definit une suite croissante de fonctions étagées tendant vers $\alpha f + \beta g$.
Par convergence monotone, $\int(\alpha f+\beta g)d\mu = \lim \int(\alpha f_n+\beta g_n)d\mu = \lim(\alpha \int f_n d\mu+\beta \int g_n d\mu) = \alpha \int f d\mu + \beta \int g d\mu$.

2. En posant $\tilde{f}_n = \sum_{k=0}^{n} f_k$ on a $\tilde{f}_n \uparrow \sum_{n=0}^{\infty} f_n$. On déduit le résultat par cv montone.

□

Definition 15. 1. Soit $N \subset E$, N est dit μ-*négligeable* si il existe $A \in \mathcal{A}$ tel que $N \subset A$ et $\mu(A) = 0$. On notera $\mathcal{N}$ l'ensemble des négligeagles.

2. Soit $\mathcal{P}$ une propriété dépendant de $x \in E$. On que $\mathcal{P}$ *est vraie presque partout*, noté $\mathcal{P}$ est vraie p.p. si $\{x \in E \mid \mathcal{P}(x)$ est fausse $\}$ est μ-négligeable.

Proposition 6. *Soient f et g positives mesurables*

1. *Si $f = 0$ p.p. alors $\int f d\mu = 0$.*
2. *Si $\mu(\{ x \in E \mid f(x) > 0 \}) > 0$ alors $\int f d\mu > 0$.*
3. *Si $\int f d\mu < \infty$ alors $f < \infty$ p.p.*
4. *Si $f \leq g$ p.p., $\int f d\mu \leq \int g d\mu$.*
5. *Si $f = g$ p.p. alors $\int f d\mu = \int g d\mu$.*

Démonstration. 1. Si $f = 0$ p.p. et $h \in \mathcal{E}^+$ telle que $h \leq f$ avec $h = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}$ et $\alpha_i > 0$ pour tout $1 \leq i \leq n$. Si il existe i tel que $\mu(A_i) > 0$ alors $f > 0$ sur A_i ce qui est absurde. Par suite $\int h d\mu = 0$.

2. Pour tout $n \geq 1$, $A_n = f^{-1}([1/n, +\infty])$ définit une suite croissante d'ensembles telle que $\cup_{n\geq 1} A_n = \{x \in E \mid f(x) > 0\}$. Par suite $\lim \mu(A_n) > 0$ et il existe n_0 tel que $\mu(A_{n_0}) > 0$. En prenant $h = n_0^{-1} \mathbb{1}_{A_{n_0}}$, on déduit que $\int f d\mu > 0$.

3. Si $\mu(\{x \in E \mid f(x) = +\infty\}) = \alpha > 0$ alors pour tout $n \geq 0$, si $A_n = f^{-1}([n, \infty])$, on a $\mu(A_n) \geq \alpha$. On considérant $h = n \mathbb{1}_{A_n}$, on a $h \leq f$ et $\int f d\mu \geq \int h_n d\mu \geq n\alpha$.

4. exo.

5. exo.

□

Théorème 5 (Lemme de Fatou). *Soit $(f_n)_{n\in\mathbb{N}}$ mesurables positives. Alors*

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

Démonstration. Soit $\tilde{f}_n = \inf_{k \geq n} f_k$ et $f = \underline{\lim} f_n$. On a $\tilde{f}_n \uparrow f$ et par convergence monotone,

$$\int f d\mu = \lim \int \tilde{f}_n d\mu .$$

Comme de plus $\tilde{f}_n \leq f_n$, on a $\int \tilde{f}_n d\mu \leq \int f_n d\mu$ et

$$\lim \int \tilde{f}_n d\mu \leq \underline{\lim} \int f_n d\mu .$$

□

2.2 Fonctions intégrables

Definition 16 (fonctions intégrables)**.** Soit $f : (E, \mathcal{A}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ mesurable. On dit que f est *intégrable* si $\int |f| d\mu < \infty$. On définit alors $\int f d\mu$ par

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

On note $\mathcal{L}^1(E, \mathcal{A}, \mu)$ l'ensemble des fonctions intégrables.

Proposition 7. *1. (Monotonie) Soient* f *et* g *intégrables telles que* $f \leq g$ *p.p. alors* $\int f d\mu \leq \int g d\mu$.

2. (Linéarité) Soient f *et* g *intégrables et* $\alpha, \beta \in \mathbb{R}$. *Alors* $\alpha f + \beta g$ *est intégrable et*

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

3. (Continuité) Soit f *intégrable, alors*

$$|\int f d\mu| \leq \int |f| d\mu .$$

Démonstration. 1. Comme $f \leq g$ p.p., on a $f^+ \leq g^+$ p.p. et $f^- \geq g^-$ p.p.. Par suite $\int f^+ d\mu - \int f^- d\mu \leq \int g^+ d\mu - \int g^- d\mu$.

2. $(f+g) = (f+g)^+ - (f+g)^- = (f^+ - f^-) + (g^+ - g^-)$ d'où $(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+$ (les fonctions sont positives de chaque coté). Par linéarité de l'intégrale pour les fonctions positives, on déduit que $\int (f+g)^+ d\mu + \int f^- d\mu + \int g^- d\mu = \int (f+g)^- d\mu + \int f^+ d\mu + \int g^+ d\mu$. En regroupant les termes, on obtient le résultat. On laisse en exercice la preuve de $\int \alpha f d\mu = \alpha \int f d\mu$.

3. Il suffit de remarquer que $-|f| \leq f \leq |f|$ puis utiliser (1) pour intégrer les inégalités et obtenir le résultat.

□

Proposition 8 (Convergence monotone p.p.)**.** *Soit* $(f_n)_{n \in \mathbb{N}}$ *une suite de fonctions mesurables telle que pour tout* $n \geq 0$ *:*

1. $f_n \geq 0$ *p.p.*
2. $f_n \leq f_{n+1}$ *p.p.*

Soit f *mesurable telle* $f = \lim f_n$ *presque partout. Alors*

$$\int f d\mu = \lim \uparrow \int f_n d\mu .$$

Démonstration. Soit N l'union *denombrable* des négligeables sur lequels on a $f_n > f_{n+1}$ ou $f_n < 0$ pour un $n \geq 0$ ou encore $f \neq \lim f_n$. Comme union dénombrable de négligeable, N est négligeable. Soit $A \in \mathcal{A}$ tel que $N \subset A$ et $\mu(A) = 0$. En considérant $\tilde{f}_n = f_n \mathbb{1}_A$ et $\tilde{f} = f \mathbb{1}_A$, on peut appliquer le théoréme de convergence monotone classique qui donne

$$\int \tilde{f} d\mu = \lim \uparrow \int \tilde{f}_n d\mu \,.$$

Comme $f = \tilde{f}$ p.p. on déduit de la proposition 6 (i), que $\int f d\mu = \int \tilde{f} d\mu$. De même, $\int f_n d\mu = \int \tilde{f}_n d\mu$ ce qui donne le résultat. □

Definition 17 (Mesure complétée)**.** Si $(E, \mathcal{A}, \mu)$ est un espace mesuré, et $\mathcal{N}$ est l'ensemble des parties μ-négligeables, alors on appelle *tribu complétée* de $\mathcal{A}$, notée $\overline{\mathcal{A}}$ la tribu :

$$\overline{\mathcal{A}} = \sigma(\mathcal{A} \cup \mathcal{N}) \,.$$

On vérifie (exo) que $\overline{\mathcal{A}} = \{A \cup N \mid A \in \mathcal{A},\ N \in \mathcal{N}\}$. On définit alors la *mesure complétée* de μ, notée $\overline{\mu}$, définie sur $\overline{\mathcal{A}}$ par

$$\overline{\mu}(A \cup N) = \mu(A)$$

pour tout $(A, N) \in \mathcal{A} \times \mathcal{N}$. Cette mesure complétée est une mesure (exo) sur $(E, \overline{\mathcal{A}})$.

Exercice 8.
- *Montrer que si* $\tilde{f} : E \to F$ *est presque partout égale à une fonction mesurable* $f : (E, \mathcal{A}) \to (F, \mathcal{F})$ *alors* $\tilde{f}$ *est mesurable pour la tribu complétée* $\overline{\mathcal{A}}$ *sur* E.
- *Si maintenant* $(F, \mathcal{F}) = (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ *et* $f \in \mathcal{L}^1(E, \mathcal{A}, \mu)$ *alors* $\tilde{f} \in \mathcal{L}^1(E, \overline{\mathcal{A}}, \overline{\mu})$ *et* $\int f d\mu = \int \tilde{f} d\overline{\mu}$.

Théorème 6 (Theoreme de convergence dominée)**.** *Soit* $(f_n)_{n\in\mathbb{N}}$ *une suite de fonctions mesurables qui converge p.p. vers* f *mesurable. Soit* g *intégrable telle que* $|f_n| \leq g$ *p.p. pour tout* $n \geq 0$. *Alors* f *est intégrable et*

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu \,.$$

Remarque 11. *On a donc l'intervertion des limites :* $\int \lim f_n d\mu = \lim \int f_n d\mu$.

Démonstration. En procédant comme dans la proposition 5, on peut éliminer les "presque partout" ce que l'on fait à partir d'ici. On verifie que f est intégrable en utilisant Fatou sur $|f_n|$ qui donne $\int |f| d\mu \leq \underline{\lim} \int |f_n| d\mu \leq \int g d\mu < \infty$. De plus, comme $g - f_n \geq 0$ et $g + f_n \geq 0$, on peut appliquer Fatou qui donne $\int g - f d\mu \leq \underline{\lim} \int g - f_n d\mu \leq \int g d\mu - \overline{\lim} \int f_n d\mu$ et $\int g + f d\mu \leq \underline{\lim} \int g + f_n d\mu \leq \int g d\mu + \underline{\lim} \int f_n d\mu$.

Par suite, on déduit que $\overline{\lim} \int f_n d\mu \leq \int f d\mu \leq \underline{\lim} \int f_n d\mu$ d'où le résultat. □

2.3 Intégrales dépendant d'un paramètre

Soit (T, d) un espace métrique et $(E, \mathcal{A}, \mu)$ un espace mesuré.

Théorème 7 (Continuité)**.** *Soit* $t_0 \in T$ *et* $f : T \times E \to \overline{\mathbb{R}}$ *telle que*

1. *pour tout* $t \in T$, $x \to f(t, x)$ *est mesurable,*
2. $t \to f(t, x)$ *est continue en* t_0, μ*-p.p. en* x,
3. *il existe* $g : E \to \mathbb{R}$ *intégrable telle que pour tout* $t \in T$, $|f(t, x)| \leq g(x)$ μ*-p.p. en* x.

Alors $t \to \int f(t, x) d\mu(x)$ *est continue en* t_0 *:*

$$\lim_{t\to t_0} \int f(t, x) d\mu(x) = \int f(t_0, x) d\mu(x) \,.$$

Démonstration. Soit $(s_n)_{n\in\mathbb{N}}$ une suite tendant vers t_0. Alors en notant $f_n(x) = f(s_n, x)$, pour tout $x \in E$ et $n \geq 0$, $(f_n)_{n\in\mathbb{N}}$ est une suite de fonctions mesurables qui cv p.p. vers $f(t_0, .)$. De plus $|f_n| \leq g$ presque partout. En appliquant le théorème de convergence dominée, on déduit que $\lim \int f_n d\mu = \int f(t_0, x) d\mu(x)$ ce qui donne le résultat. □

Exercice 9. *1. Montrer que la transformée de Fourier* $\hat{\phi} : \mathbb{R} \to \mathbb{R}$ *d'une fonction intégrable pour le mesure de Lebesgue et définie par*

$$\hat{\phi}(u) = \int \exp(iux)\phi(x)dx$$

est continue

2. Si de plus maintenant $f : \mathbb{R} \to \mathbb{R}$ *est continue bornée, alors*

$$f \star \phi(x) = \int f(x-y)\phi(y)dy \textit{ (convolution)}$$

est continue.

Théorème 8 (Intégrale dépendant d'un paramètre)**.** *Soient* I *un intervalle ouvert de* $\mathbb{R}$*,* $t_0 \in I$ *et* $f : I \times E \to \overline{\mathbb{R}}$ *tels que*

1. *pour tout* $t \in I$*,* $f(t, .)$ *est mesurable,*
2. *pour presque tout* x*,* $t \to f(t, x)$ *est dérivale en* t_0 *de dérivée notée* $\frac{\partial f}{\partial t}(t_0, x)$
3. *il existe* g *intégrable telle que pour pour tout* $t \in I$ $|f(t, x) - f(t_0, x)| \leq g(x)|t - t_0|$ *pour presque tout* x*.*

Alors $t \to \int f(t, x)d\mu(x)$ *est dérivable en* t_0 *de dérivée*

$$\int \frac{\partial f}{\partial t}(t_0, x)d\mu(x)$$

Exercice 10. *Soit* $\phi : \mathbb{R} \to \mathbb{R}$ *intégrable.*
- *Donner une condition pour la dérivabilité de* $\hat{\phi}$ *en* 0*.*
- *Montrer que si* f *est* C^∞ *à support compact, alors* $f \star \phi$ *est* C^∞*.*

Démonstration. On procède comme précédemment en considérant une suite $(s_n)_{n\in\mathbb{N}}$ tendant vers t_0 ($s_n \neq t_0$, pour tout n) et en définissant $h_n(x) = \frac{f(s_n,x)-f(t_0,x)}{s_n-t_0}$. On a alors que $h_n \to \frac{\partial f}{\partial t}(t_0, .)$ p.p. et $|h_n| \leq g$ p.p. d'où le résultat par convergence dominée. □

Chapitre 3

Existence de mesures

3.1 Mesures extérieures

Definition 18 (Mesures exterieures)**.** Soit E un ensemble. On dit que $\mu^* : \mathcal{P}(E) \to [0, +\infty]$ est une mesure extérieure sur E si

1. $\mu^*(\emptyset) = 0$;
2. si $A \subset B \subset E$ alors $\mu^*(A) \leq \mu^*(B)$;
3. si $(A_n)_{n\in\mathbb{N}}$ est une suite de sous-ensembles de E alors

$$\mu^*(\bigcup_{n\geq 0} A_n) \leq \sum_{n\geq 0} \mu^*(A_n) \ \text{ (σ-sous-additivité)}$$

Remarque 12. *L'idée derrière l'introduction de mesures extérieures pourrait être celle-ci : si* $\rho : \mathcal{C} \to [0, \infty]$ *donne la mesure de parties "simples"* $C \in \mathcal{C}$*, on peut facilement définir une mesure extérieure* ρ^* *par*

$$\rho^*(A) = \inf\{\sum_{n\geq 0} \rho(A_n) \mid (A_n)_{n\in\mathbb{N}} \in \mathcal{C}^{\mathbb{N}},\ A \subset \bigcup_{n\geq 0} A_n\}.$$

Definition 19 (Ensembles μ^*-mesurables)**.** Soit μ^* une mesure extérieure sur E. On dit que $A \subset E$ est μ^*-mesurable si pour tout $E' \subset E$,

$$\mu^*(E') = \mu^*(A \cap E') + \mu^*(A^c \cap E')\,. \tag{3.1}$$

On note $\mathcal{M}(\mu^*)$ l'ensemble des parties qui sont μ^*-mesurables.

Remarque 13. *Pour montrer l'égalité dans* (5.2)*, il suffit d'après la sous additivité de montrer que* $\mu^*(E') \geq \mu^*(A \cap E') + \mu^*(A^c \cap E')$.

Théorème 9. *Soit* μ^* *une mesure extérieure. Alors* $\mathcal{M}(\mu^*)$ *est une tribu et la restriction de* μ^* *à* $\mathcal{M}(\mu^*)$ *est une mesure.*

Démonstration

1. Montrons d'abord que $\mathcal{M}(\mu^*)$ est stable par union finie.

Il suffit de vérifier que si $A, B \in \mathcal{M}(\mu^*)$, alors $A \cup B \in \mathcal{M}(\mu^*)$. Or $\mu^*(E') = \mu^*(E' \cap A) + \mu^*(E' \cap A^c) = \mu * (E' \cap A) + \mu^*(E' \cap A^c \cap B) + \mu^*(E' \cap A^c \cap B^c)$. De plus $\mu^*(E' \cap A) + \mu^*(E' \cap A^c \cap B) \geq \mu^*(E' \cap (A \cup B))$ et $\mu^*(E' \cap A^c \cap B^c) = \mu^*(E' \cap (A \cup B)^c)$ d'où

$$\mu^*(E') \geq \mu^*(E' \cap (A \cup B)) + \mu^*(E' \cap (A \cup B)^c)$$

ce qui prouve le résultat annoncé.

2. Comme $\emptyset \in \mathcal{M}(\mu^*)$ (le vérifier) et que par symétrie de la définition, $\mathcal{M}(\mu^*)$ est stable par passage au complémentaire, il suffit, pour montrer que $\mathcal{M}(\mu^*)$ est une tribu, de montrer que si $(A_n)_{n\in\mathbb{N}}$ est une suite d'éléments disjoints de $\mathcal{M}(\mu^*)$, alors $\cup_{n\geq 0} A_n \in \mathcal{M}(\mu^*)$.

Démonstration (Suite)

3. Pour cela, on va vérifier par récurrence sur $n \geq 0$ que si $B_n = \cup_{k\leq n} A_k$:

$$\mu^*(E') = \mu^*(E' \cap B_n^c) + \sum_{k=0}^{n} \mu^*(E' \cap A_k) \tag{3.2}$$

Le cas $n = 0$ est immédiat. Si c'est vrai pour $n \geq 0$ alors en écrivant $\mu^*(E' \cap B_n^c) = \mu^*(E' \cap B_n^c \cap A_{n+1}) + \mu^*(E' \cap B_n^c \cap A_{n+1}^c)$ et remarquant que $B_n^c \cap A_{n+1} = A_{n+1}$ et $B_n^c \cap A_{n+1}^c = B_{n+1}^c$ on déduit le résultat pour $n + 1$.

Comme $B_n \subset B_\infty \doteq \cup_{k\geq 0} A_k$, on déduit $\mu^*(E') \geq \mu^*(E' \cap B_\infty^c) + \sum_{k=0}^{n} \mu^*(E' \cap A_k)$ puis en passant à la limite en n

$$\mu^*(E') \geq \mu^*(E' \cap B_\infty^c) + \sum_{k=0}^{\infty} \mu^*(E' \cap A_k)\,.$$

Comme par σ sous-additivité on a l'inégalité inverse, on déduit l'égalité.

Démonstration (Suite)

4. On tire de (3.2)

$$\mu^*(E') = \mu^*(E' \cap B_\infty^c) + \sum_{k=0}^{\infty} \mu^*(E' \cap A_k) \geq \mu^*(E' \cap B_\infty^c) + \mu^*(E' \cap B_\infty)$$

d'où $B_\infty \in \mathcal{M}(\mu^*)$ et en prenant $E' = B_\infty$:

$$\mu^*(\cup_{k\geq 0} A_k) = \sum_{k=0}^{\infty} \mu^*(A_k)$$

ce qui montre que μ^* est σ-additive sur $\mathcal{M}(\mu^*)$

Théorème 10 (Critère de Carathéodory)**.** *Si (E, d) est un espace métrique et μ^* une mesure extérieure sur E telle que $\mu^*(A) + \mu^*(B) = \mu^*(A \cup B)$ pour toutes parties $A, B \subset E$ vérifiant $d(A, B) > 0$, alors tous les boreliens sont μ^*-mesurables.*

Démonstration

Soit F un fermé et $E' \subset E$. Montrons que $\mu^*(E') \geq \mu^*(E' \cap F) + \mu^*(E' \cap F^c)$. On commence par supposer que $\mu^*(E') < \infty$ car sinon le résultat est trivial. Notons alors $F_n = \{x \in E \mid d(F, x) \geq 1/n\}$ pour tout $n \geq 1$. On a $\mu^*(E' \cap F) + \mu^*(E' \cap F_n) = \mu^*(E' \cap (F \cup F_n)) \leq \mu^*(E') \leq \mu^*(E' \cap F) + \mu^*(E' \cap F^c)$. Il suffit donc de montrer que $\mu^*(E' \cap F_n) \to \mu^*(E' \cap F^c)$. Or en posant $C_k = \{x \in E \mid 1/(k+1) \leq d(x, F) < 1/k\}$, on a

$$\mu^*(E' \cap F_n) \leq \mu^*(E' \cap F^c) \leq \mu^*(E' \cap F_n) + \sum_{k\geq n} \mu^*(E' \cap C_k)\,. \tag{3.3}$$

car $F^c = F_n \cup_{k\geq n} C_k$ puisque F^c est ouvert. Cependant comme les C_{2k} sont toutes à distance strictement positives les unes des autres, on déduit des hypothèses que $\sum_{2k\geq n} \mu^*(E' \cap C_{2k}) \leq \mu^*(E' \cap F^c)$. De même $\sum_{2k+1\geq n} \mu^*(E' \cap C_{2k+1}) \leq \mu^*(E' \cap F^c)$ ce qui donne $\sum_{k\geq n} \mu^*(E' \cap C_k) < 2\mu^*(E' \cap F^c) < \infty$. Par suite $\sum_{k\geq n} \mu^*(E' \cap C_k) \to 0$ ce qui montre avec (3.3) le résultat.

3.2 Construction de la mesure de Lebesgue

Soit $\lambda^* : \mathcal{P}(\mathbb{R}) \to [0,\infty]$ tel que

$$\lambda^*(A) = \inf_{\bigcup_{n\geq 0}]a_n,b_n[\supset A} \sum_{n\geq 0} (b_n - a_n)\,. \tag{3.4}$$

Théorème 11. *1. λ^* est une mesure extérieure*

2. La tribu $\mathcal{M}(\lambda^*)$ contient $\mathcal{B}(\mathbb{R})$. La restriction de λ^* à $\mathcal{M}(\lambda^*)$, notée λ est appelée mesure de Lebesgue

3. $\lambda([a,b]) = \lambda(]a,b[) = b - a$.

Démonstration

La partie 1. est laissée en exercice.

Pour montrer que $\mathcal{B}(\mathbb{R}) \subset \mathcal{M}(\lambda^*)$, il suffit de montrer que les intervalles fermés $]-\infty, a]$ sont tous dans $\mathcal{M}(\lambda^*)$.

Or pour tout $E \subset \mathbb{R}$ tel que $\lambda^*(E) < \infty$, on a pour tout $\epsilon > 0$, une famille d'intervalle $(]a_n, b_n[)_{n\geq 0}$ tel que $E \subset \cup_{n\geq 0}]a_n, b_n[$ et $\lambda^*(E) + \epsilon \geq \sum_{n\geq 0}(b_n - a_n)$. Par suite, en posant $a'_n = a_n \wedge a$, $b'_n = b_n \wedge a$, $a"_n = a_n \vee a$, et $b''_n = b_n \vee a$, on a $b''_n - a''_n + b'_n - a'_n = b_n - a_n$,

$E \cap]-\infty, a[\subset \cup_{n\geq 0}]a'_n, b'_n[$, $E\cap]a, +\infty[\subset \cup_{n\geq 0}]a''_n, b''_n[$

si bien que

$$\lambda^*(E) + \epsilon \geq \sum_{n\geq 0} (b''_n - a''_n + b'_n - a'_n) \geq \lambda^*(E\cap]-\infty, a[) + \lambda^*(E\cap]a, \infty[)\,.$$

Enfin, comme $\lambda^*(\{a\}) = 0$, $\lambda^*(E\cap]-\infty, a[) = \lambda^*(E\cap]-\infty, a])$ d'où en passant à la limite en $\epsilon \to 0$, on a

$$\lambda^*(E) \geq \lambda^*(E\cap]-\infty, a]) + \lambda^*(E\cap]a, \infty[)$$

ce qui prouve que $]-\infty, a] \in \mathcal{M}(\lambda^*)$ pour tout a.

démonstration (fin)

Montrons maintenant que $\lambda([a,b]) = \lambda(]a,b[) = b - a$. Comme $\lambda(\{a\}) = \lambda(\{b\}) = 0$, on a $\lambda([a,b]) = \lambda(]a,b[)$ et il suffit de montrer que $\lambda([a,b]) = b - a$. Or on a immédiatement que $\lambda([a,b]) \leq b - a$. De plus, si $\cup_{n\geq 0}]a_n, b_n[\supset]a,b[$ est un recouvrement de $[a,b]$, par compacité de $[a,b]$ on peut se ramener à un sous-recouvrement fini $\cup_{k=0}^K]a_{n_k}, b_{n_k}[\supset]a,b[$. Or en supposant les a_{n_k} croissants et en notant $a_{n_{K+1}} = b$, on a $a_{n_0} \leq a$, $b_{n_k} \geq a_{n_{k+1}}$ pour tout $0 \leq k \leq K$ (car sinon on laisse un trou) d'où

$$\sum_{k=0}^K b_{n_k} - a_{n_k} \geq \sum_{k=0}^K a_{n_{k+1}} - a_{n_k} \geq b - a\,,$$

ce prouve que $\lambda([a,b]] \geq b - a$ et termine la preuve du théorème

Remarque 14. *De la même façon, on construit la mesure de Lebesgue sur $\mathbb{R}^d$, en travaillant avec la famille $\mathcal{P}$ des pavés ouverts : $\prod_{i=1}^d]a_i, b_i[$*

$$\lambda_d^*(A) = \inf_{\cup_{n\geq 0} P_n \supset A, P_n \in \mathcal{P}\ \forall n} \sum_{n\geq 0} vol(P_n)\,,$$

où $vol(P) = \prod_{i=1}^d (b_i - a_i)$ pour tout $P = \prod_{i=1}^d]a_i, b_i[\in \mathcal{P}$.

On vérifie alors que $\mathcal{M}(\lambda_d^*) \supset \mathcal{B}(\mathbb{R}^d)$ et que la restriction λ_d de λ_d^* à $\mathcal{M}(\lambda_d^*)$ vérifie $\lambda_d(P) = vol(P)$ pour tout $P \in \mathcal{P}$.

Proposition 9.

$$\mathcal{M}(\lambda_d^*) = \overline{\mathcal{B}(\mathbb{R}^d)}$$

Démonstration

1. On commence par vérifier que $\overline{\mathcal{B}(\mathbb{R}^d)} \subset \mathcal{M}(\lambda_d^*)$:

On sait que $\mathcal{M}(\mathbb{R}^d) \subset \mathcal{M}(\lambda_d^*)$. De plus si $N \in \mathcal{N}$ est un ensemble négligeable, alors il existe $A \in \mathcal{B}(\mathbb{R}^d)$ tel que $A \supset N$ et $\lambda_d(A) = 0$. On vérifie que pour tout $E \subset \mathbb{R}^d$, on a

$$\lambda_d^*(E) \geq \lambda_d^*(E \cap N) + \lambda_d^*(E \cap N^c)$$

puisque $\lambda_d^*(E \cap N) \leq \lambda_d^*(N) \leq \lambda_d^*(A) = \lambda_d(A) = 0$. Ainsi $N \in \mathcal{M}(\lambda_d^*)$ d'où $\mathcal{N} \subset \mathcal{M}(\lambda_d^*)$ et donc

$$\overline{\mathcal{B}(\mathbb{R}^d)} = \sigma(\mathcal{N} \cup \mathcal{B}(\mathbb{R}^d)) \subset \mathcal{M}(\lambda_d^*) .$$

2. On vérifie maintenant que $\mathcal{M}(\lambda_d^*) \subset \overline{\mathcal{B}(\mathbb{R}^d)}$. Nous allons montrer que pour tout $A \in \mathcal{M}(\lambda_d^*)$ et tout $K > 0$, $A_K \doteq A\cap]-K, K[^d \in \overline{\mathcal{B}(\mathbb{R}^d)}$. Pour cela, nous allons montrer qu'il existe deux boreliens B, C tels que

$$B \subset A_K \subset C \text{ et } \lambda_d(B) = \lambda_d^*(A_K) = \lambda_d(C) . \tag{3.5}$$

En effet, on a alors $A_K = B \cup N$ avec $N = A_K \setminus B \subset C \setminus B$ ce qui montre que N est négligeable.

Or pour tout $\epsilon > 0$, il existe une famille $(P_n^\epsilon)_{n\geq 0}$ de pavés ouverts recouvrant A_K tel que $\lambda_d^*(A_K) + \epsilon \geq \sum_{n\geq 0} \lambda_d(P_n^\epsilon) \geq \lambda_d(C^\epsilon) \geq \lambda_d^*(A_K)$ où $C^\epsilon = \cup_{n\geq 0} P_n^\epsilon$. Si $C = \cap_{n\geq 0} C^{1/n}$, on a $C \in \mathcal{B}(\mathbb{R}^d)$, $C \supset A_K$ et

$$\lambda_d^*(A_K) = \lambda_d(C) . \tag{3.6}$$

On considérant $A'_K =]-K, K[^d \cap A^c$, on montre de même qu'il existe $D \in \mathcal{B}(\mathbb{R}^d)$ tel que $A'_K \subset D$ et

$$\lambda_d^*(A'_K) = \lambda_d(D) .$$

Notons qu'en considérant $C\cap]-K, K[^d$ et $D\cap]-K, K[^d$, on peut supposer que D et C sont inclus dans le pavé $]-K, K[^d$ ce que nous supposerons dorénavant. Or comme $A \in \mathcal{M}(\lambda_d^*)$, on déduit

$$\lambda_d(]-K, K[^d) = \lambda_d^*(]-K, K[^d) = \lambda_d^*(A_K) + \lambda_d^*(A'_K) = \lambda_d(D) + \lambda_d(C) .$$

Par suite, si $B \doteq]-K, K[^d \setminus D$, on a $B \subset A_K$ et

$$\lambda_d(B) = \lambda_d(]-K, K[^d) - \lambda_d(D) = \lambda_d^*(A_K) . \tag{3.7}$$

En considérant (3.6) et (3.7), on obtient (3.5) ce qui termine la démonstration.

3.3 Régularité des mesures boreliennes

Proposition 10 (Régularité des mesures boreliennes finies)**.** *soit* (E, d) *un espace métrique et* μ *une mesure borelienne finie que* E. *Alors* μ *est régulière i.e. pour tout* $A \in \mathcal{B}(E)$,

$$\mu(A) = \inf\{\, \mu(U) \mid U \textit{ ouvert}, \ U \supset A \,\} \tag{3.8}$$

$$= \sup\{\, \mu(F) \mid F \textit{ fermé}, \ F \subset A \,\} \tag{3.9}$$

Exercice 11. *On voit par là que* du point de vue de la mesure, *pour tout* $B \in \mathcal{B}(E)$, *il existe* $A \subset B \subset C$ *tels que* A *est un* F_σ *(union dénombrable de fermés),* C *est un* G_δ *(intersection dénombrable d'ouverts) et* $\mu(A) = \mu(B) = \mu(C)$. *Par suite, les boreliens généraux ne sont que des* F_σ *ou des* G_δ *à* un négligeable près *(une fois définie une mesure borelienne).*

Démonstration

Il suffit de vérifier que l'ensemble Λ des parties $A \in \mathcal{B}(E)$ étant régulières supérieurement et inférieurement forment un λ-système contenant les ouverts. Or tout ouvert est trivialement régulier supérieurement. De plus en considérant

$$F_n = \{ x \in E \mid d(x, U^c) \geq 1/n \},$$

F_n est une suite croissante de fermés qui converge vers U. Par convergence monotone on a $\mu(U) = \sup \mu(F_n)$ d'où la régularité inférieure de U.

démonstration (suite)

Montrons maintenant que Λ est un λ-système.

Comme E est ouvert, $E \subset \Lambda$.

De plus, si $A \subset B \in \Lambda$, pour tout $\epsilon > 0$, il existe un fermé F et un ouvert U tels que $F \subset A \subset B \subset U$ et $\mu(U \setminus B) \leq \epsilon$, $\mu(A \setminus F) \leq \epsilon$. Par suite, $U \setminus F \supset B \setminus A$ et $\mu((U \setminus F) \setminus (B \setminus A) \leq 2\epsilon$. On procède de la même façon, en introduisant un ouvert U' et un ferme F' tels que $F' \subset B$, $A \subset U'$, $\mu(U' \setminus A) \leq \epsilon$ et $\mu(B \setminus F) \leq \epsilon$. On a alors $F' \setminus U'$ fermé inclus dans $B \setminus A$ et $\mu((B \setminus A) \setminus (F' \setminus U')) \leq 2\epsilon$. Comme ϵ est arbitraire on déduit que $B \setminus A \in \Lambda$.

démonstration (fin)

Vérifions enfin la stabilité par union croissante. Soit $(A_n)_{n\in\mathbb{N}}$ une suite croissante d'élément de Λ. Pour tout $\epsilon > 0$ on a par convergence monotone n_0 tel que $\mu(A_\infty \setminus A_{n_0}) \leq \epsilon$ où $A_\infty = \cup_{n\geq 0} A_n$. Comme A_{n_0} est régulière inférieurement, il existe un fermé F tel que $F \subset A_{n_0}$ et $\mu(A_{n_0} \setminus F) \leq \epsilon$. Enfin, pour tout $n \geq 0$, il existe $U_n \supset A_n$ tel que U_n est ouvert et $\mu(U_n \setminus A_n) \leq 2^{-n}\epsilon$. Par suite, comme

$$(\cup_{n\geq 0} U_n) \setminus A_\infty \subset \cup_{n\geq 0} U_n \setminus A_n ,$$

on déduit $\mu((\cup_{n\geq 0} U_n) \setminus A_\infty) \leq \sum_{n\geq 0} 2^{-n}\epsilon = 2\epsilon$.

Definition 20. Soit E un espace topologique et μ une mesure borelienne sur E. On dit que μ est une *mesure de Radon* si $\mu(K) < \infty$ pour tout compact K de E.

Proposition 11 (Régularité des mesures de Radon sur $\mathbb{R}^d$)**.** *Si μ est une mesure de Radon sur $\mathbb{R}^d$ alors*

$$\begin{aligned} \mu(A) &= \inf \{ \mu(U) \mid U \textit{ ouvert, } U \supset A \} && (3.10)\\ &= \sup \{ \mu(K) \mid K \textit{ compact, } K \subset A \} && (3.11) \end{aligned}$$

Remarque 15. *On déduit évidemment la régularité de la mesure de Lebesgue λ_d.*

Démonstration. **1.** Régularité inférieure : Pour tout $n > 0$, si $\overline{B}_n$ est la boule fermée de rayon n, la restriction de μ à $\overline{B}_n$ (muni de la tribu trace) est une mesure borelienne finie (car $\overline{B}_n$ est compacte) pour laquelle on peut appliquer la proposition précédente. Par suite, si $A_n = A \cap \overline{B}_n$, la régularité inférieure par des fermés devient une régularité inférieure sur les compacts car tout fermé de $\overline{B}_n$ est un compact de $\mathbb{R}^d$.

2. Régularité supérieure : Pour la régularité supérieure, en considérant cette fois la restriction de μ à la boule ouverte B_n, on tire de la proposition précédente que pour tout $n \geq 0$, il existe un ouvert U_n de B_n (et donc de $\mathbb{R}^d$) tel que $\mu(U_n \setminus A_n) \leq 2^{-n}\epsilon$ où $\epsilon > 0$ est fixé et $A_n = A \cap B_n$. Par suite $\mu((\cup_{n\geq 0} U_n) \setminus A) \leq 2\epsilon$. □

On peut étendre la proposition 11 aux espaces métriques localement compacts séparables :

Definition 21. – On dit que qu'un espace séparé est *localement compact* si tout point admet un voisinage compact.

– On dit que l'espace est *séparable*, si il admet une suite dénombrable dense.

Exercice 12. $\mathbb{R}^d$ *est un exemple d'espace métrique localement compact séparable mais aussi tous les ouverts et tous les fermés de* $\mathbb{R}^d$.

Proposition 12 (Régularité des mesures de Radon (2))**.** *Soit* E *un espace métrique localement compact séparable Si* μ *est une mesure de Radon sur* E *alors*

$$\mu(A) = \inf\{ \mu(U) \mid U \textit{ ouvert, } U \supset A \} \tag{3.12}$$
$$= \sup\{ \mu(K) \mid K \textit{ compact, } K \subset A \} \tag{3.13}$$

Démonstration. La preuve peut se conduire exactement de la même façon que pour la proposition 11. Il nous suffit de montrer l'assertion suivante :

– il existe deux suites croissantes $(K_n)_{n\geq 0}$ de compacts et $(V_n)_{n\geq 0}$ d'ouverts telles que $V_n \subset K_n$ pour tout n et $\cup_{n\geq 0} V_n = E$.

On remplace alors dans la preuve, $\overline{B}_n$ par K_n et B_n par V_n.

Montrons l'assertion. Soit $(x_n)_{n\in\mathbb{N}}$ une suite dense de E et $D = \{ (k,l) \in \mathbb{N}\times\mathbb{N}^* \mid \overline{B(x_k, 1/l)} \text{ est compacte} \}$ où $\overline{B(x,r)}$ désigne la boule fermée de centre x et de rayon r. On construit alors pour tout $n \geq 0$, $K_n \doteq \cup_{(k,l)\in D, k,l\leq n} \overline{B(x_k, 1/l)}$ qui est compact comme réunion finie de compacts. On définit de même pour tout $n \geq 0$, l'ouvert $V_n \doteq \cup_{(k,l)\in D,k,l\leq n} B(x_k, 1/l)$. On vérifie aisément que les deux suites sont croissantes et que $\cup_{n\geq 0} V_n = E$. En effet, pour tout $x \in E$, il existe $r > 0$ tel que $\overline{B(x,r)}$ soit compacte et donc pour tout $y \in E$ et $\rho \leq r/2$, si $x \in \overline{B(y,\rho)}$ alors $\overline{B(y,\rho)} \subset \overline{B(x,r)}$ est compacte. □

3.4 Mesures de Stieltjes

Thomas Stieltjes

« Thomas-Joannes Stieltjes est un mathématicien hollandais né le 29 décembre 1856 à Zwolle, aux Pays-Bas. Durant ses études à l'école polytechnique de Dhelft, il préfère fréquenter la bibliothèque et les ouvrages de Gauss et Jacobi que les cours. Si cette activité fut sans doute profitable à son devenir de mathématicien, *la conséquence immédiate est que Stieltjes échoue par deux fois aux examens*. »
Ref : BibMath

Théorème 12. 1. *Soit* μ *une mesure de Radon sur* $\mathbb{R}$*. On définit* $F_\mu : \mathbb{R} \to \mathbb{R}$ *tel que*

$$F_\mu(x) = \begin{cases} \mu(]0,x]) & \textit{si } x \geq 0 \\ \\ -\mu(]x,0]) & \textit{si } x < 0\,. \end{cases} \tag{3.14}$$

Alors F_μ *est croissante et continue à droite (c.à.d.).*

2. *Soit* $F : \mathbb{R} \to \mathbb{R}$ *croissante et c.à.d.. Alors il existe une unique mesure de Radon,* μ *tel que* $\mu(]a,b]) = F(b) - F(a)$ *pour tout* $a < b \in \mathbb{R}$.

Démonstration

1. On vérifie facilement que pour $y > x$, $F_\mu(y) - F_\mu(x) = \mu(]x, y])$ ce qui prouve que $y \to F_\mu(y)$ est croissante. Il suffit de vérifier la continuité à droite. Or si $(y_n)_{n\in\mathbb{N}}$ est une suite décroissante tendant vers y, $y_n > y$ pour tout $n \geq 0$ alors comme $]x, y] = \cap_{n\geq 0}]x, y_n]$, et que $\mu(]x, y_0]) < \infty$, on déduit que $F_\mu(y) - F_\mu(x) = \mu(]x, y]) = \lim F_\mu(y_n) - F_\mu(x)$ ce qui prouve la continuité à droite.

2. L'unicité est une conséquence immédiate du corollaire 3 sur l'unicité des mesures. Pour l'existence, on pourrait reproduire la preuve de l'existence de la mesure de Lebesgue selon la technique d'extension de Carathéodory. On peut procéder un peu différemment : En effet si $G : \mathbb{R} \to \overline{\mathbb{R}}$ est définie par $G(y) \doteq \inf\{ x \in \mathbb{R} \mid F(x) \geq y \}$[1] alors la mesure image μ de la mesure de Lebesgue par G vérifie $\mu(]a, b]) = F(b) - F(a)$ pour tout $a < b \in \mathbb{R}$.

Il suffit pour cela de montrer que $\lambda(G^{-1}(]a, b])) = F(b) - F(a)$. Or $G(y) \leq b$ ssi $F(b) \geq y$. En effet, si $F(b) \geq y$ alors par définition de G, $G(y) \leq b$ et si $G(y) \leq b$ alors par continuité à droite $F(b) \geq y$. Par suite $G^{-1}(]a, b]) =]F(a), F(b)]$ d'où le résultat.

3.5 Théorème de représentation de Riesz

Frigyes Riesz

« As a lecturer Riesz was somewhat unpredictable. He was not always perfectly prepared for the lecture. When that happened he would ask his assistant, Laszlo Kalmar, for help. But Kalmar wasn't always available. Nevertheless, we found Riesz a first-class interpreter of science. In his lectures everything appeared naturally in historical perspective. That was highly instructive. When he was not well prepared, he often spent time on very interesting digressions. Once he gave a brilliant explanation of why scientific work is easy. »
A visit to Hungarian Mathematics, R. Hersh et V. John-Steiner, The mathematical Intelligencer, vol 15, (2), 1993, pp13–26

– Soit E un espace topologique et μ une mesure de Radon sur E. Alors l'application $J : C_c(E) \to \mathbb{R}$ définie par

$$J(f) = \int f d\mu$$

est une application linéaire positive (i.e. si $f \geq 0$ alors $J(f) \geq 0$). Elle est par ailleurs continue au sens où pour tout K compact, il existe C_K tel que $|J(f)| \leq C_K \|f\|_\infty$ pour toute f continue sur E de support K.

– La réciproque est-elle vraie ? Le théorème de représentation de Riesz répond par l'affirmative dans le cas des espaces localement compacts.

Théorème 13. *Soient E un espace métrique localement compact séparable et $J : C_c(E) \to \mathbb{R}$ une forme linéaire positive (on ne suppose pas la continuité de J).*

Alors il existe une unique mesure de Radon sur E tel que

$$J(f) = \int f d\mu$$

pour tout $f \in C_c(E)$.

[1] G est l'inverse généralisé de F qui correspond à l'inverse classique de F lorsque F est inversible. On utilise la convention habituelle : $\inf \emptyset = +\infty$

Remarque 16. *Comme pour toute mesure de Radon sur un espace localement compact séparable, la proposition 12 s'applique et* μ *est régulière extérieurement sur les ouverts et intérieurement sur les compacts.*

La preuve du théorème de représentation de Riesz n'est pas donnée ici. En gros : on définit pour tout ouvert

$$\mu^*(V) = \sup\{\, J(f) \mid \|f\|_\infty \le 1,\ \mathrm{supp}(f) \subset V \,\}$$

où $\mathrm{supp}(f)$ est le support de f puis pour tout $A \subset E$

$$\mu^*(A) = \inf\{\, \mu^*(V) \mid V \text{ ouvert }, A \subset V \,\}$$

En introduisant des partitions de l'unité (c'est la que l'on se sert de E localement compact), on montre que μ^* est une mesure extérieure telle que $\mathcal{M}(\mu^*) \supset \mathcal{B}(E)$. Il reste à vérifier que si μ est la restriction de μ^* à μ alors $J(f) = \int f d\mu$.

On trouvera une démonstration complète (et bien d'autres choses) dans l'excellent « Real and Complex Analysis »de Rudin.

Chapitre 4

Espaces L^p

4.1 Définitions

Espaces $\mathcal{L}^p$

Soit $(E, \mathcal{A}, \mu)$ un espace mesuré.

Definition 22. – Pour tout $p \geq 1$, on définit l'espace

$$\mathcal{L}^p(E, \mathcal{A}, \mu) \doteq \{ f : E \to \overline{\mathbb{R}} \text{ mesurable} \mid \int |f|^p d\mu < \infty \},$$

et pour $f \in \mathcal{L}^p(E, \mathcal{A}, \mu)$, on note $\|f\|_p \doteq (\int |f|^p d\mu)^{1/p}$.

– Pour $p = \infty$, on définit

$$\mathcal{L}^\infty(E, \mathcal{A}, \mu) \doteq \{ f : E \to \overline{\mathbb{R}} \text{ mesurable} \mid \exists c \geq 0,\ |f| \leq c\ \mu\ \text{p.p.} \},$$

et pour $f \in \mathcal{L}^\infty(E, \mathcal{A}, \mu)$, on note $\|f\|_\infty \doteq \inf\{c \geq 0 \| f| \leq c\ \mu\ \text{p.p.}\}$.

Espaces L^p

Definition 23. On définit la relation d'équivalence $f \sim g$ ssi $f = g$ μ presque partout. On définit alors pour $p \in [0, \infty]$

$$L^P(E, \mathcal{A}, \mu) \doteq \mathcal{L}^p(E, \mathcal{A}, \mu)/\sim$$

Les espaces L^p ont naturellement une structure d'espace vectoriel.

Remarque 17. *La tradition est de ne pas travailler directement avec les classes d'équivalences mais de travailler avec un représentant. On dira souvent (tjs) : Soit $f \in L^p$ bien que f soit considérée comme une fonction.*

Otto Ludwig Hölder

Inégalité de Hölder

Théorème 14 (Inégalité de Hölder)**.** *Soient* $[p,q] \in [1,+\infty]$ *tels que* $\frac{1}{p}+\frac{1}{q}=1$ *(exposants conjugués* ($\frac{1}{\infty}=0$)*). On a pour tout* $f,g : E \to \mathbb{R}$ *mesurables :*

$$\int |fg|d\mu \leq (\int |f|^p d\mu)^{1/p}(\int |g|^q d\mu)^{1/q}$$

i.e. $\|fg\|_1 \leq \|f\|_p\|g\|_q$

Remarque 18. *Pour p=q=2, on retrouve l'inégalité de Cauchy-Schwartz.*

Démonstration. Dans le cas $p=1$ et $q=\infty$, on a $|fg|(x) \leq |f|\|g\|_\infty$ p.p. ce qui donne le résultat en intégrant.

Dans le cas $p,q<\infty$, on tire de la concavité de $x \to \ln(x)$ que pour tout $x,y>0$ et $\alpha \in [0,1]$: $\ln(\alpha x+(1-\alpha)y) \geq \alpha\ln(x)+(1-\alpha)\ln(y)$ i.e. $\alpha x+(1-\alpha)y \geq x^\alpha y^{1-\alpha}$. Cette dernière inégalité est vraie pour $x,y \geq 0$. On posant $\alpha = 1/p$ on déduit

$$\frac{|f||g|}{\|f\|_p\|g\|_q} \leq \frac{1}{p}\frac{|f|^p}{\|f\|_p^p}+\frac{1}{q}\frac{|g|^q}{\|g\|_q^q}$$

puis en intégrant

$$\frac{\int|f||g|d\mu}{\|f\|_p\|g\|_q} \leq \frac{1}{p}\frac{\int|f|^p d\mu}{\|f\|_p^p}+\frac{1}{q}\frac{\int|g|^q d\mu}{\|g\|_q^q} \leq 1$$

d'où l'on déduit le résultat. □

Proposition 13. *Si la mesure est finie (*$\mu(E)<\infty$*), alors pour tout* $1 \leq p \leq p' \leq \infty$*, on a* $L^{p'} \subset L^p$.

Remarque 19. *Les inclusions sont fausses lorsque* μ *n'est plus finie.*

Démonstration. Dans le cas $p<p'=\infty$, on a $|f|^p \leq \|f\|_\infty^p$ p.p. d'où en intégrant, $\int|f|^p d\mu \leq \mu(E)\|f\|_\infty^p$.

Dans le cas $p<p'<\infty$, en posant $\tilde{p}=p'/p$ et $\tilde{q}$ l'exposant conjugué, l'inégalité de Hölder nous donne :

$$\int|f|^p d\mu = \int|f^p||1|d\mu \leq (\int|f^p|^{\tilde{p}}d\mu)^{1/\tilde{p}}(\int d\mu)^{1/\tilde{q}}$$
$$\leq (\int|f|^{p'}d\mu)^{1/\tilde{p}}\mu(E)^{1/\tilde{q}} < \infty$$

□

Hermann Minkowski

« *In the early years of his scientific career, Albert Einstein considered mathematics to be a mere tool in the service of physical intuition. In later years, he came to consider mathematics as the very source of scientific creativity. A main motive behind this change was the influence of two prominent German mathematicians : David Hilbert and Hermann Minkowski.* »

L Corry, The influence of David Hilbert and Hermann Minkowski on Einstein's views over the interrelation between physics and mathematics, Endeavor 22 (3) (1998), 95-97.

Théorème 15 (Inégalité de Minkowski)**.** *Soient* $p \in [0, +\infty]$ *et* $f, g \in L^p$. *On a alors*

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p$$

Démonstration. Si $p = 1$ ou $p = \infty$ la preuve est immédiate.

Si $p \in]1, \infty[$, on écrit

$$|f+g|^p \leq |f+g|^{p-1}(|f| + |g|)$$

puis en intégrant et en appliquant l'inégalité de Hölder pour $\tilde{p} = p/(p-1)$ et $\tilde{q} = p$:

$$\int |f+g|^p d\mu \leq \int |f+g|^{p-1}|f| d\mu + \int |f+g|^{p-1}|g| d\mu \overset{\text{Hölder}}{\leq}$$
$$\left(\int |f|^p d\mu\right)^{\frac{1}{p}} \left(\int |f+g|^p d\mu\right)^{\frac{p-1}{p}} + \left(\int |g|^p d\mu\right)^{\frac{1}{p}} \left(\int |f+g|^p d\mu\right)^{\frac{p-1}{p}} .$$

Si $\int |f+g|^p d\mu = 0$ le résultat est vrai, sinon, il suffit de diviser les membres de gauche et de droite par $\left(\int |f+g|^p d\mu\right)^{\frac{p-1}{p}}$. □

4.2 Structure d'espace de Banach des L^p

Stefan Banach

« A chance event occurred in the spring of 1916 which was to have a major impact on Banach's life. Steinhaus, who had been undertaking military service, was about to take up a post at the Jan Kazimierz University in Lvov. However he was living in Kraków in the spring of 1916, waiting to take up the appointment. He would walk through the streets of Kraków in the evenings and, as he related in his memoirs :-

« *During one such walk I overheard the words "**Lebesgue measure**". I approached the park bench and introduced myself to the two young apprentices of mathematics. They told me they had another companion by the name of Witold Wilkosz, whom they extravagantly praised. The youngsters were Stefan Banach and Otto Nikodym. From then on we would meet on a regular basis, and ... we decided to establish a mathematical society.* »

... Steinhaus told Banach of a problem which he was working on without success. After a few days Banach had the main idea for the required counterexample and Steinhaus and Banach wrote a joint paper, which they presented to Zaremba for publication. The war delayed publication but the paper, Banach's first, appeared in the Bulletin of the Kraków Academy in 1918. »

Théorème 16 (Théorème de Riesz-Fischer)**.** *Soit* $p \in [1, \infty]$. *Alors* $L^p(E, \mathcal{A}, \mu)$ *est un espace de Banach (evn complet)*

Remarque 20. *En fait, historiquement le théorème de Riesz-Fischer ne concerne « que »le cas* $p = 2$ *(cf Théorème 17).*

Ernst Sigismund Fischer

« In 1907 Ernst Fischer studied orthonormal sequences of functions and gave necessary and sufficient conditions for a sequence of constants to be the Fourier coefficients of a square integrable function. His two papers of 1907 were Sur la convergence en moyenne and Applications d'un théorème sur la convergence en moyenne both published in Comptes rendus of the Academy of Sciences in Paris. This work led to the concept of a Hilbert space. Frigyes Riesz published a similar result in the same year. The theorem, now called the Riesz-Fischer theorem, is one of the great achievements of the Lebesgue theory of integration. »

Démonstration

Cas $p = \infty$: Si $(f_n)_{n\geq 0}$ est une suite de Cauchy dans L^∞, alors, quitte à mettre à zero les f_n sur un ensemble négligeable, on peut supposer que

$$|f_r - f_n|(x) \leq \|f_r - f_n\|_\infty$$

pour tout $r, n \geq 0$ et *tout* $x \in E$. Par suite, $(f_n(x))_{n\geq 0}$ est une suite de Cauchy pour tout $x \in E$.

Si f est la limite simple des f_n, f est mesurable,

$$|f(x)| \leq \sup_{n\geq 0} \|f_n\|_\infty < \infty$$

et

$$|f_r - f_n|(x) \leq \sup_{r\geq n} \|f_r - f_n\|\infty \to 0$$

quand n tend vers $+\infty$. Cas $p \in [1, \infty[$: Comme la suite est de Cauchy, on peut extraire $g_k = f_{n_k}$, pour $k \geq 0$ tel que $\sum_{k\geq 0} \|g_k - g_{k+1}\|_p < \infty$. D'apres l'inégalité de Minkowski,

$$\|\sum_{k=1}^{n} |g_k - g_{k+1}|\|_p \leq \sum_{k=1}^{n} \|g_k - g_{k+1}\|_p \leq \sum_{k=1}^{\infty} \|g_k - g_{k+1}\|_p < \infty$$

Par convergence monotone, on déduit que $\| \sum_{k=1}^{\infty} |g_k - g_{k+1}|\|_p < \infty$ μ-p.p et donc que $\sum_{k=1}^{\infty} |g_k - g_{k+1}| < \infty$ presque partout. Il existe donc g_∞ mesurable, limite simple p.p. de la suite $(g_k)_{k\geq 0}$.

De plus, par Fatou, on a

$$\int |g_\infty - g_n|^p d\mu \leq \underline{\lim} \int |g_r - g_n|^p d\mu \leq \underline{\lim} \left(\sum_{k=n}^{r-1} \|g_{k+1} - g_k\|_p \right)^p$$

et

$$\sup_{r\geq n} \sum_{k=n}^{r-1} \|g_{k+1} - g_k\|_p \leq \sum_{k=n}^{\infty} \|g_{k+1} - g_k\|_p$$

qui tend vers 0 avec n. On termine en montrant (exo) que $\|f_n - g_\infty\|_p \to 0$.

Remarque 21. *On tire de la preuve que si $f_n \to f$ dans L^p, alors il existe une sous-suite de f_n qui converge simplement p.p. vers f.*

Théorème 17. *Si* $p = 2$*,* $L^2(E, \mathcal{A}, \mu)$ *a une structure d'espace de Hilbert réel pour le produit scalaire* $\langle f, g\rangle \doteq \int fg d\mu$.

Démonstration. On doit vérifier que le produit scalaire est bien défini. Or, d'après l'inégalité de Hölder, si $f, g \in L^2$, $fg \in L^1$. $\square$

Remarque 22. *On peut aussi considérer une structure d'espace de Hilbert complexe en considérant le produit hermitien* $\langle f, g\rangle = \int f\overline{g} d\mu$

4.3 Théorèmes de densités dans les L^p

Théorème 18. *Soit* $p \in [1, \infty[$

1. *Les fonctions étagées sont denses dans* L^P*. De plus, si* $\mathcal{A}$ *est généré par une famille dénombrable et si* μ *est* σ*-finie, alors* L^p *est séparable (il existe une famille dénombrable dense).*
2. *Si* (E, d) *est un espace métrique localement compact séparable et si* μ *est une mesure de Radon, les fonctions lipschitzienne à support compact sont denses dans* L^p*.*

(1) : On sait que pour toute fonction mesurable $f : R \to \overline{\mathbb{R}}$ positive, en considérant

$$f_n(x) = \sum_{k=1}^{n^2-1} \frac{k}{n} \mathbb{1}_{f^{-1}([\frac{k}{n}, \frac{k+1}{n}[)} \tag{4.1}$$

on a $0 \leq f_n \leq f$ et f est la limite simple des f_n. Par convergence dominée, si $f \in L^p$, alors $\int |f - f_n|^p d\mu \to 0$ d'où $f_n \overset{L^p}{\to} f$. On remarque par ailleurs que dans (4.1), on a $f_n = \sum \alpha_i \mathbb{1}_{A_i}$ avec $\mu(A_i) < \infty$. Lorsque f n'est plus supposée positive, il suffit de décomposer $f = f^+ - f^-$. Supposons maintenant que $\mathcal{A} = \sigma(\mathcal{C})$ avec $\mathcal{C}$ dénombrable. Nous savons déjà que

$$\{f = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i} | n \in \mathbb{N}^*, \alpha_i \in \mathbb{Q}, A_i \in \mathcal{A}, \mu(A_i) < \infty\}$$

est dense dans L^p. Il nous suffit donc de montrer qu'il existe une famille $\mathcal{F} \subset \mathcal{A}$ dénombrable telle que pour tout $A \in \mathcal{A}$ on a

$$\mathcal{P}_A : \quad \text{pour tout } \epsilon > 0, \text{ il existe } B \in \mathcal{F} \text{ tel que } \mu(A \Delta B) \leq \epsilon\,.$$

Nous allons montrer ce résultat lorsque μ est finie, le cas μ σ-finie est laissé en exercice.

Prenons pour $\mathcal{F}$ la plus petite algèbre d'ensemble contenant $\mathcal{C}$. On vérifie facilement que $\mathcal{F}$ est dénombrable puis que si $\tilde{\mathcal{A}}$ est l'ensemble des mesurables $A \in \mathcal{A}$ pour lesquels $\mathcal{P}_A$ est vraie, alors $\tilde{\mathcal{A}}$ est une tribu (exo). (2) : On sait que sous les hypothèses de (2), la mesure μ est régulière extérieurement sur les ouverts et intérieurement sur les compacts. Par suite, pour tout $\epsilon > 0$, il existe K compact et V ouvert tels que $K \subset A \subset V$ avec $\mu(V \setminus K) \leq \epsilon$ (on utilise ici le fait que $\mu(A) < \infty$). Comme $d(K, V^c) > 0$, on déduit que pour n assez grand, $\phi_n(x) \doteq (1 - nd(x, K)) \vee 0$ vérifie $\mathbb{1}_K \leq \phi_n \leq \mathbb{1}_V$. On a donc

$$\int |\mathbb{1}_A - \phi_n| d\mu \leq \mu(V \setminus K) \leq \epsilon$$

et on remarque que ϕ_n est lipschitzienne car $x \to d(x, K)$ l'est. Enfin, pour n assez grand, ϕ_n est à support compact. En effet, par compacité de K, K peut être recouvert par un nombre fini de boules $B(x_i, r_i)$, $1 \leq i \leq M$, d'adhérence compacte (E est localement compact). Par suite comme $d(K, (\cup_{i=1}^{M} B(x_i, r_i))^c) > 0$, on déduit que pour n assez grand

$$\text{supp}(\phi_n) \subset \cup_{i=1}^{M} \overline{B(x_i, r_i)}$$

qui est compacte.

Remarque 23. – *Si* (E, d) *métrique séparable et* $\mathcal{A} = \mathcal{B}(E)$ *alors* $\mathcal{A}$ *est dénombrablement généré.*

– *Si de plus* μ *est finie et borélienne, alors les fonctions lipschitziennes sont denses dans* L^p *pour* $p \in [1, \infty[$.
En effet μ *est alors extérieurement régulière et pour tout* $A \in \mathcal{B}(E)$ *et* $\epsilon > 0$ *il existe un ouvert* $V \supset A$ *tel que* $\mu(V \setminus A) \leq \epsilon$. *En prenant maintenant* $\phi_n(x) \doteq nd(x, V^c) \wedge 1$ *on a* ϕ_n *est une suite croissante qui converge simplement vers* $\mathbb{1}_V$. *Or*

$$\|\mathbb{1}_A - \phi_n\|_p \leq \|\mathbb{1}_A - \mathbb{1}_V\|_p + \|\mathbb{1}_V - \phi_n\|_p \leq \epsilon^{1/p} + \|\mathbb{1}_V - \phi_n\|_p$$

et on a $\|\mathbb{1}_V - \phi_n\|_p \to 0$ *par convergence dominée.*

4.4 Théorème de Radon-Nikodym

Otton Marcin Nikodym

« The Radon-Nikodym theorem (Radon proved it in 1913 for $\mathbb{R}^n$ and Nikodym in 1930 for the general case) is now a fundamental theorem in analysis
Let μ *be a* σ*-finite measure on a* σ*-algebra* Σ *of subsets of* Ω *and* ν *a countably additive set function on* Ω. *If* ν *is absolutely continuous with respect to* μ *: that is,* $\mu(A) = 0$ *implies* $\nu(A) = 0$, *then* $\nu(A) = \int_A f d\mu$ *for any* $A \in \Sigma$, *where* f *is locally integrable on* Ω.
Nikodym showed in 1927 how to produce a subset N of the unit square with area(N)= 1 such that for each point x belongs N there is a line intersecting N in the single point x. This paradoxical set in the plane, which for certain problems plays a role similar to Besicovitch sets is called a Nikodym set. »

Definition 24. Soient μ et ν deux mesure sur $(E, \mathcal{A})$

1. On dit que μ est *absolument continue* par rapport à ν, notée $\mu \ll \nu$ si pour tout $A \in \mathcal{A}$, $\nu(A)$ implique que $\mu(A) = 0$. On dira que μ est *dominée* par ν.
2. On dit que μ et ν sont *étrangères*, noté $\mu \perp \nu$ si il existe $N \in \mathcal{A}$ tel $\mu(N) = 0$ et $\nu(N^c) = 0$.

Exemple 3. δ_0 *et* λ *sont étrangères.*

Théorème 19. *Soient* μ *et* ν *deux mesures* σ*-finies sur* $(E, \mathcal{A})$. *Si* $\nu \ll \mu$ *alors il existe* $f : (E, \mathcal{A}) \to \overline{\mathbb{R}}_+$ *appelée* dérivée de Radon-Nikodym *de* ν *par rapport à* μ *tel que* $\nu = f\mu$ *i.e. pour tout* $A \in \mathcal{A}$ *:*

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

Démonstration

Nous proposons ici la démonstration dans le cas où μ et ν sont finies. Le cas σ-fini est laissé en exercice.

On commence par le cas où $\nu \leq \mu$ ie $\nu(A) \leq \mu(A)$ pour tout $A \in \mathcal{A}$. On introduit $\phi : L^2(E, \mathcal{A}, \mu) \to \mathbb{R}$ telle que $\phi(g) = \int g d\nu$. On vérifie que

$$|\phi(g)| \leq \int |g| d\mu \leq (\int |g|^2 d\mu)^{1/2} \mu(E)^{1/2} .$$

Par suite ϕ est continue et on déduit du théorème de représentation Fréchet-Riesz[1] qu'il existe $f \in L^2(E, \mathcal{A}, \mu)$ tel que $\phi(g) = \int gfd\mu$ si bien que en prenant $g = \mathbb{1}_A$, on a

$$\nu(A) = \int_A fd\mu\,.$$

En prenant $A = f^{-1}(]-\infty, 0[)$, on déduit que $\nu(A) = 0$ i.e. $f \geq 0$ ν presque partout. On montre de même que $f \leq 1$ ν presque partout. Dans le cas général, on remarque que on a toujours $\nu \leq \mu + \nu$ et que d'après ce que nous avons vu, il existe $h \in L^2(E, \mathcal{A}, \mu + \nu)$ tel que pour tout $g \in L^2(E, \mathcal{A}, \mu + \nu)$ on a

$$\int gd\nu = \int ghd\mu + \int ghd\nu\,. \tag{4.2}$$

On remarque que (4.2) s'étend à toute fonction mesurable $g \geq 0$. On sait de plus que $0 \leq h \leq 1$ $(\mu+\nu)$ presque partout. Soit $N = h^{-1}(\{1\})$. On a $\nu(N) = \mu(N) + \nu(N)$ d'où $\mu(N) = 0$ et donc $\nu(N) = 0$ puisque $\nu \ll \mu$. On peut donc supposer $h < 1$. En posant $g = \tilde{g}/(1-h)$ pour $\tilde{g} \geq 0$, on a déduit de (4.2) que

$$\int \tilde{g}d\nu = \int \tilde{g}\frac{h}{1-h}d\mu\,.$$

Théorème 20 (Décomposition de Lebesgue)**.** *Soient μ et ν deux mesures σ-finies. Alors il existe une unique décomposition $\nu = \nu_a + \nu_s$ de ν en deux mesures telles que $\nu_a \ll \mu$ et ν_s est étrangère à μ.*

Exercice 13. *Montrer que si μ est finie, alors $\nu \ll \mu$ implique que : pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que pour tout $A \in \mathcal{A}$, $\mu(A) \leq \delta \Rightarrow \nu(A) \leq \epsilon$.*

Démonstration. L'unicité est laissée en exercice. On se place à nouveau dans le cas μ et ν finies (cas général en exo). On reprend la démonstration du théorème 19 : il existe $0 \leq h \leq 1$ tel que pour tout $g \geq 0$ on a

$$\int gd\nu = \int ghd\mu + \int ghd\nu\,. \tag{4.3}$$

Si comme précédemment on introduit $N = h^{-1}(\{1\})$, on a $\mu(N) = 0$ et donc en posant $\nu_s(A) \doteq \nu(A \cap N)$ pour tout $A \in \mathcal{A}$, on a $\nu_s \perp \mu$. Si $\nu_a(A) \doteq \nu(A \cap N^c)$, on tire de (4.3) que si $\mu(A) = 0$ alors $\nu_a(A) = \int_{A\cap N^c} hd\nu$ d'où $\nu_a(A) = 0$ car $h < 1$ sur N^c. Par suite $\nu_a \ll \mu$ et le théorème est prouvé. □

4.5 Points de Lebesgue

Definition 25. Soit $f \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$. On appelle *point de Lebesgue* de f, tout point $x \in \mathbb{R}^d$ tel que

$$\lim_{r\to 0} \frac{\int_{B(x,r)} |f(y) - f(x)|dy}{\lambda_d(B(x,r))} = 0\,.$$

Sur un point de Lebesgue, f est continue dans un sens faible : $f(x)$ est la limite des moyennes locales de f sur des boules centrée en x. Il est clair que pour une application continue, tous les points sont des points de Lebesgue. Plus inattendu est le résultat suivant :

Théorème 21. *Si $f \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$, alors presque tous les points de $\mathbb{R}^d$ sont des points de Lebesgue.*

[1]Soit H est espace de Hilbert et f une forme linéaire continue sur H. Il existe $h \in H$, tel que pour tout $x \in H$, $f(x) = \langle h, x\rangle_H$ (voir Brézis Analyse fonctionnelle, p81)

Démonstration. Par densité des fonctions lipschitziennes, pour tout $\epsilon > 0$, il existe g continue tel que $\|f-g\|_1 \leq \epsilon$. Par suite, si $h = f - g$, et pour toute fonction f,

$$Tf(x) \doteq \overline{\lim_{r\to 0}} \frac{\int_{B(x,r)} |f(y) - f(x)| dy}{\lambda_d(B(x,r))},$$

on a $Tf \leq Tg + Th$ et comme $Tg = 0$ puisque g est continue, on tire $Tf \leq Th$. Or si

$$Mh(x) \doteq \overline{\lim_{r\to 0}} \frac{\int_{B(x,r)} |h(y)| dy}{\lambda_d(B(x,r))},$$

on vérifie que $Th(x) \leq Mh(x) + |h|(x)$.

Comme

$$Tf^{-1}([2t, +\infty[)) \subset \underbrace{(Mh)^{-1}([t, +\infty[)}_{A_1} \cup \underbrace{|h|^{-1}([t, +\infty[))}_{A_2}$$

pour tout $t > 0$, il nous suffit de contrôler la mesure de A_1 et A_2. Or

$$t\lambda(A_2) \leq \int t\mathbb{1}_{A_2}(x) dx \leq \|h\|_1 \leq \epsilon\,.$$

De plus, pour tout compact $K \subset A_1$, K peut être recouvert par un nombre fini de boules $B(x_i, r_i)_{i\in I}$ sur lesquelles $\int_{B(x_i,r_i)} |h|(y)dy \geq \frac{t}{2}\lambda_d(B(x_i, r_i))$. En réordonnant les r_i en ordre décroissant, on peut extraire une sous famille finie *disjointe* indexée par $J \subset I$ telle que $K \subset \cup_{j\in J} B(x_j, 3r_j)$ (exo). Par suite

$$\lambda_d(K) \leq 2 \times 3^d t^{-1} \sum_{j\in J} \int_{B(x_j,r_j)} |h|(y)dy \leq 2 \times 3^d \epsilon/t\,.$$

Par régularité intérieure, $\lambda_d(A_1) \leq 2 \times 3^d \epsilon/t$ et $\lambda_2(Tf^{-1}([2t, +\infty[)) \leq (1 + 2 \times 3^d)\epsilon/t$. Comme ϵ est arbitraire, on déduit $\lambda_2(Tf^{-1}([2t, +\infty[)) = 0$ pour tout $t > 0$. □

On obtient le corollaire suivant qui donne un sens plus concret à la dérivée de Radon-Nikodym :

Corollaire 4. *Si* μ *est une mesure de Radon,* $\mu \ll \lambda_d$ *et* $f = \frac{d\mu}{d\lambda_d} \in L^1_{loc}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ *est la dérivée de Radon-Nikodym de* μ *par rapport à* λ*, alors*

$$\lim_{r\to 0} \frac{\mu(B(x,r))}{\lambda_d(B(x,r))} = f(x)$$

pour presque tout $x \in \mathbb{R}$.

Remarque 24. *En travaillant un peu plus, on peut montrer (cf Rudin) que pour une mesure borelienne* μ *étrangère à* λ_d*, alors* $\lim_{r\to 0} \frac{\mu(B(x,r))}{\lambda_d(B(x,r))} = 0$ λ_d *presque partout. Par suite si* μ *est une mesure borelienne quelconque,* $\lim_{r\to 0} \frac{\mu(B(x,r))}{\lambda_d(B(x,r))}$ *est presque partout égale à la dérivée de Radon-Nikodym de la partie absolument continue de* μ *dans la décomposition de Lebesgue* $\mu = \mu_a + \mu_s$ *de* μ.

Un autre corollaire résout la question suivante : si $f \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ et

$$F(x) = \begin{cases} \int_{]0,x]} f(t)dt & \text{si } x \geq 0 \\ -\int_{]x,0]} f(t)dt & \text{si } x < 0\,, \end{cases} \tag{4.4}$$

on sait que pour f continue, on a F dérivable et $F' = f$. Que reste-il lorsque on ne suppose plus la continuité de f ?

Remarque 25. *Comme les singletons sont de mesure nulle pour la mesure de Lebesgue, on a donc* $\int_{]0,x]} f(t)dt = \int_{[0,x]} f(t)dt = \int_{[0,x[} f(t)dt$ *et on utilisera la notation non-ambiguë ici* $\int_0^x f(t)dt$ *qui est habituelle dans le cadre de l'intégrale de Riemann avec la convention que* $\int_0^x f(t)dt = -\int_x^0 f(t)dt$ *lorsque* $x \leq 0$*.*

Corollaire 5. *Soit* F *définie par* (4.4)*, alors* F *est dérivable presque partout et sa dérivée vaut* f *presque partout. On a donc dans ce cas* $F(b) - F(a) = \int_a^b F'(t)dt$.

Démonstration. Prenons $x \geq 0$ et $\delta > 0$, alors

$$|\frac{F(x+\delta) - F(x)}{\delta} - f(x)| \leq \frac{\int_{[x,x+\delta[} |f(y) - f(x)|dy}{\delta} \leq 2\frac{\int_{]x-\delta,x+\delta[} |f(y) - f(x)|dy}{2\delta}. \tag{4.5}$$

Or si x est un point de Lebesgue pour f, le terme de droite tend vers 0 lorsque $\delta \to 0$. On procède de même pour la limite à gauche. □

Remarque 26. *Cependant,* ***il ne faut pas croire*** *que si* F *est une fonction borélienne dérivable presque partout de dérivée presque partout égale à* f*, alors on a toujours* $F(b) - F(a) = \int_a^b f(t)dt$*. Il suffit de se rappeler le cas de l'escalier de Cantor qui est croissante continue sur* $[0, 1]$ *de dérivée presque partout nulle et pour la laquelle* $F(1) - F(0) = 1$ *!*

En fait, si F est continue croissante, elle définie une mesure de Radon. En faisant la décomposition de Lebesgue de la mesure associée notée dF, on a $dF = fd\lambda + d\mu_s$ où f est localement intégrable et μ_s est une mesure singuliere. On a donc

$$F(b) - F(a) = \int_a^b f(t)dt + \int_{]a,b]} d\mu_s(x)$$

où f est la dérivée presque sûre de F. Si la partie singulière est nulle, on dira alors que F est *absolument continue*, alors $F(b) - F(a) = \int_a^b F'(t)dt$[2]. Sinon, il manque la contribution de la partie singulière.

[2]Un traitement plus poussé des problèmes de différentiation de mesures et des fonctions absolument continues est proposé dans le chapitre 7 du Rudin "Real and Complexe Analysis"

Chapitre 5

Mesures produits

5.1 Construction de mesures produits

Soient $(E_1, \mathcal{E}_1)$ et $(E_2, \mathcal{E}_2)$ deux espaces mesurables. On rappelle que

$$\mathcal{E}_1 \otimes \mathcal{E}_2 \doteq \sigma(\mathcal{E}_1 \times \mathcal{E}_2)$$

Proposition 14. *On suppose que* μ_1 *et* μ_2 *sont* σ*-finies sur* $(E_1, \mathcal{E}_1)$ *et* $(E_2, \mathcal{E}_2)$*. Alors, il existe au plus une mesure* ν *sur* $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ *telle que*

$$\nu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)\,.$$

On appelle ν la mesure produit *de* μ_1 *avec* μ_2.

Démonstration. On utilise le corollaire sur l'unicité des mesures du chap. 1. Comme les mesures sont σ-finies, il existe deux suites croissantes de mesurables $(A_n^1)_{n \geq 0} \in \mathcal{E}_1^{\mathbb{N}}$ et $(A_n^2)_{n \geq 0} \in \mathcal{E}_2^{\mathbb{N}}$ telles que $E_1 = \cup_{n \geq 0} A_n^1$, $E_2 = \cup_{n \geq 0} A_n^2$ et $\mu_1(A_n^1) < \infty$ et $\mu_2(A_n^2) < \infty$ pour tout $n \geq 0$. Par suite $A_n^1 \times A_n^2 \uparrow E_1 \times E_2$ et $\nu(A_n^1 \times A_n^2) = \mu_1(A_n^1)\mu_2(A_n^2) < \infty$. Comme de plus $\mathcal{E}_1 \times \mathcal{E}_2$ est un π-système sur lesquel les valeurs de ν sont fixées, on déduit l'unicité. $\square$

Le problème central est celui de l'existence bien sûr. On considère pour tout $x \in E_1$ et $y \in E_2$:

$$T_x : \begin{array}{ccc} E_2 & \to & E_1 \times E_2 \\ y & \to & (x, y) \end{array} \text{ et } T^y : \begin{array}{ccc} E_1 & \to & E_1 \times E_2 \\ x & \to & (x, y) \end{array}$$

On remarque que T_x et T^y sont mesurables. En effet

$$T_x^{-1}(A_1 \times A_2) = \begin{cases} \emptyset \text{ si } x \notin A_1 \\ A_2 \text{ sinon}\,. \end{cases}$$

Comme les pavés définissent un π-système générateur, on déduit la mesurabilité de T_x. On procède de même pour T^y.

Théorème 22. *1. Soit* $C \in \mathcal{E}_1 \otimes \mathcal{E}_2$*. Alors pour tout* $(x, y) \in E_1 \times E_2$, $C_x = T_x^{-1}(C)$ *et* $C^y = T^{y^{-1}}(C)$ *sont des mesurables (resp. de* $\mathcal{E}_2$ *et* $\mathcal{E}_1$*).*

2. *Si* $f : E_1 \times E_2 \to F$ *mesurable, alors* $f_x \doteq f \circ T_x$ *et* $f^y = f \circ T^y$ *sont mesurables.*
3. *Si* μ_1 *est* σ*-finie sur* $(E_2, \mathcal{E}_2)$*, alors*

$$x \to \mu_2(C_x) \text{ est mesurable par rapport à } \mathcal{E}_1. \tag{5.1}$$

Démonstration. Les points (1) et (2) sont évidents. Pour le point (3), il suffit d'introduire $\Lambda = \{C \in \mathcal{E}_1 \otimes \mathcal{E}_2 |$ la prop (5.1) est vraie $\}$ puis de montrer que c'est un Λ-système qui contient le π-système générateur $\mathcal{E}_1 \times \mathcal{E}_2$. Le théorème de Dynkin donne alors le résultat. $\square$

Théorème 23. *Soient μ_1 et μ_2 deux mesures σ-**finies**. Il existe une unique mesure, notée $\mu_1 \otimes \mu_2$ sur $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ telle que*

$$\mu_1 \otimes \mu_2(A \times B) = \mu_1(A)\mu_2(B)\,.$$

Cette mesure est σ-finie et pour tout $C \in \mathcal{E}_1 \otimes \mathcal{E}_2$:

$$\mu_1 \otimes \mu_2(C) = \int \mu_1(C^y)d\mu_2(y) = \int \mu_2(C_x)d\mu_1(x)\,.$$

Démonstration. L'unicité a déjà été vue, il suffit donc de considérer seulement ici l'existence. On définit alors pour tout $C \in \mathcal{E}_1 \otimes \mathcal{E}_2$:

$$\nu(C) \doteq \int \mu_2(C_x)d\mu_1(x)\,.$$

Il nous suffit de montrer que ν ainsi définie est une mesure positive. Or si $(C_n)_{n\geq 0}$ est une famille disjointe de mesurables, en remarquant que $(\cup_{n\geq 0}C_n)_x = \cup_{n\geq 0}C_{n,x}$ où $C_{n,x} = T_x^{-1}(C_n)$ et que les $C_{n,x}$ sont disjoints, on déduit

$$\begin{aligned}
\nu((\cup_{n\geq 0}C_n) &= \int \mu_2(\cup_{n\geq 0}C_{n,x})d\mu_1(x) && (5.2)\\
&= \int (\sum_{n\geq 0} \mu_2(C_{n,x}))d\mu_1(x) && (5.3)\\
&\overset{cm}{=} \sum_{n\geq 0}\int \mu_2(C_{n,x}))d\mu_1(x) && (5.4)\\
&= \sum_{n\geq 0} \nu(C_n)\,. && (5.5)
\end{aligned}$$

De plus, si on procède en utilisant la formule $\int \mu_1(C^y)d\mu_2(y)$, on vérifie que l'on coïncide sur les pavés $A \times B$ ce qui prouve l'égalité pour tout C. $\square$

On rappelle que si E_1 et E_2 sont deux espaces métriques séparables, alors $\mathcal{B}(E_1 \times E_2) = \mathcal{B}(E_1) \otimes \mathcal{B}(E_2)$.

Remarque 27.
- *Si λ_d est la mesure de Lebesgue sur $\mathbb{R}^d$, alors $\lambda_{d+1} = \lambda_d \times \lambda$ sur $\mathcal{B}(\mathbb{R}^{d+1})$.*
- *Le théorème est faux si on enlève l'hypothèse de σ-finitude : il suffit de prendre pour μ_1 la mesure de comptage sur $\mathbb{R}$ et pour μ_2 la mesure de Lebesgue sur $\mathbb{R}$. Si Δ est la diagonale de $\mathbb{R}^2$, on a*

$$\int_{\mathbb{R}} \lambda(\Delta_x)d\mu_1(x) = 0$$

et

$$\int \mu_1(\Delta^y)d\lambda(y) = \infty\,.$$

5.2 Théorèmes de Fubini

Théorème 24 (Fubini-Tonelli). *On suppose que μ_1 et μ_2 sont deux mesures σ-finies. Soit $f : E_1 \times E_2 \to [0, +\infty]$ mesurable. Alors,*

1. *$x \to \int f(x, y)d\mu_2(y)$ est $\mathcal{E}_1$ mesurable et $y \to \int f(x, y)d\mu_1(x)$ est $\mathcal{E}_2$ mesurable.*

2. *On a*

$$\int f(x,y)d(\mu_1 \otimes \mu_2)(x,y) = \int_{E_1} \left(\int_{E_2} f(x,y) d\mu_2(y) \right) d\mu_1(x)$$
$$= \int_{E_2} \left(\int_{E_1} f(x,y) d\mu_1(x) \right) d\mu_2(y)$$

Remarque 28. *Le théorème justifie la notation*

$$\int\int f(x,y)d\mu_1(x)d\mu_2(y) \doteq \int f(x,y)d(\mu_1 \otimes \mu_2)(x,y)\,.$$

Démonstration. Preuve de (1) : Pour $f = \mathbb{1}_C$ avec $C \in \mathcal{E}_1 \otimes \mathcal{E}_2$, le résultat est vrai d'apres le théorème précédent. Par linéarité, c'est vrai pour f étagée. Enfin, si $f \geq 0$ quelconque, alors il existe une suite croissante $(f_n)_{n\geq 0}$ de fonctions étagées qui convergent simplement vers f. Par convergence monotone, $x \to \int f(x,y)d\mu_2(y)$ est limite simple de fonctions $\mathcal{E}_1$ mesurables donc est $\mathcal{E}_1$ mesurable. Preuve de (2) : On procède de même. Le résultat est vrai pour $f = \mathbb{1}_C$ puis par linéarité pour f étagée puis par convergence monotone pour toutes les limites simples croissantes de fonctions étagées. □

Théorème 25 (Fubini-Lebesgue)**.** *On suppose toujours μ_1 et μ_2 σ-finie. Soit $f : E_1 \times E_2 \to \overline{\mathbb{R}}$ mesurable et intégrable (i.e. $\int |f| d(\mu_1 \otimes \mu_2) < \infty$). Alors,*

1. $x \to \int |f|(x,y)d\mu_2(y) < \infty$ μ_1 *p.p. et* $y \to \int |f|(x,y)d\mu_1(x) < \infty$ μ_2 *presque partout.*
2. *On a*

$$\int f(x,y)d(\mu_1 \otimes \mu_2)(x,y) = \int_{E_1} \left(\int_{E_2} f(x,y) d\mu_2(y) \right) d\mu_1(x)$$
$$= \int_{E_2} \left(\int_{E_1} f(x,y) d\mu_1(x) \right) d\mu_2(y)$$

Remarque 29. *La seule hypothèse pour (2) est l'intégrabilité de f que l'on peut montrer en utilisant Fubini-Tonelli sur $|f|$ pour calculer $\int |f| d(\mu_1 \otimes \mu_2)$ en intégrant par rapport à une variable puis par rapport à l'autre.*

Démonstration. Preuve de (1). Comme $\int |f| d(\mu_1 \otimes \mu_2) < \infty$ et que par Fubini-Tonelli on a $\int |f| d(\mu_1 \otimes \mu_2) = \int_{E_1} \left(\int_{E_2} |f(x,y)| d\mu_2(y) \right) d\mu_1(x)$ on déduit que $x \to \int |f|(x,y)d\mu_2(y) < \infty$ μ_1 presque partout. On procède de même pour $y \to \int |f|(x,y)d\mu_1(x) < \infty$ μ_2 presque partout. Preuve de (2) : Il suffit d'utiliser la partie (2) de Fubini-Tonelli sur un décomposition de $f = f^+ - f^-$. □

Attention, si f n'est pas intégrable, Fubini-Lebesgue ne s'applique plus. On construit un contre-exemple en prenant $E_1 = E_2 = \mathbb{N}$ et $\mu_1 = \mu_2$ mesure de comptage sur $\mathbb{N}$ et

$$f(x,y) = \begin{cases} 1 & \text{si } x = y \\ -1 & \text{si } x = y+1 \\ 0 & \text{sinon }. \end{cases}$$

5.3 Application à l'intégration par parties

Si f et g sont deux fonctions boréliennes localement intégrables sur $\mathbb{R}$ pour la mesure de Lebesgue. On définit alors $F(x) = \int_0^x f(t)dt$ et $G(x) = \int_0^x g(t)dt$. On sait que F et G sont continues [1], dérivables presque partout de dérivées presque partout égales respectivement à f et g. On a alors le théorème d'intégration par partie suivant :

Théorème 26 (Intégration par parties)**.**

$$\int_a^b f(t)G(t)dt = (F(b)G(b) - F(a)G(a)) - \int_a^b F(t)g(t)dt\,.$$

Démonstration. En effet, fG et gF sont localement intégrables sur $\mathbb{R}$ puisque f,g le sont et F et G sont continues. De plus

$$\begin{aligned}\int_a^b f(t)G(t)dt &= \int_a^b f(t)(G(a) + \int_a^t g(u)du)dt \\ &= (F(b) - F(a))G(a) + \int_a^b f(t)(\int_a^t g(u)du)dt\,.\end{aligned}$$

Or que $(t,u) \to f(t)g(u)$ est intégrable sur $[a,b]^2$ puisque par Fubini-Tonelli $\int_a^b \int_a^b |f(t)g(u)|dtdu = \int_a^b |f(t)|dt \int_a^b |(g(u)|du < \infty$. On déduit par Fubini-Lebesgue que

$$\begin{aligned}\int_a^b f(t)(\int_a^t g(u)du)dt &= \int\int_{a\leq u\leq t\leq b} f(t)g(u)dudt \\ &= \int_a^b g(u)(\int_u^b f(t)dt)du \\ &= (G(b) - G(a))F(b) - \int_a^b g(u)F(u)du\,.\end{aligned}$$

□

[1] en fait absolument continue au sens où pour tout $\epsilon > 0$, il existe δ tel que pour toute famille finie d'intervalles ouverts disjoints telle que $\sum_{i=1}^n (b_i - a_i) \leq \delta$, on a $\sum_{i=1}^n |f(b_i) - f(a_i)| \leq \epsilon$

Chapitre 6

Espaces probabilisés, variables aléatoires

6.1 Fragments d'histoire des probabilités

L'émergence au $16^{\text{ième}}$ siècle

- **Une apparition assez tardive :** La théorie des probabilités apparaît plutôt tardivement comme « science du hasard » à partir du $16^{\text{ième}}$ siècle alors que les jeux de hasard existent depuis la plus haute antiquité.
- **Les débuts de la Renaissance :**

Les premiers pas significatifs sont fait par Gerolamo Cardano (connu en France sous le nom de Jérôme Cardan, celui des formules de Cardan) dans *Liber de Ludo Aleae* (achevé en 1563, publié en 1663 !).

Il jette les bases de la méthode « classique »

$$\text{probabilité de gain} = \frac{\#\text{issues favorables}}{\#\text{issues possibles}}$$

chaque issue ayant un « poids » équivalent. Pour un dé, l'ensemble des issues possibles est $\Omega = \{1, 2, 3, 4, 5, 6\}$. La probabilité d'obtenir un nombre ≥ 3 : $4/6 = 2/3$

Remarque 30. – *La mise en œuvre de cette méthode n'est pas sans difficulté. Comme définir l'ensemble des issues possibles ? On lance deux dés : ensemble des paires ordonnées ou non ordonnées ?*
- *Erreur classique est de considérer les paires non ordonnées. On trouve alors* 21 *paires non ordonnées. La probabilité d'obtenir* 6 *est* $p_6 = 3/21$ *(paires* $1-5$, $2-4$, $3-3$*) comme celle d'obtenir* 7 *(paires* $1-6$, $2-5$, $3-4$*) ce qui est contredit par l'expérience : la somme de deux dés vaut plus souvent* 7 *que* 6.
- *Cardano donne la bonne solution : prendre les paires ordonnées* (i, j) *pour lesquelles on obtient* $p_6 = \frac{5}{36} < p_7 = \frac{6}{36}$.
- *Au* $18^{\text{ième}}$ *siècle, d'Alembert tombera à nouveau dans ce piège dans un article de l'encyclopédie de Diderot : la probabilité que le nombre de « face » sur deux lancers de pièces soit* k *vaut* $1/3$ *pour* $k = 0, 1, 2$ *!*

Le travail de Cardano reste sans prolongement jusqu'au $17^{\text{ième}}$ siècle.

La correspondance de Pascal et Fermat

Pascal est en contact avec le Chevalier de Méré, noble et écrivain, à la cour de Louis XIV et grand joueur. Il pose plusieurs problèmes à Pascal qui va les résoudre avec Fermat dans une série de cinq lettres entre 1651 et 1654.

- **Premier problème, le « grand scandale » :** Méré a remarqué que dans le jeu suivant : on lance 4 dés et on gagne si au moins un « 6 » sort est un jeu assez favorable avec lequel il a gagné pas mal d'argent. Pour trouver des partenaires il invente un nouveau jeu censé lui être toujours aussi favorable mais différent[1] : On lance 24 fois un paire de dés et l'on gagne si on sort un « double six ».

  Le calcul de Méré est le suivant : la probabilité d'obtenir un « double 6 » est $1/36$, soit 6 fois moins que celle d'obtenir un « 6 ». En jouant 6 fois plus longtemps ($6 \times 4 = 24$) on doit avoir la même chance de gagner que dans le jeu précédent. Pourtant il remarque que ce nouveau jeu à l'air défavorable d'où le *grand scandale*.

Extrait de la lettre du 29 juillet 1654 de Pascal à Fermat, mentionnant le problème du chevalier de Méré :

« Je n'ai pas eu le temps de vous envoyer la démonstration d'une difficulté qui étonnait fort M. de Méré, car il a très bon esprit, mais il n'est pas géomètre (c'est, comme vous savez, un grand défaut) et même il ne comprend pas qu'une ligne mathématique soit divisible à l'infini et croit fort bien entendre qu'elle est composée de points en nombre fini, et je n'ai jamais pu l'en tirer. Si vous pouviez le faire, on le rendrait parfait.

Il me disait donc qu'il avait trouvé fausseté dans les nombres par cette raison :

Si on entreprend de faire un six avec un dé, il y a avantage de l'entreprendre en 4, comme de 671 à 625.

Si on entreprend de faire Sonnez avec deux dés, il y a désavantage de l'entreprendre en 24.

Et néanmoins 24 est à 36 (qui est le nombre des faces de deux dés) comme 4 à 6 (qui est le nombre des faces d'un dé).

Voilà quel était son grand scandale qui lui faisait dire hautement que les propositions n'étaient pas constantes et que l'arithmétique se démentait : mais vous en verrez bien aisément la raison par les principes où vous êtes. » Pascal et Fermat montrent que le raisonnement du Chevalier est faux et que le jeu devient favorable avec 25 lancers. Il utilisent la méthode classique : $\frac{\#\text{issues favorables}}{\#\text{issues possibles}}$

Exercice 14. *Retrouver le résultat de Pascal et Fermat*

- **Le deuxième problème, « le problème des points » ou « problème des paris » :** Dans un jeu d'argent se faisant en plusieurs manches (disons N), comment répartir les gains si la partie est interrompue avant la fin en fonction des manches déjà remportées. La solution la plus courante à l'époque (c'était celle proposée par Cardan) est de répartir les gains au *pro rata* des manches remportées par les adversaires.

Exemple : Deux joueurs A et B et trois manches pour gagner la partie ($N = 3$). Supposons à l'interruption, A a gagné $n_A = 2$ manches et B $n_B = 1$ manche.

Idée de Fermat

Si A et B jouaient virtuellement encore deux parties, le gagnant serait déterminé de façon sûre. Toutes les sorties possibles sont équiprobables (jeux de hasard).

3 séquences gagnantes pour A : $(AB), (AA), (BA)$; 1 séquence gagnante pour B : (BB).

Conclusion : A doit ramasser $3/4$ et B $1/4$ des mises.

Il s'agit là encore d'une utilisation de la méthode classique dans un cadre non conventionnel et difficile pour l'époque.

Idée de Pascal

Dans la même situation, au coup suivant, si A gagne alors la partie est terminée et si B gagne alors on est à égalité et chacun on doit partager la mise.

[1]On voit en regardant l'activité de la Française des Jeux que les choses n'ont pas beaucoup évolué depuis !

Conclusion, si chacun a mis 32 pistoles, A doit récupérer $\frac{1}{2} \times 64 + \frac{1}{2} \times 32 = 48$ pistoles et B, 16 pistoles.

Prémisses d'un calcul d'espérance. Sa méthode peut être mise sous une forme algorithmique comme le remplissage d'un tableau par propagation à partir de valeurs connues (en gras) donnant en fonction de (n_A, n_B) la proportion des mises que doit prendre A.

	$n_A = 0$	$n_A = 1$	$n_A = 2$	$n_A = 3$
$n_B = 0$	**1/2**	11/16	7/8	**1**
$n_B = 1$	5/16	**1/2**	3/4	**1**
$n_B = 2$	1/8	1/4	**1/2**	**1**
$n_B = 3$	**0**	**0**	**0**	

– Pascal et Fermat arrivent aux mêmes résultats (« Je vois bien que la vérité est la même à Toulouse et à Paris », Lettre de Pascal à Fermat en 1654) mais ils n'explicitent pas vraiment les principes qu'ils suivent.
– Les premiers pas vers une formalisation seront le fait de Christian Huygens en 1657 dans son livre *De ratiocinii in ludo alea* (La théorie des jeux de dès) et peut être à ce titre considéré comme le père de la théorie des probabilités.

Les développements du 18[ième] siècle

A cette approche classique qui ramène les probabilités à la combinatoire ($\frac{\text{\# cas favorables}}{\text{\# cas possibles}}$) s'oppose une approche fréquentiste pour laquelle la probabilité d'une issue particulière dans une expérience aléatoire est la fréquence limite d'occurrences dans la répétition de cette expérience dans des conditions inchangées :

Probabilité « *a priori* » versus *probabilité* « *a posteriori* »

C'est la compréhension profonde des liens entre ces deux aspects qui est l'un des grands succès des développements ultérieurs de la théorie des probabilités.
– Jacques Bernoulli introduit dans « Ars Conjectandis » publié en 1713 (pub. posthume) le concept d'indépendance et de loi faible des grands nombres. Utilisation de méthodes d'analyse. Il y développe également les applications « aux affaires civiles, morales et économiques ».
– Abraham De Moivre, dans « The Doctrine Of Chance » publié en 1718, utilise l'approche analytique (fonctions génératrices) pour étudier les fluctuations autour de la loi de grands nombres (cf Théorème central limite)
– Thomas Bayes dans « An essay Toward Solving a Problem in the Doctrine of Chance » en 1763 développe ce qui sera le champ énorme de l'approche « bayesienne » en statistique. Il introduit les probabilités conditionnelles pour résoudre le problème de la probabilités des causes. Outil fondamental dans toutes les méthodes modernes de diagnostic et de reconnaissance des formes.

Vers la théorie moderne des probabilités

– Pierre Simon Laplace dans son livre « Théorie Analytique des Probabilités » publié en 1812 propose une exposition claire et précise de la théorie des probabilités dans ces composantes analytiques et dans ses applications et qui fera autorité au XIX ième :
« *Ce calcul délicat s'étend aux questions les plus importantes de la vie, qui ne sont, en effet, pour la plupart, que des problèmes de probabilité. Il doit, sous ce rapport, intéresser votre Majesté dont le génie sait si bien apprécier et si dignement encourager tout ce qui peut contribuer au progrès des lumières, et de la prospérité publique.* » (Dédicace à Napoléon)[2]

[2] Laplace changera très opportunément cette dédicace dans les éditions ultérieures après la chute de Napoléon...

– Une certaine stagnation au niveau conceptuel jusqu'à Andrey Kolmogorov qui publie dans « Grundbegriffe der Wahrscheinlichkeitsrechnung » en 1933, l'axiomatique moderne des probabilités en se fondant sur la théorie de la mesure.

6.2 Axiomatique de Kolmogorov

Espaces probabilisés

Definition 26. On appelle *espace probabilisé* la donnée d'un triplet $(\Omega, \mathcal{F}, P)$ où Ω est un ensemble, $\mathcal{F}$ une tribu sur Ω et P une mesure de probabilité sur $(\Omega, \mathcal{F})$.

Remarque 31. *Les trois éléments du triplet jouent un rôle fondamental.*

Univers

L'ensemble Ω représente *l'univers* de l'expérience aléatoire, tous les états possibles dans l'expérience.

– On fait des expériences sur le lancer de 4 dès. Dans ce cas,

$$\Omega_1 = \{1, \cdots, 6\}^4$$

représente toutes les issues possibles de l'expérience aléatoire.

– On lance un dè jusqu'à obtenir 6. On peut prendre

$$\Omega_2 = \{(x_i)_{1 \le i \le n} \,|\, n \in \mathbb{N}_*,\ x_i \in \{1, \cdots, 5\},\ 1 \le i < n,\ x_n = 6\}$$

qui modélise exactement les issues qui nous intéressent. On peut prendre aussi un ensemble plus gros (cf approche de Fermat) en prenant

$$\Omega_3 = \{1, \cdots, 6\}^{\mathbb{N}_*}$$

qui modélise une suite infinie de lancers.

Tribu des événements

– La tribu $\mathcal{F}$ est appelée *tribu des événements*.

C'est une nouveauté par rapport à l'approche historique.

– Du point de vue de la théorie de la mesure, c'est l'ensemble des mesurables.
– On verra plus tard que cette tribu représentera *l'information disponible* sur une expérience. En général, on n'a pas accès directement à l'aléa ω mais aux valeurs de fonctions dépendant de ω (variable aléatoires).
– Les plus simples sont du type $\mathbb{1}_A(\omega)$ qui dit si $\omega \in A$ (on dit alors que *l'événement* A *est réalisé*), ou au contraire si $\omega \notin A$.

Probabilité

– Pour un événement $A \in \mathcal{F}$, $P(A)$ représente la probabilité d'occurrence de l'événement A à interpréter comme la probabilité que l'aléa ω soit dans A.

Remarque 32. *On remarque que l'axiomatique des mesures donne ici*

1. $P(\emptyset) = 0$ *(l'événement vide est de probabilité nulle),* $P(\Omega) = 1$ *(l'événement* Ω *est certain)*

2. $P(A \cup B) = P(A) + P(B)$ *(les probabilités des événements disjoints s'ajoutent)*

La théorie des probabilités ne nous renseigne pas sur la nature *profonde du hasard. Celle-ci lui échappe. Elle propose cependant une axiomatique efficace de calcul sur le hasard. L'appui sur la théorie de la mesure la rend très riche et opérationnelle.*

- Dans le cas de Ω_1, on peut prendre $\mathcal{F}_1 = \mathcal{P}(\Omega)$ et $P_1(\{\omega\}) = 1/6^4$ pour tout ω. L'événement « on obtient au moins un six » sera $A = \{\omega = (x_1, \cdots, x_4) \in \Omega_1 \,|\, \exists i\, x_i = 6\}$ de probabilité $P(A) = 1 - (5/6)^4 = 0,5177$.
- Dans le cas de Ω_2 qui est dénombrable, on peut prendre encore $\mathcal{F}_2 = \mathcal{P}(\Omega_2)$. Pour P, on prendra $P_2(\{(x_1, \cdots, x_n)\}) = 1/6^n$ (et pourquoi donc ?). On vérifie que $P(\Omega_2) = 1$.
- Dans le cas de $\Omega_3 = \{1, \cdots, 6\}^{\mathbb{N}_*}$, on ne peut pas prendre pour $\mathcal{F}$ l'ensemble des parties (pas de probabilités intéressantes sur une si grosse tribu). On introduit pour tout $i \geq 1$: $X_i : \Omega_3 \to \{1, \cdots, 6\}$ telle que $X_i(\omega) = x_i$ (appelée *application coordonnée*). On peut prendre $\mathcal{F}_3 = \sigma(X_i,\, i \geq 1)$ la tribu produit rendant les X_i mesurables et pour P_3

$$P_3(\{\omega \in \Omega_3 \,|\, X_1(\omega) = x_1, \cdots, X_n(\omega) = x_n\}) = 1/6^n \tag{6.1}$$

Exercice 15. *Montrer que* P_3 *est définie de façon unique par (6.1) et que* $\mathcal{F}_2$ *s'identifie avec une sous tribu de* $\mathcal{F}_3$ *sur laquelle* P_2 *et* P_3 *coïncident. (*$(\Omega_3, \mathcal{F}_3, P_3)$ *est un sur-modèle de* $(\Omega_2, \mathcal{F}_2, P_2)$*).*

Remarque 33. *Il y a donc plusieurs façon de modéliser l'aléa pour une même expérience* concrète. *On verra que le bon point de vue est de souvent de prendre des univers assez gros (des ensembles produits) mais de modéliser des expériences aléatoires par des fonctions mesurables de l'aléa.*

Definition 27. Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé et $(E, \mathcal{E})$ un espace mesurable. Si $X : \Omega \to E$ est mesurable, alors X est appelée *variable aléatoire* (v.a. en abrégé) à valeurs dans E.

Remarque 34.
1. *On voit que les variables aléatoires non rien d'aléatoire ! Cependant, l'intuition sous-jacente est que l'aléa* ω *lui-même est* tiré *selon la probabilité* P *et n'est pas connu de l'observateur. Ainsi,* $X(\omega)$ *révèle une information sur* ω*, prenant des valeurs différentes (aléatoire pour l'observateur) selon la valeur de l'aléa.*
2. *On voit que la connaissance de* P *n'est pas nécessaire pour définir les variables aléatoires à valeurs dans* E

Reprenons nos exemples avec le langage des variables aléatoires.
- Exemple 1 : $X_i : \Omega_1 \to \{1, \cdots, 6\}$ avec $X_i(\omega) = x_i$ pour $\omega = (x_1, \cdots, x_4)$. Si $B_i = \mathbb{1}_{X_i=6}$ et $M = \mathbb{1}_{\sum_i B_i \geq 1}$ alors l'événement « on obtient au moins un six »est défini par $A = \{\omega \in \Omega \,|\, M(\omega) = 1\}$
- Exemple 3 : Si $\tau(\omega) = \inf\{n \geq 1 \,|\, X_n(\omega) = 6\}$ où $X_n(\omega) = x_n$ pour $\omega = (x_i)_{i \geq 1} \in \Omega_3$, τ est une v.a. à valeurs dans $\overline{\mathbb{N}}$ (le vérifier). On a $P_3(\{\omega \in \Omega \,|\, \tau(\omega) = k\}) = (\frac{5}{6})^{k-1}\frac{1}{6}$ si bien que $P_3(\{\omega \in \Omega \,|\, \tau(\omega) = \infty\}) = 0$.

Definition 28. Soit X une variable aléatoire à valeur dans E. On appelle loi de X la mesure image de X, notée P_X, de P par X et vérifiant

$$P_X(A) = P(X^{-1}(A)) = P(\{\omega \in \Omega \,|\, X(\omega) \in A\}) \,.$$

Remarque 35.
- *Les probabilistes préfèrent utiliser la notation* $P(X \in A)$ *pour* $P(\{\omega \in \Omega \,|\, X(\omega) \in A\})$*. On écrira ainsi* $P(\tau = k)$ *pour* $P(\{\omega \in \Omega \,|\, \tau(\omega) = k\})$ *dans l'exemple 3 ou encore* $P(X_1 = x_1, \cdots X_n = x_n)$ *pour* $P(\cap X_i^{-1}(\{x_i\}))$
- *La disparition de* ω *dans la notation* $P(X \in A)$ *s'accorde bien avec l'intuition qui entoure le concept de variable aléatoire (voir remarque précédente).*

6.3 Espérance , moments d'une v.a.r

6.3.1 Espérance d'une v.a.r.

Definition 29 (Espérance)**.** Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé et X une variable aléatoire à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On dit que X est intégrable si $\int |X(\omega)| dP(\omega) < \infty$. On note alors

$$E(X) \doteq \int X(\omega) dP(\omega)$$

appelée *espérance* de X.

Remarque 36. *1. La définition s'étend à X à valeurs dans $\mathbb{R}^d$ en calculant l'espérance coordonnées par coordonnées.*

2. *On note $L^1(\Omega, \mathcal{F}, P)$ l'ensemble des variables aléatoires intégrables définies à p.p. près (classes d'équivalences déjà vues dans le chapitre sur les espaces L^p). Lorsque $(\Omega, \mathcal{F}, P)$ sera fixé, on notera souvent $L^1 = L^1(\Omega, \mathcal{F}, P)$.*
3. *Dans le contexte probabiliste, $P(\Omega) = 1$ et on utilise l'expression « presque sûrement » (abrégé p.s.) à la place de presque partout. Presque sûrement veut donc dire avec probabilité 1.*
4. *Si X est une v.a.r. presque sûrement positive, on peut alors toujours calculer son espérance qui peut valoir $+\infty$ le cas échéant.*

Proposition 15 (Changement de variable)**.** *Soit X une variable aléatoire à valeur dans E (muni d'une tribu $\mathcal{E}$) et $f : E \to \overline{\mathbb{R}}$ mesurable tel que $f \geq 0$ p.p. ou $E(|f(X)|) < \infty$. Alors on a*

$$E(f(X)) = \int_E f(x) dP_X(x) .$$

Remarque 37. *La proposition montre que la donnée de la loi P_X permet de calculer toutes les espérance des fonctions dépendant de X sans avoir besoin de connaître Ω et P.*

Démonstration. Il suffit de montrer le résultat pour $f \geq 0$, le cas général s'obtenant par décomposition $f = f^+ - f^-$.

Si $f = \mathbb{1}_A$ pour $A \in \mathcal{E}$. Alors

$$E(f(X)) = P(X \in A) = P_X(A) = \int \mathbb{1}_A(x) dP_X(x) = \int f(x) dP_X(x) .$$

Par linéarité, c'est vrai pour f étagée. Pour $f \geq 0$ mesurable quelconque, il existe une suite croissante f_n de fonctions étagées telle que $f = \lim f_n$. Par convergence monotone, on obtient le résultat. □

6.3.2 Moments, inégalités de moments

Definition 30 (Moments d'ordre p)**.** Soit $p \in \mathbb{N}_*$ et $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle. Si $E(|X|^p) < \infty$ on dit que $X \in L^p(\Omega, \mathcal{F}, P)$ (abrégé en $X \in L^p$) et on appelle *moment d'ordre* p la quantité $E(X^p)$.

Remarque 38. *Dans le cas des mesures finies (c'est évidemment le cas ici), nous savons que les espaces L^p sont emboîtés ie $L^\infty \subset \cdots \subset L^1$.*

Definition 31. Soit $X \in L^2$. On appelle *variance* de X, notée $V(X)$ la quantité

$$V(X) \doteq E((X - E(X))^2) = E(X^2) - E(X)^2 .$$

On appelle *écart-type*, $\sigma = \sqrt{V(X)}$.

L'écart type est une mesure de dispersion autour de l'espérance. On établit ici une inégalité souvent extrêmement utile :

Théorème 27 (Inégalité de Jensen)**.** *Soit* $X \in L^1$ *et* $\phi : \mathbb{R} \to \mathbb{R}$ *convexe telle que* $E(|\phi(X)|) < \infty$*. Alors* $\phi(E(X)) \leq E(\phi(X))$.

Démonstration. Soit $(a, b) \in \mathbb{R}^2$ tels que $ax + b \leq \phi(x)$ pour tout $x \in \mathbb{R}$. On a alors $aE(X) + b \leq E(\phi(X))$. Or

$$\phi(y) = \sup_{a,b \,|\, l_{ab} \leq \phi} l_{ab}(y) \text{ avec } l_{ab}(x) = ax + b$$

d'où $\phi(E(X)) \leq E(\phi(X))$. □

Exercice 16. *Retrouver en utilisant l'inégalité de Jensen que* $\|X\|_p \leq \|X\|_q$ *pour tous* $1 \leq p \leq q \leq \infty$ *où* $\|X\|_r = E(|X|^r)^{1/r}$.

Voici deux inégalités, très simples mais également extrêmement utiles.

Proposition 16. *1. Soit* $X \geq 0$ *p.s. une variable aléatoire. Alors pour tout* $\alpha > 0$ *on a*

$$P(X \geq \alpha) \leq \frac{E(X)}{\alpha} \text{ (Inégalité de Markov).}$$

2. Soit $X \in L^2$ *et* $m = E(X)$*. Alors pour tout* $\alpha > 0$*, on a*

$$P(|X - m| \geq \alpha) \leq \frac{V(X)}{\alpha^2} \text{ (Inégalité de Bienaymé-Chebyshev).}$$

Démonstration. 1. $\alpha \mathbb{1}_{X \geq \alpha} \leq X$ p.s. d'où le résultat en prenant l'espérance.

2. $P(|X - m| \geq \alpha) = P(|X - m|^2 \geq \alpha^2)$ puis inégalité de Markov.

□

6.4 Quelques lois classiques

Les distributions ou lois classiques des variables aléatoires sont organisées en lois discrètes et lois continues.

- Une loi discrète est associée à une variable aléatoire qui prend ses valeurs dans un ensemble dénombrable E. On note souvent alors $p_x \doteq P(X = x)$ pour tout $x \in E$ et on a $P_X = \sum_{x \in E} p_x \delta_x$. C'est le cadre des probabilités dites *discrètes*. L'espérance de $g(X)$ se calcule alors par la sommation :
$$E(g(X)) = \sum_{x \in E} p_x g(x)$$
- Le cas des lois continues (on devrait plutôt dire absolument continues) correspond aux lois sur $\mathbb{R}^d$ absolument continues par rapport à la mesure de Lebesgue λ_d. Il existe alors par le théorème de Radon-Nikodym une dérivée $f \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ telle que $P_X = f\lambda_d$ appelée *densité* de la loi. On a alors pour tout $g : \mathbb{R}^d \to \mathbb{R}$ positive ou tel que $E(|g(X)|) < \infty$:
$$E(g(X)) = \int_\Omega g(X(\omega))dP(\omega) = \int_{\mathbb{R}^d} g(x)f(x)dx\,.$$

Remarque 39. *Cette dichotomie discret-continu n'épuise pas évidemment toutes les possibilités. On peut avoir des mélanges discret-continu, des lois non-discrètes mais singulières par rapport à Lebesgue.*

Lois discrètes

La plus simple :

Loi de Bernoulli

$B(p)$ à valeurs dans $\{0, 1\}$: $P(X = 1) = p$ $P(X = 0) = 1 - p$. Modélise les phénomènes aléatoires à issue binaire (pile ou face, réponse oui/non dans à une question dans un sondage, absence ou présence d'un défaut dans un objet, etc). $E(X) = p,\ V(X) = p(1 - p)$.

Les variables de Bernoulli apparaissent de façon massive dans les probabilités ne serait-ce qu'en seuillant une v.a.r : $\mathbb{1}_{X\geq\lambda}$ à pour loi $B(p)$ avec $p = P(X \geq \lambda)$. Elles constituent des sortes de briques élémentaires.

Exercice 17. *La loi de Bernoulli de paramètre* $1/2$ *modélise évidemment le lancer d'une pièce de monnaie équilibré. Lorsque le paramètre* p *est quelconque, trouver un algorithme construit sur des lancers successifs d'une telle pièce de monnaie dont l'issue* 0 *ou* 1 *serait modélisable par une* $B(p)$ *(autrement on cherche un modèle physique pour une Bernoulli de paramètre* p *pour un* p *quelconque). En cherchant bien, on peut trouver un algorithme qui nécessite en moyenne seulement deux lancers de la pièce...*

Exercice 18. *Trouver au moins une façon de choisir au hasard et uniformément* $\omega \in \Omega = \{1, \cdots, n\}$ *à l'aide d'une pièce de monnaie. On pourra chercher par exemple la façon la plus économe en lancers de pièce ou encore une façon décentralisée (*n *participants ayant chacun une pièce veulent tirer au sort l'un d'entre eux).*

Loi binomiale

$B(n, p)$ à valeurs dans $\{1, \cdots, n\}$: $P(S = i) = C_n^i p^i (1 - p)^{n-i}$, pour $i = 1, \cdots, n$. Correspond au nombre de pile obtenus dans n lancers d'une pièce équilibrée. $E(S) = np,\ V(S) = np(1 - p)$.

Remarque 40. *Si* $\Omega = \{0, 1\}^n$ *et* $P(\{\omega\}) = \prod_{i=1}^n p^{x_i}(1 - p)^{1-x_i}$ *pour* $\omega = (x_1, \cdots, x_n)$*, alors* $P(X_1 = x_1, \cdots, X_n = x_n) = \prod_{i=1}^n p^{x_i}(1 - p)^{1-x_i}$ *où* X_i *est la* $i^{\text{ième}}$ *application coordonnée.*

On a donc $P(X_i = x_i) = p^{x_i}(1 - p)^{1-x_i}$ *d'où* $X_i \sim B(p)$*. Or si* $S = \sum_{i=1}^n X_i$*,* S *est de loi* $B(n, p)$ *car*

$$P(S = k) = \sum_{\sum x_i = k} p^k (1 - p)^{n-k} = C_n^k p^k (1 - p)^{n-k}$$

On retrouve alors $E(S) = \sum_{i=1}^n E(X_i) = np$.

Loi de Poisson $\mathcal{P}(\lambda)$

de paramètre $\lambda > 0$: $P(X = k) = e^{-\lambda}\lambda^k/k!$ pour $k \in \mathbb{N}$. Modélise le nombre d'événements (rares) se produisant dans une période de temps donnée, le nombre de "clients" arrivant dans une file d'attente. $E(X) = \lambda,\ V(X) = \lambda$

- Nb de photons reçu par un détecteur en scintigraphie (bruit Poissonien)
- Nb de fois que l'on obtient un « triple 6 » en lançant $6 \times 6 \times 6 = 216$ fois les 3 dès ! (Cf Chevalier de Méré) est très bien approché par une loi de Poisson : On montre en exercice que $B(n, \lambda_n) \xrightarrow{\text{loi}} \mathcal{P}(\lambda)$ dans le sens que pour tout $k \geq 0$:

$$C_n^k \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}$$

On déduit d'ailleurs que la façon de raisonner du Chevalier de Méré n'était pas si fausse (raisonnement proportionnel) au moins asymptotiquement...

Loi géométrique
$\mathcal{G}(p)$ de paramètre $p \in]0,1[$ à valeurs dans $\mathbb{N}^*$: $P(X = i) = (1-p)^{i-1}p$ pour $i \in \mathbb{N}^*$. Correspond au temps du premier succès dans une suite d'épreuves de Bernoulli indépendantes (cf chapitre suivant) de paramètre p. $E(X) = 1/p, V(X) = (1-p)/p^2$.

C'est la loi obtenue dans l'exemple 2 de l'introduction avec $p = 1/6$.

Loi exponentielle $\mathcal{E}(\lambda)$ de paramètre $\lambda > 0$
Loi continue de densité

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{x \in [0,+\infty[} .$$

Modélise bien certaines durées d'événements aléatoires (temps de désintégration d'un atome, la durée d'une connexion à un serveur, le temps de vie d'une machine,...). Le paramètre λ est la fréquence moyenne (nombre d'événements par unité de temps). Possède la propriété d'absence de mémoire ou de non vieillissement : $P(X \geq t+s)/P(X \geq t) = P(X \geq s)$. $E(X) = \frac{1}{\lambda},\ V(X) = \frac{1}{\lambda^2}$.

Loi gaussienne ou normale $\mathcal{N}(\mu, \sigma^2)$ (avec $\mu \in \mathbb{R}$ et $\sigma > 0$)
Loi continue de densité

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) .$$

Bien adaptée à la modélisation du résultat d'une mesure d'un objet variable et/ou entachée d'une erreur de mesure. Rôle central en statistique (TCL). $E(X) = \mu,\ V(X) = \sigma^2, \Phi(t) \doteq E(e^{itX}) = \exp(it\mu - \frac{t^2\sigma^2}{2})$.

6.5 Fonctions de répartitions, fonctions caractéristiques, fonctions génératrices

Definition 32 (fonction de répartition)**.** Pour toute v.a.r., on appelle *fonction de répartition* de X la donnée de $F_X : \mathbb{R} \to [0,1]$ définie par $F_X(t) = P(X \leq t) = P_X(]-\infty, t])$.

Proposition 17. *La fonction F_X est continue à gauche limitée à droite (càglàg), croissante, admet la limite 0 en $-\infty$, 1 en $+\infty$ et détermine de façon unique P_X.*

Démonstration. Déjà vu dans la partie intégration (cf Théorème 12). □

Definition 33 (Fonction caractéristique)**.** Soit X une v.a. à valeur dans $\mathbb{R}^d$. On appelle *fonction caractéristique* de X, notée ϕ_X, la fonction $\phi_X : \mathbb{R}^d \to \mathbb{C}$ définie par

$$\phi_X(\xi) \doteq \int_{\mathbb{R}^d} e^{i\langle x,\xi\rangle} dP_X(x) = E(e^{i\langle X,\xi\rangle}).$$

Remarque 41.
1. *ϕ_X n'est rien d'autre que la transformée de Fourier de la loi P_X.*
2. $\phi_X(0) = 1$ *et* $|\Phi_X(\xi)| \leq E(|e^{i\langle\xi,X\rangle}|) = 1$.
3. $|\phi_X(\xi+h) - \phi_X(\xi)| \leq E(|e^{i\langle h,X\rangle} - 1|)$. *Or* $|e^{i\langle h,X\rangle} - 1| \leq 2$ *et* $e^{i\langle h,X\rangle} \to 1$ *lorsque* $h \to 0$. *Par suite, par convergence dominée,* $h \to E(|e^{i\langle h,X\rangle} - 1|)$ *est continue en* $h = 0$ *et* $\phi_X(\xi)$ *est uniformément continue.*

Proposition 18. *Si* $X \sim \mathcal{N}(\mu, \sigma^2)$ *alors* $\phi_X(\xi) = \exp(i\xi\mu - \frac{\xi^2\sigma^2}{2})$.

Démonstration. Par changement de variable $x' = \frac{(x-\mu)}{\sigma}$,

$$\begin{aligned}
\phi_X(\xi) &\doteq \int_{\mathbb{R}} e^{i\xi x} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \frac{dx}{\sqrt{2\pi\sigma^2}} \\
&= \exp(i\xi\mu) \int_{\mathbb{R}} e^{i\sigma\xi x} e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \\
&= \exp(i\xi\mu - \frac{\sigma^2\xi^2}{2}) I(\xi)
\end{aligned}$$

avec $I(\xi) \doteq \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-i\sigma\xi)^2}{2}} dx$.
Or, en appliquant le théorème de dérivation sous le signe somme, on montre facilement que I est dérivable et $I'(\xi) = \int i(x - i\sigma\xi) e^{-\frac{(x-i\sigma\xi)^2}{2}} \frac{dx}{\sqrt{2\pi}} = 0$. Par suite I est constante et pour $\xi = 0$ on a

$$I(0)^2 \overset{\text{fub.}}{=} \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-(x^2+y^2)/2} \frac{dx dy}{2\pi} \overset{\text{pol.}}{=} \int_{\mathbb{R}^+} r e^{-r^2/2} dr = 1 ,$$

la dernière égalité s'obtenant par un passage en coordonnées polaires. □

Théorème 28. *Soient X_1 et X_2 deux variables aléatoires à valeurs dans $\mathbb{R}^d$. Si $\phi_{X_1} = \phi_{X_2}$ alors $P_{X_1} = P_{X_2}$.*

Démonstration. On fait la preuve dans le cas $d = 1$, le cas général se faisant de façon similaire.

Pour tout $\sigma > 0$, on note $f_{\sigma,j}(x) \doteq \int \frac{\exp(-\frac{(x-y)^2}{2\sigma^2})}{\sqrt{2\pi\sigma^2}} dP_{X_j}(y)$ pour $1 \le j \le 2$. En notant $P^\sigma_{X_j} \doteq f_{\sigma,j}\lambda$, l'idée de la preuve est de montrer que

1. $P^\sigma_{X_1} = P^\sigma_{X_2}$ pour tout $\sigma > 0$
2. pour tout $h \in C_c(\mathbb{R}, \mathbb{R})$, $\int h dP_{X_1} = \lim_{\sigma\to 0} \int h dP^\sigma_{X_1} = \int h dP_{X_2}$.

Montrons le point 1) en vérifiant que $f_{\sigma,1} = f_{\sigma,2}$. En effet,

$$\begin{aligned}
f_{\sigma,1}(y) &= \int_{\mathbb{R}} (\int_{\mathbb{R}} e^{i\frac{(x-y)}{\sigma}t} e^{-t^2/2} \frac{dt}{2\pi\sigma}) dP_{X_1}(x) \\
&\overset{\text{F.-L.}}{=} \int_{\mathbb{R}} e^{-i\frac{y}{\sigma}t} e^{-t^2/2} \phi_{X_1}(\frac{t}{\sigma}) \frac{dt}{2\pi\sigma} \\
&= \int_{\mathbb{R}} e^{-i\frac{y}{\sigma}t} e^{-t^2/2} \phi_{X_2}(\frac{t}{\sigma}) \frac{dt}{2\pi\sigma} \\
&= f_{\sigma,2}(y) .
\end{aligned}$$

Pour le dernier point, on a pour $h \in C_c(\mathbb{R}, \mathbb{R})$:

$$|\int_{\mathbb{R}} h(x) f_{\sigma,1}(x) dx - \int_{\mathbb{R}} h(y) dP_{X_1}(y)| \overset{\text{F.-L.}}{=} |\int_{\mathbb{R}} (\int_{\mathbb{R}} (h(x) - h(y)) \exp(-\frac{(x-y)^2}{2\sigma^2}) \frac{dx}{\sqrt{2\pi\sigma^2}}) dP_{X_1}(y)| \to 0$$

par convergence dominée lorsque $\sigma \to 0$. □

Remarque 42. *On a formellement $E(e^{i\langle \xi, X\rangle}) = E(\sum_{n=0}^{\infty} \frac{(i\langle \xi, X\rangle)^n}{n!}) = \sum_{n=0}^{\infty} \frac{i^n}{n!} E(\langle \xi, X\rangle^n)$. Evidemment il faut des conditions de finitude des moments pour que cela ait des chances d'être vérifié. Le théorème suivant va dans cette direction.*

Théorème 29. *On suppose que $E(|X|^2) < \infty$. Alors, $\xi \to \phi_X(\xi)$ est C^2 et*

$$\phi_X(\xi) = 1 + iE(\langle \xi, X\rangle) - \frac{1}{2} E(\langle \xi, X\rangle^2) + o(|\xi|^2) .$$

Démonstration. On notant $\xi = (\xi_1, \cdots \xi_d)$ on a $\frac{\partial}{\partial \xi_k} e^{i\langle \xi, X\rangle} = iX_k e^{i\langle \xi, X\rangle}$ d'où $|\frac{\partial}{\partial \xi_k} e^{i\langle \xi, X\rangle})| \le |X|$ qui est intégrable (puisque $E(|X|)^2 \le E(|X|^2)$). Par théorème de dérivation sous le signe somme et le théorème de continuité des intégrales dépendant d'un paramètre, on déduit que ϕ_X est C^1. En dérivant à nouveau, on obtient $\frac{\partial^2}{\partial \xi_k \partial \xi_l} e^{i\langle \xi, X\rangle} = -X_k X_l e^{i\langle \xi, X\rangle}$. D'après la domination $|\frac{\partial^2}{\partial \xi_k \partial \xi_l} e^{i\langle \xi, X\rangle}| \le |X_k X_l|$, on déduit que ϕ_X est deux fois dérivable, puis C^2. Le résultat s'obtient alors par un développement de Taylor à l'ordre 2. □

Definition 34. Soit X une v.a. à valeurs dans $\mathbb{N}$. On appelle fonction génératrice de X, la fonction $G_X : [0,1] \to \mathbb{R}^+$ définie par

$$G_X(t) \doteq E(t^X) = \sum_{n \ge 0} t^n P(X = n)\,.$$

Remarque 43. *On remarque que*

- G_X *est croissante sur* $[0,1]$, $G_X(0) = P(X = 0)$, $G_X(1) = 1$ *;*
- *Comme* G_X *est une série entière de rayon de convergence au moins* 1, G_X *est* C^∞ *sur* $[0,1[$.
- G_X *admet une dérivée à gauche en* 1 *par convergence monotone puisque* $t \to \frac{1-t^n}{1-t}$ *est croissante. Cette limite vaut* $E(X)$.

Plus généralement on a le résultat suivant :

Théorème 30. *Soit* X *une v.a. à valeurs dans* $\mathbb{N}$ *de fct génératrice* G_X*. Alors pour tout* $k \ge 1$*,*

$$E(\prod_{i=0}^{k-1}(X - i)) = \lim_{t \to 1, t<1} G_X^{(k)}(t)$$

Remarque 44. *Tous les moments de* X *sont donc déterminés par la donnée de* G_X*. Enfin, comme* $G_X^{(k)}(0) = k!P(X = k)$*, on déduit que la loi de* X *est entièrement déterminée par* G_X*.*

Exercice 19. *Calculer* G_X *pour* $X \sim B(n,p)$*,* $X \sim \mathcal{P}(\lambda)$*, et* $X \sim \mathcal{G}(p)$*.*

Démonstration. On vérifie facilement que $G_X^{(k)}(t) = \sum_{n \ge k} \prod_{i=0}^{k-1}(n - i) t^{n-k} P(X = n)$ qui converge par convergence monotone vers

$$\sum_{n \ge k} \prod_{i=0}^{k-1}(n - i) P(X = n) = \sum_{n \ge k} E\left(\prod_{i=0}^{k-1}(X - i)\mathbb{1}_{X=n}\right) \stackrel{\text{cm}}{=} E(\prod_{i=0}^{k-1}(X - i))\,.$$

□

Chapitre 7

Evenements, tribus, variables aléatoires indépendantes

7.1 Définitions

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

Definition 35. Soit $(A_i)_{i\in I}$ une famille d'événements. On dit que la famille est une famille d'*événements indépendants* si pour tout $J \in \mathcal{P}_f(I)$ on a

$$P(\cap_{j\in J} A_j) = \prod_{j\in J} P(A_j)$$

où $\mathcal{P}_f(I)$ est l'ensemble des parties finies de J.

Remarque 45. *Dans le cas fondamental de deux événements A_1 et A_2, $P(A_1 \cap A_2) = P(A_1)P(A_2)$ i.e. $P(A_2) = P(A_1 \cap A_2)/P(A_1)$ (en supposant $P(A_1) > 0$). Autrement dit, sachant que A_1 est réalisé, la probabilité* conditionnelle *que A_2 soit réalisé est la même que la probabilité de A_2* a priori. *La réalisation de A_1 n'apporte pas d'information sur la réalisation de A_2.*

Exemple 4. *On lance deux dès :*
- *La réalisation d'un événement concernant le premier lancé n'affecte pas la probabilité d'un événement concernant le deuxième (contrairement à une croyance assez repandue). En effet, si $\Omega = \{1, \cdots, 6\}^2$, X_1 et X_2 sont les deux projections canoniques et P est la loi uniforme sur Ω alors pour $A_1 = (X_1 \in B_1)$ et $A_2 = (X_2 \in B_2)$ on a $P(A_2 \cap A_2) = |B_1||B_2|/36 = P(A_1)P(A_2)$.*
- *Par contre, les deux événements $A_1 = (X_1 \in B_1)$ et $A_3 = (X_1 + X_2 \in B_2)$ ne sont généralement pas indépendants (prendre $B_1 = \{3\}$ et $B_2 = \{3, 4\}$). Ici la connaissance de A_1 donne des informations sur la réalisation de A_3.*

Definition 36. Soit $(\mathcal{A}_i)_{i\in I}$ une famille de sous tribus de $\mathcal{F}$. On dit que cette famille est une famille de tribus indépendantes si pour tout $(A_i)_{i\in I}$ telle que $A_i \in \mathcal{A}_i$, les événements sont indépendants.

Remarque 46. *Il suffit bien sûr de montrer que pour tout $J \in \mathcal{P}_f(I)$, pour toute famille $(A_j)_{j\in J}$ avec $A_j \in \mathcal{A}_j\ \forall j$, on a $P(\cap_{j\in J} A_j) = \prod_{j\in J} P(A_j)$.*

Definition 37. Soit $(X_i)_{i\in I}$ une famille de variable aléatoire sur $(\Omega, \mathcal{F}, P)$ à valeurs dans des espaces mesurables $(E_i, \mathcal{E}_i)$. On dit que les variables aléatoires sont indépendantes si la famille $(\sigma(X_i))_{i\in I}$ des tribus l'est.

Remarque 47. *1. L'indépendance des X_i porte sur les tribus engendrées (sur Ω) et non sur les valeurs proprement dites de ces variables aléatoires. Par suite, si $\Phi_i : (E_i, \mathcal{E}_i) \to (E'_i, \mathcal{E}'_i)$ l'indépendance des variables aléatoires de la famille $(X_i)_{i\in I}$ implique celle des $(\phi(X_i))_{i\in I}$.*

2. La vérification de l'indépendance de la famille se ramène à montrer que pour toute sous famille finie $(X_j)_{j\in J}$, on a $P(\cap_{j\in J}(X_j \in A_j)) = \prod_{j\in J} P(X_j \in A_j)$ avec $A_j \in \mathcal{E}_j$, pour tout $j \in J$.

Exemple 5. *On reprend l'exemple de $\Omega = \{0,1\}^{\mathbb{N}^*}$, $\mathcal{F} = \sigma(X_n, n \geq 1)$ où X_n désigne la nième projection canonique pour tout $n \geq 1$ et P est l'unique mesure de probabilité sur $(\Omega, \mathcal{F})$ telle que $P(X_1 = b_1, \cdots X_n = b_n) = 2^{-n}$ avec $(b_1, \cdots, b_n) \in \{0,1\}^n$. On vérifie facilement que la famille $(X_n)_{n\geq 0}$ est une famille de variables aléatoires, indépendantes. En effet, $P(X_1 \in B_1, \cdots, X_n \in B_n) = |B_1| \times \cdots \times |B_n|/2^n = \prod_{i=1}^n P(X_i \in B_i)$.*

7.2 Caractérisations

Proposition 19. *Soit $(X_i)_{1\leq i\leq n}$ une famille de variables aléatoires avec $X_i : (\Omega, \mathcal{F}) \to (E_i, \mathcal{E}_i)$. Soit $X \doteq (X_1, \cdots, X_n) : (\Omega, \mathcal{F}) \to (\prod_{i=1}^n E_i, \otimes_{i=1}^n \mathcal{E}_i)$ telle que $X(\omega) = (X_1(\omega), \cdots, X_n(\omega))$. Alors $(X_i)_{1\leq i\leq n}$ est une famille de v.a. indépendantes ssi $P_X = \otimes_{i=1}^n P_{X_i}$.*

Démonstration. Soit $(A_i)_{1\leq i\leq n} \in \prod_{i=1}^n \mathcal{E}_i$.

($\Rightarrow$) Si $(A_i)_{1\leq i\leq n} \in \prod_{i=1}^n \mathcal{E}_i$, alors $P_X(\prod_{i=1}^n A_i) = P(\cap_{i=1}^n (X_i \in A_i)) = \prod_{i=1}^n P(X_i \in A_i) = \prod_{i=1}^n P_{X_i}(A_i)$.

($\Leftarrow$) $P(\cap_{i=1}^n (X_i \in A_i)) = P_X(A_1 \times \cdots A_n) = \prod_{i=1}^n P_{X_i}(A_i) = \prod_{i=1}^n P(X_i \in A_i)$.

□

Corollaire 6. *1. Soit $(X_i)_{1\leq i\leq n}$ une famille de v.a.r. . On suppose que pour tout $1 \leq i \leq n$, $P_{X_i} \ll \lambda$ de densité f_{X_i}. Alors si $X = (X_1, \cdots, X_n)$, $P_X \ll \lambda_n$ de densité $f_X(x_1, \cdots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$ (noté parfois $f_X = \otimes_{i=1}^n f_{X_i}$) pour tout $(x_1 \cdots, x_n) \in \mathbb{R}^n$.*

2. Soit $X = (X_1, \cdots, X_n)$ v.a. à valeurs dans $\mathbb{R}^n$ et $P_X \ll \lambda_n$ de densité f_X se décomposant en $f_X(x_1, \cdots, x_n) = f_1(x_1) \cdots f_n(x_n)$ où f_i est une densité de v.a.r. (i.e. $f_i \geq 0$ λ p.p. et $\int f_i(x_i)dx_i = 1$). Alors $(X_i)_{1\leq i\leq n}$ est une famille de v.a.r. indépendantes et pour tout $1 \leq i \leq n$, $P_{X_i} = f_i\lambda$.

Démonstration. 1. En effet pour tout $(A_i)_{1\leq i\leq n} \in \mathcal{B}(\mathbb{R})^n$,

$$\begin{aligned} P_X(\prod_{i=1}^n A_i) &= \prod_{i=1}^n P_{X_i}(A_i) \\ &= \prod_{i=1}^n \int_{A_i} f_{X_i}(x_i)dx_i \\ &\overset{\text{fub}}{=} \int_{A_1\times\cdots\times A_n} \prod_{i=1}^n f_{X_i}(x_i) \prod_{i=1}^n dx_i \end{aligned}$$

Par suite P_X coïncide avec $\otimes_{i=1}^n f_{X_i}\lambda_n$ sur un π-sytème génératif, ce qui donne le résultat par unicité.

2. On vérifie que pour tout $A_i \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned} P(X_i \in A_i) &= \int_{\mathbb{R}^n} \mathbb{1}_{x_i\in A_i} \prod_{j=1}^n f_j(x_j)dx_1 \cdots dx_n \\ &\overset{\text{fub}}{=} \int_{A_i} f_i(x_i)dx_i \end{aligned}$$

Par suite $P_{X_i} = f_i\lambda$. On déduit alors

$$\begin{aligned} P(X_1 \in A_1, \cdots, X_n \in A_n) &= \int_{A_1\times\cdots\times A_n} \prod_{i=1}^n f_i(x_i)dx_1\cdots dx_n \\ &\overset{\text{fub}}{=} \prod_{i=1}^n P(X_i \in A_i)\,. \end{aligned}$$

□

Exercice 20. *On dit que le vecteur aléatoire X à valeurs dans $\mathbb{R}^n$ suit une loi $\mathcal{N}(0,\Gamma)$ où Γ est symétrique, définie, positive si*

$$f_X(x) = \frac{\exp(-\langle\Gamma^{-1}x, x\rangle/2)}{\sqrt{det(\Gamma)(2\pi)^n}}\,.$$

1. *Vérifier que f_X est bien une densité sur $\mathbb{R}^n$;*
2. *verifier que pour toute racine carré $\Gamma^{1/2}$ de Γ, les coordonnées $(Y_1, \cdots, Y_n)$ du vecteur aléatoire $Y = \Gamma^{-1/2}X$ sont indépendantes et de loi $\mathcal{N}(0,1)$.*

Corollaire 7. *Soit $(X_i)_{1\le i\le n}$ une famille de v.a.r. . Les assertions suivantes sont équivalentes :*

1. *$(X_i)_{1\le i\le n}$ est une famille de variables aléatoires indépendantes ;*
2. *pour toute famille $(f_i)_{1\le i\le n}$ de fonctions boréliennes positives $E(\prod_{i=1}^n f_i(X_i)) = \prod_{i=1}^n E(f_i(X_i))$;*
3. *$\phi_X(\xi) = E(e^{i\langle\xi,X\rangle}) = \prod_{i=1}^n \phi_{X_i}(\xi_i)$, pour tout $\xi = (\xi_1, \cdots, \xi_n) \in \mathbb{R}^n$ i.e. $\phi_X = \otimes_{i=1}^n \phi_{X_i}$.*

Démonstration.
– (1 ⇒ 2) On a directement le résultat pour $f_i = \mathbb{1}_{A_i}$ puis par linéarité, on a le résultat pour les f_i étagées positives. On termine par densité en utilisant une convergence monotone.
– (2 ⇒ 3) En découpant en partie positive et en partie négative, on étend immédiatement le résultat à une famille $(f_i)_{1\le i\le n}$ de fonctions telle que $E(|f_i(X_i)|) < \infty$ pour tout $1 \le i \le n$. En décomposant en partie réelle et en partie imaginaire, on obtient aussi le résultat pour f_i à valeurs complexe et P_{X_i}-intégrable. Par suite 3) est vraie.
– (3 ⇒ 1) On vérifie que si $\mu \doteq \otimes_{i=1}^n P_{X_i}$ alors la transformée de Fourier de μ coïncide avec celle de P_X :

$$\begin{aligned} \int_{\mathbb{R}^n} e^{i\langle\xi,x\rangle}d\mu(x) &= \int_{\mathbb{R}^n} e^{i\langle\xi,x\rangle}d(\otimes_{i=1}^n P_{X_i})(x) \\ &\overset{\text{fub}}{=} \prod_{i=1}^n \int_{\mathbb{R}} e^{i\xi_i x_i}dP_{X_i}(x_i) \\ &= \prod_{i=1}^n \phi_{X_i}(x_i)\,. \end{aligned}$$

Par injectivité de la transformée de Fourier on a le résultat.

□

Proposition 20. *Soit $(\mathcal{A}_i)_{i\in I}$ une famille de sous tribus de $\mathcal{F}$ telle que pour tout $i \in I$, $\mathcal{A}_i = \sigma(\mathcal{C}_i)$ où $\mathcal{C}_i$ est un π-système générateur. On suppose que pour tout $J \in \mathcal{P}_f(I)$, et tout $(A_j)_{j\in J} \in \prod_{j\in J}\mathcal{C}_j$ on a*

$$P(\cap_{j\in J}A_j) = \prod_{j\in J} P(A_j)\,. \tag{7.1}$$

Alors, $(\mathcal{A}_i)_{i\in I}$ est une famille de tribus indépendantes.

Démonstration. Soit $J = \{j_1, \cdots, j_n\} \in \mathcal{P}_f(I)$. Pour tout $0 \le k \le n$, on note $J_0 = \emptyset$ et $J_k = \{j_1, \cdots, j_k\}$ pour $1 \le k \le n$. Montrons par récurrence sur $0 \le k \le n$ que si $A_j \in \mathcal{A}_j$ pour tout $j \in J_k$ et $A_j \in \mathcal{C}_j$ pour $j \in J \setminus J_k$ alors (7.1) reste vrai.

Le cas $n = 0$ correspond à l'hypothèse. Si c'est vrai pour $k < n$, alors en fixant $A_j \in \mathcal{A}_j$ pour $j \in J_k$ et $A_j \in \mathcal{C}_j$ pour $j \in J \setminus (J_k \cup \{j_{k+1}\})$ on a que $\Lambda = \{A_{j_{k+1}} \in \mathcal{A}_{j_{k+1}} \mid$ (7.1) est vraie $\}$ est un λ-système qui contient le π-système $\mathcal{C}_{j_{k+1}}$. Par suite le résultat est vrai par le théorème de Dynkin. La proposition se déduit du cas $k = n$. □

Corollaire 8 (Théorème des coalitions)**.** *Soient* $(X_i)_{i \in I}$ *une famille de variables aléatoires indépendantes et* $(I_k)_{k \in K}$ *une partition de* I *en sous ensembles disjoints. On note* $\mathcal{B}_k = \sigma(X_i,\ i \in I_k)$. *Alors la famille* $(\mathcal{B}_k)_{k \in K}$ *est une famille de tribus indépendantes (regroupement par blocs).*

Démonstration. On notant $\mathcal{C}_k = \{\cap_{j \in J}(X_j \in A_j) \mid J \in \mathcal{P}_f(I_k), A_j \in \mathcal{E}_j, \forall j \in J\}$, $\mathcal{C}_k$ est un π-système qui engendre $\mathcal{B}_k$. Or pour tout $H \in \mathcal{P}_f(K)$ et tout $C_h = \cap_{j \in J_h}(X_j \in A_j) \in \mathcal{C}_h$ pour $h \in H$, on a $P(\cap_{h \in H} C_h) = P(\cap_{h \in H} \cap_{j \in J_h} (X_j \in A_j)) = \prod_{h \in H, j \in J_h} P(X_j \in A_j) = \prod_{h \in H} P(C_h)$. On déduit donc le résultat de la proposition précédente. □

7.3 Sommes de variables aléatoires indépendantes

Proposition 21. *Soient* X *et* Y *deux v.a. indépendantes à valeurs dans* $\mathbb{R}^d$. *Alors*

1. *Pour tout* $A \in \mathcal{B}(\mathbb{R}^d)$, $P_{X+Y}(A) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{1}_{x+y \in A} dP_X(x) dP_Y(y)$. *On note* $P_{X+Y} = P_X \star P_Y$.
2. *Si* $P_X = f_X \lambda_d$ *et* $P_Y = f_Y \lambda_d$ *alors* $P_{X+Y} = (f_X \star f_Y)\lambda_d$ *avec* $f_X \star f_Y(z) = \int f_X(z-y) f_Y(y) dy = \int f_X(x) f_Y(z-x) dx$.
3. $\phi_{X+Y}(\xi) = \phi_X(\xi)\phi_Y(\xi)$ *i.e.* $\phi_{X+Y} = \phi_X \phi_Y$.

Démonstration.

1. On sait que $P_{(X,Y)} = P_X \otimes P_Y$. Par suite $P_{X+Y}(A) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{1}_{x+y \in A} dP_{(X,Y)}(x, y) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{1}_{x+y \in A} dP_X(x) dP_Y(y)$.
2. On écrit $P_{X+Y}(A) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{1}_{x+y \in A} f_X(x) f_Y(y) dx dy$. Par le changement de variable $(z, y) = (x + y, y)$, on obtient le résultat.
3. $E(e^{i\langle \xi, X+Y\rangle}) = E(e^{i\langle \xi, X\rangle} e^{i\langle \xi, Y\rangle})$. Comme X et Y sont indépendantes, $E(e^{i\langle \xi, X\rangle} e^{i\langle \xi, Y\rangle}) = \phi_X(\xi)\phi_Y(\xi)$. □

Exemple 6.

- *Soit* X *et* Y *deux v.a.r. gaussiennes indépendantes, alors* $X + Y$ *est une variable aléatoire gaussienne de loi* $\mathcal{N}(E(X) + E(Y), V(X) + V(Y))$ *: en effet, on sait que si* $Z \sim \mathcal{N}(\mu, \sigma^2)$ *alors* $\phi_Z(t) = e^{it\mu - t^2\sigma^2/2}$. *Or* $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t) = \exp(itE(X) - t^2V(X)/2)\exp(itE(Y) - t^2V(Y)/2)$ *d'où* $\phi_{X+Y}(t) = \exp(it(E(X) + E(Y)) - t^2(V(X) + V(Y))/2)$.
- *Faux si pas indépendance : prendre* X *et* $Y = -X$ *avec* $V(X) > 0$. *Mieux encore, en prenant* B *telle que* $P(B = 1) = P(B = -1) = 1/2$ *avec* B *indépendant de* X *puis* $Y = BX$, *alors* $Y \overset{\text{loi}}{=} X$ *mais* $P(X + Y = 0) = 1/2$ *(quelle est exactement la loi de* $X + Y$ *?).*
- *Si* $X \sim \mathcal{P}(\lambda)$ *et* $Y \sim \mathcal{P}(\mu)$, *avec* X *et* Y *indépendantes, alors* $X + Y \sim \mathcal{P}(\mu + \lambda)$. *En effet, il suffit de remarquer que si* $Z \sim \mathcal{P}(\lambda)$ *alors* $\phi_Z(t) = \exp(\lambda(e^{it} - 1))$ *puis d'appliquer la proposition précédente.*

Exercice 21. *Montrer que si* $(X_i)_{1 \le i \le n}$ *est une famille de variables aléatoires indépendantes et identiquement distribuée (ie de même loi) de loi commune* $B(p)$, *alors* $\sum_{i=1}^n X_i \sim B(n, p)$.

Proposition 22. *Soient* X *et* Y *deux v.a.r. indépendantes et dans* L^2. *Alors* $X + Y \in L^2$ *et* $V(X + Y) = V(X) + V(Y)$.

Démonstration. En écrivant $(X+Y)^2 \leq 2(X^2+Y^2)$ on déduit que $X+Y \in L^2$. De plus $E((X+Y-E(X+Y))^2) = V(X) + V(Y) + 2E((X-E(X))(Y-E(Y))$. Or par indépendance, $E((X-E(X))(Y-E(Y)) = E(X-E(X))E(Y-E(Y)) = 0$. □

Corollaire 9 (Loi faible des grands nombres)**.** *Soit* $(X_n)_{n\geq 1}$ *une suite de v.a.r. indépendantes et dans* L^2 *telle que* $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n E(X_n) = \mu$ *et* $\sup_{n\geq 1} V(X_n) \leq \sigma^2 < \infty$. *Alors,*

1. $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i \overset{L^2}{=} \mu$ *;*

2. $P(|\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n E(X_i))| > \alpha) \leq \frac{\sigma^2}{n\alpha^2}$ *pour tout* $\alpha > 0$.

Démonstration. 1. On a vérifié que si $X \in L^2$, alors $E((X-a)^2) = V(X) + (E(X)-a)^2$ (il suffit de développer le carré). Par suite

$$E((\frac{1}{n}\sum_{i=1}^n X_i - \mu)^2) = V(\frac{1}{n}\sum_{i=1}^n X_i) + (\frac{1}{n}\sum_{i=1}^n E(X_i) - \mu)^2 .$$

Or par indépendance,

$$V(\frac{1}{n}\sum_{i=1}^n X_i) = \frac{1}{n^2}\sum_{i=1}^n V(X_i) \leq \sigma^2/n .$$

2. Par l'inégalité de Markov,

$$P(|\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n E(X_i))| > \alpha) \leq \frac{1}{\alpha^2}V(\frac{1}{n}\sum_{i=1}^n X_i) \leq \frac{\sigma^2}{n\alpha^2} .$$

□

Remarque 48. *1. C'est un résultat très important. Il montre que le fait de moyenner un grand nombre de quantité aléatoires* indépendantes *de variances bornées* concentre *les valeurs mayennes observées autour de la moyenne des espérances. Dans le cas d'une suite i.i.d. de variable aléatoires* L^2 *d'espérance commune* m*, la limite est* m*. Dans le cas d'une suite de lancers de dés,*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{X_i=4}$$

converge dans L^2 *vers* $1/6$.

2. Le résultat précise la vitesse de convergence ; en effet, si n *est fixé, alors en choisissant* $\epsilon > 0$ *assez petit et* α *tel que* $\frac{\sigma^2}{n\alpha^2} = \epsilon$ *ie* $\alpha = \frac{\sigma}{\sqrt{n\epsilon}}$ *alors on sait avec proabilité* $1-\epsilon$ *que la moyenne empirique* $\overline{X}_n \doteq \frac{1}{n}\sum_{i=1}^n X_i$ *est dans l'intervalle* $[\frac{1}{n}\sum_{i=1}^n E(X_i) \pm \frac{\sigma}{\sqrt{n\epsilon}}]$. *La quantité,* $\frac{1}{n}\sum_{i=1}^n E(X_i)$ *est estimable empiriquement avec une précision donnée en observant une réalisation de* $\overline{X}_n(\omega)$.

"On one occasion he was assigned (by some unkown and unfathomable process) to teach an introductory service course in statistics to a class of undergraduates comprising basketball players, footballers, and other assorted attendees, none of whom was attending UNC, Chapel Hill with the primary intent of learning statistics. It is not clear who received the greater shock from the semester's experience : the students or Wassily."

Proposition 23 (Inégalité de Hoeffding (1963 !))**.** *Soit* $(X_i)_{1\leq i\leq n}$ *une famille de v.a. indépendantes à valeurs dans l'intervalle* $[a,b]$. *Alors pour tout* $\epsilon > 0$

$$P(\overline{X}_n \geq E(\overline{X}_n) + \epsilon) \leq \exp(-2\frac{n\epsilon^2}{(b-a)^2})$$

et

$$P(\overline{X}_n \leq E(\overline{X}_n) - \epsilon) \leq \exp(-2\frac{n\epsilon^2}{(b-a)^2})$$

où $\overline{X}_n \doteq \frac{1}{n}\sum_{i=1}^n X_i$.

Démonstration. En posant $Y_i = X_i - E(X_i)$, on a pour tout $s > 0$,

$$\begin{aligned} P(\sum_{i=1}^n Y_i \geq n\epsilon) &\leq \exp(-sn\epsilon)E(\exp(s\sum_{i=1}^n Y_i)) \\ &= \prod_{i=1}^n E(\exp(sY_i - s\epsilon)) \end{aligned}$$

Or en notant $\phi_i(s) \doteq \log(E(\exp(sY_i)))$, on a $\phi_i(0) = 0$, $\phi_i'(s) = \int Y_i \frac{e^{sY_i}}{E(e^{sY_i})} dP = \int Y_i dP_s = E_s(Y_i)$ où $P_s = \exp(sY_i - \Phi_i(s))P$ est une nouvelle probabilité sur $(\Omega, \mathcal{F})$. De plus

$$\phi"_i(s) = \int Y_i^2 \frac{e^{sY_i}}{E(e^{sY_i})} dP - (\int Y_i \frac{e^{sY_i}}{E(e^{sY_i})} dP)^2 = V_s(Y_i)$$

où $V_s(Y_i)$ désigne la variance de Y_i calculée sous la probabilité P_s. Or comme

$$V_s(Y_i) \leq E_s((Y_i - \frac{a+b}{2} - E(X_i))^2) \leq (b-a)^2/4$$

on déduit que $\phi_i"(s) \leq \frac{(b-a)^2}{4}$ pour tout $s > 0$. Par un développement de Taylor en 0, on déduit finalement $\phi_i(s) \leq \frac{s^2(b-a)^2}{8}$ ie $E(e^{sY_i}) \leq \exp(s^2\frac{(b-a)^2}{8})$.
On a donc $P(\sum_{i=1}^n Y_i \geq n\epsilon) \leq \exp(-n(s\epsilon - \frac{s^2(b-a)^2}{8}))$. En optimisant en s, on obtient le résultat pour $s = \frac{\epsilon}{4(b-a)^2}$. □

– Ce type d'inégalité est un exemple d'inégalité de concentration qui donne des bornes explicites à n fixé de la probabilité de déviation par rapport à la l'espérance d'une quantité aléatoire constituée d'un grand nombre d'éléments indépendants (inégalités de Bennett, Bernstein, Mc Diarmid...).

– Il est remarquable que les hypothèses sur la loi des (X_i) sont très faibles : $P(a \leq X_i \leq b) = 1$ et indépendance. Cela permet d'être applicable pour calibrer des tests statistiques pour l'analyse de données.

Exercice 22. *Pierre Simon Laplace dans son traité de "Théorie analytique des probabilités" rapporte page 577 :" Dans l'espace des 95 années écoulées depuis le commencement de 1664 jusqu'à la fin 1758, il est né à Londres 737629 garçons et 698958 filles ; ce qui donne 19/18 à peu près pour le rapport des naissances des garçons à celle des filles". On veut tester si la supériorité du nombre de naisances de garçons sur celui des filles est significatif. Utiliser l'inégalité de Hoeffding pour cela.*

Chapitre 8

Théorèmes limites

8.1 Lemmes de Borel-Cantelli et loi du 0-1

Lemme 1 (Borel-Cantelli)**.** *Soit* $(A_n)_{n\geq 0}$ *une suite d'évènements. On note* $\overline{\lim} A_n = \cap_{n\geq 0} \cup_{k\geq n} A_k$. *On a alors :*

1. *si* $\sum_{n\geq 0} P(A_n) < \infty$*, alors* $P(\overline{\lim} A_n) = 0$ *i.e.*

$$\sum_{n\geq 0} \mathbb{1}_{A_n} < \infty, \ P \text{ p.s.} ;$$

2. *si de plus les événements sont indépendants, alors si* $\sum_{n\geq 0} P(A_n) = \infty$ *on a* $P(\overline{\lim} A_n) = 1$ *i.e.*

$$\sum_{n\geq 0} \mathbb{1}_{A_n} = \infty, \ P \text{ p.s.} .$$

Démonstration. 1. Il suffit de remarquer que $E(\sum_{n\geq 0} \mathbb{1}_{A_n}) = \sum_{n\geq 0} P(A_n) < \infty$, ce qui donne le résultat.

2. On remarque que par indépendance,

$$P(\cap_{k=n}^{p} A_k^c) = \prod_{k=n}^{p} P(A_k^c) = \prod_{k=n}^{p} (1 - P(A_k)) \leq \exp(-\sum_{k=n}^{p} P(A_k)) \to 0$$

lorsque $p \to \infty$. Par suite, $P(\cap_{k\geq n} A_k^c) = 0$ pour tout $n \geq 0$ d'où $P((\overline{\lim} A_n)^c) = 0$ ce qui donne le résultat.

□

Applications immédiates

Singe dactylographe
Si $(X_n)_{n\geq 1}$ est une suite i.i.d. de loi $B(p)$. Pour tout motif $(\epsilon_1, \cdots, \epsilon_p) \in \{0,1\}^p$, en notant $A_n = (X_n = \epsilon_1, \cdots, X_{n+p-1} = \epsilon_p)$, on a $\sum_{n\geq 1} P(A_n) = \infty$ et avec probabilité 1, la suite $(X_n(\omega))_{n\geq 1}$ contient une infinité de fois le motif.

Loi forte des grand nombres pour les v.a. bornées
Si $(X_n)_{n\geq 1}$ est une suite de v.a.r. indépendantes telles que $P(|X_n| \leq M) = 1$ pour tout $n \geq 1$, alors avec probabilité 1 on a $\sum_{i=1}^{n} (X_i - E(X_i))/n \to 0$

En effet, par le théorème de Hoeffding, pour tout $\epsilon > 0$ on a : $\sum_{n=1}^{\infty} P(|\sum_{i=1}^{n}(X_i - E(X_i))/n| \geq \epsilon) \leq 2\sum_{n=1}^{\infty} \exp(-2n\epsilon^2/(2M)^2) < +\infty$. En appliquant Borel-Cantelli 1), on déduit que P p.s. $|\sum_{i=1}^{n}(X_i - E(X_i))/n| < \epsilon$ pour n assez grand.

Moyenne empirique de v.a.r. non intégrables
Si $(X_n)_{n\geq 1}$ une suite de v.a.r.i.i.d. positives telles que $E(X_1) = \infty$ alors presque sûrement, si $\overline{X}_n \doteq \frac{1}{n}\sum_{i=1}^{n} X_i$, $\overline{\lim}\, \overline{X}_n = +\infty$.

– Il suffit de remarquer que pour tout $M \geq 0$, $\sum_{n\geq 1} P(X_n \geq Mn) = \sum_{n\geq 1} P(X_1 \geq Mn) \geq E(X_1 - 1)/M = +\infty$. En utilisant le point 2) de Borel-Cantelli, on déduit $\overline{\lim}_{n\to\infty} \overline{X}_n \geq M$.

Théorème 31 (Loi du 0-1 ou du “tout” ou “rien”)**.** *Soit* $(X_n)_{n\geq 1}$ *une famille de variables aléatoires indépendantes. On note* $\mathcal{F}_n^+ \doteq \sigma(X_k,\ k \geq n)$ *pour tout* $n \geq 1$*, et* $\mathcal{F}_\infty^+ \doteq \cap_{n\geq 1}\mathcal{F}_n^+$ *appelée tribu asymptotique. Alors, pour tout* $A \in \mathcal{F}_\infty^+$

$$P(A) \in \{0, 1\}.$$

Remarque 49. *$\mathcal{F}_\infty^+$ est la tribu des événements asymptotiques (qui ne dépendent que du comportement asymptotique de X_n). C'est un résultat en apparence assez paradoxal : c'est parce que les variables sont indépendantes (et donc imprévisibles sachant le passé) que les événements asymptotiques sont tous déterministes.*

Remarque 50. *Une conséquence importante est que toute variable aléatoire $\mathcal{F}_\infty^+$-mesurable à valeurs dans $\overline{\mathbb{R}}$ est presque sûrement constante (il suffit de considérer* $c = \inf\{t \in \mathbb{R} \,|\, P(X \leq t) = 1\}$*).*

Démonstration. On montre que pour tout $A \in \mathcal{F}_\infty^+$, A est indépendant de lui-même d'où $P(A) = P(A \cap A) = P(A)^2$ ce qui donne le résultat.

En effet, si $\mathcal{F}_n \doteq \sigma(X_k,\ k \leq n)$, alors $\mathcal{F}_n$ et $\mathcal{F}_{n+1}^+$ sont indépendantes (thm des coalitions) et donc $\mathcal{F}_n$ et $\mathcal{F}_\infty^+$ sont indépendantes.

Comme $\cup_{n\geq 1}\mathcal{F}_n$ est un π-système générant $\mathcal{F}_\infty \doteq \sigma(X_k,\ k \geq 1) = (\mathcal{F}_1^+)$, on déduit que $\mathcal{F}_\infty$ est indépendant de $\mathcal{F}_\infty^+$. Comme $\mathcal{F}_\infty^+ \subset \mathcal{F}_\infty$, on obtient le résultat. □

Limite d'une suite i.i.d.
Si $(X_n)_{n\geq 0}$ est une suite i.i.d. de v.a.r. alors $\overline{\lim}\, X_n$ et $\underline{\lim}\, X_n$ sont $\mathcal{F}_\infty^+$-mesurables et donc constantes presque sûrement. Par suite, soit il existe $c \in \overline{\mathbb{R}}$ tel que $P(\lim X_n = c) = 1$ soit la limite n'existe presque sûrement jamais. On a le résultat analogue pour la limite de $\sum_{k=1}^{n} X_k$.

Comportement limite d'une marche aléatoire symétrique
Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. intégrables i.i.d. telles que $P(X_1 = 1) = P(X_1 = -1) = 1/2$. On note pour tout $n \geq 0$, $S_n = \sum_{i=1}^{n} X_i$. Alors P presque sûrement : $\overline{\lim}\, S_n = +\infty$ et $\underline{\lim}\, S_n = -\infty$.

En effet, si $P(\overline{\lim}\, S_n < +\infty) = 1$, par symétrie, $P(\underline{\lim}\, S_n > -\infty) = 1$ et donc il existe $c > 0$ tel que que $P(\overline{\lim}\, |S_n| \leq c) \geq 1/2$. Or en prenant $M \geq 2c + 1$, on a pour tout n, $P(X_n = \cdots = X_{n+M-1} = 1) = 2^{-M}$ et (singe dactylographe) pour presque tout $\omega \in \Omega$ il existe une infinité de valeurs de n pour lesquelles $X_n(\omega) = \cdots = X_{n+M-1}(\omega) = 1$ ce qui entraîne $\overline{\lim}\, |S_n| > c$. Par conséquent, $P(\overline{\lim}\, |S_n| \leq c) = 0$ ce qui est une contradiction.

8.2 Loi forte des grands nombres

Théorème 32 (Inégalité maximale de Kolmogorov)**.** *Soit* $(X_n)_{n\geq 1}$ *une suite de variables aléatoires* centrées *(i.e.* $E(X_n) = 0$*) indépendantes et de carré intégrable (i.e.* $E(X_n^2) < \infty$ *pour tout* $n \geq 1$*). On pose* $S_0 = 0$ *et* $S_n = \sum_{i=1}^{n} X_i$ *pour tout* $n \geq 1$*. Alors pour tout* $x > 0$*, on a*

$$P\left(\max_{0\leq k\leq n} |S_k| \geq x\right) \leq \frac{E(S_n^2)}{x^2}.$$

Démonstration. Soit $\tau_x = \inf\{k \geq 0 \mid |S_k| \geq x\}$. On vérifie que τ_x est une v.a. à valeurs dans $\mathbb{N}$ puisque pour tout $k \geq 0$, $(\tau = k) \in \sigma(X_i,\ i \leq k)$. De plus on a $E(S_n^2) \geq E(S_n^2 \sum_{k=0}^n \mathbf{1}_{\tau=k})$. Or

$$E(S_n^2 \mathbf{1}_{\tau=k}) = E((S_k + S_n - S_k)^2 \mathbf{1}_{\tau=k}) \geq E(S_k^2 \mathbf{1}_{\tau=k}) + 2E((S_n - S_k) S_k \mathbf{1}_{\tau=k}) \,.$$

Comme $S_n - S_k$ et $S_k \mathbf{1}_{\tau=k}$ sont indépendants, on a $2E((S_n - S_k) S_k \mathbf{1}_{\tau=k}) = 2E(S_n - S_k) E(S_k \mathbf{1}_{\tau=k}) = 0$ puisque $E(S_n - S_k) = 0$. Par suite

$$E(S_n^2 \mathbf{1}_{\tau=k}) \geq E(S_k^2 \mathbf{1}_{\tau=k}) \geq x^2 E(\mathbf{1}_{\tau=k})$$

et $E(S_n^2) \geq x^2 E(\sum_{k=0}^n \mathbf{1}_{\tau=k}) = x^2 P\left(\max_{0 \leq k \leq n} |S_n| \geq x\right)$ ce qui donne le résultat. $\square$

Théorème 33 (Convergence de séries aléatoires)**.** *Soit $(X_n)_{n \geq 0}$ une suite i.i.d. de v.a.r. telle que $E(X_n) = 0$ pour tout $n \geq 0$ et $\sum_{n=0}^\infty E(X_n^2) < \infty$. Alors p.s. la série $\sum_{n=0}^\infty X_n$ est convergente.*

Remarque 51.
- *On étend directement le résultat au cas où $\sum E(X_n)$ est convergente.*
- *Si on considère un terme général aléatoire du type $X_n = B_n / n^\alpha$ où $(B_n)_{n \geq 0}$ est une suite i.i.d. de variables de Rademacher (i.e. $P(B_n = 1) = P(B_n = -1) = 1/2$, alors pour $\alpha > 1/2$ la série convergence p.s. On remarque que la série n'est absolument convergente que si $\alpha > 1$.*

Démonstration. Soit $x > 0$. Pour tout $m \in \mathbb{N}$, on a en utilisant l'inégalité maximale de Kolmogorov que

$$P(\sup_{n \geq m} |S_n - S_m| \geq x) \leq \frac{1}{x^2} \sum_{k=m+1}^{\infty} E(X_k^2) \to 0 \,.$$

Ainsi, $P(\sup_{p,q \geq m} |S_p - S_q| \geq 2x) \to 0$ et comme $A_m = (\sup_{p,q \geq m} |S_p - S_q| \geq 2x)$ est une famille décroissante d'événement, $P(\cap A_m) = 0$ ce qui implique que p.s. il existe $m \geq 0$ tel que $\sup_{p,q \geq m} |S_p - S_q| < 2x$. Comme x est arbitraire, en considérant une famille décroissante x_p tendant vers 0, nous avons montré que p.s. la suite $n \to S_n$ est de Cauchy. $\square$

Pour obtenir notre première preuve de la loi des grands nombres, il nous faut le

Lemma 38 (Lemme de Kronecker)**.** *Soient $(a_n)_{n \geq 1}$ et $(x_n)_{n \geq 1}$ deux suites telles que $(a_n)_{n \geq 1}$ est strictement positive, croissante et tendant vers $+\infty$. Alors*

$$\sum_{n \geq 1} \frac{x_n}{a_n} \; cv \; \Rightarrow \lim_{n \to \infty} \frac{1}{a_n} \sum_{i=1}^n x_i = 0 \,.$$

Démonstration. Notons $A_n = \sum_{k=1}^n \frac{x_k}{a_k}$ pour tout $n \geq 0$ (en particulier $A_0 = 0$) et A_∞ sa limite. On a alors

$$\sum_{i=1}^n x_i = \sum_{i=1}^n a_i (A_i - A_{i-1}) = A_n a_n - \sum_{i=1}^{n-1} A_i (a_{i+1} - a_i) \,.$$

Or $A_n \to A_\infty$ et par moyenne de Césaro, $\sum_{i=1}^{n-1} A_i (a_{i+1} - a_i) / a_n \to A_\infty$. $\square$

Théorème 34 (Loi forte des grands nombres)**.** *Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. i.i.d. telle que $E(|X_1|) < \infty$. Alors P p.s.*

$$\frac{1}{n} \sum_{i=1}^n X_i \to E(X_1) \,.$$

Remarque 52.
- *La loi forte des grands nombres est un résultat clef pour "montrer" la pertinence du formalisme des variables aléatoires pour calculer sur le hasard. Il fait le lien, comme la loi faible, entre moyenne empirique (accessible à l'expérience) et espérance c'est à dire entre données et modèles ou fréquences et probabilités.*

– *Ce très joli résultat est un outil formidable pour comprendre les phénomènes limites. Une limitation cependant est qu'il n'est pas "testable" en lui-même du point de vue pratique car il est asymptotique. Les résultats à horizon fini du type loi faible des grands nombres ou Hoeffding sont plus "opérationnels" de ce point de vue.*

Démonstration. On remarque que $\sum_{i=1}^{\infty} P(|X_i| \geq i) \leq E(|X_i|) < \infty$. Par suite, en notant $Y_n = X_n \mathbf{1}_{|X_n|\leq n}$, on déduit du premier lemme de Borel-Cantelli que p.s. les suites Y_n et X_n coïncident à partir d'un certain rang. Montrons maintenant que

$$\sum_{n\geq 0} E(Y_n^2/n^2) < \infty \tag{8.1}$$

ce qui donnera en utilisant le théorème de convergence des séries aléatoires que $\sum_{n\geq 1} \frac{Y_n - E(Y_n)}{n}$ est convergente p.s. et donc en utilisant le lemme de Kronecker que

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - E(Y_i)) \to 0\,.$$

Comme par convergence dominée on a $E(Y_n) = E(X_1 \mathbf{1}_{|X_1|\leq n}) \to E(X_1)$, nous obtiendrons le résultat. Pour prouver (8.1), on remarque en utilisant Fubini-Tonelli que $E(Y_n^2) = E(\int \mathbf{1}_{y\leq |Y_n|} 2y\,dy) = \int 2yP(|Y_n| \geq y)dy = \int_0^n 2yP(|Y_n| \geq y)dy$. Par suite, en utilisant encore Fubini-Tonelli et le fait que pour $y \leq n$ on a $P(|Y_n| \geq y) = P(|X_n| \geq y) = P(|X_1| \geq y)$, on déduit

$$\sum_{n\geq 1} E(\frac{Y_n^2}{n^2}) = \sum_{n=1}^{\infty}\int_0^{\infty} y\frac{\mathbf{1}_{y\leq n}}{n^2}P(|X_n| \geq y)dy = \int_0^{\infty}\sum_{n\geq y\vee 1}\frac{y}{n^2}P(|X_1| \geq y)dy \leq CE(|X_1|)\,,$$

puisque $\sup_{y\geq 1}\sum_{n\geq y}\frac{y}{n^2} < +\infty$. □

Une démonstration directe qui n'utilise pas le théorème de convergence des séries aléatoires est possible comme proposé par exemple dans le poly de Jean-François Le Gall. Nous la donnons ici :

Démonstration. On pose $S_n \doteq \sum_{i=1}^{n} X_i$ et on va montrer que pour tout $a > E(X_1)$, on a

$$P(\sup_{n\geq 1}(S_n - na) < \infty) = 1$$

ce qui assure que $\overline{\lim}\, S_n/n \leq E(X_1)$ et le résultat par symétrie.

En notant

$$M_n \doteq \sup_{1\leq k\leq n}(S_k - ka) \text{ et } M_n' \doteq \sup_{2\leq k\leq n+1}(\sum_{i=2}^{k} X_i - (k-1)a)\,,$$

les deux familles $M \doteq (M_n)_{n\geq 1}$ et $M' \doteq (M_n')_{n\geq 1}$ ont même loi. Supposons que $M_n \to +\infty$ p.s., alors $M_n' \to +\infty$ p.s. De plus, comme

$$M_{n+1} = (X_1 - a) \vee (X_1 - a + M_n')$$

on en déduit

$$M_{n+1} = (X_1 - a) + (M_n')^+\,.$$

Par suite $E(M_{n+1} - (M_n')^+) = E(X_1 - a) < 0$. Comme $E((M_n')^+) = E(M_n^+)$ et

$$M_{n+1} - (M_n)^+ \geq (X_1 - a)\mathbb{1}_{M_n\leq 0}$$

(en effet, si $M_n > 0$, alors $M_{n+1} - (M_n)^+ = M_{n+1} - M_n \geq 0$ et si $M_n \leq 0$, alors $M_{n+1} - (M_n)^+ = M_{n+1} \geq (X_1 - a)$) on déduit

$$\underline{\lim}\, E(M_{n+1} - (M'_n)^+) \geq \underline{\lim}\, E((X_1 - a)\mathbb{1}_{M_n \leq 0}) \stackrel{\text{cd}}{=} 0$$

car $M_n \to +\infty$ ce qui contredit $E(M_{n+1} - (M'_n)^+) = E(X_1 - a) < 0$. Par suite $P(\sup_{n\geq 1} M_n < +\infty) = 1$ ce qui était annoncé. □

8.3 Convergence en loi et théorème de Levy

Definition 39. Soit E un espace métrique et $\mathcal{B}(E)$ la tribu des boréliens. Soit $(\mu_n)_{n\geq 0}$ une suite de probabilités boréliennes sur E. On dit que μ_n converge étroitement vers $\mu_\infty \in \mathcal{M}_1(E)$ si pour toute fonction continue bornée $f \in C_b(E, \mathbb{R})$ on a $\int f d\mu_n \to \int f d\mu_\infty$. On notera alors $\mu_n \Longrightarrow \mu_\infty$.

Remarque 53. *On étend la notion à convergence en loi de variables aléatoires en considérant la convergence étroite des lois : Une suite de variables aléatoires $(X_n)_{n\geq 0}$ à valeurs dans E converge en loi vers X_∞, noté $X_n \xrightarrow{\mathcal{L}} X_\infty$, si pour tout $f \in C_b(E, \mathbb{R})$,*

$$E(f(X_n)) \to E(f(X_\infty)) .$$

Proposition 24. *Les assertions suivantes sont équivalentes :*

1. $\mu_n \Longrightarrow \mu_\infty$
2. *pour tout ouvert U, $\mu_\infty(U) \leq \underline{\lim}\, \mu_n(U)$*
3. *pour tout fermé F, $\mu_\infty(F) \geq \overline{\lim}\, \mu_n(F)$*
4. *pour tout $A \in \mathcal{B}(E)$ tel que $\mu_\infty(\partial A) = 0$, on a $\mu_\infty(A) = \lim_{n\to\infty} \mu_n(A)$.*

Remarque 54.
- *La perte de masse sur les ouverts s'illustre facilement en considérant $A = \mathbb{R}^*$ et $\mu_n \doteq \delta_{1/n}$ pour tout $n \geq 1$.*
- *Dans le cas discret, il suffit de montrer $\mu_n(x) \to \mu_\infty(x)$ pour tout $x \in E$. Ainsi, au sens de la convergence en loi, on a $B(n, \lambda/n) \xrightarrow{\mathcal{L}} \mathcal{P}(\lambda)$.*

Démonstration. On vérifie facilement que 2) ⇔ 3) et que 2) + 3) ⇒ 4). Il nous suffit donc de montrer 1) ⇒ 3) et 4) ⇒ 1).

1) ⇒ 3) : Soit F un fermé et $f_p(x) \doteq (1 - p d(F, x)) \vee 0$ une approximation continue de $\mathbb{1}_F$ par valeurs supérieures. On a

$$\mu_\infty(f_p) = \lim_{n\to\infty} \mu_n(f_p) \geq \overline{\lim}\, \mu_n(F)$$

car $f_p \geq \mathbb{1}_F$. Comme par c.d. on a $\mu_\infty(F) = \lim_{p\to\infty} \mu_\infty(f_p)$, on déduit le résultat.

4) ⇒ 1) : Soit $f \in C_b(E, \mathbb{R})$. On peut supposer que $0 \leq f \leq 1$ et on définit pour tout $t \in [0, 1]$, $A_t \doteq f^{-1}([t, 1])$. Par application de Fubini-Tonelli, on a $\mu_n(f) = \int_0^1 \mu_n(A_t)dt$. Or $t \to \mu_\infty(A_t)$ est décroissante sur $[0, 1]$. Elle est donc continue sauf sur un nombre dénombrable N de points (les sauts). Comme

$$\mu_\infty(\partial A_t) \leq \mu_\infty(A_t) - \lim_{s>t} \mu_\infty(A_s) ,$$

on déduit que sur $[0, 1] \setminus N$, $\mu_\infty(\partial A_t) = 0$. En appliquant 4) et la c.d., on déduit $\int_0^1 \mu_n(A_t)dt \to \int_0^1 \mu_\infty(A_t)dt$ ce qui donne le résultat. □

Definition 40 (Convergence en probabilité)**.** On dit que la suite de variables aléatoires $(X_n)_{n\geq 0}$ à valeurs dans $\mathbb{R}^d$ tend en probabilité vers une variable aléatoire X_∞ à valeurs dans $\mathbb{R}^d$ si pour tout $\epsilon > 0$, on a $P(|X_n - X_\infty| \geq \epsilon) \to 0$ lorsque $n \to \infty$. On écrit alors $X_n \xrightarrow{P} X_\infty$.

Remarque 55. *Remarquons que la convergence en probabilité n'a de sens, contrairement à la convergence en loi, que si les variables aléatoires sont définies sur le* même *espace probabilisé. C'est une forme de convergence plus forte que la convergence en loi comme le montre la proposition suivante :*

Proposition 25. *Si* $X_n \xrightarrow{P} X_\infty$ *alors* $X_n \xrightarrow{\mathcal{L}} X_\infty$

Démonstration. Soient $f \in C_b(\mathbb{R}^d, \mathbb{R})$ et $\epsilon > 0$. Pour M assez grand, on peut supposer que

$$2\|f\|_\infty P(|X_\infty| > M) \leq \epsilon\,,$$

puis comme f est uniformément continue sur $|x| \leq M+1$, on peut choisir $\eta > 0$ tel que $|f(x)-f(y)| \leq \epsilon$ pour tout $|x-y| \leq \eta$, $|x| \vee |y| \leq M+1$. Or

$$\begin{aligned}|E(f(X_n) - f(X_\infty))| \leq\ & E(|f(X_\infty) - f(X_n)|\mathbb{1}_{|X_\infty|>M}) \\ & + E(|f(X_\infty) - f(X_n)|\mathbb{1}_{|X_\infty|\vee|X_n|\leq M+1, |X_n - X_\infty|<\eta}) \\ & + E(|f(X_\infty) - f(X_n)|\mathbb{1}_{|X_n - X_\infty|\geq\eta})\,.\end{aligned}$$

Par suite $|E(f(X_n) - f(X_\infty))| \leq 2\epsilon + 2\|f\|_\infty P(|X_n - X_\infty| \geq \eta)$ et donc $\overline{\lim}\,|E(f(X_n) - f(X_\infty))| \leq 2\epsilon$. Comme ϵ est arbitraire, on a le résultat. □

- On montre facilement que si $X_n \to X_\infty$ P p.s. alors $X_n \xrightarrow{P} X_\infty$ car $P(|X_n - X_\infty| \geq \epsilon) \leq E(\mathbb{1}_{|X_n - X_\infty|\geq\epsilon}) \to 0$ par convergence dominée.
- En appliquant Borel-Cantelli 1), on a que si $X_n \xrightarrow{P} X_\infty$ alors il existe une sous suite $(X_{n_k})_{k\geq 1}$ telle que $X_{n_k} \to X_\infty$ P p.s. En effet, on peut extraire une sous-suite (n_k) telle que $P(|X_{n_k} - X_\infty| \geq 1/k) \leq 1/k^2$ si bien que P p.s. on a $|X_{n_k} - X_\infty| < 1/k$ pour k assez grand.
- Au final, en ajoutant l'inégalité de Markov, on a la proposition suivante :

Proposition 26. *On a les implications suivantes :*

$$\begin{array}{ccccc} X_n \xrightarrow{L^1} X_\infty & \Rightarrow & X_n \xrightarrow{P} X_\infty & \Rightarrow & X_n \xrightarrow{\mathcal{L}} X_\infty \\ & & \textit{s-suite} \downarrow \Uparrow & & \\ & & X_n \xrightarrow{p.s.} X_\infty & & \end{array}$$

Théorème 35 (Théorème de Levy)**.** *Soit* $(X_n)_{n\geq 1}$ *une suite de variable aléatoire à valeurs dans* $\mathbb{R}^d$ *: Alors* $X_n \xrightarrow{\mathcal{L}} X_\infty$ *ssi* $\phi_{X_n} \xrightarrow{simp.} \phi_{X_\infty}$ *où* ϕ_X *désigne la fonction caractéristique d'une va* X.

Pour la preuve du thm on commence par montrer le résultat suivant :

Lemme 2. $X_n \xrightarrow{\mathcal{L}} X_\infty$ *ssi pour tout* $f \in C_c(\mathbb{R}^d, \mathbb{R})$*, on a* $E(f(X_n)) \to E(f(X_\infty))$.

Démonstration. En effet, si (ϕ_k) est une suite de fonctions dans $C_c(\mathbb{R}^d, \mathbb{R})$ telle que $0 \leq \phi_k \leq 1$ et $\phi_k \uparrow 1$, alors pour tout $f \in C_b(\mathbb{R}^d, \mathbb{R})$, on a $f\phi_k \in C_c(\mathbb{R}^d, \mathbb{R})$ et $\mu_n(f\phi_k) \to \mu_\infty(f\phi_k)$ où $\mu_n = \mathcal{L}(X_n)$ et $\mu_\infty = \mathcal{L}(X_\infty)$. Comme $|f(1-\phi_k)| \leq \|f\|_\infty(1-\phi_k)$, on déduit :

$$|\mu_\infty(f) - \mu_n(f)| \leq \|f\|_\infty(1 - \mu_\infty(\phi_k)) + |\mu_\infty(f\phi_k) - \mu_n(f\phi_k)| + \|f\|_\infty(1 - \mu_n(\phi_k))$$

et donc

$$\overline{\lim_{n\to\infty}}\,|\mu_\infty(f) - \mu_n(f)| \leq 2\|f\|_\infty(1 - \mu_\infty(\phi_k))\,.$$

Par c.d., $\mu_\infty(1-\phi_k)) \to 0$ lorsque $k \to \infty$ d'où le résultat. □

Preuve du Thm 35. On se ramène donc par le lemme à montrer le résultat pour $f \in C_c(\mathbb{R}^d, \mathbb{R})$. On procède par régularisation comme pour l'injectivité de la transformée de Fourier. On note pour tout $\sigma > 0$,

$$f_\sigma \doteq f \star g_\sigma$$

où

$$g_\sigma(x) \doteq \exp(-|x|^2/(2\sigma^2))(2\pi\sigma^2)^{-d/2}$$

est la densité d'un vecteur $(U_1, \cdots, U_d)$ de coordonnées i.i.d. $\mathcal{N}(0, \sigma^2)$. On a alors que

$$\begin{aligned}
\int f_\sigma(x) d\mu_n(x) &= \int (\int f(y) \exp(-|x-y|^2/(2\sigma^2))/(2\pi\sigma^2)^{d/2} dy) \mu_n(dx) \\
&= \int (\int f(y) (\int \exp(i\langle x - y/\sigma, z) \exp(-|z|^2/2)/(2\pi\sigma)^d dz) dy) \mu_n(dx) \\
&\overset{\text{Fub-Leb}}{=} \int \hat{f}(-z/\sigma) \phi_{X_n}(z/\sigma) \exp(-|z|^2/2)/(2\pi\sigma)^d dz
\end{aligned}$$

où $\hat{f}$ est la transformée de Fourier de f. Par convergence dominée ($\hat{f}$ et ϕ_{X_n} sont bornées uniformément en n et $z \to \exp(-|z|^2/2)$ est intégrable), on a

$$\mu_n(f) \to \int \hat{f}(-z/\sigma) \phi_{X_\infty}(z/\sigma) \exp(-|z|^2/2)/(2\pi\sigma)^d dz = \mu_\infty(f)\,.$$

Enfin,

$$|\mu_\infty(f) - \mu_n(f)| \leq 2\|f - f_\sigma\|_\infty + |\mu_n(f_\sigma) - \mu_\infty(f_\sigma)|$$

d'où

$$\overline{\lim}\, |(\mu_\infty - \mu_n)(f)| \leq 2\|f - f_\sigma\|_\infty \to 0$$

pour $\sigma \to 0$ (cette dernière égalité se prouvant comme d'habitude en utilisant l'uniforme continuité de f). □

8.4 Théorème central limite

La principale conséquence du théorème de Levy est le théorème central limite souvent appelé le TCL.

Théorème 36 (Théorème central limite)**.** *Soit* $(X_n)_{n\geq 1}$ *une suite de v.a.i.i.d. tel que* $E(X_1^2) < \infty$. *On note* $\sigma^2 = V(X_1)$. *Alors si* $m = E(X_1)$ *et* $\overline{X}_n = \sum_{i=1}^n X_i/n$ *:*

$$\sqrt{n}(\overline{X}_n - m) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)\,,$$

- Le théorème central limite permet l'analyse des fluctuations autour de l'espérance. En utilisant la proposition 24 4) on déduit que $P(\overline{X}_n \in [m \pm a\frac{\sigma}{\sqrt{n}}]) \to \int_{-a}^{a} \exp(-x^2/2)/\sqrt{2\pi}dx$.
- Par symétrie, on a donc $P([\overline{X}_n \pm a\frac{\sigma}{\sqrt{n}}] \ni m) = P(|U| \leq a)$ où $U \sim \mathcal{N}(0,1)$. Ainsi, si a est tel que $P(|U| \geq a) = \alpha$, alors l'intervalle aléatoire $[\overline{X}_n \pm a\frac{\sigma}{\sqrt{n}}]$ contient $m = E(X_1)$ avec une probabilité $1-\alpha$. La probabilité de fournir un résultat faux est donc contrôlée (asymptotiquement) et inférieure à α.

Démonstration. On peut se ramener au cas $E(X_1) = 0$. On a alors en posant $Z_n = \sum_{i=1}^n X_i/\sqrt{n}$ que

$$\phi_{Z_n}(t) = E(\exp(it\sum_{i=1}^n X_i/\sqrt{n})) = \phi_{X_1}(t/\sqrt{n})^n\,.$$

Or comme X_1 admet un moment d'ordre 2, on déduit du thm 29 le développement de Taylor

$$\phi_{X_1}(t/\sqrt{n}) = 1 - t^2\sigma^2/2n + o(1/n)$$

ce qui donne

$$\phi_{Z_n}(t) = \exp(n \log(1 - t^2\sigma^2/2n + o(1/n)))$$

où log est une détermination du log complexe dans la boule $\{z \in \mathbb{C} \mid |z-1| < 1\}$ (cf cours fonction holomorphe). Comme $\log(1+z) = z + o(z)$ on déduit que

$$\phi_{Z_n}(t) \to \exp(-\frac{t^2\sigma^2}{2}) .$$

On reconnait la fonction caractéristique d'une gaussienne $\mathcal{N}(0, \sigma^2)$, ce qui donne le résultat par le théorème de Levy. $\square$