

Sommaire

Chapitre I. Topologie, Convergence	2
I.1 - Espaces métriques	2
I.2 - Espaces vectoriels normés	3
I.3 - Espaces topologiques	5
Chapitre II. Espaces de Hilbert, Séries de Fourier	8
II.1 - Produit scalaire	8
II.2 - Espaces de Hilbert	8
II.3 - Séries de Fourier	12
Chapitre III. Mesurabilité	13
III.1 - Tribus	13
III.2 - Mesures	15
Chapitre IV. Intégration	18
IV.1 - Intégrale par rapport à une mesure	18
IV.2 - Intégrale de Lebesgue	22
IV.3 - Mesure de densité	24
Chapitre V. Espaces L^p	26
V.1 - Relations d'équivalence	26
V.2 - Construction de l'e.v.n. L^p	27
V.3 - Propriétés de l'e.v.n. L^p	29
V.4 - L'espace $L^2_{\mathbb{C}}$	30
Chapitre VI. Introduction aux probabilités	33
VI.1 - Mesure de probabilité	33
VI.2 - Probabilité conditionnelle	33
VI.3 - Variables aléatoires	34
VI.4 - Moments	35
VI.5 - Fonction de répartition	37
VI.6 - Quelques lois remarquables	38
Chapitre VII. Mesure produit, Convolution	41
VII.1 - Espace produit	41
VII.2 - Intégrales multiples	43
VII.3 - Indépendance des variables aléatoires	45
VII.4 - Convolution	46
Chapitre VIII. Vecteurs aléatoires	50
VIII.1 - Fonctions de répartition, Copules	50
VIII.2 - Moments, Covariance	52
Chapitre IX. Transformée de Fourier, Fonction caractéristique	55
IX.1 - Transformée de Fourier d'une mesure	55
IX.2 - Transformée de Fourier d'une fonction	55
IX.3 - Fonction caractéristique	60
Chapitre X. Vecteurs Gaussiens	62
X.1 - Définition d'un vecteur gaussien	62
X.2 - Caractérisation d'un vecteur gaussien	62
X.3 - Loi d'un vecteur gaussien	63
Chapitre XI. Convergence de variables aléatoires	64
XI.1 - Les différents modes de convergence d'une v.a.	64
XI.2 - Lois des grands nombres	65
XI.3 - Convergence en loi	66
XI.4 - Théorème Central Limite (TCL)	68
Chapitre XII. Introduction aux processus stochastiques	69
XII.1 - Espérance conditionnelle	69
XII.2 - Processus stochastiques	71

Chapitre I. Topologie, Convergence

Section I.1 - Espaces métriques

Définition

Soit E un ensemble et $d : E \times E \rightarrow \mathbb{R}^+$ une fonction.

d est une **distance** sur E ssi :

1. $\forall (x, y) \in E \times E, d(x, y) = 0 \Leftrightarrow x = y$
2. $\forall (x, y) \in E \times E, d(x, y) = d(y, x)$
3. $\forall (x, y, z) \in E \times E \times E, d(x, z) \leq d(x, y) + d(y, z)$

On dit alors que (E, d) est un **espace métrique**.

Exemples : Sur n'importe quel ensemble E , on peut définir une distance : la distance triviale, pour laquelle $d(x, y) = 0$ si $x = y$ et $d(x, y) = 1$ sinon.

Sur \mathbb{R}^n , on note $d_p(X, Y) = (\sum_{i=1}^n |y_i - x_i|^p)^{\frac{1}{p}}$.

Sur $\mathcal{C}([0, 1], \mathbb{R})$, $d(f, g) = \sup_{x \in [0, 1]} |g(x) - f(x)|$ définit une distance.

Définition

Soit (E, d) un espace métrique et $l \in E$.

(u_n) **tend vers** l ssi $\forall \epsilon > 0, \exists N \in \mathbb{N}, n \geq N \Rightarrow d(u_n, l) < \epsilon$.

Remarque : \mathbb{R} peut être muni de distances différentes, qui peuvent mener à des convergences différentes. La suite $u_n = \frac{1}{n}$ tend vers 0 avec les distances d_p , mais pas avec la distance triviale.

Définition

Soit (E, d) un espace métrique, $a \in E$ et $r \geq 0$.

La **boule ouverte** centrée en a de rayon r est :

$$B(a, r) = \{x \in E \mid d(x, a) < r\}$$

Proposition

Soit (E, d) un espace métrique.

(u_n) tend vers l ssi $\forall \epsilon > 0, \exists N \in \mathbb{N}, n \geq N \Rightarrow u_n \in B(l, \epsilon)$

Définition

Soit E un ensemble, d_a et d_b deux distances sur E .

On dit que d_a est **plus fine** que d_b si $\exists C > 0, d_b \leq C d_a$.

Si d_a est plus fine que d_b et d_b est plus fine que d_a , alors on dit que d_a et d_b sont **équivalentes**.

Exemple : Sur \mathbb{R}^n , toutes les distances d_p sont équivalentes.

Définition

Soit E un ensemble, $A \subset E$ non vide et $x \in E$. La **distance** du point x à A est :

$$d(x, A) = \inf\{d(x, a), a \in A\}$$

Définition

Soit (u_n) une suite réelle majorée.

On définit sa **limite supérieure** par :

$$\limsup_{n \rightarrow +\infty} u_n = \lim_{n \rightarrow +\infty} \sup_{m \geq n} u_m$$

Soit (u_n) une suite réelle minorée.

On définit sa **limite inférieure** par :

$$\liminf_{n \rightarrow +\infty} u_n = \lim_{n \rightarrow +\infty} \inf_{m \geq n} u_m$$

Remarque : Si (u_n) converge, limite, limite supérieure et limite inférieure sont des quantités égales.

Définition

Une suite (u_n) est de Cauchy si :

$$\forall \epsilon > 0, \exists N \in \mathbb{N}, \forall (p, q) \in \mathbb{N}^2, q > p > N \Rightarrow d(u_q, u_p) < \epsilon$$

Proposition

Toute suite convergente est de Cauchy.

Remarque : La réciproque est fautive : la suite de \mathbb{Q} définie par $u_n = \frac{\lfloor \sqrt{2}10^n \rfloor}{10^n}$ ne converge pas dans \mathbb{Q} , mais est de Cauchy car si $q > p > N$ alors $|u_q - u_p| < \frac{1}{10^N}$.

Définition

Soit E un ensemble. On dit que E est **complet** si toute suite de Cauchy de E converge.

Théorème

\mathbb{R} est complet.

Exemple : $\mathcal{C}([0, 1], \mathbb{R})$ muni de la distance $d(f, g) = \int_0^1 |g(x) - f(x)| dx$ n'est pas complet. En effet, la suite de fonctions définies par :

$$f_n(x) = \begin{cases} 0 & \text{si } x < \frac{1}{2} - \frac{1}{n} \\ \frac{n}{2}x + \frac{1}{2} - \frac{n}{4} & \text{si } \frac{1}{2} - \frac{1}{n} \leq x \leq \frac{1}{2} + \frac{1}{n} \\ 1 & \text{si } x > \frac{1}{2} + \frac{1}{n} \end{cases}$$

vérifie, pour $q > p$, $d(f_q, f_p) = \frac{1}{2p} - \frac{1}{2q}$, et est donc de Cauchy, mais ne converge pas dans $\mathcal{C}([0, 1], \mathbb{R})$.

Section I.2 - Espaces vectoriels normés

Définition

Soit E un espace vectoriel et $N : E \times E \rightarrow \mathbb{R}^+$ une fonction.

N est une **norme** sur E ssi :

1. $\forall x \in E, N(x) = 0 \Leftrightarrow x = 0$
2. $\forall x \in E, \lambda \in \mathbb{R}, N(\lambda x) = |\lambda|N(x)$
3. $\forall (x, y) \in E \times E, N(x + y) \leq N(x) + N(y)$

On dit alors que (E, d) est un **espace vectoriel normé**.

Exemple : Soit $p \in [1, +\infty[$.

Sur \mathbb{R}^n , on définit la norme $N_p(f) = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$.

Sur $\mathcal{C}([0, 1], \mathbb{R})$, on définit la norme $N_p(f) = (\int_0^1 |f(x)|^p dx)^{\frac{1}{p}}$.

Proposition

Soit (E, N) un espace vectoriel normé.

$d(x, y) = N(x - y)$ est une distance sur E , appelée **distance induite** par N .

Remarque : Toute distance n'est pas forcément induite par une norme ; par exemple, la distance triviale ne l'est jamais.

Définition

Soit E un espace vectoriel, N_a et N_b deux normes sur E .

On dit que N_a est **plus fine** que N_b si $\exists C > 0, N_b \leq CN_a$.

Si N_a est plus fine que N_b et N_b est plus fine que N_a , alors on dit que N_a et N_b sont **équivalentes**.

Proposition

La relation "être plus fine que" est réflexive et transitive. On dit que c'est un **pré-ordre** et on note $N_b \preceq N_a$.

Théorème

Soit (E, N) un espace vectoriel sur \mathbb{R} ou \mathbb{C} .

E est de dimension finie ssi toutes ses normes sont équivalentes.

Définition

Soit (E, N) un espace vectoriel normé, $a \in E$ et $r \geq 0$.

La **boule ouverte** centrée en a de rayon r est :

$$B(a, r) = \{x \in E \mid N(x - a) < r\}$$

Proposition

Deux normes sont équivalentes si et seulement si leurs boules unité peuvent être incluses l'une dans l'autre après application d'une homothétie.

Définition

Soit (E, N) un espace métrique et $l \in E$.

(u_n) **tend vers** l ssi $\forall \epsilon > 0, \exists N \in \mathbb{N}, n \geq N \Rightarrow N(u_n - l) < \epsilon$.

Définition

On appelle espace de Banach tout espace vectoriel normé complet.

Exemples : $\mathcal{C}([0, 1], \mathbb{R})$ muni de N_p est un espace vectoriel normé, mais pas un espace de Banach.

\mathbb{R}^3 est un espace de Banach (peu importe la norme choisie : cf. proposition suivante)

Proposition

Deux normes équivalentes conduisent à la même convergence.

Remarque : En dimension infinie, il faut toujours préciser la norme lorsque l'on parle de convergence. Par exemple dans $\mathcal{C}([0, 1], \mathbb{R})$, la suite de fonctions définies par :

$$f_n(x) = \begin{cases} 1 - nx & \text{si } 0 \leq x \leq \frac{1}{n} \\ 0 & \text{si } \frac{1}{n} \leq x \leq 1 \end{cases}$$

converge vers 0 pour N_1 (car $N_1(f_n) = \frac{1}{n}$) mais converge vers 1 pour N_∞ (car $N_\infty(f_n) = 1$).

Section I.3 - Espaces topologiques

Définition

Soit E un ensemble. \mathcal{T} est une **topologie** sur E ssi :

1. $\emptyset \in \mathcal{T}$ et $E \in \mathcal{T}$.
2. Toute union d'éléments de \mathcal{T} est dans \mathcal{T} .
3. Toute intersection finie d'éléments de \mathcal{T} est dans \mathcal{T} .

(E, \mathcal{T}) est alors un **espace topologique** et les éléments de \mathcal{T} sont appelés les ouverts.

Exemples : Pour $E = \{1, 2, 3, 4, 5\}$, $\mathcal{T} = \{\emptyset, \{1, 2\}, \{3, 4\}, \{1, 2, 3, 4\}, E\}$ est une topologie.

Pour E un ensemble quelconque, les topologies $\mathcal{T} = \{\emptyset, E\}$ et $\mathcal{T} = \mathcal{P}(E)$ sont toujours des topologies sur E , qu'on appelle respectivement **topologie grossière** et **topologie discrète**.

Définition

Soit \mathcal{T}_a et \mathcal{T}_b deux topologies sur E . On dit que \mathcal{T}_b est **plus fine** que \mathcal{T}_a si $\mathcal{T}_a \subset \mathcal{T}_b$.

On dit alors que \mathcal{T}_a est **plus grossière** que \mathcal{T}_b .

Définition

Soit (E, \mathcal{T}) un espace topologique. $X \subset E$ est un **fermé** si $E \setminus X$ est un ouvert.

Définition

Soit (E, \mathcal{T}) un espace topologique et $x \in E$. On dit que $V \subset E$ est un **voisinage** de x si $\exists U \in \mathcal{T}$ tel que $x \in U$ et $U \subset V$.

On note $\mathcal{V}(x)$ l'ensemble des voisinages de x .

On appelle **base de voisinages** de x toute partie $\mathcal{B} \subset \mathcal{V}(x)$ telle que $\forall V \in \mathcal{V}(x), \exists B \in \mathcal{B}, B \subset V$.

Remarque : Si \mathcal{T}_a et \mathcal{T}_b sont deux topologies sur E telles que \mathcal{T}_b est plus fine que \mathcal{T}_a , alors tout voisinage de x pour \mathcal{T}_a sera un voisinage de x pour \mathcal{T}_b .

Proposition

Soit (E, \mathcal{T}) un espace topologique.

$U \subset E$ est un ouvert ssi il est voisinage de chacun de ses points.

Démonstration : Si U est un ouvert, alors pour chaque point x de U , on a $x \in U \subset U$ et donc U est voisinage de chacun de ses points.

Réciproquement, si U est voisinage de chacun de ses points, alors pour tout x de U , on choisit un ouvert A_x qui contient x inclus dans U . Alors, $A = \cup_{x \in U} A_x$ est un ouvert (union d'ouverts), tel que $U \subset A$ car tous les éléments de x sont dans A et $A \subset U$ car chaque A_x est inclus dans U . On a donc $U = A$ ouvert.

Proposition

Soit (E, d) un espace métrique.

$\mathcal{T} = \{\text{unions de } B(x, r), x \in E, r > 0\}$ est une topologie sur E , on parle de **topologie induite** par la distance.

Démonstration : Vérifions que l'on a effectivement une topologie.

- $\emptyset = B(0, 0) \in \mathcal{T}$ et $E = \cup_{r \geq 0} B(0, r) \in \mathcal{T}$.
- \mathcal{T} est par définition stable par union.
- Soit $U, V \in \mathcal{T}$. On écrit $U = \cup_{i \in I} B(x_i, r_i)$ et $V = \cup_{j \in J} B(x_j, r_j)$. Alors $U \cap V = \cup_{(i,j) \in I \times J} (B(x_i, r_i) \cap B(x_j, r_j))$. Soit $B(x_i, r_i) \cap B(x_j, r_j)$ est vide, et alors c'est un ouvert, soit elle est non vide et alors on considère, pour tout $z \in B(x_i, r_i) \cap B(x_j, r_j)$, $\rho_z = \min(r_i - d(z, x_i), r_j - d(z, x_j))$ de sorte que $B(z, \rho_z) \subset B(x_i, r_i) \cap B(x_j, r_j)$. On a alors $\cup_{z \in B(x_i, r_i) \cap B(x_j, r_j)} B(z, \rho_z) \subset B(x_i, r_i) \cap B(x_j, r_j)$, et puisque $B(x_i, r_i) \cap B(x_j, r_j) \subset \cup_{z \in B(x_i, r_i) \cap B(x_j, r_j)} B(z, \rho_z)$, on en déduit que $B(x_i, r_i) \cap B(x_j, r_j) = \cup_{z \in B(x_i, r_i) \cap B(x_j, r_j)} B(z, \rho_z)$. Donc $B(x_i, r_i) \cap B(x_j, r_j)$ est un ouvert, et on étend le résultat par récurrence à une intersection finie, ce qui conclut.

Exemple : La topologie induite par la distance triviale est la topologie discrète.

Remarque : Si d_a et d_b sont deux distances sur E telles que d_b est plus fine que d_a , alors la topologie induite par d_b est plus fine que la topologie induite par d_a .

Définition

Sur \mathbb{R} , la distance $d(x, y) = |y - x|$ induit la topologie suivante : $\mathcal{T} = \{\text{unions d'intervalles ouverts}\}$.
On l'appelle la **topologie usuelle** de \mathbb{R} .

Définition

Soit (E, \mathcal{T}) un espace topologique et $l \in E$.
 (u_n) tend vers l ssi $\forall V \in \mathcal{V}(l), \exists N \in \mathbb{N}, n \geq N \Rightarrow u_n \in V$.

Remarque : Dans les espaces métriques, on peut prendre $V = B(l, \epsilon)$, ce qui nous ramène à la définition de la convergence dans un espace métrique.

Définition

Un espace topologique E est dit de **Hausdorff** (ou T_2) si :

$$\forall (x, y) \in E^2, x \neq y, \exists U \in \mathcal{V}(x), \exists V \in \mathcal{V}(y), U \cap V = \emptyset$$

Proposition

Dans un espace de Hausdorff, la limite, si elle existe, est unique.

Démonstration : Soit $(u_n)_{n \in \mathbb{N}}$ une suite d'un espace de Hausdorff et l sa limite. Supposons par l'absurde que $l' \neq l$ soit une autre limite de $(u_n)_{n \in \mathbb{N}}$. Alors il existe $U \in \mathcal{V}(l)$ et $V \in \mathcal{V}(l')$ tel que $U \cap V = \emptyset$. Or par définition de la limite, il existe $N \in \mathbb{N}$ tel que $u_N \in U$ et $u_N \in V$, d'où la contradiction.

Proposition

Toute topologie induite par une distance est de Hausdorff.

Démonstration : Soit x et y deux points de l'espace topologique. En posant $U = B(x, \frac{d(x,y)}{2})$ et $V = B(y, \frac{d(x,y)}{2})$, on a $U \in \mathcal{V}(x), V \in \mathcal{V}(y)$ et $U \cap V = \emptyset$.

Définition

Soit (E, \mathcal{T}_E) et (F, \mathcal{T}_F) deux espaces topologiques.
Une fonction $f : E \rightarrow F$ est **continue** ssi $\forall U \in \mathcal{T}_F, f^{-1}(U) \in \mathcal{T}_E$.

Proposition

Soit (E, \mathcal{T}_E) et (F, \mathcal{T}_F) deux espaces topologiques, $f : E \rightarrow F$ une fonction continue et $(u_n)_{n \in \mathbb{N}}$ une suite d'éléments de E convergente vers l .
Alors $\lim_{n \rightarrow +\infty} f(u_n) = f(l)$.

Démonstration : Soit W un voisinage de $f(l)$. Il existe un ouvert U tel que $f(l) \in U$ et $U \subset W$. On a alors $l \in f^{-1}(U)$, $f^{-1}(U) \subset f^{-1}(W)$ et $f^{-1}(U)$ ouvert car f est continue et U ouvert. Ainsi, $f^{-1}(W)$ est un voisinage de l . Or $(u_n)_{n \in \mathbb{N}}$ converge vers l , donc $\exists N \in \mathbb{N}, n \geq N \Rightarrow u_n \in f^{-1}(W) \Rightarrow f(u_n) \in W$. Ceci vaut quelque soit le voisinage de $f(l)$ considéré, et donc on conclut que $f(u_n)$ tend vers $f(l)$.

Définition

Soit (E, \mathcal{T}_E) un espace topologique.
 $K \in E$ non vide est **compact** ssi pour tout recouvrement de K par des ouverts, on peut extraire un sous-recouvrement fini.

Exemple : Pour $E = \mathbb{R}$ avec la topologie usuelle, \mathbb{N} n'est pas compact : en considérant $U_i =]i - \frac{1}{10}, i + \frac{1}{10}[$, on a bien $\mathbb{N} \subset \bigcup_{i \in \mathbb{N}} U_i$ mais on ne peut pas trouver de sous-recouvrement fini de \mathbb{N} . (enlever un des U_i ne recouvre plus \mathbb{N})

Théorème (Borel-Lebesgue)

Lorsque $E = \mathbb{R}^n$ est muni de la topologie usuelle, les compacts sont les fermés bornés.

Théorème (Bolzano-Weierstrass)

Soit E un espace topologique métrisable (dont la topologie est induite par une distance).
 $K \subset E$ est compact ssi toute suite d'éléments de K admet une sous-suite convergente (dans K).

Définition

Soit (E, \mathcal{T}_E) un espace topologique, $A \subset E$ et $x \in E$.
On dit que x est **adhérent** à A ssi $\forall V \in \mathcal{V}(x), V \cap A \neq \emptyset$.
On dit que x est un **point isolé** de A ssi $\exists V \in \mathcal{V}(x), V \cap A = \{x\}$.
On dit que x est un **point d'accumulation** de A ssi $\forall V \in \mathcal{V}(x), V \cap A \setminus \{x\} \neq \emptyset$

Définition

Soit (E, \mathcal{T}_E) un espace topologique et $D \subset E$.
On dit que D est **discret** ssi tout point de D est isolé.

Définition

Soit (E, \mathcal{T}_E) un espace topologique.
On appelle **adhérence** de A , et on note \overline{A} , l'ensemble des points adhérents à A .

Définition

Soit (E, \mathcal{T}_E) un espace topologique et (u_n) une suite de E .
On dit que $a \in E$ est une **valeur d'adhérence** de (u_n) si $\forall N \in \mathbb{N}, a \in \overline{\{u_n, n \geq N\}}$.

Définition

Soit (u_n) une suite majorée (resp. minorée).
 $\liminf u_n$ (resp. $\limsup u_n$) est la plus petite (resp. plus grande) valeur d'adhérence de (u_n) .

Remarque : Dans le cas réel, la définition donnée ci-dessus coïncide bien avec celle donnée au début du chapitre.

Définition

Soit (E, \mathcal{T}_E) un espace topologique.
On dit que A est **dense** dans E si $\overline{A} = E$.

Exemple : Pour la topologie usuelle, \mathbb{Q} est dense dans \mathbb{R} .

Définition

Soit (E, \mathcal{T}_E) un espace topologique, $A \subset E$
On appelle **intérieur** de A , et on note $\overset{\circ}{A}$, l'ensemble des points dont A est le voisinage.

Proposition

$\overline{\overset{\circ}{A}}$ est le plus petit fermé contenant A .
 $\overset{\circ}{\overline{A}}$ est le plus grand ouvert contenu dans A .

Définition

Soit (E, \mathcal{T}_E) un espace topologique, $A \subset E$
On appelle **frontière** de A , et on note ∂A , l'ensemble $\overline{A} \setminus \overset{\circ}{A}$.

Chapitre II. Espaces de Hilbert, Séries de Fourier

Section II.1 - Produit scalaire

Définition

Soit E un espace vectoriel sur \mathbb{C} .

On dit que $\phi : E \times E \rightarrow \mathbb{C}$ est une **forme sesquilinéaire** si :

$$\forall (x, y, z) \in E \times E \times E, \forall \lambda \in \mathbb{C}, \begin{cases} \phi(x + \lambda z, y) = \phi(x, y) + \lambda \phi(z, y) \\ \phi(x, y + \lambda z) = \phi(x, y) + \bar{\lambda} \phi(x, z) \end{cases}$$

On dit alors que cette forme est :

- **hermitienne** ssi $\forall (x, y) \in E \times E, \phi(x, y) = \overline{\phi(y, x)}$
- **positive** ssi $\forall x \in E, \phi(x, x) \in \mathbb{R}^+$
- **définie** ssi $\phi(x, x) = 0 \Rightarrow x = 0$.

Définition

Soit E un espace vectoriel sur \mathbb{C} .

On appelle **produit scalaire** sur E toute forme sesquilinéaire ϕ hermitienne définie positive.

On dit alors que (E, ϕ) est un **espace préhilbertien**.

Lorsque E est de dimension finie, on dit que (E, ϕ) est un **espace hermitien**.

Exemples : \mathbb{C}^2 muni de $\phi : (x, y) \mapsto 2x_1\bar{y}_1 + x_2\bar{y}_2$ est un espace hermitien.

$\mathcal{C}([0, 1], \mathbb{C})$ muni de $\phi : (f, g) \mapsto \int_0^1 f(x)\overline{g(x)}dx$ est un espace préhilbertien.

Proposition (Identité du parallélogramme)

Soit E un espace préhilbertien et $x, y \in E$. Alors :

$$||x + y||^2 + ||x - y||^2 = 2||x||^2 + 2||y||^2$$

Proposition (Pythagore)

Soit E un espace préhilbertien et $x, y \in E$. Alors :

$$x \perp y \Rightarrow ||x + y||^2 = ||x||^2 + ||y||^2$$

Remarque : On veillera bien au fait que dans \mathbb{C} , il n'y a qu'une implication.

Proposition (Identité de polarisation)

Soit E un espace préhilbertien et $x, y \in E$. Alors :

$$\langle x, y \rangle = \frac{1}{4} (||x + y||^2 + i||x + iy||^2 - ||x - y||^2 - i||x - iy||^2)$$

Section II.2 - Espaces de Hilbert

Définition

On appelle **espace de Hilbert** tout espace préhilbertien complet.

Exemples : $l^2 = \{(u_n)_{n \in \mathbb{N}} \mid \sum_{n \geq 0} u_n^2 \text{ converge}\}$ muni de $\langle (u_n)_{n \in \mathbb{N}}, (v_n)_{n \in \mathbb{N}} \rangle = \sum_{n=0}^{+\infty} u_n v_n$ est un espace de Hilbert.

$\mathcal{C}([0, 1], \mathbb{C})$ muni de $\langle f, g \rangle = \int_0^1 f(x)\overline{g(x)}dx$ n'est pas un espace de Hilbert (car non complet)

Définition

Soit H un espace de Hilbert.

On dit que $\{e_i\}_{i \in I}$ est une **base hilbertienne** de H ssi :

- $\forall (i, j) \in I \times I, \langle e_i, e_j \rangle = \delta_{ij}$
- $\text{Vect}\{e_i, i \in I\} = H$

Remarque : Une base hilbertienne est donc une base orthonormale totale.

Définition

On dit qu'un espace de Hilbert H est séparable s'il existe $E \subset H$ dénombrable et dense dans H .

Proposition

Tout espace de Hilbert séparable admet une base hilbertienne au plus dénombrable.

Démonstration : Soit $(v_n)_{n \in \mathbb{N}}$ une suite d'éléments de H telle que $\overline{\{v_n, n \in \mathbb{N}\}} = H$. Pour $N \in \mathbb{N}$, on note $F_N = \text{Vect}(\{v_n, n \in [1, N]\})$; la suite $(F_N)_{N \in \mathbb{N}}$ est une suite croissante d'espaces vectoriels de dimension finie. On construit alors une base orthonormée pour F_1 , qu'on complète pour F_2 ... etc, ce qui conclut puisque $\cup_{N \in \mathbb{N}} F_N$ est dense dans H .

Exemple : Une base hilbertienne de l^2 est $\{(u_n^i)_{n \in \mathbb{N}}, i \in \mathbb{N}\}$ où $u_n^i = \delta_{i,n}$.

Théorème (Projection sur un convexe fermé)

Soit H un espace de Hilbert et $A \subset H$ un convexe fermé non vide.

Pour tout x dans H , il existe un unique $x_0 \in A$ tel que $d(x, x_0) = \min_{a \in A} d(x, a)$.

On note alors $x_0 = P_A(x)$, qu'on appelle **projection orthogonale** de x sur A .

De plus, $x_0 = P_A(x) \Leftrightarrow \forall u \in A, \langle x - x_0, u - x_0 \rangle \leq 0$.

Remarque : Dans le cas complexe, on aurait $\forall u \in A, \text{Re}(\langle x - x_0, u - x_0 \rangle) \leq 0$.

Démonstration : On a défini $d(x, A) = \inf_{a \in A} d(x, a)$. Soit $(u_n)_{n \in \mathbb{N}}$ une suite de A telle que $(d_n)_{n \in \mathbb{N}}$ définie par $d_n = d(x, u_n)$ soit décroissante et tende vers $d(x, A)$ (on dit que $(d_n)_{n \in \mathbb{N}}$ est une suite minimisante). On va montrer que $(u_n)_{n \in \mathbb{N}}$ est de Cauchy.

Soit $\epsilon > 0$ et $q > p$ deux entiers. On applique l'inégalité du parallélogramme avec $x - u_p$ et $x - u_q$:

$$\begin{aligned} \|(x - u_p) + (x - u_q)\|^2 + \|(x - u_p) - (x - u_q)\|^2 &= 2\|x - u_p\|^2 + 2\|x - u_q\|^2 \\ \Leftrightarrow \|u_q - u_p\|^2 &= 2\|x - u_p\|^2 + 2\|x - u_q\|^2 - 4\|x - \frac{u_p + u_q}{2}\|^2 \end{aligned}$$

Or A est convexe donc $\frac{u_p + u_q}{2} \in A$; on a donc

$$\begin{aligned} \|u_q - u_p\|^2 &\leq 2d(x, u_p)^2 + 2d(x, u_q)^2 - 4d(x, A)^2 \\ \Leftrightarrow \|u_q - u_p\|^2 &\leq 2(d(x, u_p)^2 - d(x, A)^2) + 2(d(x, u_q)^2 - d(x, A)^2) \end{aligned}$$

Or d_p et d_q tendent vers $d(x, A)$; on peut donc écrire qu'il existe $N_1 \in \mathbb{N}$ tel que $p \geq N_1 \Rightarrow d_p^2 - d(x, A)^2 < \epsilon$ et $N_2 \in \mathbb{N}$ tel que $q \geq N_2 \Rightarrow d_q^2 - d(x, A)^2 < \epsilon$. Alors, pour $q > p > N = \max(N_1, N_2)$, on a $\|u_q - u_p\|^2 < 4\epsilon$, et on en déduit que $(u_n)_{n \in \mathbb{N}}$ est de Cauchy. Puisque $(u_n)_{n \in \mathbb{N}}$ est une suite de Cauchy d'un ensemble fermé et complet, on sait qu'il existe $x_0 \in A$ tel que $\lim_{n \rightarrow +\infty} u_n = x_0$. D'où $d(x, A) = \inf_{a \in A} d(x, a) = \min_{a \in A} d(x, a) = d(x, x_0)$.

Soit $u \in A$ et $t \in]0, 1]$. On pose $v = (1 - t)x_0 + tu \in A$. Alors :

$$\begin{aligned} \|x - x_0\| &\leq \|x - v\| = \|x - x_0 + t(u - x_0)\| \\ \Leftrightarrow \|x - x_0\|^2 &\leq \langle (x - x_0) - t(u - x_0), (x - x_0) - t(u - x_0) \rangle \\ \Leftrightarrow \|x - x_0\|^2 &\leq \|x - x_0\|^2 - 2t\langle x - x_0, u - x_0 \rangle + t^2\|u - x_0\|^2 \\ \Leftrightarrow \langle x - x_0, u - x_0 \rangle &\leq \frac{t}{2}\|u - x_0\|^2 \end{aligned}$$

Lorsque $t \rightarrow 0$, on obtient alors $\langle x - x_0, u - x_0 \rangle \leq 0$.

Réciproquement, on suppose que $\forall u \in A, \langle x - x_0, u - x_0 \rangle \leq 0$. On a alors $2\langle x - x_0, u - x_0 \rangle - \|x_0 - u\|^2 \leq 0$. Or, $2\langle u - x_0, x - x_0 \rangle - \|x_0 - u\|^2 = \langle 2x - 2x_0, u - x_0 \rangle + \langle x_0 - u, u - x_0 \rangle = \langle 2x - x_0 - u, u - x_0 \rangle = 2\langle x, u \rangle - 2\langle x, x_0 \rangle + \|x_0\|^2 - \|u\|^2 = (\|x_0\|^2 - 2\langle x, x_0 \rangle + \|x\|^2) - (\|u\|^2 - 2\langle x, u \rangle + \|x\|^2) = \|x_0 - x\|^2 - \|x - u\|^2$. Ainsi, on a $\|x_0 - x\|^2 \leq \|u - x\|^2$ soit $d(x_0, x) \leq d(u, x)$: x_0 est donc bien égal à $P_A(x)$, puisqu'il minimise la distance de x à A .

On termine par vérifier l'unicité de x_0 : si il existe $x_1 \in A$ tel que $\forall u \in A, \langle x - x_1, u - x_1 \rangle \leq 0$, alors $\langle x - x_1, x_0 - x_1 \rangle \leq 0$ et $\langle x - x_0, x_1 - x_0 \rangle \leq 0$ implique $\langle x_1 - x + x - x_0, x_1 - x_0 \rangle = \|x_1 - x_0\|^2 \leq 0$, d'où $x_0 = x_1$.

Proposition

Soit H un espace de Hilbert et $A \subset H$ un convexe fermé non vide.
Soit $x, y \in H$, et x_0, y_0 leurs projections orthogonales sur A respectives.
Alors $\|x_0 - y_0\| \leq \|x - y\|$.

Démonstration : $\forall u \in A, \langle x - x_0, u - x_0 \rangle \leq 0$ et $\langle y - y_0, u - y_0 \rangle \leq 0$. On a donc $\langle x - x_0, y_0 - x_0 \rangle \leq 0$ et $\langle y - y_0, x_0 - y_0 \rangle \leq 0 \Rightarrow \langle x - y + y_0 - x_0, y_0 - x_0 \rangle \leq 0 \Rightarrow \|y_0 - x_0\|^2 \leq \langle x - y, x_0 - y_0 \rangle \leq \|x - y\| \|x_0 - y_0\| \Rightarrow \|x_0 - y_0\| \leq \|x - y\|$.

Remarque : En particulier, l'application P_A est 1-lipschitzienne, donc continue.

Proposition

Soit H un espace de Hilbert et $A \subset H$ un sev fermé. Soit $x \in H$.
Alors $x_0 = P_A(x) \Leftrightarrow x_0 \in A$ et $\forall u \in A, \langle x - x_0, u \rangle = 0$

Démonstration : Supposons que $x_0 = P_A(x)$, et soit $u \in A$. Puisque $u + x_0 \in A$, on a $\langle x - x_0, (u + x_0) - x_0 \rangle \leq 0$ donc $\langle x - x_0, u \rangle \leq 0$. Or $-u \in A$, donc on a aussi $\langle x - x_0, -u \rangle \leq 0$ soit $\langle x - x_0, u \rangle \geq 0$. Ainsi $\langle x - x_0, u \rangle = 0$. La réciproque est immédiate.

Proposition

Soit H un espace de Hilbert et $A \subset H$ un sev fermé.
Alors P_A est un opérateur linéaire.

Démonstration : Soit $x, y \in H, \lambda \in \mathbb{R}$. $\forall u \in A, \langle x - P_A(x), u \rangle = 0$ et $\langle y - P_A(y), u \rangle = 0$. Donc $\forall u \in A, \langle x + \lambda y - (P_A(x) + \lambda P_A(y)), u \rangle = 0$, et on en déduit que $P_A(x) + \lambda P_A(y) = P_A(x + \lambda y)$.

Théorème (Parseval)

Soit H un espace de Hilbert séparable, et $\{e_n, n \in \mathbb{N}\}$ une base hilbertienne de H .
Pour tout x dans H , on a :

$$x = \sum_{n=0}^{+\infty} \langle x, e_n \rangle e_n \text{ et } \|x\|^2 = \sum_{n=0}^{+\infty} |\langle x, e_n \rangle|^2$$

Démonstration : Soit $N \in \mathbb{N}$ et $E_N = \text{Vect}(\{e_n, n \in [0, N]\})$. E_N est un sev fermé de H , donc P_{E_N} est un opérateur linéaire de H dans H . Soit $x \in H$, alors :

$$\begin{aligned} P_{E_N}(x) &= \sum_{n=0}^N \langle x, e_n \rangle e_n \\ \Rightarrow \|P_{E_N}(x)\|^2 &= \left\| \sum_{n=0}^N \langle x, e_n \rangle e_n \right\|^2 = \sum_{n=0}^N |\langle x, e_n \rangle|^2 \end{aligned}$$

On remarque par ailleurs que $\langle x, e_n \rangle e_n = |\langle x, e_n \rangle|^2$, et donc que $\langle x, P_{E_N}(x) \rangle = \sum_{n=0}^N |\langle x, e_n \rangle|^2$. On a alors $\|P_{E_N}(x)\|^2 = \langle x, P_{E_N}(x) \rangle \leq \|P_{E_N}(x)\| \|x\|$. Ainsi, pour tout $x \in H, \|P_{E_N}(x)\| \leq \|x\|$. On note désormais $F = \cup_{N \in \mathbb{N}} E_N$, et on considère $y \in H$ et $\epsilon > 0$. F est dense dans H , donc il existe $y' \in F$ tel que $\|y - y'\| < \epsilon$. Comme $y' \in F$, on sait qu'il existe n_0 tel que $y' \in E_{n_0} \Rightarrow P_{E_{n_0}}(y') = y'$. Alors : $\|P_{E_{n_0}}(y) - y'\| = \|P_{E_{n_0}}(y) - P_{E_{n_0}}(y') - y + y'\| \leq \|P_{E_{n_0}}(y - y')\| + \|y - y'\| \leq 2\|y - y'\| \leq 2\epsilon$. On conclut alors que $y = \lim_{N \rightarrow +\infty} P_{E_N}(y)$. En passant à la limite dans les égalités $P_{E_N}(x) = \sum_{n=0}^N \langle x, e_n \rangle e_n$ et $\|P_{E_N}(x)\|^2 = \sum_{n=0}^N |\langle x, e_n \rangle|^2$, on obtient donc le résultat recherché.

Définition

Soit E un espace vectoriel.

On appelle **dual algébrique** de E l'ensemble des formes linéaires. On le note E^* .

Si de plus E est muni d'une topologie, on appelle **dual topologique** de E l'ensemble des formes linéaires continues. On le note E' .

Remarque : Si E est de dimension finie, alors bien entendu $E^* = E'$.

Théorème (Représentation de Riesz)

Soit H un espace de Hilbert.

Pour tout $\phi \in H'$, il existe un unique $u \in H$ tel que $\phi = x \mapsto \langle x, u \rangle$.

On a par ailleurs $\|\phi\|_{H'} = \|u\|_H$.

Démonstration : Soit $M = \text{Ker } \phi$. Si $M = H$, alors $\phi = 0$; on peut donc prendre $u = 0$. Sinon, on suppose $M \neq H$. Soit $z \in H \setminus M$. On pose $g = \frac{z - P_M(z)}{\|z - P_M(z)\|}$ puis $u = \phi(g)g$. On remarque qu'on a $\|g\| = 1$. Soit $x \in H$; on note $\lambda = \frac{\phi(x)}{\phi(g)}$ et $m = x - \lambda g$. Ainsi, $x = \lambda g + m$ avec $g \in M^\perp$ et $m \in M$ (car $\phi(m) = 0$). $\langle g, m \rangle = 0 \Rightarrow \langle g, x - \lambda g \rangle = 0 \Rightarrow \langle g, x \rangle = \lambda \langle g, g \rangle = \lambda = \frac{\phi(x)}{\phi(g)}$. D'où $\phi(x) = \langle u, x \rangle$.
Pour l'unicité, si il existe $v \in H$ tel que $\forall x \in E, \phi(x) = \langle x, u \rangle = \langle x, v \rangle$, alors pour $x = u - v$, on a $\langle u - v, u - v \rangle = 0$ soit $u = v$.

Remarque : L'application $\phi \mapsto u$ est un isomorphisme isométrique ; on peut donc **identifier** H et H' , et on notera (un peu abusivement) $H = H'$.

Définition

Soit E un espace vectoriel normé.

On appelle **bidual** de E le dual de son dual, c'est-à-dire E'' .

Lorsque $E = E''$ (au sens de l'identification), on dit que E est **réflexif**.

Proposition (Prolongement de H' dans V')

Soit H un espace de Hilbert, $V \subset H$ un espace de Banach dense dans H .

Soit $\phi \in H'$ et $u \in H$ sa représentation au sens du théorème de Riesz.

On définit $T\phi : V \rightarrow \mathbb{R}$ telle que $T\phi = (v \mapsto \langle v, u \rangle)$. $T\phi \in V'$; on peut donc définir $T : H' \rightarrow V'$ telle que $T = (\phi \mapsto T\phi)$. T est linéaire, injective et continue, et $T(H)$ est dense dans V' .

On dit qu'on a **injecté** H' dans V' . On identifie H et H' qu'on appelle **espace pivot**, et on écrira $V \subset H = H' \subset V'$. (ou $V \subset H \subset V'$)

Exemple : $l_1 = \{(x_n)_{n \in \mathbb{N}}, \sum_{n=0}^{+\infty} |x_n| < \infty\}$ est un espace de Banach mais pas un espace de Hilbert, et $l^2 = \{(x_n)_{n \in \mathbb{N}}, \sum_{n=0}^{+\infty} x_n^2 < \infty\}$ est un espace de Hilbert. On admet ici que $l_1 \subset l_2$ et que l_1 est dense dans l_2 . Alors en posant $H = l^2$ et $V = l^1$, on a par ce qui précède $V \subset H \subset V'$. On a cependant pas $V' = H'$; par exemple ϕ définie par $\phi((u_n)_{n \in \mathbb{N}}) = \sum_{n=0}^{+\infty} u_n$ appartient à V' , mais pas à H' .

Section II.3 - Séries de Fourier

Définition

Soit $f : \mathbb{R} \rightarrow \mathbb{C}$ continue par morceaux et 2π -périodique.

On appelle **coefficient de Fourier** de f les coordonnées de f dans la base hilbertienne $\{e_n : x \mapsto e^{inx}, n \in \mathbb{Z}\}$ avec le produit scalaire :

$$\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \overline{g(x)} dx$$

On note ces coefficients c_n et on a :

$$\forall n \in \mathbb{Z}, c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} dx$$

On appelle **série de Fourier** la série $\sum_{n \in \mathbb{Z}} c_n e^{inx}$.

Définition

Soit $f : \mathbb{R} \rightarrow \mathbb{C}$ continue par morceaux et 2π -périodique.

On appelle **coefficients de Fourier trigonométriques** les coefficients $a_n = c_n + c_{-n}$ et $b_n = i(c_n - c_{-n})$.

On a alors :

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx) dx \text{ et } b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx) dx$$

La série de Fourier s'écrit $\frac{a_0}{2} + \sum_{n=1}^{+\infty} (a_n \cos(nx) + b_n \sin(nx))$.

Remarque : Cette écriture permet, lorsque f est à valeurs dans \mathbb{R} , de ne travailler qu'avec des nombres réels.

Définition

Soit f une fonction continue par morceaux. On note \tilde{f} la fonction définie pour tout x du domaine de f par :

$$\tilde{f}(x) = \begin{cases} f(x) & \text{si } f \text{ est continue en } x \\ \frac{1}{2}(\lim_{x^-} f + \lim_{x^+} f) & \text{sinon} \end{cases}$$

Théorème (Dirichlet)

Soit $f : \mathbb{R} \rightarrow \mathbb{C}$ de classe \mathcal{C}^1 par morceaux et 2π -périodique.

Alors la série de Fourier de f converge simplement vers \tilde{f} .

Si de plus f est continue, alors la convergence est normale.

Chapitre III. Mesurabilité

Section III.1 - Tribus

Définition

Soit E un ensemble.

On dit que $\mathcal{E} \subset \mathcal{P}(E)$ est une **tribu** ssi :

1. $\emptyset \in \mathcal{E}$
 2. \mathcal{E} est stable par complémentarité ($A \in \mathcal{E} \Rightarrow E \setminus A \in \mathcal{E}$)
 3. \mathcal{E} est stable par union dénombrable ($\forall n \in \mathbb{N}, A_n \in \mathcal{E} \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{E}$)
- (E, \mathcal{E}) est alors un **espace mesurable**, et les ensembles de \mathcal{E} sont les **ensembles mesurables**.

Exemples : Pour $E = \{1, 2, 3, 4\}$, $\mathcal{E} = \{\emptyset, \{1, 2\}, \{3, 4\}, E\}$ est une tribu.

Pour $E = \mathbb{R}$, $\mathcal{E} = \{\emptyset, \mathbb{R}^{-*}, \mathbb{R}^{+}, \mathbb{R}\}$ est une tribu. Par contre, l'ensemble des ouverts de \mathbb{R} pour la topologie usuelle n'en est pas une, car il n'est pas stable par complémentarité.

Pour E un ensemble quelconque, $\mathcal{E} = \{\emptyset, E\}$ et $\mathcal{E} = \mathcal{P}(E)$ sont toujours des tribus sur E , qu'on appelle respectivement **tribu grossière** et **tribu discrète**.

Proposition

Soit (E, \mathcal{E}) un espace mesurable. La définition d'une tribu entraîne :

- $E \in \mathcal{E}$
- La stabilité de \mathcal{E} par différence ensembliste ($A, B \in \mathcal{E} \Rightarrow A \setminus B \in \mathcal{E}$)
- La stabilité de \mathcal{E} par intersection dénombrable ($\forall n \in \mathbb{N}, A_n \in \mathcal{E} \Rightarrow \bigcap_{n \in \mathbb{N}} A_n \in \mathcal{E}$)

Proposition

Soit E un ensemble, et $(\mathcal{E}_i)_{i \in I}$ une famille de tribus sur E .

Alors $\bigcap_{i \in I} \mathcal{E}_i$ est une tribu sur E .

Démonstration : 1. $\forall i \in I, \emptyset \in \mathcal{E}_i \Rightarrow \emptyset \in \bigcap_{i \in I} \mathcal{E}_i$

2. Soit $A \in \bigcap_{i \in I} \mathcal{E}_i$. Alors $\forall i \in I, A \in \mathcal{E}_i \Rightarrow \forall i \in I, E \setminus A \in \mathcal{E}_i \Rightarrow E \setminus A \in \bigcap_{i \in I} \mathcal{E}_i$.

3. Soit $(A_n)_{n \in \mathbb{N}}$ des éléments de $\bigcap_{i \in I} \mathcal{E}_i$, alors $\forall n \in \mathbb{N}, \forall i \in I, A_n \in \mathcal{E}_i \Rightarrow \forall i \in I, \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{E}_i \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \bigcap_{i \in I} \mathcal{E}_i$.

Définition

Soit E un ensemble, et $C \subset \mathcal{P}(E)$ une famille de sous-ensembles de E .

On appelle **tribu engendrée** par C , et on note $\sigma(C)$, l'intersection de toutes les tribus de E contenant C .

Il s'agit de la plus petite tribu de E contenant C .

Exemple : Si $E = \{1, 2, 3, 4\}$, alors $\sigma(\{1\}) = \{\emptyset, \{1\}, \{2, 3, 4\}, E\}$.

Définition

Soit (E, \mathcal{T}) un espace topologique.

La **tribu de Borel** de (E, \mathcal{T}) est la tribu engendrée par \mathcal{T} .

On note $\mathcal{B}(\mathcal{T}) = \sigma(\mathcal{T})$. Lorsqu'il y a une topologie usuelle sur E , on note aussi $\mathcal{B}(E)$.

Les éléments de cette tribu sont appelés les **boréliens**.

Exemples : $\mathcal{B}(\mathbb{R})$ est la tribu engendrée par les intervalles ouverts. Elle contient les ouverts, les fermés donc les singletons, tous les ensembles dénombrables...

$\mathcal{B}(\mathbb{N}) = \mathcal{P}(\mathbb{N})$, la topologie usuelle sur \mathbb{N} étant $\mathcal{P}(\mathbb{N})$.

Définition

On note $\overline{\mathbb{R}^+} = [0, +\infty]$ l'ensemble $\mathbb{R}^+ \cup \{+\infty\}$.

On peut définir une addition et une multiplication qui étend les opérations de \mathbb{R}^+ :

- $\forall a \in \mathbb{R}^+, a + (+\infty) = +\infty$
- $(+\infty) + (+\infty) = +\infty$
- $\forall a \in \mathbb{R}^{+*}, a \times (+\infty) = +\infty$
- $0 \times (+\infty) = 0$
- $(+\infty) \times (+\infty) = +\infty$

Définition

On munit $\overline{\mathbb{R}^+}$ de la topologie obtenue par union des ensembles :

- $\forall a, b \in \mathbb{R}^+,]a, b[$
- $\forall a \in \mathbb{R}^+,]a, +\infty]$
- $\forall b \in \mathbb{R}^+, [0, b[$

Cette topologie s'appelle **topologie de l'ordre**.

Définition

Soit (E, \mathcal{E}) et (F, \mathcal{F}) deux espaces mesurables.

La fonction $f : E \rightarrow F$ est **mesurable** ssi $f^{-1}(\mathcal{F}) \subset \mathcal{E}$, c'est-à-dire si pour tout ensemble mesurable B inclus dans F , son image réciproque $\{x \in E, f(x) \in B\}$ est mesurable.

Proposition

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

1_A est mesurable ssi A est mesurable.

Démonstration : Si 1_A est mesurable, alors $1_A^{-1}(\{1\}) = A$ donc A est mesurable.

Réciproquement soit A mesurable, et soit $B \in \mathcal{B}(\mathbb{R})$. Il y a 4 cas à considérer :

- B ne contient ni 0 ni 1 ; alors $1_A^{-1}(B) = \emptyset$.
- B contient 1, mais pas 0, alors $1_A^{-1}(B) = A$
- B contient 0, mais pas 1, alors $1_A^{-1}(B) = E \setminus A$
- B contient 0 et 1, alors $1_A^{-1}(B) = E$

Dans tous les cas $1_A^{-1}(B)$ est mesurable, ce qui conclut.

Proposition

Soit (E, \mathcal{E}) et (F, \mathcal{F}) deux espaces mesurables avec $\mathcal{F} = \sigma(C)$ pour $C \in \mathcal{P}(F)$.

$f : E \rightarrow F$ est mesurable ssi $f^{-1}(C) \in \mathcal{E}$

Démonstration : Le sens direct est immédiat ; montrons la réciproque.

Vérifions que $\mathcal{F}' = \{B \subset F, f^{-1}(B) \in \mathcal{E}\}$ est une tribu.

1. $f^{-1}(\emptyset) = \emptyset \in \mathcal{E}$ donc $\emptyset \in \mathcal{F}'$.
2. Soit $B \in \mathcal{F}'$, alors $f^{-1}(B) \in \mathcal{E} \Rightarrow E \setminus f^{-1}(B) \in \mathcal{E} \Rightarrow f^{-1}(F \setminus B) \in \mathcal{E} \Rightarrow F \setminus B \in \mathcal{F}'$.
3. Soit $(B_n)_{n \in \mathbb{N}}$ des éléments de \mathcal{F}' , alors $\cup_{n \in \mathbb{N}} f^{-1}(B_n) \in \mathcal{E} \Rightarrow f^{-1}(\cup_{n \in \mathbb{N}} B_n) \in \mathcal{E} \Rightarrow \cup_{n \in \mathbb{N}} B_n \in \mathcal{F}'$.

Si $C \subset \mathcal{F}'$, alors $\mathcal{F} = \sigma(C) \subset \mathcal{F}'$. Ainsi $\forall B \in \mathcal{F}, f^{-1}(B) \in \mathcal{E}$. Donc f est mesurable.

Définition

Soit (E, \mathcal{T}) et (F, \mathcal{U}) deux espaces topologiques, qu'on équipe de leurs tribus de Borel $\mathcal{E} = \sigma(\mathcal{T})$ et $\mathcal{F} = \sigma(\mathcal{U})$.

On appelle **fonction borélienne** toute fonction mesurable $f : (E, \mathcal{E}) \rightarrow (F, \mathcal{F})$.

Proposition

Toute fonction continue est borélienne.

Démonstration : Les ouverts engendrent la tribu, et l'image réciproque des ouverts sont des ouverts.

Proposition

Soit (E, \mathcal{E}) , (F, \mathcal{F}) et (G, \mathcal{G}) trois espaces mesurables.
 Soit $f : E \rightarrow F$ et $g : F \rightarrow G$ deux fonctions mesurables. Alors $g \circ f$ est mesurable.

Proposition

Soit (E, \mathcal{E}) un espace mesurable.
 Soit f et g deux fonctions mesurables de E dans \mathbb{R} , \mathbb{R}^+ ou $\overline{\mathbb{R}^+}$. Alors $f + g$, fg , $\max(f, g)$, $\min(f, g)$ et $|f|$ sont mesurables.
 Soit $(f_n)_{n \in \mathbb{N}}$ des fonctions mesurables de E dans \mathbb{R} , \mathbb{R}^+ ou $\overline{\mathbb{R}^+}$. Alors $\sup_{n \in \mathbb{N}} f_n$, $\inf_{n \in \mathbb{N}} f_n$, $\limsup_{n \rightarrow +\infty} f_n$, $\liminf_{n \rightarrow +\infty} f_n$, $\lim_{n \rightarrow +\infty} f_n$ et $\sum_{n=0}^{+\infty} f_n$ sont mesurables lorsqu'elles existent.

Définition

Soit (E, \mathcal{E}) un espace mesurable.
 Une fonction $f : E \rightarrow \mathbb{R}$ est dite **étagée** ssi elle est mesurable et prend un nombre fini de valeurs.

Remarque : Une fonction est étagée si et seulement si elle est combinaison linéaire de fonctions indicatrices.

Théorème

Soit (E, \mathcal{E}) un espace mesurable.
 Toute fonction mesurable $f : E \rightarrow \overline{\mathbb{R}^+}$ est la limite simple d'une suite croissante de fonctions étagées.

Section III.2 - Mesures

Définition

Soit (E, \mathcal{E}) un espace mesurable.
 On dit que $\mu : \mathcal{E} \rightarrow [0, +\infty]$ est une **mesure** ssi :

- $\mu(\emptyset) = 0$
- Pour toute famille dénombrable $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{E} deux-à-deux disjoints, $\mu(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n=0}^{+\infty} \mu(A_n)$

(E, \mathcal{E}, μ) est alors un **espace mesuré**.

Exemples : Pour $E = \mathbb{N}$, $\mathcal{E} = \mathcal{P}(\mathbb{N})$, on définit la mesure $\mu : \mathcal{E} \rightarrow [0, +\infty]$ telle que pour $A \subset \mathbb{N}$, on a :

$$\mu(A) = \begin{cases} \text{Card}(A) & \text{si } A \text{ est fini} \\ +\infty & \text{sinon.} \end{cases}$$

Cette mesure s'appelle **mesure de comptage**.

Pour E quelconque, \mathcal{E} une tribu et $x_0 \in E$, on définit la mesure $\mu : \mathcal{E} \rightarrow [0, +\infty]$ telle que pour $A \in \mathcal{E}$, $\mu(A) = 1_A(x_0)$.

Cette mesure s'appelle **mesure de Dirac** au point x_0 , qu'on note δ_{x_0} .

Pour $E = \mathbb{R}^3$, $\mathcal{E} = \mathcal{P}(\mathbb{R}^3)$, on sait par le théorème de Banach-Tarski qu'il ne peut pas exister de mesure $\mu : \mathcal{E} \rightarrow [0, +\infty]$ qui généralise la notion de volumes.

Proposition

Soit (E, \mathcal{E}, μ) un espace mesuré.
 Soit $A \in \mathcal{E}$ et $B \in \mathcal{E}$, alors :

- $A \subset B \Rightarrow \mu(A) \leq \mu(B)$
- $A \subset B$ et $\mu(B) < +\infty \Rightarrow \mu(B \setminus A) \leq \mu(B) - \mu(A)$
- $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$

Soit $(A_n)_{n \in \mathbb{N}}$ une famille d'éléments de \mathcal{E} , alors :

- $\mu(\bigcup_{n \in \mathbb{N}} A_n) \leq \sum_{n=0}^{+\infty} \mu(A_n)$
- $A_n \subset A_{n+1} \Rightarrow \mu(\bigcup_{n \in \mathbb{N}} A_n) = \lim_{n \rightarrow +\infty} \mu(A_n) = \sup_{n \in \mathbb{N}} \mu(A_n)$
- $A_{n+1} \subset A_n$ et $\mu(A_0) < +\infty \Rightarrow \mu(\bigcap_{n \in \mathbb{N}} A_n) = \lim_{n \in \mathbb{N}} \mu(A_n) = \inf_{n \in \mathbb{N}} \mu(A_n)$

Définition

Soit (E, \mathcal{E}, μ) un espace mesuré.
 Si $\mu(E) < +\infty$, on dit que μ est une **mesure finie**.
 Si $\mu(E) = 1$, on dit que μ est une **mesure de probabilité**.

Exemple : La mesure de Dirac est une mesure de probabilité. La mesure de comptage de \mathbb{N} n'est pas une mesure finie.

Définition

Soit (E, \mathcal{E}, μ) un espace mesuré.
 $x \in E$ est un **atome** si $\{x\} \in \mathcal{E}$ et $\mu(\{x\}) > 0$.
 Si μ est sans atome, on dit que c'est une **mesure diffuse**.

Définition

Soit (E, \mathcal{E}, μ) un espace mesuré. On dit que μ est **discrète** s'il existe une suite $(a_i)_{i \in I}$ dans E , avec I au plus dénombrable, telle que $\mu(E \setminus \cup_{i \in I} \{a_i\}) = 0$.

Remarque : Si les singletons appartiennent à la tribu, alors μ se décompose comme une combinaison linéaire de mesures de Dirac.

Définition

Soit (E, \mathcal{E}, μ) un espace mesuré.
 Si E est une réunion dénombrable d'ensembles de mesures finies, on dit que μ est **σ -finie**.

Exemple : La mesure de comptage est σ -finie sur \mathbb{N} , mais pas sur \mathbb{R} .

Définition

Soit (E, \mathcal{E}, μ) un espace mesuré.
 On dit que $A \in \mathcal{E}$ est **négligeable** si $\mu(A) = 0$.
 Lorsqu'une proposition logique est vraie, sauf sur un ensemble négligeable, on dit qu'elle est vraie **presque partout** (p.p.).

Définition

Soit (E, \mathcal{E}, μ) un espace mesuré.
 On dit que μ est une **mesure complète** si tout sous-ensemble d'un ensemble mesurable négligeable est lui-même mesurable (et donc négligeable).

Remarque : Si μ n'est pas une mesure complète, on peut toujours "compléter" \mathcal{E} afin qu'elle le devienne : en considérant l'ensemble $N = \{S \subset E, \exists A \in \mathcal{E}, \mu(A) = 0, S \subset A\}$, la tribu complétée est $\bar{\mathcal{E}} = \sigma(\mathcal{E} \cup N)$. μ s'étend de manière unique de \mathcal{E} à $\bar{\mathcal{E}}$, et cette extension est une mesure complète.

Objectif : On cherche désormais à définir une mesure μ sur \mathbb{R}^n telle que :

$$\mu([a_1, b_1] \times \dots \times [a_n, b_n]) = \prod_{i=1}^n |b_i - a_i|$$

Nous n'y parviendrons pas sur la tribu $\mathcal{P}(\mathbb{R})$, et nous allons donc devoir accepter une tribu (légèrement) plus petite.

Définition

On définit l'application λ^* pour $A \in \mathcal{P}(\mathbb{R})$ par :

$$\lambda^*(A) = \inf \left\{ \sum_{i \in \mathbb{N}} (b_i - a_i), A \subset \bigcup_{i \in \mathbb{N}}]a_i, b_i[, a_i \leq b_i \right\}$$

Proposition

Soit $\mathcal{M} = \{B \in \mathcal{P}(\mathbb{R}), \forall X \in \mathcal{P}(\mathbb{R}), \lambda^*(B) = \lambda^*(B \cap X) + \lambda^*(B \setminus X)\}$.
Alors \mathcal{M} est une tribu, et $\mathcal{B}(\mathbb{R}) \subset \mathcal{M}$.

Proposition

La restriction de λ^* à \mathcal{M} est une mesure ; on la note λ et on l'appelle **mesure de Lebesgue**.
 $\mathcal{M} = \overline{\mathcal{B}(\mathbb{R})}$ est le complété de la tribu de Borel, on l'appelle la **tribu de Lebesgue**.

Proposition

La mesure de Lebesgue λ a la propriété suivante :

$$\forall A \in \overline{\mathcal{B}(\mathbb{R})}, \lambda(A) = \inf\{\lambda(U), A \subset U, U \text{ ouvert}\}$$

$$\forall A \in \overline{\mathcal{B}(\mathbb{R})}, \lambda(A) = \sup\{\lambda(K), K \subset A, K \text{ compact}\}$$

On dit qu'elle est **régulière**.

Démonstration : Soit $A \in \overline{\mathcal{B}(\mathbb{R})}$. Clairement $\lambda(A) \leq \inf\{\lambda(U), A \subset U, U \text{ ouvert}\}$. Supposons $\lambda(A) < +\infty$ (le cas échéant, c'est trivial). Pour tout $\epsilon > 0$, il existe un recouvrement de A par des $]a_i, b_i[$ tels que $\lambda(A) \geq \sum_{i \in \mathbb{N}} (b_i - a_i) - \epsilon$. En notant $U = \cup_{i \in \mathbb{N}}]a_i, b_i[$, on a donc $\lambda(A) \geq \lambda(U) - \epsilon$. Ainsi $\lambda(A) \geq \inf\{\lambda(U), A \subset U, U \text{ ouvert}\}$, puis $\lambda(A) = \inf\{\lambda(U), A \subset U, U \text{ ouvert}\}$.

Montrons la seconde proposition ; clairement $\lambda(A) \geq \sup\{\lambda(K), K \subset A, K \text{ compact}\}$. On suppose d'abord qu'il existe un compact C tel que $A \subset C$. Pour tout $\epsilon > 0$, il existe U ouvert contenant $C \setminus A$ tel que $\lambda(C \setminus A) \geq \lambda(U) - \epsilon$. Or $C \setminus U = (A \cup (C \setminus A)) \setminus U \subset (A \cup (C \setminus A)) \setminus (C \setminus A) = A$. On note donc $K = C \setminus U$ tel que K soit compact et inclus dans A ; on a alors $\lambda(K) = \lambda(C \setminus U) \geq \lambda(C) - \lambda(U) \geq \lambda(C) - \lambda(C \setminus A) - \epsilon \geq \lambda(A) - \epsilon$. En conclusion, pour tout $\epsilon > 0$, il existe un compact K tel que $\lambda(K) \geq \lambda(A) - \epsilon$, ce qui montre que $\lambda(A) \leq \sup\{\lambda(K), K \subset A, K \text{ compact}\} \Rightarrow \lambda(A) = \sup\{\lambda(K), K \subset A, K \text{ compact}\}$.

Supposons maintenant qu'il n'existe pas de compact C tel que $A \subset C$. On se ramène au cas précédent en faisant entrer $A \cap [-n, n]$ dans un compact ; on a alors $\forall n \in \mathbb{N}^*, \lambda(A \cap [-n, n]) \leq \sup\{\lambda(K), K \subset A \cap [-n, n], K \text{ compact}\}$, d'où le résultat en passant à la limite lorsque $n \rightarrow +\infty$.

Proposition

Soit μ une mesure sur \mathbb{R}^d invariante par translations, et telle que $0 < \mu([0, 1]^d) < +\infty$.
Alors, μ est proportionnelle à la mesure de Lebesgue λ .

Remarque : La mesure de Lebesgue est elle-même invariante par translations, et telle que $0 < \lambda([0, 1]^d) = 1 < +\infty$.

Chapitre IV. Intégration

Section IV.1 - Intégrale par rapport à une mesure

Définition

Soit (E, \mathcal{E}) un espace mesurable, μ une mesure sur (E, \mathcal{E}) et $f : E \rightarrow \overline{\mathbb{R}}^+$ une fonction étagée.

On note α_i les n valeurs distinctes prises par f qu'on ordonne $(\alpha_1 < \dots < \alpha_n)$, et $A_i = f^{-1}(\alpha_i)$. On a alors $f = \sum_{i \in I} \alpha_i 1_{A_i}$

L'intégrale de la fonction étagée positive f par rapport à μ est :

$$\int_E f(x) \mu(dx) = \sum_{i=1}^n \alpha_i \mu(A_i)$$

On la note également $\int f d\mu$.

Remarque : Si f est exprimée sous forme d'une autre combinaison linéaire de fonction indicatrices $f = \sum_{i \in I} \beta_i 1_{B_i}$, alors $\sum_{i \in I} \beta_i \mu(B_i) = \int_E f(x) \mu(dx) = \sum_{i=1}^n \alpha_i \mu(A_i)$. En effet, pour tout $i \in I$, on peut définir un ensemble fini J_i tel que $\forall j \in J_i, \beta_j = \alpha_i$ et $A_i = \cup_{j \in J_i} B_j$.

Proposition

Soit (E, \mathcal{E}, μ) un espace mesuré, f, g deux fonctions étagées à valeurs dans $\overline{\mathbb{R}}^+$ et $\lambda \in \overline{\mathbb{R}}^+$. Alors :

$$\int (f + \lambda g) d\mu = \int f d\mu + \lambda \int g d\mu$$

Démonstration : On écrit $f = \sum_{i=1}^n \alpha_i 1_{A_i}$ et $g = \sum_{j=1}^m \beta_j 1_{B_j}$. On a $A_i = \cup_{j=1}^m (A_i \cap B_j)$ et $B_j = \cup_{i=1}^n (A_i \cap B_j)$ donc $f = \sum_{i=1}^n \sum_{j=1}^m \alpha_i 1_{A_i \cap B_j}$, $g = \sum_{i=1}^n \sum_{j=1}^m \beta_j 1_{A_i \cap B_j}$ et $f + \lambda g = \sum_{i=1}^n \sum_{j=1}^m (\alpha_i + \lambda \beta_j) 1_{A_i \cap B_j}$. Ainsi $\int (f + \lambda g) d\mu = \sum_{i=1}^n \sum_{j=1}^m (\alpha_i + \lambda \beta_j) \mu(1_{A_i \cap B_j}) = \sum_{i=1}^n \sum_{j=1}^m \alpha_i \mu(1_{A_i \cap B_j}) + \lambda \sum_{i=1}^n \sum_{j=1}^m \beta_j \mu(1_{A_i \cap B_j}) = \int f d\mu + \lambda \int g d\mu$.

Proposition

Soit (E, \mathcal{E}, μ) un espace mesuré et f, g deux fonctions étagées à valeurs dans $\overline{\mathbb{R}}^+$ telles que $f \leq g$. Alors :

$$\int f d\mu \leq \int g d\mu$$

Démonstration : $g - f \geq 0$, donc $\int g d\mu = \int f d\mu + \int (g - f) d\mu \geq \int f d\mu$.

Proposition

Soit (E, \mathcal{E}, μ) un espace mesuré et f une fonction étagée à valeurs dans $\overline{\mathbb{R}}^+$ nulle presque partout. Alors :

$$\int f d\mu = 0$$

Démonstration : On écrit $f = \sum_{i=1}^n \alpha_i 1_{A_i}$ avec $\alpha_1 < \dots < \alpha_n$ et $A_i = f^{-1}(\alpha_i)$. Si $\alpha_1 = 0$, alors $\forall i \in \llbracket 2, n \rrbracket, A_i = \{x \in E; f(x) = \alpha_i\} \subset \{x \in E; f(x) > 0\}$. Si $\alpha_1 > 0$, alors $\forall i \in \llbracket 1, n \rrbracket, A_i = \{x \in E; f(x) = \alpha_i\} \subset \{x \in E; f(x) > 0\}$. Dans tous les cas $\int f d\mu = \sum_{i=1}^n \alpha_i \mu(A_i) = 0$.

Définition

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

On note $\mathcal{S}(\mathcal{E})$ l'ensemble des fonctions étagées de (E, \mathcal{E}) .

On note $\mathcal{S}^+(\mathcal{E})$ l'ensemble des fonctions étagées positives de (E, \mathcal{E}) .

Définition

Soit $f : (E, \mathcal{E}, \mu) \rightarrow ([0, +\infty], \mathcal{B}([0, +\infty]))$ une fonction mesurable.

L'intégrale de f par rapport à la mesure μ est définie par :

$$\int_E f(x) \mu(dx) = \sup_{h \in \mathcal{S}^+(\mathcal{E}), h \leq f} \int_E h(x) \mu(dx)$$

On la note également $\int f d\mu$.

Proposition

Soit (E, \mathcal{E}, μ) un espace mesuré et f, g deux fonctions mesurables de (E, \mathcal{E}, μ) à valeurs dans $([0, +\infty], \mathcal{B}([0, +\infty]))$ telles que $f \leq g$. Alors :

$$\int f d\mu \leq \int g d\mu$$

Démonstration : Si $h \in \mathcal{S}^+(\mathcal{E})$ et $h \leq f$, alors $h \in \mathcal{S}^+(\mathcal{E})$ et $h \leq f$. Donc $\sup_{h \in \mathcal{S}^+(\mathcal{E}), h \leq f} \int h d\mu \leq \sup_{h \in \mathcal{S}^+(\mathcal{E}), h \leq g} \int h d\mu$.

Proposition

Soit (E, \mathcal{E}, μ) un espace mesuré et f une fonction mesurable de (E, \mathcal{E}, μ) à valeurs dans $([0, +\infty], \mathcal{B}([0, +\infty]))$ nulle presque partout. Alors :

$$\int f d\mu = 0$$

Démonstration : Soit h une fonction étagée à valeurs dans $[0, +\infty]$ inférieure à f . Alors $\mu(\{x \in E, f(x) > 0\}) = 0 \Rightarrow \mu(\{x \in E, h(x) > 0\}) = 0$. Donc $\int h d\mu = 0$, et ceci valant quelque soit h , $\int f d\mu = 0$.

Remarque : L'intégrale de f peut être nulle sans que f ne soit nulle (elle ne le sera seulement que presque partout).

Théorème (Convergence monotone)

Soit $(f_n)_{n \in \mathbb{N}}$ une suite croissante de fonctions mesurables $f_n : E \rightarrow \overline{\mathbb{R}^+}$ convergeant simplement vers $f : E \rightarrow \overline{\mathbb{R}^+}$. Alors :

$$\int f d\mu = \lim_{n \rightarrow +\infty} \int f_n d\mu$$

Démonstration : f est mesurable donc $\int f d\mu = \sup_{h \in \mathcal{S}^+(\mathcal{E}), h \leq f} \int h d\mu$. $\forall n \in \mathbb{N}, f_n \leq f_{n+1}$ donc $\int f_n d\mu \leq \int f_{n+1} d\mu$. La suite $(\int f_n d\mu)_{n \in \mathbb{N}}$ est croissante. Par ailleurs, $f_n \leq f$ donc $\int f_n d\mu \leq \int f d\mu$. La suite $(\int f_n d\mu)_{n \in \mathbb{N}}$ est également majorée ; ainsi, elle converge et $\lim_{n \rightarrow +\infty} \int f_n d\mu \leq \int f d\mu$.

Soit $h \in \mathcal{S}^+(\mathcal{E})$ telle que $h \leq f$. On écrit $h = \sum_{i=1}^m \alpha_i 1_{A_i}$. Soit $a \in]0, 1[$. Pour tout $n \in \mathbb{N}$, on définit $E_n^a = \{x \in E; ah(x) \leq f_n(x)\}$. Comme f_n et h sont mesurables, E_n^a est mesurable et on a :

$$\int f_n d\mu \geq \int ah 1_{E_n^a} d\mu = a \sum_{i=1}^m \alpha_i \mu(A_i \cap E_n^a)$$

Or $(f_n)_{n \in \mathbb{N}}$ est croissante donc $E_n^a \subset E_{n+1}^a \Rightarrow A_i \cap E_n^a \subset A_i \cap E_{n+1}^a$. Supposons qu'il existe $x \in E$ tel que $x \notin \cup_{n \in \mathbb{N}} E_n^a$, alors $\forall n \in \mathbb{N}, ah(x) > f_n(x)$ donc $h(x) > ah(x) \geq f(x)$ impossible. Ainsi $E = \cup_{n \in \mathbb{N}} E_n^a$, soit $A_i = \cup_{n \in \mathbb{N}} (A_i \cap E_n^a)$. On a donc $\lim_{n \rightarrow +\infty} \mu(A_i \cap E_n^a) = \mu(A_i)$, soit :

$$\lim_{n \rightarrow +\infty} \int f_n d\mu \geq a \sum_{i=1}^m \alpha_i \mu(A_i) = a \int h d\mu$$

Ceci vaut pour tout $a < 1$; on a donc $\lim_{n \rightarrow +\infty} \int f_n d\mu \geq \int f d\mu$, et en conclusion, $\int f d\mu = \lim_{n \rightarrow +\infty} \int f_n d\mu$.

Proposition

Soit (E, \mathcal{E}, μ) un espace mesuré, f, g deux fonctions mesurables de (E, \mathcal{E}, μ) à valeurs dans $([0, +\infty], \mathcal{B}([0, +\infty]))$ et $\lambda \in [0, +\infty]$. Alors :

$$\int (f + \lambda g) d\mu = \int f d\mu + \lambda \int g d\mu$$

Démonstration : Il existe une suite de fonctions étagées positives $(f_n)_{n \in \mathbb{N}}$ qui converge simplement vers f , et il existe une suite de fonctions étagées positives $(g_n)_{n \in \mathbb{N}}$ qui converge simplement vers g (cf. théorème du chapitre précédent). Alors $\forall n \in \mathbb{N}$, $\int (f_n + \lambda g_n) d\mu = \int f_n d\mu + \lambda \int g_n d\mu$, soit en passant à la limite par le théorème de convergence monotone : $\int (f + \lambda g) d\mu = \int f d\mu + \lambda \int g d\mu$.

Proposition

Soit (E, \mathcal{E}, μ) un espace mesuré, $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables de (E, \mathcal{E}, μ) à valeurs dans $([0, +\infty], \mathcal{B}([0, +\infty]))$. Alors :

$$\int \left(\sum_{n=0}^{+\infty} f_n \right) d\mu = \sum_{n=0}^{+\infty} \int f_n d\mu$$

Démonstration : On applique le théorème de convergence monotone à la suite des sommes partielles $S_N = \sum_{n=0}^N f_n$:

$$\int \left(\sum_{n=0}^{+\infty} f_n \right) d\mu = \lim_{N \rightarrow +\infty} \int S_N d\mu = \lim_{N \rightarrow +\infty} \sum_{n=0}^N \int f_n d\mu = \sum_{n=0}^{+\infty} \int f_n d\mu$$

Proposition (Inégalité de Markov)

Soit (E, \mathcal{E}, μ) un espace mesuré, f une fonction mesurable de (E, \mathcal{E}, μ) à valeurs dans $([0, +\infty], \mathcal{B}([0, +\infty]))$. Alors :

$$\forall a > 0, \mu(\{x \in E; f(x) \geq a\}) \leq \frac{1}{a} \int f d\mu$$

Démonstration : Soit $A = \{x \in E; f(x) \geq a\}$. Alors $f \geq a 1_A \Rightarrow \int f d\mu \geq \int a 1_A d\mu = a \mu(A)$.

Proposition

Soit (E, \mathcal{E}, μ) un espace mesuré, f une fonction mesurable de (E, \mathcal{E}, μ) à valeurs dans $([0, +\infty], \mathcal{B}([0, +\infty]))$. Alors :

$$f = 0 \text{ p.p.} \Leftrightarrow \int f d\mu = 0$$

Démonstration : On a déjà traité le sens direct ; pour la réciproque, on pose $B_n = \{x \in E; f(x) \geq \frac{1}{n}\}$. Alors $\mu(B_n) \leq \frac{1}{n} \int f d\mu = 0$. Or $B_n \subset B_{n+1}$ et $\cup_{n \in \mathbb{N}^*} B_n = \{x \in E; f(x) > 0\}$ donc $\mu(\{x \in E; f(x) > 0\}) = \lim_{n \rightarrow +\infty} \mu(B_n) = 0$. Ainsi $f = 0$ presque partout.

Proposition

Soit (E, \mathcal{E}, μ) un espace mesuré, f, g deux fonctions mesurables de (E, \mathcal{E}, μ) à valeurs dans $([0, +\infty], \mathcal{B}([0, +\infty]))$. Alors :

$$f = g \text{ p.p.} \Leftrightarrow \int f d\mu = \int g d\mu$$

Démonstration : $f - \min(f, g) = 0$ p.p. et $f - \min(f, g) \geq 0$. Par la proposition précédente, on a donc $\int (f - \min(f, g)) d\mu = 0$, soit $\int f d\mu = \int \min(f, g) d\mu$. De la même manière, on montre que $\int g d\mu = \int \min(f, g) d\mu$, et donc $\int f d\mu = \int g d\mu$.

Proposition

Soit (E, \mathcal{E}, μ) un espace mesuré, f une fonction mesurable de (E, \mathcal{E}, μ) à valeurs dans $([0, +\infty], \mathcal{B}([0, +\infty]))$. Alors :

$$\int f d\mu < +\infty \Rightarrow f < +\infty \text{ p.p.}$$

Démonstration : Soit $A_n = \{x \in E; f(x) \geq n\}$ et $A_\infty = \{x \in E; f(x) = +\infty\}$. $\mu(A_n) \leq \frac{1}{n} \int f d\mu$ donc $\lim_{n \rightarrow +\infty} \mu(A_n) = 0$. Comme $A_{n+1} \subset A_n$, $\mu(A_0) < \infty$ et $\cap_{n \in \mathbb{N}^*} A_n = A_\infty$, on a $\mu(A_\infty) = \mu(\cap_{n \in \mathbb{N}^*} A_n) = \lim_{n \rightarrow +\infty} \mu(A_n) = 0$. Ainsi $f < +\infty$ p.p.

Proposition (Lemme de Fatou)

Soit (E, \mathcal{E}, μ) un espace mesuré, $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables de (E, \mathcal{E}, μ) à valeurs dans $([0, +\infty], \mathcal{B}([0, +\infty]))$. Alors :

$$\int (\liminf f_n) d\mu \leq \liminf \int f_n d\mu$$

Démonstration : On applique le théorème de la convergence monotone à $(\inf_{m \geq n} f_m)_{n \in \mathbb{N}}$: on a donc $\lim_{n \rightarrow +\infty} \int (\inf_{m \geq n} f_m) d\mu = \int (\lim_{n \rightarrow +\infty} (\inf_{m \geq n} f_m)) d\mu$. Or $p \geq n \Rightarrow \int (\inf_{m \geq n} f_m) d\mu \leq \int f_p d\mu$. On en déduit que $\int (\inf_{m \geq n} f_m) d\mu \leq \inf_{p \geq n} \int f_p d\mu$. En passant à la limite quand $n \rightarrow +\infty$, on obtient donc $\int (\liminf f_n) d\mu = \lim_{n \rightarrow +\infty} \int (\inf_{m \geq n} f_m) d\mu \leq \lim_{n \rightarrow +\infty} \inf_{p \geq n} \int f_p d\mu$.

Définition

Soit (E, \mathcal{E}, μ) un espace mesuré, $f : (E, \mathcal{E}, \mu) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une fonction mesurable. On dit que f est **intégrable** par rapport à la mesure μ ssi :

$$\int |f| d\mu < +\infty$$

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

Lorsque f est intégrable par rapport à la mesure μ , on note $f^+ = \max(f, 0)$ et $f^- = -\min(f, 0)$. On définit l'intégrale de f par :

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

Proposition

Soit (E, \mathcal{E}, μ) un espace mesuré.

Pour tout $f \in \mathcal{L}^1(E, \mathcal{E}, \mu)$, $|\int f d\mu| \leq \int |f| d\mu$.

Démonstration : Puisque $|f| = f^+ + f^-$, on a :

$$\left| \int f d\mu \right| = \left| \int f^+ d\mu - \int f^- d\mu \right| \leq \left| \int f^+ d\mu \right| + \left| \int f^- d\mu \right| = \int f^+ + f^- d\mu = \int |f| d\mu$$

Proposition

Soit (E, \mathcal{E}, μ) un espace mesuré. $\mathcal{L}^1(E, \mathcal{E}, \mu)$ est un espace vectoriel et l'application $f \mapsto \int f d\mu$ est une forme linéaire sur $\mathcal{L}^1(E, \mathcal{E}, \mu)$.

Démonstration : Soit $f, g \in \mathcal{L}^1(E, \mathcal{E}, \mu)$. $0 \leq |f + g| \leq |f| + |g|$ donc $\int |f + g| d\mu \leq \int |f| d\mu + \int |g| d\mu < +\infty$ et donc $f + g \in \mathcal{L}^1(E, \mathcal{E}, \mu)$. Par ailleurs, $f + g = (f + g)^+ - (f + g)^- = f^+ + g^+ - f^- - g^- \Rightarrow (f + g)^+ + f^- + g^- = (f + g)^- + f^+ + g^+ \Rightarrow \int (f + g)^+ d\mu - \int (f + g)^- d\mu = \int f^+ d\mu - \int f^- d\mu + \int g^+ d\mu - \int g^- d\mu \Rightarrow \int f + g d\mu = \int f d\mu + \int g d\mu$. Soit $f \in \mathcal{L}^1(E, \mathcal{E}, \mu)$ et $\lambda \in \mathbb{R}$. $\int |\lambda f| d\mu \leq |\lambda| \int |f| d\mu < +\infty$ donc $\lambda f \in \mathcal{L}^1(E, \mathcal{E}, \mu)$. Par ailleurs, si $\lambda \geq 0$, alors $\int \lambda f d\mu = \int \lambda f^+ d\mu - \int \lambda f^- d\mu = \lambda (\int f^+ d\mu - \int f^- d\mu) = \lambda \int f d\mu$ et si $\lambda < 0$, alors $\int \lambda f d\mu = \int (\lambda f)^+ - \int (\lambda f)^- d\mu = -\lambda \int f^- d\mu + \lambda \int f^+ d\mu = \lambda (\int f^+ d\mu - \int f^- d\mu) = \lambda \int f d\mu$ (en utilisant le fait que pour $a > 0$, $(-af)^+ = -af^-$ et $(-af)^- = -af^+$).

Remarque : L'application $f \mapsto \int |f|$ n'est pas une norme sur $\mathcal{L}^1(E, \mathcal{E}, \mu)$.

Proposition

Soit (E, \mathcal{E}, μ) un espace mesuré, $f, g \in \mathcal{L}^1(E, \mathcal{E}, \mu)$. Alors :

- $f \leq g \Rightarrow \int f d\mu \leq \int g d\mu$
- $f = g$ p.p. $\Rightarrow \int f d\mu = \int g d\mu$

Théorème (Convergence dominée)

Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables de $\mathcal{L}^1(E, \mathcal{E}, \mu)$. On suppose que :

- Il existe une fonction mesurable f tel que $\lim_{n \rightarrow +\infty} f_n(x) = f(x)$ pour presque tout x dans E
 - Il existe une fonction mesurable g à valeurs positives tel que $\forall n \in \mathbb{N}, |f_n| \leq g$ p.p. et $\int g d\mu < +\infty$
- Alors $f \in \mathcal{L}^1(E, \mathcal{E}, \mu)$, $\lim_{n \rightarrow +\infty} \int f_n d\mu = \int f d\mu$ et $\lim_{n \rightarrow +\infty} \int |f_n - f| d\mu = 0$.

Démonstration : On commence par supposer les hypothèses partout (et pas seulement presque partout). En faisant tendre n vers $+\infty$ dans $|f_n| \leq g$, on a $|f| \leq g$ donc $f \in \mathcal{L}^1(E, \mathcal{E}, \mu)$. On a aussi $|f_n - f| \leq 2g$, soit $2g - |f_n - f| \geq 0$; en appliquant le lemme de Fatou, on trouve alors $\liminf_{n \rightarrow +\infty} \int (2g - |f_n - f|) d\mu \geq \int 2g d\mu$. Or $\liminf_{n \rightarrow +\infty} \int (2g - |f_n - f|) d\mu = 2 \int g d\mu - \limsup_{n \rightarrow +\infty} \int |f_n - f| d\mu$; ceci est donc équivalent à $\int 2g d\mu - \limsup_{n \rightarrow +\infty} \int |f_n - f| d\mu \geq \int 2g d\mu \Leftrightarrow \limsup_{n \rightarrow +\infty} \int |f_n - f| d\mu \leq 0$. Par positivité de l'intégrale, on a donc $\lim_{n \rightarrow +\infty} \int |f_n - f| d\mu = 0$. Ceci implique aussi $\lim_{n \rightarrow +\infty} \int f_n d\mu = \int f d\mu$.

On suppose désormais les hypothèses telles quelles. Soit $\tilde{E} = \{x \in E; \lim_{n \rightarrow +\infty} f_n(x) = f(x) \text{ et } \sup_{n \in \mathbb{N}} f_n(x) \leq g(x)\}$. Les fonctions $\tilde{f} = f1_{\tilde{E}}$ et $\tilde{f}_n = f_n1_{\tilde{E}}$ satisfont les hypothèses partout ; par ailleurs, $\mu(E \setminus \tilde{E}) = 0$ donc $f = \tilde{f}$ et $f_n = \tilde{f}_n$ p.p. soit $\int |f_n - f| d\mu = \int |\tilde{f}_n - \tilde{f}| d\mu$, ce qui conclut.

Définition

Pour tout $p \in [1, +\infty[$, on définit :

$$\mathcal{L}^p(E, \mathcal{E}, \mu) = \{f : E \rightarrow \mathbb{R} \text{ mesurable ; } \int |f|^p d\mu < +\infty\}$$

et $\mathcal{L}^\infty(E, \mathcal{E}, \mu) = \{f : E \rightarrow \mathbb{R} \text{ mesurable ; } \exists C > 0, |f| \leq C \text{ p.p.}\}$

Remarque : Lorsque μ est une mesure finie, alors $p < q \Rightarrow \mathcal{L}^p(E, \mathcal{E}, \mu) \subset \mathcal{L}^q(E, \mathcal{E}, \mu)$. Attention, cela est faux dans le cas général.

Section IV.2 - Intégrale de Lebesgue

Définition

Considérons $E = \mathbb{R}^d$ muni de la tribu de Lebesgue et de la mesure de Lebesgue $\lambda^{(d)}$.

On appelle **intégrale de Lebesgue** l'intégrale par rapport à $\lambda^{(d)}$.

Soit $f \in \mathcal{L}^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^{(d)})$. L'intégrale de f est notée :

$$\int_{\mathbb{R}^d} f d\lambda^{(d)} \text{ ou } \int_{\mathbb{R}^d} f(x) \lambda^{(d)}(dx) \text{ ou } \int_{\mathbb{R}^d} f(x_1, \dots, x_d) \lambda^{(d)}(dx_1, \dots, dx_d)$$

Définition

Soit $f \in \mathcal{L}^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^{(d)})$ et $U \subset \mathbb{R}^d$ mesurable.

$|f1_U| \leq |f|$ donc $f1_U \in \mathcal{L}^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^{(d)})$.

On note alors $\int_U f d\lambda^{(d)} = \int_U f1_U d\lambda^{(d)}$

Définition

Une fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ est dite **localement intégrable** si pour tout compact $K \subset \mathbb{R}^d, f1_K \in \mathcal{L}^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^{(d)})$.

On note $\mathcal{L}_{\text{loc}}^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^{(d)})$ l'ensemble de ces fonctions.

Remarque : $\mathcal{L}_{\text{loc}}^1(\mathbb{R}^d, \overline{\mathcal{B}(\mathbb{R}^d)}, \lambda^{(d)}) \subset \mathcal{L}^1(\mathbb{R}^d, \overline{\mathcal{B}(\mathbb{R}^d)}, \lambda^{(d)})$, mais l'inclusion est stricte (on peut par exemple considérer la fonction de \mathbb{R} dans \mathbb{R} constante égale à 1, qui est intégrable sur tout compact mais pas sur \mathbb{R}).

Définition

Soit a et b deux réels tels que $a < b$.

On dit que $f : [a, b] \rightarrow \mathbb{R}$ est une **fonction en escalier** s'il existe une subdivision de $[a, b] : a = x_0 < x_1 < \dots < x_J = b$ et des réels y_1, \dots, y_J tels que $\forall i \in \llbracket 1, J \rrbracket, \forall x \in]x_{i-1}, x_i[, f(x) = y_i$.

L'ensemble de ces fonctions se note $\mathcal{R}([a, b])$.

Pour $h \in \mathcal{R}([a, b])$, on note $I(h) = \sum_{i=1}^J (x_i - x_{i-1})y_i$.

Remarque : $\mathcal{R}([a, b]) \subset \mathcal{S}([a, b])$.

Définition

Une fonction $f : [a, b] \rightarrow \mathbb{R}$ est dite **Riemann-intégrable** ssi :

$$\sup_{h \in \mathcal{R}([a, b]), h \leq f} I(h) = \inf_{h \in \mathcal{R}([a, b]), h \geq f} I(h)$$

On note alors $\int_a^b f(x)dx$ cette valeur.

Proposition

Soit $h \in \mathcal{R}([a, b])$. Alors $I(h) = \int_{[a, b]} h d\lambda$.

Démonstration : $\int_{[a, b]} h d\lambda = \sum_{i=1}^J y_i \lambda([x_{i-1}, x_i]) = I(h)$

Théorème

Soit $f : [a, b] \rightarrow \mathbb{R}$ une fonction Riemann-intégrable.

Alors f est mesurable pour la tribu de Lebesgue, et les intégrales de Riemann et de Lebesgue coïncident i.e.

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

.

Démonstration : Il existe deux fonctions en escalier $(h_n^+)_{n \in \mathbb{N}}$ et $(h_n^-)_{n \in \mathbb{N}}$ deux suites de fonctions en escalier telles que $h_n^- \leq f \leq h_n^+$ et $\lim_{n \rightarrow +\infty} I(h_n^+) = \lim_{n \rightarrow +\infty} I(h_n^-)$. On peut extraire une sous-suite croissante de $(h_n^-)_{n \in \mathbb{N}}$ et une sous-suite décroissante de $(h_n^+)_{n \in \mathbb{N}}$. Elles sont bornées. On pose par ailleurs h_∞^+ et h_∞^- les limites simples de $(h_n^+)_{n \in \mathbb{N}}$ et $(h_n^-)_{n \in \mathbb{N}}$. Elles sont mesurables.

On applique le théorème de convergence dominée à h_n^+ et à h_n^- :

$$\begin{aligned} \int_{[a, b]} h_\infty^+ d\lambda &= \lim_{n \rightarrow +\infty} \int_{[a, b]} h_n^+ d\lambda = \lim_{n \rightarrow +\infty} I(h_n^+) = \int_a^b f(x)dx \\ \int_{[a, b]} h_\infty^- d\lambda &= \lim_{n \rightarrow +\infty} \int_{[a, b]} h_n^- d\lambda = \lim_{n \rightarrow +\infty} I(h_n^-) = \int_a^b f(x)dx \end{aligned}$$

On a donc $\int_{[a, b]} h_\infty^+ d\lambda = \int_{[a, b]} h_\infty^- d\lambda$, soit $\int_{[a, b]} h_\infty^+ - h_\infty^- d\lambda = 0$ ou encore $h_\infty^+ = h_\infty^-$ presque partout (car $h_\infty^+ - h_\infty^- \geq 0$).

Puisque $h_\infty^- \leq f \leq h_\infty^+$, on a donc $f = h_\infty^+$ presque partout soit :

$$\int_{[a, b]} f d\lambda = \int_{[a, b]} h_\infty^+ d\lambda = \int_a^b f(x)dx$$

Remarque : Certaines fonctions peuvent être Lebesgue-intégrables sans être Riemann-intégrables, par exemple $f = 1_{\mathbb{Q}}$

Définition

Soit $a \in \mathbb{R}$ et $b \in]a, +\infty]$ (respectivement $b \in \mathbb{R}$ et $a \in [-\infty, a[$).

La fonction $f : [a, b[\rightarrow \mathbb{R}$ (respectivement $f :]a, b] \rightarrow \mathbb{R}$) est **localement Riemann-intégrable** si f est intégrable sur tout compact de $[a, b[$ (respectivement $]a, b]$).

Théorème

Toute fonction localement Riemann-intégrable est Lebesgue-intégrable si et seulement si elle est Riemann-absolument convergente (i.e. $\int_a^b |f(x)|dx$ existe et est finie).
 Dans ce cas les deux intégrales coïncident.

Conséquence : Les intégrales impropres absolument convergentes sont dans \mathcal{L}^1 , mais les intégrales impropres semi-convergentes ne sont pas dans \mathcal{L}^1 .

Théorème

Soit $f \in \mathcal{L}_{\text{loc}}^1$ et $a \in \mathbb{R}$. On définit :

$$F(x) = \int_{[a,x]} f d\lambda$$

Alors F est continue et dérivable presque partout, et $F' = f$ p.p.

Théorème

Soit $F : \mathbb{R} \rightarrow \mathbb{R}$ une fonction dérivable en tout point de \mathbb{R} .

Supposons $f = F' \in \mathcal{L}_{\text{loc}}^1$. Alors pour tous réels a et b tels que $a < b$:

$$\int_{[a,b]} f d\lambda = F(b) - F(a)$$

Proposition

Considérons $E = \mathbb{N}$, $\mathcal{E} = \mathcal{P}(\mathbb{N})$ et $\mu = \sum_{n \in \mathbb{N}} \delta_n$.

Soit $u : E \rightarrow \mathbb{R}$; on note $u_n = u(n)$. Si la série de terme général $(u_n)_{n \in \mathbb{N}}$ est absolument convergente, alors :

$$\int u(x) \mu(dx) = \sum_{n=0}^{+\infty} u_n$$

Définition

On note :

$$\ell^p = \mathcal{L}^p \left(\mathbb{N}, \mathcal{P}(\mathbb{N}), \sum_{n \in \mathbb{N}} \delta_n \right)$$

$$\ell^\infty = \mathcal{L}^\infty \left(\mathbb{N}, \mathcal{P}(\mathbb{N}), \sum_{n \in \mathbb{N}} \delta_n \right)$$

Section IV.3 - Mesure de densité

Proposition

Soit $f : (E, \mathcal{E}, \mu) \rightarrow ([0, +\infty], \mathcal{B}([0, +\infty]))$ une fonction mesurable.

L'application ν définie pour tout $A \in \mathcal{E}$ par $\nu(A) = \int_A f d\mu = \int_E f 1_A d\mu$ est une mesure sur (E, \mathcal{E}) .

Définition

On dit que ν est la **mesure de densité f** par rapport à μ .

Exemple : Considérons $E = \mathbb{R}$ équipé de la tribu de Lebesgue et de la mesure de Lebesgue λ . Soit f définie sur \mathbb{R} par :

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ e^{-x} & \text{si } x \geq 0 \end{cases}$$

On a alors, par exemple, $\nu([0, 1]) = 1 - \frac{1}{e}$, $\nu([-69, 420]) = 1 - e^{-420}$ et $\nu(\mathbb{R}) = 1$, ce qui fait par ailleurs de ν une mesure de probabilité.

Remarque : Si A est de mesure nulle pour μ alors $\nu(A) = 0$ donc A est de mesure nulle pour ν .
On dit que ν est absolument continue par rapport à μ et on note $\nu \ll \mu$.

Théorème

Une fonction borélienne $f : \mathbb{R} \rightarrow \mathbb{R}^+$ est la densité d'une mesure de probabilité \mathbb{P} ssi

$$\int_{\mathbb{R}} f(x) \lambda(dx) = 1$$

Dans ce cas, $\mathbb{P}(A) = \int_A f(x) \lambda(dx)$ et on dit que f est la **dérivée de Radon-Nikodym** de \mathbb{P} par rapport à λ .

Démonstration : Immédiate en prenant $A = \mathbb{R}$ dans la définition d'une mesure de densité f .

Théorème (Continuité des intégrales dépendant d'un paramètre)

Soit (E, \mathcal{E}, μ) un espace mesuré, (U, d) un espace métrique, $f_u : E \rightarrow \mathbb{R}$ une fonction dépendant d'un paramètre $u \in U$ et $u_0 \in U$.

On suppose que :

- Pour presque tout $u \in U$, la fonction $x \mapsto f_u(x)$ est mesurable.
 - Pour presque tout $x \in E$, la fonction $u \mapsto f_u(x)$ est continue en u_0 .
 - Il existe une fonction positive $g \in \mathcal{L}^1(E, \mathcal{E}, \mu)$ telle que $\forall u \in U, |f_u(x)| \leq g(x)$ pour presque tout x .
- Alors $u \mapsto \int_E f_u(x) \mu(dx)$ est définie pour presque tout $u \in U$ et continue en u_0 .

Théorème (Dérivabilité sous le signe somme)

Soit (E, \mathcal{E}, μ) un espace mesuré, $U \subset \mathbb{R}$ muni de sa tribu de Borel et $I \subset U$ un intervalle, $f : I \times E \rightarrow \mathbb{R}$ une fonction dépendant d'un paramètre et $u_0 \in I$. On suppose que :

- Pour presque tout $u \in U$, la fonction $x \mapsto f_u(x) \in \mathcal{L}^1(E, \mathcal{E}, \mu)$
- Pour presque tout $x \in E$, la fonction $u \mapsto f_u(x)$ est dérivable en u_0 .
- Il existe une fonction positive $g \in \mathcal{L}^1(E, \mathcal{E}, \mu)$ telle que $\forall u \in U, |f_u(x) - f(u_0, x)| \leq g(x)|u - u_0|$ pour presque tout x .

Alors $u \mapsto \int_E f_u(x) \mu(dx)$ est dérivable en u_0 , de dérivée $\int_E \frac{\partial f}{\partial u}(u_0, x) \mu(dx)$.

Chapitre V. Espaces L^p

Section V.1 - Relations d'équivalence

Définition

Soit E un ensemble. On dit qu'une relation \sim est une **relation d'équivalence** ssi :

- Elle est **réflexive** ($\forall x \in E, x \sim x$)
- Elle est **symétrique** ($\forall x, y \in E, x \sim y \Rightarrow y \sim x$)
- Elle est **transitive** ($\forall x, y, z \in E, x \sim y \wedge y \sim z \Rightarrow x \sim z$)

Définition

Soit \sim une relation d'équivalence sur un ensemble E et $x \in E$.

On appelle **classe d'équivalence de x** l'ensemble $\{y \in E, y \sim x\}$.

On le note \dot{x} ou $[x]$.

Définition

Soit \sim une relation d'équivalence sur un ensemble E et $x \in E$.

On appelle **l'ensemble quotient de E par \sim** l'ensemble des classes d'équivalences des éléments de E , qu'on note E/\sim .

Proposition

E/\sim forme une partition de E .

Démonstration : $\forall x \in E, x \in \dot{x}$ donc $E = \cup_{x \in E} \dot{x}$.

Si $\dot{x} \cap \dot{y} \neq \emptyset$, soit $z \in \dot{x} \cap \dot{y}$. Soit $a \in \dot{x}$ et $b \in \dot{y}$, alors $a \sim z \sim b$ donc $\dot{x} = \dot{a} = \dot{b} = \dot{y}$. On a donc une partition de E . (les classes d'équivalences sont deux-à-deux disjointes et leur réunion forme E)

Exemple : Soit $E = \mathcal{L}^p(E, \mathcal{E}, \mu)$ et $p \in [1, +\infty]$.

La relation \sim définie par $f \sim g \Leftrightarrow f - g = 0$ p.p est une relation d'équivalence.

On aura alors, par exemple, $1_{\mathbb{Q}} \sim 0$ si μ est la mesure de Lebesgue.

Définition

Soit \sim une relation d'équivalence sur un ensemble E .

Une application $f : E \rightarrow E$ est **compatible** avec \sim ssi

$$\forall x \in E, \forall y \in E, x \sim y \Rightarrow f(x) \sim f(y)$$

On peut alors définir une fonction f/\sim sur l'ensemble quotient E/\sim . Pour $C \in E/\sim$, on considère un représentant $x \in C$ et on pose :

$f/\sim(C) = f(\dot{x})$. On notera souvent f au lieu de f/\sim .

Définition

Soit \sim une relation d'équivalence sur un ensemble E .

Une loi interne $*$ est **compatible** avec \sim ssi

$$\forall x_1, x_2, y_1, y_2 \in E, x_1 \sim x_2 \text{ et } y_1 \sim y_2 \Rightarrow x_1 * y_1 \sim x_2 * y_2$$

On définit alors la loi quotient $*/\sim$ sur E/\sim en associant aux classes d'équivalences de x et y la classe d'équivalence de $x * y$. On notera souvent $*$ au lieu de $*/\sim$.

Section V.2 - Construction de l'e.v.n. L^p

Définition

Soit $p \in [1, +\infty]$. On note $L^p(E, \mathcal{E}, \mu)$ le quotient de l'espace $\mathcal{L}^p(E, \mathcal{E}, \mu)$ par la relation d'égalité μ -presque partout. On note $L^p(\mathbb{R}^d) = L^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^{(d)})$.

Remarque : $L^p(\mathbb{N}, \mathcal{P}(\mathbb{N}), \sum_{n \in \mathbb{N}} \delta_n) = \mathcal{L}^p(\mathbb{N}, \mathcal{P}(\mathbb{N}), \sum_{n \in \mathbb{N}} \delta_n) = \ell^p$, puisque l'égalité presque partout pour la mesure $\sum_{n \in \mathbb{N}} \delta_n$ sur \mathbb{N} est l'égalité (chaque classe d'équivalence contient un unique élément, donc les ensembles $L^p(\mathbb{N}, \mathcal{P}(\mathbb{N}), \sum_{n \in \mathbb{N}} \delta_n)$ et $\mathcal{L}^p(\mathbb{N}, \mathcal{P}(\mathbb{N}), \sum_{n \in \mathbb{N}} \delta_n)$ sont en bijection ; on les identifie).

Proposition

Les opérations $+$ et \times de $\mathcal{L}^p(E, \mathcal{E}, \mu)$ sont compatibles avec la relation d'équivalence μ -pp.

Démonstration : Soit f_1, f_2, g_1, g_2 dans $\mathcal{L}^p(E, \mathcal{E}, \mu)$ avec $f_1 \sim f_2$ et $g_1 \sim g_2$. Alors $(f_1 + g_1) - (f_2 + g_2) = (f_1 - f_2) + (g_1 - g_2) = 0$ presque partout, donc $f_1 + g_1 \sim f_2 + g_2$, et $f_1 g_1 - f_2 g_2 = f_1(g_1 - g_2) + (f_1 - f_2)g_2 = 0$ presque partout, donc $f_1 g_1 \sim f_2 g_2$.

Proposition

$L^p(E, \mathcal{E}, \mu)$ est un espace vectoriel.

Remarque : Soit $x_0 \in E$. La fonction d'évaluation en x_0 (appelée également trace sur $\{x_0\}$) de $L^p(E, \mathcal{E}, \mu) \rightarrow \mathbb{R}$ qui à f associe $f(x_0)$ n'est pas compatible avec la relation d'équivalence égalité μ -pp. En d'autres termes, la valeur des éléments de $L^p(E, \mathcal{E}, \mu)$ en un point n'a pas de sens.

Proposition

La forme linéaire $f \mapsto \int_E f^p d\mu$ sur $\mathcal{L}^p(E, \mathcal{E}, \mu)$ est compatible avec la relation d'équivalence μ -pp.

Démonstration : Soit f, g dans $\mathcal{L}^p(E, \mathcal{E}, \mu)$ avec $f \sim g$, alors $\int_E f^p d\mu = \int_E g^p d\mu$.

Proposition

Dans $L^p(E, \mathcal{E}, \mu)$:

$$\int_E |f|^p d\mu = 0 \Leftrightarrow f = 0$$

Définition

On dit que $M \in \mathbb{R}$ est un presque majorant de $f : E \rightarrow \mathbb{R}$ si $f(x) \leq M$ pour presque tout $x \in E$.

Définition

Soit $f : E \rightarrow \mathbb{R}$. Si f admet un ou plusieurs presque majorants, on appelle **borne supérieure essentielle** le plus petit d'entre eux et on le note $\sup_{\text{ess}} f$

Définition

Soit $p \in [1, +\infty]$.

- Si $p \in]1, +\infty[$, son conjugué est $\frac{p}{p-1}$ i.e. le réel q tel que $\frac{1}{p} + \frac{1}{q} = 1$.
- Si $p = 1$, son conjugué est $+\infty$.
- Si $p = \infty$, son conjugué est 1.

Théorème (Inégalité de Young)

Soit p et q dans $]1, +\infty[$ conjugués. Alors :

$$\forall (a, b) \in \mathbb{R}^+ \times \mathbb{R}^+, ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

Démonstration : Par concavité de $x \mapsto \ln(x)$ sur $]0, +\infty[$, on a $\forall t \in [0, 1], \ln(ta^p + (1-t)b^q) \geq t \ln(a^p) + (1-t) \ln(b^q)$. En posant $t = \frac{1}{p}$, alors $1-t = \frac{1}{q}$ et :

$$\ln\left(\frac{a^p}{p} + \frac{b^q}{q}\right) \geq \frac{1}{p} \ln(a^p) + \frac{1}{q} \ln(b^q) = \ln(ab)$$

d'où le résultat en passant à l'exponentielle strictement croissante.

Définition

Pour $f \in L^p(E, \mathcal{E}, \mu)$ avec $p \in [1, +\infty]$, on note :

$$\|f\|_p = \left(\int_E |f|^p d\mu \right)^{\frac{1}{p}} \text{ si } p < +\infty \text{ et } \|f\|_\infty = \sup \text{ess } |f|$$

Théorème (Inégalité de Hölder)

Soit p et q dans $]1, +\infty[$ conjugués. Soit $f \in L^p(E, \mathcal{E}, \mu)$ et $g \in L^q(E, \mathcal{E}, \mu)$. Alors :

$$fg \in L^1(E, \mathcal{E}, \mu) \text{ et } \|fg\|_1 \leq \|f\|_p \|g\|_q$$

Démonstration : Si $p = 1$ ou $q = 1$ alors le résultat est trivial, si $f = 0$ ou $g = 0$ aussi. On élimine donc ces cas, et on suppose $p \in]1, +\infty[$. L'inégalité de Young donne :

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

Ainsi $fg \in L^1(E, \mathcal{E}, \mu)$ et, en intégrant :

$$\|fg\|_1 \leq \frac{1}{p} \|f\|_p^p + \frac{1}{q} \|g\|_q^q$$

Pour $\lambda > 0$, le même raisonnement sur les fonctions λf et g conduisent à l'inégalité :

$$\|fg\|_1 \leq \frac{\lambda^{p-1}}{p} \|f\|_p^p + \frac{1}{\lambda q} \|g\|_q^q$$

On pose alors $\lambda = \frac{\|g\|_q^{\frac{q}{p}}}{\|f\|_p}$, ce qui nous permet d'obtenir :

$$\|fg\|_1 \leq \frac{1}{p} \left(\frac{\|g\|_q^{\frac{q}{p}}}{\|f\|_p} \right)^{p-1} \|f\|_p^p + \frac{1}{q} \frac{\|f\|_p}{\|g\|_q^{\frac{q}{p}}} \|g\|_q^q = \frac{1}{p} \|g\|_q^{\frac{q(p-1)}{p}} \|f\|_p + \frac{1}{q} \|f\|_p \|g\|_q^{\frac{q(p-1)}{p}}$$

Or $\frac{q(p-1)}{p} = 1$ et $\frac{1}{p} + \frac{1}{q} = 1$, d'où :

$$\|fg\|_1 \leq \|f\|_p \|g\|_q$$

Théorème (Inégalité de Minkowski)

Soit $p \in [1, +\infty]$. Soit f et g dans $L^p(E, \mathcal{E}, \mu)$. Alors :

$$f + g \in L^p(E, \mathcal{E}, \mu) \text{ et } \|f + g\|_p \leq \|f\|_p + \|g\|_p$$

Démonstration : Puisque $p \in [1, +\infty]$, $x \mapsto x^p$ est convexe sur \mathbb{R}^+ donc :

$$\left| \frac{1}{2}f + \frac{1}{2}g \right|^p \leq \left| \frac{1}{2}|f| + \frac{1}{2}|g| \right|^p \leq \frac{1}{2}|f|^p + \frac{1}{2}|g|^p$$

$$\Leftrightarrow |f + g|^p \leq 2^{p-1}|f|^p + 2^{p-1}|g|^p$$

On a donc $f + g \in L^p(E, \mathcal{E}, \mu)$. Alors :

$$\|f + g\|_p^p = \int_E |f + g|^{p-1} |f + g| d\mu \leq \int_E |f + g|^{p-1} |f| + \int_E |f + g|^{p-1} |g|$$

Appliquons l'inégalité de Hölder :

$$\begin{aligned} \int_E |f + g|^{p-1} |f| &\leq \| |f + g|^{p-1} \|_{\frac{p}{p-1}} \|f\|_p = \left(\int_E (|f + g|^{p-1})^{\frac{p}{p-1}} d\mu \right)^{\frac{p-1}{p}} \|f\|_p \\ &= \left(\left(\int_E (|f + g|^p) d\mu \right)^{\frac{1}{p}} \right)^{p-1} \|f\|_p = (\|f + g\|_p)^{p-1} \|f\|_p \end{aligned}$$

De manière équivalente, on a aussi :

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

Ainsi :

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

Proposition

Soit $p \in [1, +\infty]$.

Alors $L^p(E, \mathcal{E}, \mu)$ est un espace vectoriel normé, de norme $\|f\|_p$.

Démonstration : Clairement $\|f\|_p = 0 \Leftrightarrow f = 0$ et $\|\lambda f\|_p = |\lambda| \|f\|_p$. L'inégalité triangulaire n'est autre que l'inégalité de Minkowski démontrée ci-dessus.

Remarque : Il ne faut pas confondre " f une fonction continue presque partout " et " f est égale presque partout à une fonction continue ".

Définition

Soit $f : E \rightarrow \mathbb{R}$ une classe de fonctions.

S'il y a une fonction continue dans cette classe, on dira que f est **continue**.

Dans ce cas, pour $x_0 \in E$ on donnera à $f(x_0)$ la valeur de son représentant continu en x_0 .

Section V.3 - Propriétés de l'e.v.n. L^p

Théorème (Fischer-Riesz)

Soit $p \in [1, +\infty]$. $L^p(E, \mathcal{E}, \mu)$ est un espace de Banach.

Démonstration : On commence par traiter le cas $p = +\infty$. Soit $(f_n)_{n \in \mathbb{N}}$ une suite de Cauchy d'éléments de $L^\infty(E, \mathcal{E}, \mu)$. $\forall k \in \mathbb{N}^*, \exists N \in \mathbb{N}, \forall (m, n) \in \mathbb{N}^2, m > n > N \Rightarrow \|f_m - f_n\|_\infty < \frac{1}{k}$. Il existe Z_k de mesure nulle tel que $\forall k \in \mathbb{N}^*, \exists N \in \mathbb{N}, \forall (m, n) \in \mathbb{N}^2, \forall x \in E \setminus Z_k, m > n > N \Rightarrow |f_m(x) - f_n(x)| < \frac{1}{k}$. $Z = \cup_{k \in \mathbb{N}^*} Z_k$ est de mesure nulle, alors $\forall k \in \mathbb{N}^*, \exists N \in \mathbb{N}, \forall (m, n) \in \mathbb{N}^2, \forall x \in E \setminus Z, m > n > N \Rightarrow |f_m(x) - f_n(x)| < \frac{1}{k}$. On en déduit que $\forall x \in E \setminus Z, (f_n(x))_{n \in \mathbb{N}}$ est une suite de Cauchy d'éléments de \mathbb{R} , qui converge car \mathbb{R} est complet. Notons $f(x)$ sa limite ; $(f_n)_{n \in \mathbb{N}}$ converge simplement vers f sur $E \setminus Z$. Ainsi $\forall x \in E \setminus Z, \forall k \in \mathbb{N}^*, \exists N \in \mathbb{N}, \forall n \in \mathbb{N}, |f_n(x) - f(x)| < \frac{1}{k}$. Ainsi $f \in L^\infty(E, \mathcal{E}, \mu)$ et $\lim_{n \rightarrow +\infty} \|f_n - f\|_\infty = 0$, donc $(f_n)_{n \in \mathbb{N}}$ converge dans $L^\infty(E, \mathcal{E}, \mu)$.

Désormais, soit $p \in [1, +\infty[$, et $(f_n)_{n \in \mathbb{N}}$ une suite de Cauchy d'éléments de $L^p(E, \mathcal{E}, \mu)$. On extrait (f_{n_k}) telle que $\|f_{n_{k+1}} - f_{n_k}\|_p < \frac{1}{2^k}$. On note $g_n = \sum_{k=1}^n |f_{n_{k+1}} - f_{n_k}|$. Alors :

$$\|g_n\|_p = \left\| \sum_{k=1}^n |f_{n_{k+1}} - f_{n_k}| \right\|_p \leq \sum_{k=1}^n \|f_{n_{k+1}} - f_{n_k}\|_p \leq 1 - \frac{1}{2^n} \leq 1$$

Ainsi $(g_n(x))$ converge vers $g(x)$ presque partout. Soit s et t deux entiers avec $s > t$. Par télescopage, $|f_{n_s} - f_{n_t}| \leq g - g_{t-1}$ donc $(f_{n_k}(x))$ est de Cauchy pour presque tout x . Ainsi, elle converge, et on note $f(x)$ sa limite. Lorsque

$s \rightarrow +\infty$, on $|f - f_{n_t}| \leq g - g_{t-1} \leq g$, ce qu'on réécrit $|f_{n_k} - f(x)|^p < g^p(x)$, soit $|f_{n_k}(x) - f(x)|^p \rightarrow 0$ lorsque $n_k \rightarrow +\infty$. D'après le théorème de convergence dominée, $f \in L^p(E, \mathcal{E}, \mu)$ et $\lim_{k \rightarrow +\infty} \|f_{n_k}(x) - f(x)\|_p = 0$.

Proposition

$L^2(E, \mathcal{E}, \mu)$ est un espace de Hilbert.

Démonstration : $\langle f, g \rangle = \int_E f g d\mu$ est un produit scalaire sur $L^2(E, \mathcal{E}, \mu)$. C'est un espace préhilbertien, et il est complet pour la norme induite par le produit scalaire par la proposition précédente.

Théorème (Riesz)

Soit $p \in]1, +\infty[$ et q son conjugué.

Pour tout $\phi \in (L^p(E, \mathcal{E}, \mu))'$, il existe un unique $g \in L^q(E, \mathcal{E}, \mu)$ tel que $\phi = f \mapsto \int f g d\mu$.

En outre, $\|\phi\|_{(L^p)'} = \|g\|_q$.

Remarque : On identifie $(L^p)'$ et L^q : $(L^p)' = L^q$.

Attention cependant, on a exclu $p = 1$: $(L^\infty)' \neq L^1$. Le dual de L^∞ contient strictement L^1 .

Définition

Soit $p \in]1, +\infty[$ et q son conjugué.

Pour $f \in L^p$ et $g \in L^q$, on note $\langle f, g \rangle = \int_E f g d\mu$.

$\langle \cdot, \cdot \rangle$ s'appelle un **crochet de dualité**.

Avec ces notations, pour $\phi \in (L^p)'$, il lui correspond un unique $g \in L^q$ par Riesz. On a alors $\phi(f) = \langle f, g \rangle$.

Remarque : Dans l'espace de Hilbert L^2 , le crochet de dualité est le produit scalaire.

Théorème

Soit $p \in [1, +\infty[$.

L'ensemble $C_c(E)$ des fonctions continues à support compact de E dans \mathbb{R} est dense dans $L^p(E, \mathcal{E}, \mu)$.

Mieux encore, l'ensemble $C_c^\infty(E)$ des fonctions infiniment dérivables à support compact de E dans \mathbb{R} est dense dans $L^p(E, \mathcal{E}, \mu)$.

Section V.4 - L'espace $L^2_{\mathbb{C}}$

Définition

Soit $f : (E, \mathcal{E}, \mu) \rightarrow (\mathbb{C}, \mathcal{B}(\mathbb{C}))$ une fonction mesurable.

On dit que f est **intégrable** par rapport à la mesure μ ssi

$$\int |f| d\mu < +\infty$$

On note $\mathcal{L}^1_{\mathbb{C}}(E, \mathcal{E}, \mu)$ l'ensemble des fonctions intégrables par rapport à μ .

Lorsque f est intégrable par rapport à la mesure μ , on définit l'**intégrale** de f par

$$\int f d\mu = \int \operatorname{Re} f d\mu + i \int \operatorname{Im} f d\mu$$

Définition

Pour tout $p \in [1, +\infty[$, on définit :

$$\mathcal{L}_{\mathbb{C}}^p(E, \mathcal{E}, \mu) = \{f : E \rightarrow \mathbb{C} \text{ mesurable ; } \int |f|^p d\mu < +\infty\}$$

$$\mathcal{L}_{\mathbb{C}}^{\infty}(E, \mathcal{E}, \mu) = \{f : E \rightarrow \mathbb{C} \text{ mesurable ; } \exists C > 0, |f| \leq C \text{ p.p.}\}$$

Pour tout $p \in [1, +\infty]$, on définit

$$L_{\mathbb{C}}^p(E, \mathcal{E}, \mu) = \mathcal{L}_{\mathbb{C}}^p(E, \mathcal{E}, \mu) / \sim$$

où \sim est la relation d'égalité presque partout.

Proposition

Pour tout $p \in [1, +\infty]$, $L_{\mathbb{C}}^p(E, \mathcal{E}, \mu)$ est un espace de Banach.
 $L_{\mathbb{C}}^p(E, \mathcal{E}, \mu)$ est un espace de Hilbert.

Proposition

On considère l'espace de Hilbert $H = L_{\mathbb{C}}^2([0, 2\pi], \overline{\mathcal{B}([0, 2\pi])}, \frac{1}{2\pi}\lambda)$.

Le produit scalaire est $\langle f, g \rangle = \frac{1}{2\pi} \int_{[0, 2\pi]} f g d\lambda$.

Alors, H admet la base hilbertienne $\{e_n, n \in \mathbb{Z}\}$, où e_n est défini par $e_n(x) = e^{inx} = \cos(nx) + i \sin(nx)$

Démonstration : Soit $n, m \in \mathbb{Z}$. Par le calcul :

$$\langle e_n, e_m \rangle = \frac{1}{2\pi} \int_{[0, 2\pi]} e^{inx} e^{-imx} dx = \delta_{n,m}$$

Soit $f \in H$, et $\epsilon > 0$. $C_c([0, 2\pi], \mathbb{C})$ est dense dans $H = L_{\mathbb{C}}^2([0, 2\pi], \overline{\mathcal{B}([0, 2\pi])}, \frac{1}{2\pi}\lambda)$. Ainsi il existe $u \in C_c([0, 2\pi], \mathbb{C})$ tel que $\|u - f\|_2 < \frac{\epsilon}{2}$. On pose :

$$D_k(x) = \sum_{n=-k}^k e^{inx} = \frac{\sin((k + \frac{1}{2})x)}{\sin \frac{x}{2}}$$

qu'on appelle le k -ième noyau de Dirichet. On pose alors :

$$F_K(x) = \frac{1}{K} \sum_{k=0}^{K-1} D_k(x) = \frac{1}{K} \left(\frac{\sin \frac{Kx}{2}}{\sin \frac{x}{2}} \right)$$

qu'on appelle K -ième terme du noyau de Fejér. Alors $\forall K \in \mathbb{N}^*, F_K(x) \geq 0, \frac{1}{2\pi} \int_{[-\pi, \pi]} F_K(x) d\lambda = 1$ et $\forall h > 0, \frac{1}{2\pi} \int_{[-\pi, -\frac{h}{2}] \cup [\frac{h}{2}, \pi]} F_K(x) d\lambda \rightarrow 0$ lorsque $K \rightarrow +\infty$. En notant $Z_K = \frac{1}{2\pi} \int_{[\frac{h}{2}, 2\pi - \frac{h}{2}]} F_K(x) d\lambda$, ceci implique que $\forall h > 0, \lim_{K \rightarrow +\infty} Z_K = 0$. On pose maintenant :

$$u_K(x) = \frac{1}{2\pi} \int_{[0, 2\pi]} u(x-t) F_K(t) d\lambda(t)$$

On a alors :

$$\|u_K - u\|_2 = \|x \mapsto \frac{1}{2\pi} \int_{[0, 2\pi]} [u(x-t) - u(x)] F_K(t) d\lambda(t)\|_2$$

Soit $h > 0$ tel que $|y_2 - y_1| < \frac{h}{2} \Rightarrow |u(y_1) - u(y_2)| \leq \frac{\epsilon}{4}$. Sur $[0, \frac{h}{2}] \cup [2\pi - \frac{h}{2}, 2\pi]$, on a $|u(x-t) - u(x)| < \frac{\epsilon}{4}$. Donc :

$$\|x \mapsto \frac{1}{2\pi} \int_{[0, \frac{h}{2}] \cup [2\pi - \frac{h}{2}, 2\pi]} [u(x-t) - u(x)] F_K(t) d\lambda(t)\|_2 < \frac{\epsilon}{4}$$

$$\text{et } \exists K \in \mathbb{N}^*, \|x \mapsto \frac{1}{2\pi} \int_{[\frac{h}{2}, 2\pi - \frac{h}{2}]} [u(x-t) - u(x)] F_K(t) d\lambda(t)\|_2 \leq M Z_K < \frac{\epsilon}{4}$$

$$\Rightarrow \|u_K - u\|_2 = \|x \mapsto \frac{1}{2\pi} \int_{[0, 2\pi]} [u(x-t) - u(x)] F_K(t) d\lambda(t)\|_2 < \frac{\epsilon}{2}$$

Or, on a :

$$u_K(x) = \sum_{k=0}^{K-1} \sum_{n=-k}^k \left(\frac{1}{2\pi} \frac{1}{K} \int_{[0,2\pi]} u(x-t) d\lambda(t) \right) e^{int}$$

Donc u_K est une combinaison linéaire de e_n . Il existe un $N \in \mathbb{N}$ et des c_n tels que $\|u - \sum_{n=-N}^N c_n e_n\|_2 < \frac{\epsilon}{2}$. Alors :

$$\|f - \sum_{n=-N}^N c_n e_n\|_2 \leq \|f - u\|_2 + \|u - \sum_{n=-N}^N c_n e_n\|_2 < \epsilon$$

On conclut que $H = \overline{\text{Vect}\{e_n, n \in \mathbb{Z}\}}$.

Chapitre VI. Introduction aux probabilités

Section VI.1 - Mesure de probabilité

Définition

On appelle espace probabilisé un espace mesuré pour lequel la mesure \mathbb{P} est une mesure de probabilité. ($\mathbb{P}(\Omega) = 1$)

Définition

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé.
On appelle **espace d'états** l'ensemble Ω .
On appelle **événements** les éléments de \mathcal{F} .
La mesure \mathbb{P} associe à chaque événement une **probabilité**.

Définition

Les singletons de \mathcal{F} sont appelés **événements élémentaires**.

Définition

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé avec Ω fini et $\mathcal{F} = \mathcal{P}(\Omega)$.
On dit qu'il y a **équiprobabilité** si la mesure \mathbb{P} est définie par

$$\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1], \mathbb{P}(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}$$

Les événements élémentaires sont dits **équiprobables**. Ils ont tous la même probabilité $\frac{1}{\text{Card}(\Omega)}$.
On dit également que \mathbb{P} est la **mesure uniforme discrète** sur l'ensemble Ω .

Théorème

Soit $\Omega = \{\omega_i; i \in I\}$ un ensemble fini ou dénombrable. Soit $\mathcal{F} = \mathcal{P}(\Omega)$.
Toute mesure de probabilité \mathbb{P} est caractérisée par sa valeur sur les atomes : $p_i = \mathbb{P}(\omega_i)$ pour tout $i \in I$.
Réciproquement, soit $(p_i)_{i \in I}$ une suite de réels positifs de nombres réels positifs tels que $\sum_{i \in I} p_i = 1$ alors il existe une mesure de probabilité \mathbb{P} telle que $\forall i \in I, \mathbb{P}(\omega_i) = p_i$

Démonstration : Soit $\Omega = \{\omega_i; i \in I\}$ un ensemble fini ou dénombrable. Supposons connaître $p_i = \mathbb{P}(\omega_i)$ pour tout $i \in I$.

Soit $A \in \mathcal{F} = \mathcal{P}(\Omega)$. $A = \cup_{i \in I, \omega_i \in A} \omega_i$ donc $\mathbb{P}(A) = \sum_{i \in I, \omega_i \in A} p_i$ est définie de manière unique.

Soit $(p_i)_{i \in I}$ une suite de réels positifs tels que $\sum_{i \in I} p_i = 1$. On suppose $\mathbb{P}(\omega_i) = p_i$. Soit $A = \cup_{i \in I, \omega_i \in A} \omega_i \in \mathcal{F}$. Alors on définit la mesure \mathbb{P} par $\mathbb{P}(A) = \sum_{i \in I, \omega_i \in A} p_i$.

Définition

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Soit $A \in \mathcal{F}$.
On dit que A est **presque sûr** ssi $\mathbb{P}(A) = 1$.

Section VI.2 - Probabilité conditionnelle

Définition

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, et A et B deux événements avec $\mathbb{P}(B) > 0$. La **probabilité conditionnelle** de A sachant B est définie par

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

Remarque : $A \mapsto \mathbb{P}(A|B)$ définit une mesure de probabilité sur (Ω, \mathcal{F}) .

Proposition (Formule des probabilités totales)

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Soit $(E_i)_{i \in I}$ une partition des évènements de mesure non nulle, avec I fini ou dénombrable.

Pour tout évènement A , on a :

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A|E_i)\mathbb{P}(E_i)$$

Théorème (Bayes)

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Soit $(E_i)_{i \in I}$ une partition des évènements de mesure non nulle, avec I fini ou dénombrable.

Soit A un évènement et $n \in I$. Alors :

$$\mathbb{P}(E_n|A) = \frac{\mathbb{P}(A|E_n)\mathbb{P}(E_n)}{\sum_{i \in I} \mathbb{P}(A|E_i)\mathbb{P}(E_i)}$$

Définition

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

On dit que deux évènements A et B sont **indépendants** ssi :

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

Remarque : Si $\mathbb{P}(B) > 0$, alors A et B sont indépendants ssi $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Définition

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, $(A_i)_{i \in I}$ une famille d'évènements.

Les A_i sont **mutuellement indépendants** ssi :

$$\forall J \subset I, J \text{ fini}, \mathbb{P}(\cap_{i \in J} A_i) = \prod_{i \in J} \mathbb{P}(A_i)$$

Remarque : L'indépendance mutuelle entraîne l'indépendance deux-à-deux, mais la réciproque est fautive. Prenons $\Omega = \llbracket 1, 6 \rrbracket^2$, $\mathcal{F} = \mathcal{P}(\Omega)$ et \mathbb{P} la mesure d'équiprobabilité. Alors les évènements $A_1 = \{6\} \times \llbracket 1, 6 \rrbracket$, $A_2 = \llbracket 1, 6 \rrbracket \times \{6\}$ et $A_3 = \{(x, x); x \in \llbracket 1, 6 \rrbracket\}$ sont deux-à-deux indépendants, mais pas mutuellement indépendants.

Section VI.3 - Variables aléatoires

Définition

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

On appelle **variable aléatoire** (de Ω à valeurs dans E) toute fonction mesurable de Ω dans E .

Définition

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, (E, \mathcal{E}) un espace mesuré et X une variable aléatoire.

L'application P_X définie de \mathcal{E} dans $[0, 1]$ par $P_X(A) = \mathbb{P}(X^{-1}(A))$ est une mesure de probabilité sur (E, \mathcal{E}) , que l'on appelle **loi de X** .

On ne peut que recommander d'aller voir [la vidéo de John Cagnol](#), qui introduit les variables aléatoires par l'exemple du jeu de l'oie.

Exemples : Pour $A \in \mathcal{E}$, $\mathbb{P}(X \in A)$ signifie $\mathbb{P}(X^{-1}(A))$.

Pour $E = \mathbb{R}$, $\mathcal{E} = \overline{\mathcal{B}(\Omega)}$ et $a \in \mathbb{R}$, $\mathbb{P}(X \geq a)$ signifie $\mathbb{P}(X^{-1}([a, +\infty[))$. $\mathbb{P}(X^2 + 1 \geq a)$ signifie $\mathbb{P}(\{\omega \in \Omega; X^2(\omega) + 1 \geq a\})$.

$\mathbb{P}(X = Y)$ signifie $\mathbb{P}(\{\omega \in \Omega; X(\omega) = Y(\omega)\})$.

Remarque : Supposons E au plus dénombrable et prenons $\mathcal{E} = \mathcal{P}(E)$. Puisqu'une variable aléatoire X est une fonction mesurable de (Ω, \mathcal{F}) dans (E, \mathcal{E}) , il y a équivalence entre " X est une variable aléatoire" et " $\forall e \in E, X^{-1}(\{e\}) \in \mathcal{F}$ ".

Définition

Soit X une variable aléatoire.

On appelle **tribu engendrée par la variable aléatoire**, et on note $\sigma(X)$, la tribu $\sigma(X^{-1}(\mathcal{E}))$.

Section VI.4 - Moments

Définition

Soit X une variable aléatoire sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans \mathbb{R} .

On dit que X admet un **moment** d'ordre $n \in \mathbb{N}^*$ si $X \in L^n(\Omega, \mathcal{F}, \mathbb{P})$. Dans ce cas, on note :

$$m_n = \int_{\Omega} X^n d\mathbb{P}$$

Le moment d'ordre 1 est appelé **espérance** de la variable aléatoire et noté $\mathbb{E}(X)$.

Remarque : $p \leq q \Rightarrow L^q(\Omega, \mathcal{F}, \mathbb{P}) \subset L^p(\Omega, \mathcal{F}, \mathbb{P})$ puisque \mathbb{P} est une mesure finie.

Proposition

Soit X une variable aléatoire sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans (E, \mathcal{E}) et $h : E \rightarrow \mathbb{R}$ une fonction mesurable.

Alors $h(X)$ est une variable aléatoire sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans \mathbb{R} .

Théorème (de transfert)

Soit X une variable aléatoire sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans (E, \mathcal{E}) . Alors pour toute fonction mesurable bornée $h : E \rightarrow \mathbb{R}$:

$$\mathbb{E}(h(X)) = \int_E h dP_X$$

Démonstration : Soit $A \in \mathcal{E}$ et $h = 1_A$.

$$\mathbb{E}(h(X)) = \int_{\Omega} 1_A(X) d\mathbb{P} = \mathbb{P}(X^{-1}(A)) = P_X(A) = \int_E 1_A dP_X = \int_E h dP_X$$

On a donc l'égalité pour toute fonction indicatrice, et par linéarité de l'intégrale, cela s'étend pour toute fonction étagée $h : E \rightarrow \mathbb{R}^+$.

Soit $h : E \rightarrow [0, +\infty]$ une fonction mesurable. Il existe une suite croissante $(h_n)_{n \in \mathbb{N}}$ de fonctions étagées positives convergeant simplement vers h . Le théorème de transfert s'applique aux (h_n) , et le théorème de convergence monotone permet d'obtenir le théorème pour h .

Soit $h : E \rightarrow \mathbb{R}$ une fonction mesurable. Le théorème de transfert s'applique à $|h|$:

$$\mathbb{E}(|h(X)|) = \int_E |h| dP_X$$

Ainsi $h \in \mathcal{L}^1(E, \mathcal{E}, P_X) \Leftrightarrow h(X) \in \mathcal{L}^1(E, \mathcal{E}, \mathbb{P})$. On décompose $h = h^+ - h^-$ et on applique le théorème de transfert à h^+ et h^- , ce qui conclut.

Remarque : Si il existe une mesure μ telle que pour toute fonction mesurable bornée $h : E \rightarrow \mathbb{R}$, $\mathbb{E}(h(X)) = \int_E h d\mu$, alors $\mu = P_X$ est la loi de X .

Théorème

Soit X une variable aléatoire sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans E dont la loi P_X admet une densité f_X et soit $h : E \rightarrow \mathbb{R}$ une fonction mesurable telle que :

$$\int_{\mathbb{R}} |h(x)| f_X(x) \lambda(dx) < +\infty$$

Alors, X admet un moment d'ordre 1 et :

$$\mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x) f_X(x) \lambda(dx)$$

Proposition

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans \mathbb{R} .

- Si (X_n) est une suite croissante et positive, alors $\lim_{n \rightarrow +\infty} \mathbb{E}(X_n) = \mathbb{E}(\lim_{n \rightarrow +\infty} X_n)$ (théorème de la convergence monotone)
- Si $(X_n)_{n \in \mathbb{N}}$ est une suite positive alors $\mathbb{E}(\liminf X_n) \leq \liminf \mathbb{E}(X_n)$ (lemme de Fatou)
- Si $\forall n \in \mathbb{N}, X_n \leq Z$ avec $Z \in \mathcal{L}^1$ alors $\lim_{n \rightarrow +\infty} \mathbb{E}(X_n) = \mathbb{E}(\lim_{n \rightarrow +\infty} X_n)$ (théorème de la convergence dominée)

Proposition

Pour un événement A , $\mathbb{E}(1_A) = \mathbb{P}(A)$.

Pour deux variables aléatoires X et Y , et un réel a , $\mathbb{E}(aX + Y) = a\mathbb{E}(X) + \mathbb{E}(Y)$.

Remarque : Cela nous permet d'en déduire par exemple que pour tout réel a , $\mathbb{E}(a) = a$ ou encore que $\mathbb{E}(X - \mathbb{E}(X)) = 0$.

Définition

Soit X une variable aléatoire sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans \mathbb{R} .

On dit que X admet un **moment centré** d'ordre $n \in \mathbb{N}^*$ si $X - \mathbb{E}(X) \in L^n(\Omega, \mathcal{F}, \mathbb{P})$. Dans ce cas, on note :

$$\mu_n = \int_{\Omega} (X - \mathbb{E}(X))^n d\mathbb{P}$$

Le moment centré d'ordre 2 est appelé **variance** de la variable aléatoire et noté $\text{Var}(X)$.

Remarque : $\mu_2 = m_2 - m_1^2$ c'est-à-dire $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 > 0$.

Proposition

Pour deux variables aléatoires X et Y , et un réel a , $\text{Var}(aX) = a^2 \text{Var}(X)$ et $\text{Var}(X + a) = \text{Var}(X)$.

Définition

Soit X une variable aléatoire sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans \mathbb{R} admettant un moment d'ordre 2.

On appelle **écart-type** le réel positif $\sigma(X) = \sqrt{\text{Var}(X)}$.

Remarque : On a la propriété $\sigma(aX) = a\sigma(X)$. Attention cependant, contrairement à l'espérance on a généralement pas $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ ou $\sigma(X + Y) = \sigma(X) + \sigma(Y)$.

Théorème (Inégalité de Chebyshev)

Soit X une variable aléatoire sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans \mathbb{R} admettant un moment d'ordre 2. Alors :

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a\sigma) \leq \frac{1}{a^2}$$

Démonstration : On utilise l'inégalité de Markov :

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a\sigma) = \mathbb{P}(|X - \mathbb{E}(X)|^2 \geq a^2 \text{Var}(X)) \leq \frac{\mathbb{E}(|X - \mathbb{E}(X)|^2)}{a^2 \text{Var}(X)} = \frac{1}{a^2}$$

Remarque : Ceci implique que $\text{Var}(X) = 0 \Leftrightarrow X = \mathbb{E}(X)$ presque partout.

Définition

Le moment d'ordre 3 donne une indication sur la symétrie. On utilise souvent le **coefficient d'asymétrie** $\frac{\mu_3}{\mu_2^{3/2}}$.
Le moment d'ordre 4 donne une indication sur les queues de distribution. On utilise souvent le **kurtosis** $\frac{\mu_4}{\mu_2^2}$ (et l'excès de kurtosis : $\frac{\mu_4}{\mu_2^2} - 3$).

Section VI.5 - Fonction de répartition

Définition

Considérons \mathbb{R} muni d'une tribu contenant la tribu de Borel, et muni d'une mesure de probabilité \mathbb{P} .
On appelle fonction de **répartition** l'application $F : \mathbb{R} \rightarrow [0, 1]$ définie par $F(x) = \mathbb{P}(] - \infty, x])$.

Exemple : Pour la modélisation du lancé d'un dé, $\mathbb{P} = \sum_{i=1}^6 \frac{1}{6} \delta_i$. La fonction de répartition est alors $f(x) = \sum_{i=1}^6 1_{[i, +\infty[}(x)$.

Proposition

Soit F une fonction de répartition. Alors, F est croissante et continue à droite, et vérifie $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$.

Démonstration : $x < y \Rightarrow] - \infty, x] \subset] - \infty, y] \Rightarrow \mathbb{P}(] - \infty, x]) \leq \mathbb{P}(] - \infty, y]) \Rightarrow F(x) \leq F(y)$.

Soit (x_n) une suite décroissante convergeant vers x . On a $] - \infty, x_{n+1}] \subset] - \infty, x_n]$ et $\bigcap_{n \in \mathbb{N}}] - \infty, x_n] =] - \infty, x]$. Donc $\lim_{n \rightarrow +\infty} F(x_n) = F(x)$, ce qui montre la continuité à droite. Les limites en $-\infty$ et $+\infty$ se démontrent de façon analogue.

Théorème

Soit F une fonction de \mathbb{R} dans \mathbb{R} croissante, continue à droite et vérifiant $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$.
Alors il existe une mesure de probabilité dont elle est la fonction de répartition.

Définition

On appelle π -système sur Ω toute collection \mathcal{J} de parties de Ω stable par intersection finie.

Exemple : L'ensemble $\{] - \infty, x]; x \in \mathbb{R}\}$ est un π -système.

Proposition (Lemme de classe monotone)

Deux mesures de probabilité qui coïncident sur un π -système \mathcal{J} coïncident également sur $\sigma(\mathcal{J})$, la tribu engendrée par \mathcal{J} .

Théorème

Considérons \mathbb{R} muni de la tribu de Borel. Soit \mathbb{P}_1 et \mathbb{P}_2 deux mesures, F_1 et F_2 leurs fonctions de répartition respectives. Alors :

$$F_1 = F_2 \Leftrightarrow \mathbb{P}_1 = \mathbb{P}_2$$

Démonstration : Le sens \Leftarrow est immédiat. Pour le sens \Rightarrow , on suppose que $\forall x \in \mathbb{R}, \mathbb{P}_1(] - \infty, x]) = \mathbb{P}_2(] - \infty, x])$. On a $\sigma(\{] - \infty, x]; x \in \mathbb{R}\}) \subset \mathcal{B}(\mathbb{R})$ car les fermés sont dans $\mathcal{B}(\mathbb{R})$, et $\mathcal{B}(\mathbb{R}) \subset \sigma(\{] - \infty, x]; x \in \mathbb{R}\})$ car les $]a, b[$ sont une

base de la topologie de \mathbb{R} et $]a, b[= (\cup_{n \in \mathbb{N}^*}] - \infty, b - \frac{1}{n}]) \cap] - \infty, a]$. Ainsi $\sigma(\{] - \infty, x]; x \in \mathbb{R} \}) = \mathcal{B}(\mathbb{R})$ ce qui conclut que $\mathbb{P}_1 = \mathbb{P}_2$ par le lemme de classe monotone.

Proposition

Considérons \mathbb{R} muni de la tribu de Borel, \mathbb{P} une mesure et F sa fonction de répartition. Alors :

$$\forall x \in \mathbb{R}, \mathbb{P}(\{x\}) = F(x) - \lim_{x^-} F$$

Démonstration : $\mathbb{P}(] - \infty, x[) = \lim_{x^-} F$ et $] - \infty, x] =] - \infty, x[\cup \{x\}$ donc $\mathbb{P}(] - \infty, x]) = \mathbb{P}(] - \infty, x[) + \mathbb{P}(\{x\})$. Ainsi $\mathbb{P}(\{x\}) = F(x) - \lim_{x^-} F$.

Proposition

La fonction de répartition est continue si et seulement si la mesure de probabilité associée est diffuse (i.e. sans atomes)

Démonstration : Il s'agit d'un corollaire de la proposition précédente.

Proposition

Si \mathbb{P} est une mesure de probabilité de densité f alors sa fonction de répartition est :

$$F : x \mapsto \int_{]-\infty, x]} f d\lambda$$

Section VI.6 - Quelques lois remarquables

Définition

Soit $n \in \mathbb{N}^*$. On considère $E = \{e_1, \dots, e_n\}$.

Une variable aléatoire X suit **la loi uniforme discrète** signifie :

$$P_X = \sum_{i=1}^n \frac{1}{n} \delta_{e_i}$$

Cette loi permet de modéliser des situations où il y a un nombre fini de résultats équiprobables.

Définition

Soit $p \in]0, 1[$. On considère $E = \{e_1, e_2\}$.

X suit une **loi de Bernoulli** de paramètre p signifie

$$P_X = p\delta_{e_1} + (1-p)\delta_{e_2}$$

Cette loi permet de modéliser des expériences aléatoires dont l'issue est le succès ou l'échec.

Définition

Soit $n \in \mathbb{N}^*, p \in]0, 1[$. On considère $E = \llbracket 0, n \rrbracket$.

X suit une **loi binomiale** de paramètres n, p signifie

$$P_X = \sum_{k=0}^n p^k (1-p)^{n-k} \delta_k$$

Cette loi permet de modéliser le nombre de succès lors de la répétition de n expériences aléatoires identiques et indépendantes dont la probabilité de succès est p . On note $X \sim B(n, p)$.

Proposition

Soit $n \in \mathbb{N}^*$, $p \in]0, 1[$ et $X \sim B(n, p)$.

Alors $\mathbb{E}(X) = np$, $\text{Var}(X) = np(1-p)$ et le coefficient d'asymétrie de X vaut $\frac{1-2p}{\sqrt{np(1-p)}}$.

Définition

Soit $\lambda \in]0, +\infty[$. On considère $E = \mathbb{N}$.

X suit une **loi de Poisson** de paramètre λ signifie

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

Cette loi permet de modéliser le nombre de fois où un évènement se produit dans un intervalle, lorsque l'on sait que le nombre moyen d'occurrences est habituellement de λ dans cet intervalle. On note $X \sim \text{Pois}(\lambda)$.

Proposition

Soit $\lambda \in]0, +\infty[$ et $X \sim \text{Pois}(\lambda)$.

Alors $\mathbb{E}(X) = \lambda$, $\text{Var}(X) = \lambda$ et le coefficient d'asymétrie de X vaut $\frac{1}{\sqrt{\lambda}}$.

Définition

Soit $p \in]0, 1[$ et $E = \mathbb{N}^*$.

X suit une **loi géométrique** de paramètre p signifie

$$P_X = \sum_{k=1}^{+\infty} p^k (1-p) \delta_k$$

Cette loi est utile pour modéliser le nombre de succès consécutifs avant un échec lorsque l'on répète des expériences identiques et indépendantes de probabilité de succès p . On note $X \sim G(p)$.

Proposition

Soit $p \in]0, 1[$ et $X \sim G(p)$.

Alors $\mathbb{E}(X) = \frac{p}{1-p}$, $\text{Var}(X) = \frac{p}{(1-p)^2}$ et le coefficient d'asymétrie de X vaut $\frac{1+p}{\sqrt{p}}$.

Définition

Soit $a, b \in \mathbb{R}$ tels que $a < b$. On considère $E = \mathbb{R}$.

X suit une **loi uniforme continue** de paramètres a et b signifie P_X a pour densité

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

Cette loi est utile pour modéliser le nombre de succès consécutifs avant un échec lorsque l'on répète des expériences identiques et indépendantes de probabilité de succès p . On note $X \sim \mathcal{U}(a, b)$.

Proposition

Soit $a, b \in \mathbb{R}$ tels que $a < b$ et $X \sim \mathcal{U}(a, b)$.

Alors $\mathbb{E}(X) = \frac{a+b}{2}$, $\text{Var}(X) = \frac{(b-a)^2}{12}$ et le coefficient d'asymétrie de X vaut 0.

Définition

Soit $\lambda \in]0, +\infty[$. On considère $E = \mathbb{R}^+$.

X suit une **loi exponentielle** de paramètre λ signifie P_X a pour densité

$$f_X(x) = \lambda e^{-\lambda x}$$

Cette loi permet de modéliser la durée entre les occurrences d'un évènement.

Proposition

Soit $\lambda \in]0, +\infty[$ et X qui suit une loi exponentielle.

Alors $\mathbb{E}(X) = \frac{1}{\lambda}$, $\text{Var}(X) = \frac{1}{\lambda^2}$ et le coefficient d'asymétrie de X vaut 2.

Définition

Soit $p \in]0, +\infty[$ et $\lambda \in]0, +\infty[$. On considère $E = \mathbb{R}^+$.

X suit une **loi Gamma** de paramètres p et λ signifie P_X a pour densité

$$f_X(x) = \frac{\lambda}{\Gamma(p)} (\lambda x)^{p-1} e^{-\lambda x}$$

où $\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt$. Lorsque $p = 1$, on retrouve une loi exponentielle. On note parfois $\theta = \frac{1}{\lambda}$. On note $X \sim \gamma(p, \lambda)$.

Proposition

Soit $p, \lambda \in]0, +\infty[$ et $X \sim \gamma(p, \lambda)$.

Alors $\mathbb{E}(X) = \frac{p}{\lambda}$, $\text{Var}(X) = \frac{p}{\lambda^2}$ et le coefficient d'asymétrie de X vaut $\frac{2}{\sqrt{p}}$.

Définition

Soit $m \in \mathbb{R}$ et $\sigma \in]0, +\infty[$. On considère $E = \mathbb{R}$.

X suit une **loi normale** de paramètres m et σ^2 signifie P_X a pour densité

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

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

Proposition

Soit $m \in \mathbb{R}, \sigma \in]0, +\infty[$ et $X \sim \mathcal{N}(m, \sigma^2)$.

Alors $\mathbb{E}(X) = m$, $\text{Var}(X) = \sigma^2$ et le coefficient d'asymétrie de X vaut 0.

Chapitre VII. Mesure produit, Convolution

Section VII.1 - Espace produit

Définition

Soit (E, \mathcal{E}) et (F, \mathcal{F}) deux espaces mesurables.
On appelle **tribu produit** sur $E \times F$ la tribu $\sigma(\mathcal{E} \times \mathcal{F})$. On la note $\mathcal{E} \otimes \mathcal{F}$.

Exemple : Si $E = F = \mathbb{R}, \mathcal{E} = \mathcal{F} = \mathcal{B}(\mathbb{R})$, alors $\mathcal{E} \times \mathcal{F}$ n'est pas directement une tribu (la réunion de deux rectangles n'est pas un rectangle). $\mathcal{E} \otimes \mathcal{F}$ est la plus petite tribu contenant $\mathcal{E} \times \mathcal{F}$; on part de $\mathcal{E} \times \mathcal{F}$, et on espère s'arrêter avant $\mathcal{P}(\mathcal{E} \times \mathcal{F})$.

Remarque : En général, \otimes n'est pas commutatif.

Proposition

Soit $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$ des espaces mesurables.
Pour tout $k \in \llbracket 1, n \rrbracket$, la projection canonique $\pi_k : \prod_{i=1}^n E_i \rightarrow E_k$ définie par $\pi_k = (x_1, \dots, x_n) \mapsto x_k$ est mesurable.
La tribu produit est la plus petite tribu rendant mesurable les n projections canoniques.

Démonstration : Soit $A \in \mathcal{E}_k$. Alors $\pi_k^{-1}(A) = (\prod_{i=1}^{k-1} E_i) \times A \times (\prod_{i=k+1}^n E_i) \in \prod_{i=1}^n \mathcal{E}_i \subset \otimes_{i=1}^n \mathcal{E}_i$, donc π_k est mesurable.

Supposons π_1, \dots, π_n mesurables et $A = \prod_{i=1}^n A_i$ où $A_i \in \mathcal{E}_i$. Alors $\forall k \in \llbracket 1, n \rrbracket, \pi_k^{-1}(A_k) = (\prod_{i=1}^{k-1} E_i) \times A_k \times (\prod_{i=k+1}^n E_i) \in \otimes_{i=1}^n \mathcal{E}_i$. Donc $A = \cap_{k=1}^n (\pi_k^{-1}(A_k)) \in \otimes_{i=1}^n \mathcal{E}_i$. Ainsi $\otimes_{i=1}^n \mathcal{E}_i$ contient $\sigma(\prod_{i=1}^n \mathcal{E}_i)$. La plus petite tribu rendant les π_k mesurables est $\sigma(\prod_{i=1}^n \mathcal{E}_i)$.

Proposition

On a $\mathcal{B}(\mathbb{R}^n) = (\mathcal{B}(\mathbb{R}))^{\otimes n} = \mathcal{B}(\mathbb{R}) \otimes \dots \otimes \mathcal{B}(\mathbb{R})$.

Démonstration : Soit l'ensemble \mathcal{R} des produits cartésiens mesurables de $(\mathcal{B}(\mathbb{R}))^{\otimes n}$. Soit $A \in \mathcal{B}(\mathbb{R}^n)$, alors $A = \prod_{i=1}^n A_i$ avec $A_i \in \mathcal{B}(\mathbb{R})$. Pour tout $k \in \llbracket 1, n \rrbracket, \pi_k^{-1}(A_k) \in \mathcal{B}(\mathbb{R}^n)$. Or $\cap_{i=1}^n \pi_i^{-1}(A_i) = A$ donc $A \in \mathcal{B}(\mathbb{R}^n)$. Donc $\mathcal{R} \subset \mathcal{B}(\mathbb{R}^n)$, donc $\sigma(\mathcal{R}) \subset \mathcal{B}(\mathbb{R}^n)$. Or $\sigma(\mathcal{R}) = (\mathcal{B}(\mathbb{R}))^{\otimes n}$ d'où $(\mathcal{B}(\mathbb{R}))^{\otimes n} \subset \mathcal{B}(\mathbb{R}^n)$.

On considère désormais \mathcal{C} l'ensemble des pavés de \mathbb{R}^n :

$$\mathcal{C} = \left\{ \prod_{i=1}^n]a_i, b_i[; \forall i \in \llbracket 1, n \rrbracket, a_i \in \mathbb{R}, b_i \in \mathbb{R} \text{ et } a_i < b_i \right\}$$

et l'ensemble \mathbb{R} des produits cartésiens mesurables de $(\mathcal{B}(\mathbb{R}))^{\otimes n}$. $\mathcal{C} \subset \mathbb{R}$ donc $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{C}) \subset \sigma(\mathcal{R}) = (\mathcal{B}(\mathbb{R}))^{\otimes n}$, car les ouverts s'expriment comme réunion dénombrable de pavés. D'où le résultat.

Définition

Soit E et F deux ensembles et $A \subset E \times F$.
Pour $e \in E$, on appelle la **x-section** de A l'ensemble

$$A_e = \{y \in F ; (e, y) \in A\}$$

Pour $f \in F$, on appelle la **y-section** de A l'ensemble

$$A^f = \{x \in E ; (x, f) \in A\}$$

Proposition

Soit E et F deux ensembles.
Pour tout $A \subset E \times F, (E \times F \setminus A)_e = F \setminus A_e$.
Pour toute famille $(A_i)_{i \in I}$ d'ensembles inclus dans $E \times F, (\cup_{i \in I} A_i)_e = \cup_{i \in I} (A_i)_e$ et $(\cap_{i \in I} A_i)_e = \cap_{i \in I} (A_i)_e$.
Les propriétés sont analogues pour les y -sections.

Proposition

Soit (E, \mathcal{E}) et (F, \mathcal{F}) deux espaces mesurables, et $C \in \mathcal{E} \otimes \mathcal{F}$.
Alors, $\forall e \in E, C_e \in \mathcal{F}$ et $\forall f \in F, C^f \in \mathcal{E}$.

Démonstration : Soit $\mathcal{C}(e) = \{C \in \mathcal{E} \otimes \mathcal{F}; C_e \in \mathcal{F}\}$. C'est une tribu. Soit $C = A \times B$ où $A \in \mathcal{E}$ et $B \in \mathcal{F}$. Si $e \in A$ alors $C_e = B$, sinon $C = \emptyset$. Dans les deux cas, $C_e \in \mathcal{F}$. Donc $C \in \mathcal{C}(e)$: cette tribu contient les $A \times B$ où $A \in \mathcal{E}$ et $B \in \mathcal{F}$. Ainsi $\mathcal{E} \otimes \mathcal{F} \subset \mathcal{C}(e)$, d'où $\mathcal{E} \otimes \mathcal{F} = \mathcal{C}(e)$.

Proposition

Soit (E, \mathcal{E}) , (F, \mathcal{F}) et (G, \mathcal{G}) des espaces mesurables.
Soit $e \in E, f \in F$ et $\phi : E \times F \rightarrow G$ mesurable.
Alors les applications partielles $\phi_{y=f} : E \rightarrow G$ telle que $\phi_{y=f} : x \mapsto \phi(x, f)$ et $\phi_{x=e} : F \rightarrow G$ telle que $\phi_{x=e} : y \mapsto \phi(e, y)$ sont mesurables.

Démonstration : Soit $C \in \mathcal{G}$. Alors $\phi_{y=f}^{-1}(C) = \{x \in E; \phi(x, f) \in C\} = \{(x, f) \in E \times F; (x, f) \in \phi^{-1}(C)\} = \phi^{-1}(C)^f$, ce qui conclut puisque $\phi^{-1}(C)$ est mesurable, donc sa y-section aussi.

Proposition (Lemme de classe monotone, généralisation)

Soit μ et ν deux mesures finies sur un espace mesurable (E, \mathcal{E}) .
Soit \mathcal{J} un π -système sur E .
Si μ et ν coïncident sur \mathcal{J} alors elles coïncident sur $\sigma(\mathcal{J})$. De plus, s'il existe dans \mathcal{J} une suite $(A_n)_{n \in \mathbb{N}}$ telle que $\forall n \in \mathbb{N}, \mu(A_n) < +\infty$ et $\cup_{n \in \mathbb{N}} A_n = E$, alors le résultat persiste même lorsque $\mu(E) = +\infty$.

Définition

On dit qu'une collection \mathcal{J} de parties de E est un λ -système ssi :

1. $E \in \mathcal{J}$
2. $A \in \mathcal{J} \Rightarrow E \setminus A \in \mathcal{J}$
3. Pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments disjoints de \mathcal{J} , $\cup_{n \in \mathbb{N}} A_n \in \mathcal{J}$.

Théorème (Dynkin)

Tout λ -système qui contient un π -système contient également la tribu engendrée par ce π -système.

Théorème

Soit (E, \mathcal{E}, μ) et (F, \mathcal{F}, ν) . On suppose que μ et ν sont σ -finies. Alors :

- Il existe une unique mesure m sur $(E \times F, \mathcal{E} \otimes \mathcal{F})$ telle que

$$\forall A \in \mathcal{E}, \forall B \in \mathcal{F}, m(A \times B) = \mu(A)\nu(B)$$

- m est σ -finie, on l'appelle mesure produit de μ et ν et on note

$$m = \mu \otimes \nu$$

- Pour tout $C \in \mathcal{E} \otimes \mathcal{F}$:

$$(\mu \otimes \nu)(C) = \int_E \nu(C_x) \mu(dx) = \int_F \mu(C^y) \nu(dy)$$

Démonstration : On se place dans le cas où μ et ν sont finies. On définit la fonction m de $\mathcal{E} \otimes \mathcal{F}$ dans $[0, +\infty]$ par :

$$\forall C \in \mathcal{E} \otimes \mathcal{F}, m(C) = \int_E \nu(C_x) \mu(dx)$$

$\nu(C_x)$ est bien défini puisque C_x est la x -section d'un ensemble mesurable. On pose

$$\mathcal{G} = \{C \in \mathcal{E} \otimes \mathcal{F}; h_C \text{ est borélienne}\}$$

\mathcal{G} contient tous les produits cartésiens $A \times B$ où $A \in \mathcal{E}$ et $B \in \mathcal{F}$. En effet, soit $A \in \mathcal{E}$ et $B \in \mathcal{F}$. $(A \times B)_x = B$ si $x \in A$, $(A \times B)_x = \emptyset$ sinon. Ainsi $\nu((A \times B)_x) = 1_A(x)\nu(B)$, donc $h_{A \times B}$ est borélienne (A est mesurable). Par

ailleurs, c'est un λ -système ; $\emptyset \in \mathcal{G}$ car h_\emptyset est la fonction nulle, donc mesurable. Soit $C \in \mathcal{G}$. Alors $h_{(E \times F) \setminus C} = \nu((E \times F) \setminus C)_x = \nu(F \setminus C_x) = \nu(F) - \nu(C_x)$. Ainsi $h_{(E \times F) \setminus C} = \nu(F) - h_C$ est borélienne, et $(E \times F) \setminus C \in \mathcal{G}$. Si C_1 et C_2 sont deux ensembles disjoints de \mathcal{G} alors $h_{C_1 \cup C_2} = h_{C_1} + h_{C_2}$ est borélienne donc $C_1 \cup C_2 \in \mathcal{G}$. \mathcal{G} est stable par union disjointe finie. Soit $(C_n)_{n \in \mathbb{N}}$ une suite d'ensembles disjoints de \mathcal{G} . On pose $Y_N = \bigcup_{n=0}^N C_n$ et $Z = \bigcup_{n=0}^{\infty} C_n$. h_{Y_N} est borélienne et croissante. Elle converge vers h_Z , qui est donc borélienne. Ainsi $Z \in \mathcal{G}$. En conséquence, \mathcal{G} est un λ -système, qui contient le π -système de l'ensemble des produits cartésiens $A \times B$ où $A \in \mathcal{E}$ et $B \in \mathcal{F}$. D'après le théorème de Dynkin, \mathcal{G} contient la tribu engendrée par ce π -système, donc contient $\mathcal{E} \otimes \mathcal{F}$, soit $\mathcal{G} = \mathcal{E} \otimes \mathcal{F}$. Ainsi $h_C : x \mapsto \nu(C_x)$ est bien borélienne, et m est bien définie, et

$$\begin{aligned} m(A \times B) &= \int_E \nu((A \times B)_x) \mu(dx) = \int_E 1_A(x) \nu(B) \mu(dx) \\ &= \nu(B) \int_E 1_A(x) \mu(dx) = \mu(A) \nu(B) \end{aligned}$$

Vérifions maintenant que m est une mesure. On a $m(\emptyset) = \int_E \nu(\emptyset_x) \mu(dx) = 0$, et pour toute suite $(C_n)_{n \in \mathbb{N}}$ d'éléments deux-à-deux disjoints de $\mathcal{E} \otimes \mathcal{F}$:

$$\begin{aligned} m(\bigcup_{n \in \mathbb{N}} C_n) &= \int_E \nu((\bigcup_{n \in \mathbb{N}} C_n)_x) \mu(dx) = \int_E \nu(\bigcup_{n \in \mathbb{N}} (C_n)_x) \mu(dx) \\ &= \int_E \left(\sum_{n=0}^{+\infty} \nu((C_n)_x) \right) \mu(dx) = \sum_{n=0}^{+\infty} \int_E \nu((C_n)_x) \mu(dx) = \sum_{n=0}^{+\infty} m(C_n) \end{aligned}$$

m est donc bien une mesure. Celle-ci est par ailleurs unique, car si m et m' sont deux mesures telles que $\forall A \in \mathcal{E}, \forall B \in \mathcal{F}, m(A \times B) = \mu(A) \nu(B) = m'(A \times B)$, alors m et m' coïncident sur un π -système et donc d'après le lemme de classe monotone, m et m' coïncident sur la tribu engendrée par les produits cartésiens d'ensembles de \mathcal{E} et \mathcal{F} , c'est-à-dire $\mathcal{E} \otimes \mathcal{F}$; donc $m = m'$.

Exemple : Considérons la mesure de Lebesgue λ sur \mathbb{R} muni de la tribu de Lebesgue. Soit a_1, a_2, b_1, b_2 quatre réels avec $a_1 < b_1$ et $a_2 < b_2$. Alors $(\lambda \otimes \lambda)(]a_1, b_1[\times]a_2, b_2[) = \lambda(]a_1, b_1[) \lambda(]a_2, b_2[) = (b_1 - a_1)(b_2 - a_2)$. On a donc bien généralisé le fait que l'aire d'un rectangle est le produit du longeur par la largeur. De façon analogue, $\lambda^{(n)} = \lambda^{\otimes n}$.

Section VII.2 - Intégrales multiples

Théorème (Fubini-Tonelli)

Soit (E, \mathcal{E}, μ) et (F, \mathcal{F}, ν) deux espaces mesurés, avec μ et ν σ -finies. Soit $f : E \times F \rightarrow [0, +\infty]$ mesurable. Alors :

- $x \mapsto \int_F f(x, y) \nu(dy)$ est μ -mesurable
- $y \mapsto \int_E f(x, y) \mu(dx)$ est ν -mesurable
- $\int_E \left(\int_F f(x, y) \nu(dy) \right) \mu(dx) = \int_F \left(\int_E f(x, y) \mu(dx) \right) \nu(dy)$
 $= \int_{E \times F} f(x, y) (\mu \otimes \nu)(dx, dy)$

Démonstration : Soit $C \in \mathcal{E} \otimes \mathcal{F}$. Pour $f = 1_C$, on a $x \mapsto \int_F 1_C(x, y) \nu(dy) = \int_F 1_{C_x}(y) \nu(dy) = \nu(C_x)$ qui est μ -mesurable, et $y \mapsto \int_E 1_C(x, y) \mu(dx) = \mu(C^y)$ est ν -mesurable. Par linéarité, on obtient la mesurabilité pour toute fonction étagée positive, puis par limite croissante, pour tout f positive.

Soit $C \in \mathcal{E} \otimes \mathcal{F}$. Pour $f = 1_C$, l'égalité demandée est $(\mathcal{E} \otimes \mathcal{F})(C) = \int_E \nu(C_x) \mu(dx) = \int_F \mu(C^y) \nu(dy)$ que l'on sait vraie. On l'obtient ensuite par linéarité pour toute f étagée positive, puis, par limite croissante, pour toute fonction f positive.

Théorème (Fubini-Lebesgue)

Soit (E, \mathcal{E}, μ) et (F, \mathcal{F}, ν) deux espaces mesurés, avec μ et ν σ -finies. Soit $f \in \mathcal{L}^1(E \times F, \mathcal{E} \otimes \mathcal{F}, \mu \otimes \nu)$. Alors :

- $x \mapsto f(x, y)$ est dans $\mathcal{L}^1(E, \mathcal{E}, \mu)$ pour ν -presque tout y , $y \mapsto f(x, y)$ est dans $\mathcal{L}^1(F, \mathcal{F}, \nu)$ pour μ -presque tout x .
- $y \mapsto \int_E f(x, y) \mu(dx)$ est ν -mesurable, définie presque partout et dans $\mathcal{L}^1(F, \mathcal{F}, \nu)$, et $x \mapsto \int_F f(x, y) \nu(dy)$ est μ -mesurable, définie presque partout et dans $\mathcal{L}^1(E, \mathcal{E}, \mu)$.

$$\begin{aligned} \bullet \int_E \left(\int_F f(x, y) \nu(dy) \right) \mu(dx) &= \int_F \left(\int_E f(x, y) \mu(dx) \right) \nu(dy) \\ &= \int_{E \times F} f(x, y) (\mu \otimes \nu)(dx, dy) \end{aligned}$$

Démonstration : $|f|$ est mesurable est positive, donc d'après le théorème de Fubini-Tonelli, $\int_E (\int_F |f(x, y)| \mu(dx)) \nu(dy) = \int_{E \times F} |f(x, y)| (\mu \otimes \nu)(dx, dy) < +\infty$ par hypothèse, donc $\int_F |f(x, y)| \mu(dx) < +\infty$ presque partout. Ainsi $y \mapsto \int_E f(x, y) \mu(dx)$ est dans $\mathcal{L}^1(F, \mathcal{F}, \nu)$ presque partout. De même, on montre que $x \mapsto \int_F f(x, y) \nu(dy)$ est dans $\mathcal{L}^1(E, \mathcal{E}, \mu)$ presque partout. $x \mapsto \int_F f(x, y) \nu(dy)$ est bien définie sauf sur un ensemble négligeable. On a alors $\int_E |\int_F f(x, y) \nu(dy)| \mu(dx) \leq \int_E (\int_F |f(x, y)| \nu(dy)) \mu(dx) \leq \int_{E \times F} |f| d(\mu \otimes \nu)$. Ainsi $x \mapsto \int_F f(x, y) \nu(dy)$ est dans $\mathcal{L}^1(E, \mathcal{E}, \mu)$ et de la même manière, $y \mapsto \int_E f(x, y) \mu(dx)$ est dans $\mathcal{L}^1(F, \mathcal{F}, \nu)$. Enfin, en décomposant $f = f^+ - f^-$, et en appliquant le théorème de Fubini-Tonelli à f^+ et à f^- , on obtient le dernier résultat.

Exemples : Pour calculer $\int_{[2,3] \times [0,1]} xy \lambda^{(2)}(dx, dy)$, on peut remarquer que la mesure de Lebesgue est σ -finie et que $(x, y) \mapsto xy 1_{[2,3] \times [0,1]} \in \mathcal{L}^1(\mathbb{R}^2)$ donc le théorème de Fubini-Lebesgue s'applique.

$$\int_{[2,3] \times [0,1]} xy \lambda^{(2)}(dx, dy) = \int_{[0,1]} y (\int_{[2,3]} x \lambda(dx)) \lambda(dy) = \frac{5}{4}.$$

Pour calculer $\sum_{n \in \mathbb{N}, m \in \mathbb{N}} \frac{1}{2^n 3^m}$, on peut remarquer que la mesure de comptage sur \mathbb{N} est σ -finie sur \mathbb{N} , et donc le théorème de Fubini-Tonelli s'applique : $\sum_{n \in \mathbb{N}, m \in \mathbb{N}} \frac{1}{2^n 3^m} = \sum_{n \in \mathbb{N}} \frac{1}{2^n} \sum_{m \in \mathbb{N}} \frac{1}{3^m} = 3$.

Proposition (Changement de variable linéaire)

Soit $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ une application linéaire bijective.
Soit f une application intégrable sur \mathbb{R}^d . Alors :

$$\int_{\mathbb{R}^d} f(\phi(x)) |\det \phi| \lambda^{(d)}(dx) = \int_{\mathbb{R}^d} f(y) \lambda^{(d)}(dy)$$

et pour tout borélien A :

$$\int_A f(\phi(x)) |\det \phi| \lambda^{(d)}(dx) = \int_{\phi(A)} f(y) \lambda^{(d)}(dy)$$

Exemple : $\int_{B(0,1)} (y_1^2 + y_2^2) \lambda^{(2)}(dy_1, dy_2) = \int_{B(0, \frac{1}{2})} ((2x_1)^2 + (2x_2)^2) 4 \lambda^{(2)}(dx_1, dx_2)$.

Définition

Soit U et V deux ouverts non vides de \mathbb{R}^d , $\phi : U \rightarrow V$ un difféomorphisme \mathcal{C}^1 et $x = (x_1, \dots, x_n) \in \mathbb{R}^d$. On appelle **matrice jacobienne** de ϕ la matrice :

$$D\phi(x) = \begin{pmatrix} \frac{\partial \phi_1}{\partial x_1}(x_1, \dots, x_n) & \dots & \frac{\partial \phi_1}{\partial x_d}(x_1, \dots, x_d) \\ \vdots & \ddots & \vdots \\ \frac{\partial \phi_d}{\partial x_1}(x_1, \dots, x_d) & \dots & \frac{\partial \phi_d}{\partial x_d}(x_1, \dots, x_d) \end{pmatrix}$$

On appelle alors **Jacobien** de ϕ en x le nombre réel $J\phi(x) = \det(D\phi(x))$.

Proposition (Changement de variable linéaire)

Soit U et V deux ouverts non vides de \mathbb{R}^d et $\phi : U \rightarrow V$ un difféomorphisme \mathcal{C}^1 .

Soit f une application borélienne sur U . Alors f est intégrable sur V ssi $(f \circ \phi)|J\phi|$ est intégrable sur U . Dans ce cas :

$$\int_U f(\phi(x))|J\phi(x)|\lambda^{(d)}(dx) = \int_V f(y)\lambda^{(d)}(dy)$$

Exemple : En prenant $\phi : (r, \theta) \mapsto (r \cos \theta, r \sin \theta)$, on a $J\phi = r$ et :

$$\int_{B(0,1)} e^{x_1^2 + x_2^2} \lambda^{(2)}(dx_1, dx_2) = \int_{]0,1[\times]0,2\pi[} r e^{r^2} \lambda(dr, d\theta) = \int_{]0,2\pi[} \left(\int_{]0,1[} r \exp(r^2) \lambda(dr) \right) \lambda(d\theta) = \pi(e - 1).$$

Section VII.3 - Indépendance des variables aléatoires

Définition

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $X : \Omega \rightarrow (E, \mathcal{E})$, $Y : \Omega \rightarrow (F, \mathcal{F})$ deux variables aléatoires.

La construction de la tribu produit et de la mesure produit permet de définir une variable aléatoire $Z : \Omega \rightarrow (E \times F, \mathcal{E} \otimes \mathcal{F})$ telle que $Z(\omega) = (X(\omega), Y(\omega))$. Z sera notée (X, Y) .

La loi $P_{(X,Y)}$ de (X, Y) est la mesure définie sur $\mathcal{E} \otimes \mathcal{F}$ par $\forall C \in \mathcal{E} \otimes \mathcal{F}, P_{(X,Y)}(C) = \mathbb{P}((X, Y) \in C)$.

Remarque : Pour $C = A \times B$ avec $A \in \mathcal{E}$ et $B \in \mathcal{F}$, on a $P_{(X,Y)}(C) = \mathbb{P}((X, Y) \in A \times B) = \mathbb{P}(X \in A, Y \in B)$.

Définition

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $X : \Omega \rightarrow (E, \mathcal{E})$, $Y : \Omega \rightarrow (F, \mathcal{F})$ deux variables aléatoires.

On note P_X la loi de X , P_Y la loi de Y et $P_{(X,Y)}$ la loi jointe de (X, Y) . On dit alors que X et Y sont **indépendantes** ssi $P_{(X,Y)} = P_X \otimes P_Y$. Les lois P_X et P_Y sont appelées **lois marginales** de (X, Y) .

Proposition

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $X : \Omega \rightarrow (E, \mathcal{E})$, $Y : \Omega \rightarrow (F, \mathcal{F})$ deux variables aléatoires. X et Y sont indépendantes ssi :

$$\forall A \in \mathcal{E}, \forall B \in \mathcal{F}, \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B)$$

Démonstration : Le sens \Rightarrow découle directement de la définition de l'indépendance, et le sens \Leftarrow repose sur le fait que $P_{(X,Y)}$ et $P_X \otimes P_Y$ sont finies et coïncident sur un π -système, donc sont égales par le lemme de classe monotone.

Proposition

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, $X : \Omega \rightarrow (E, \mathcal{E})$ et $Y : \Omega \rightarrow (F, \mathcal{F})$ deux variables aléatoires.

X et Y sont indépendantes ssi pour toutes fonctions bornées mesurables $f : E \rightarrow \mathbb{R}$ et $g : F \rightarrow \mathbb{R}$, $\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$.

Démonstration : Le sens direct se démontre en remarquant qu'on a l'égalité pour les fonctions indicatrices, puis on procède comme habituellement : on étend l'égalité aux fonctions étagées positives, puis aux fonctions positives, puis à toute fonction bornée mesurable. Le sens indirect se montre en choisissant, pour $A \in \mathcal{E}$ et $B \in \mathcal{F}$, les fonctions $f = 1_A$ et $g = 1_B$.

Remarque : X et Y sont indépendantes si et seulement si pour toutes fonctions bornées mesurables $f : E \rightarrow \mathbb{R}$ et $g : F \rightarrow \mathbb{R}$, $f(X)$ et $g(Y)$ sont indépendantes.

Définition

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(X_i)_{i \in I}$ une famille de variables aléatoires.

On dit que $(X_i)_{i \in I}$ est une famille **indépendante** ssi :

$$\text{Pour tout } J \subset I \text{ fini, } P_{(X_i)_{i \in J}} = \otimes_{i \in J} P_{X_i}$$

Définition

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(\mathcal{A}_i)_{i \in I}$ une famille de sous-tribus de \mathcal{A} .
On dit que $(\mathcal{A}_i)_{i \in I}$ est une famille de sous-tribus **indépendante** ssi :

$$\forall J \subset I \text{ fini}, \forall A_i \in \mathcal{A}_i, \mathbb{P}(\cap_{i \in J} A_i) = \prod_{i \in J} \mathbb{P}(A_i)$$

Proposition

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, $X : \Omega \rightarrow (E, \mathcal{E})$ et $Y : \Omega \rightarrow (F, \mathcal{F})$ deux variables aléatoires.
 X et Y sont indépendantes ssi pour toutes fonctions bornées mesurables $f : E \rightarrow \mathbb{R}$ et $g : F \rightarrow \mathbb{R}$,
 $\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$.

Proposition

Une famille de variables aléatoires $(X_i)_{i \in J}$ est indépendante ssi les tribus $\sigma(X_i)$ le sont.

Section VII.4 - Convolution

Définition

Soit $(E, +)$ un groupe commutatif, et \mathcal{T} une topologie rendant l'application $(x, y) \mapsto x + y$ continue. On munit E de sa tribu borélienne $\mathcal{B}(\mathcal{T})$. Soit λ et μ deux mesures σ -finies sur $(E, \mathcal{B}(\mathcal{T}))$. On appelle **produit de convolution** de la mesure μ par la mesure ν la mesure $\mu * \nu$ définie par :

$$\forall A \in \mathcal{B}(\mathcal{T}), (\mu * \nu)(A) = \int_{E \times E} 1_A(x + y)(\mu \otimes \nu)(dx, dy)$$

Remarque : $\mu * \nu$ est bien définie puisque $(x, y) \mapsto x + y$ est borélienne. $\mu * \nu$ est la mesure image de $\mu \otimes \nu$ par $(x, y) \mapsto x + y$.

Proposition

Si μ et ν sont des mesures de probabilité, alors $\mu * \nu$ est une mesure de probabilité.

Démonstration : $(\mu * \nu)(E) = \int_{E \times E} 1_E(x + y)(\mu \otimes \nu)(dx, dy) = (\mu \otimes \nu)(E) = 1$.

Proposition

Si X et Y sont deux variables aléatoires indépendantes, alors $P_X * P_Y = P_{X+Y}$.

Démonstration : $P_{X+Y}(A) = \mathbb{P}((X+Y) \in A) = \int_{E \times E} 1_A(X+Y)dP_{X,Y} = \int_{E \times E} 1_A(x+y)(P_X \otimes P_Y)(dx, dy) = P_X * P_Y$.

Proposition

La mesure de Dirac en 0 est élément neutre pour la convolution.

Démonstration : $(\delta * \nu)(A) = \int_E (\int_E 1_A(x + y)\delta(dx))\nu(dy) = \int_E 1_A(y)\nu(dy) = \nu(A)$.

Proposition

Le produit de convolution est commutatif.

Démonstration : $\mu * \nu$ est la mesure image de $\mu \otimes \nu$ par $(x, y) \mapsto x + y$. L'addition étant commutative, on a donc $\mu * \nu = \nu * \mu$.

Définition

Soit f et g deux fonctions mesurables de \mathbb{R}^d dans \mathbb{R} .

Lorsque $\int_{\mathbb{R}^d} |f(x-y)g(y)|\lambda^{(d)}(dy) < +\infty$, on définit le **produit de convolution** de la fonction f par la fonction g par :

$$f * g = x \mapsto \int_{\mathbb{R}^d} f(x-y)g(y)\lambda^{(d)}(dy)$$

Remarque : Si f et g sont positives et si μ et ν sont des mesures de densité f et g par rapport à la mesure de Lebesgue, alors $\mu * \nu$ est une mesure de densité $f * g$ par rapport à la mesure de Lebesgue.

Proposition

- $f * g = g * f$
- $(f * g) * h = f * (g * h)$
- $\forall a \in \mathbb{R}, f * (g + ah) = f * g + a(f * h)$

Théorème

Soit f et g dans $L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^{(d)})$. Alors :

- $(f * g)(x)$ est définie pour presque tout $x \in \mathbb{R}$
- $f * g \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^{(d)})$
- $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$

Démonstration : On note par commodité $\lambda = \lambda^{(d)}$. D'après le théorème de Fubini-Tonelli :

$$\begin{aligned} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f(x-y)| |g(y)| \lambda(dy) \right) \lambda(dx) &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f(x-y)| |g(y)| \lambda(dx) \right) \lambda(dy) \\ \int_{\mathbb{R}^d} |g(y)| \left(\int_{\mathbb{R}^d} |f(t-y)| |g(y)| \lambda(dt) \right) \lambda(dy) &= \int_{\mathbb{R}^d} |g(y)| \left(\int_{\mathbb{R}^d} |f(x)| \lambda(dx) \right) \lambda(dy) \\ &= \int_{\mathbb{R}^d} |f(x)| \lambda(dx) \int_{\mathbb{R}^d} |g(y)| \lambda(dy) < +\infty \end{aligned}$$

Théorème

Soit p et q dans $[1, +\infty]$ conjugués, soit $f \in L^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^{(d)})$ et $g \in L^q(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^{(d)})$. Alors $f * g$ est bien définie, uniformément continue et bornée.

Démonstration : D'après Hölder :

$$\|(y \mapsto f(x-y)) \times g\|_1 \leq \|y \mapsto f(x-y)\|_p \|g\|_q = \|f\|_p \|g\|_q$$

Ainsi $f * g$ est bien définie. On ne démontrera pas ici les autres propriétés.

Théorème

Soit $f \in \mathcal{C}_C^1(\mathbb{R}^d)$ et $g \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^{(d)})$.

Alors $f * g$ est bien définie, de classe \mathcal{C}^1 et $\forall i \in \llbracket 1, d \rrbracket, \partial_i(f * g) = (\partial_i f) * g$.

Démonstration : On traite le cas $d = 1$. On a $\mathcal{C}_C^1(\mathbb{R}) \subset \mathcal{C}_C^0(\mathbb{R}) \subset L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^{(d)})$ donc $f * g$ et $f' * g$ sont bien définies. Soit $x \in \mathbb{R}$ et $\epsilon > 0$. Soit $z \in \mathbb{R}$ et $h > 0$. Alors :

$$\begin{aligned} f(z+h) - f(z) &= \int_{[z, z+h]} f'(u) \lambda(du) \\ \Rightarrow f(z+h) - f(z) - hf'(z) &= \int_{[z, z+h]} (f'(u) - f'(z)) \lambda(du) \\ &= h \int_{[0,1]} (f'(z+hv) - f'(z)) \lambda(dv) \end{aligned}$$

$$\Rightarrow \frac{f(z+h) - f(z)}{h} - f'(z) = \int_{[0,1]} (f'(z+hv) - f'(z))\lambda(dv)$$

f' est continue sur un compact, donc uniformément continue (Heine), donc il existe $\eta > 0$ tel que $|z_1 - z_2| < \eta \Rightarrow |f'(z_1) - f'(z_2)| < \frac{\epsilon}{\|g\|_1}$. Pour $h < \eta$, on a donc $|f'(z+hv) - f'(z)| < \frac{\epsilon}{\|g\|_1}$ d'où $|\frac{f(z+h) - f(z)}{h} - f'(z)| < \frac{\epsilon}{\|g\|_1}$. On a alors, en multipliant par $g(y)$ et en intégrant :

$$\begin{aligned} \int_{\mathbb{R}} \left(\frac{f(x-y+h) - f(x-y)}{h} - f'(x-y) \right) g(y) \lambda(dy) &< \epsilon \\ \Rightarrow \frac{(f * g)(x+h) - (f * g)(x)}{h} - (f' * g)(x) &< \epsilon \end{aligned}$$

D'où le résultat.

Proposition

Soit $f \in \mathcal{C}_c^k(\mathbb{R}^d)$ et $g \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^{(d)})$.
Alors $f * g$ est bien définie, de classe \mathcal{C}^k et :

$$\partial_1^{n_1} \dots \partial_d^{n_d} (f * g) = (\partial_1^{n_1} \dots \partial_d^{n_d} f) * g$$

où $n_1 + \dots + n_d \leq k$.

Démonstration : C'est un corollaire du théorème précédent.

Définition

Soit $u = (u_n)_{n \in \mathbb{Z}^d}$ et $v = (v_n)_{n \in \mathbb{Z}^d}$ deux suites.
La suite $u * v$ dont le n -ième terme vaut

$$\sum_{k \in \mathbb{Z}^d} u_{n-k} v_k$$

est le **produit de convolution** de $u = (u_n)_{n \in \mathbb{Z}^d}$ par $v = (v_n)_{n \in \mathbb{Z}^d}$.

Remarques : On prendra garde au fait que n et k sont des multi-indices : $n = (n_1, \dots, n_d)$ et $k = (k_1, \dots, k_d)$.

Si u et v sont positives et si μ et ν sont des mesures de densité u et v par rapport à la mesure de comptage alors $\mu * \nu$ est une mesure de densité $u * v$ par rapport à la mesure de comptage.

Si u et v sont absolument convergentes alors $u * v$ est bien défini.

La mesure de Dirac δ en 0 est une mesure de densité $u = (u_n)_{n \in \mathbb{Z}}$ par rapport à la mesure de comptage pour $u_0 = 1$ et $\forall n \in \mathbb{Z} \setminus \{0\}, u_n = 0$. Cette suite $(u_n)_{n \in \mathbb{Z}}$ est donc élément neutre pour la convolution des suites.

Exemple : Soit $u_n = \frac{\alpha^n}{n!} e^{-\alpha}$ et $v_n = \frac{\beta^n}{n!} e^{-\beta}$, où α et β sont des réels strictement positifs. Soit la suite $(w_n)_{n \in \mathbb{N}} = (u_n)_{n \in \mathbb{N}} * (v_n)_{n \in \mathbb{N}}$. Alors :

$$\begin{aligned} w_n &= \sum_{k=0}^n u_{n-k} v_k = \sum_{k=0}^n \frac{\alpha^{n-k}}{(n-k)!} e^{-\alpha} \frac{\beta^k}{k!} e^{-\beta} \\ &= \frac{e^{-(\alpha+\beta)}}{n!} \sum_{k=0}^n \binom{n}{k} \alpha^{n-k} \beta^k = \frac{(\alpha + \beta)^n}{n!} e^{-(\alpha+\beta)} \end{aligned}$$

On vient ici de montrer que la somme de deux variables indépendantes suivant des lois de Poisson de paramètre α et β est une loi de Poisson de paramètre $\alpha + \beta$.

Définition

On appelle **noyau de sommabilité** toute suite $(k_n)_{n \in \mathbb{N}}$ de fonctions intégrables vérifiant :

1. $\int_E k_n d\mu = 1$
2. $\sup_{n \in \mathbb{N}} \int |k_n| d\mu < +\infty$
3. Pour tout $F \subset E \setminus \{0\}$ fermé, $\lim_{n \rightarrow +\infty} \int_F k_n d\mu = 0$

Proposition

Soit $f \in L^p(E, \mathcal{E}, \mu)$ et $(k_n)_{n \in \mathbb{N}}$ un noyau de sommabilité. Alors :

$$\lim_{n \rightarrow +\infty} \|k_n * f - f\|_p = 0$$

Exemple : On considère $D_k(x) = \sum_{n=-k}^k e^{inx}$ et $F_K(x) = \frac{1}{K} \sum_{k=0}^{K-1} D_k(x)$. On rappelle que nous avons déjà vu dans le chapitre V que :

$$F_K(x) = \frac{1}{K} \left(\frac{\sin \frac{Kx}{2}}{\sin \frac{x}{2}} \right)^2 \text{ prolongé par } K \text{ en }]0, 2\pi]$$

$(F_K)_{K \in \mathbb{N}^*}$ est un noyau de sommabilité pour $E = [0, 2\pi]$. Dans le chapitre V, la démonstration effectuée pour démontrer que $\{x \mapsto e^{inx}, n \in \mathbb{Z}\}$ est une base hilbertienne de $L^2_{\mathbb{C}}([0, 2\pi], \mathcal{B}([0, 2\pi]), \frac{1}{2\pi}\lambda)$ revient fondamentalement à appliquer cette proposition.

On se place désormais dans \mathbb{R}^d avec $d \in \mathbb{N}^*$. Pour tout $n \in \mathbb{N}^*$, on définit :

$$k_n(x) = x \mapsto n^d \exp(-\pi n^2 \|x\|^2)$$

Il s'agit d'un noyau de sommabilité, qu'on appelle noyau de Gauss.

Définition

On appelle **suite régularisante** toute suite $(\rho_n)_{n \in \mathbb{N}}$ satisfaisant pour tout n :

1. $\int_E \rho_n d\mu = 1$
2. $\rho_n \geq 0$
3. $\text{Supp } \rho_n \subset B(0, \epsilon_n)$ avec $\lim_{n \rightarrow +\infty} \epsilon_n = 0$
4. $\rho_n \in C^{+\infty}(\mathbb{R}^d)$

Exemple : On pose :

$$\Psi(x) = \begin{cases} \exp\left(\frac{-1}{1-\|x\|^2}\right) & \text{si } \|x\| < 1 \\ 0 & \text{sinon} \end{cases}$$

et on note $c = \int_{\mathbb{R}^d} \Psi d\lambda$. Un exemple de suite régularisante est alors :

$$\rho_n(x) = \frac{n^d}{c} \Psi(nx) = \begin{cases} \frac{n^d}{c} \exp\left(\frac{-1}{1-n^2\|x\|^2}\right) & \text{si } \|x\| < \frac{1}{n} \\ 0 & \text{sinon} \end{cases}$$

Proposition

Soit $(\rho_n)_{n \in \mathbb{N}}$ une suite régularisante, $p \in [1, +\infty[$ et $f \in L^p(\mathbb{R}^d)$. Alors $\rho_n * f \rightarrow f$ dans L^p et $\rho_n * f \rightarrow f$ uniformément sur tout compact.

Théorème

Pour tout ouvert connexe Ω de \mathbb{R}^d et pour tout $p \in [1, +\infty[$, $\mathcal{D}(\Omega) = C_c^\infty(\Omega)$ est dense dans $L^p(\Omega, \mathcal{B}(\Omega), \lambda)$.

Chapitre VIII. Vecteurs aléatoires

Section VIII.1 - Fonctions de répartition, Copules

Définition

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, X_1, \dots, X_d des variables aléatoires définies sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On dit que $X : \Omega \mapsto \mathbb{R}^d$ telle que

$$\forall \omega \in \Omega, X(\omega) = \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_d(\omega) \end{pmatrix}$$

est un **vecteur aléatoire**.

On parle aussi de **variable aléatoire multidimensionnelle**.

Exemple : Le lancer de 2 dés peut être modélisée par un vecteur aléatoire de \mathbb{R}^2 .

Définition

Soit $X : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire ($X = (X_1, \dots, X_d)$).

La **fonction de répartition (multivariée)** de X est la fonction $F_X : \mathbb{R}^d \rightarrow [0, 1]$ définie par :

$$F_X(x_1, \dots, x_d) = \mathbb{P}(X_i \leq x_i \text{ pour } i \in \llbracket 1, d \rrbracket)$$

Proposition

Soit $X : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire. Notons $F = F_X$. Alors :

- F est croissante dans chacune de ses variables.
- F est continue à droite dans chacune de ses variables.
- Pour tout $i \in \llbracket 1, d \rrbracket$, $\lim_{x_i \rightarrow -\infty} F(x_1, \dots, x_d) = 0$.
- $\lim_{(x_1, \dots, x_d) \rightarrow (+\infty, \dots, +\infty)} F(x_1, \dots, x_d) = 1$.

Proposition

Soit $X = (X_1, \dots, X_d) : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire dont la fonction de répartition est $F : \mathbb{R}^d \rightarrow \mathbb{R}$, alors :

$$\mathbb{P}(X \in [x_1^i, x_2^i], i \in \llbracket 1, d \rrbracket) = \sum_{(i_1, \dots, i_d) \in \{1, 2\}^d} (-1)^{(\sum_{j=1}^d i_j)} F(x_{i_1}^1, \dots, x_{i_d}^d)$$

Démonstration : Par récurrence sur d .

Proposition

Soit $X : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire. Sa loi $P_X : \mathbb{R}^d \rightarrow \mathbb{R}$ est caractérisée par sa fonction de répartition $F_X : \mathbb{R}^d \rightarrow \mathbb{R}$.

Démonstration : P_X caractérise F_X par construction de F_X , et F_X caractérise P_X par coïncidence sur le π -système des pavés.

Définition

Soit $X = (X_1, \dots, X_d) : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire et F_X sa fonction de répartition.

On appelle **lois marginales** de X les lois :

- des X_i prises séparément

$$F_{X_i}(x_i) = \mathbb{P}(X_i \leq x_i) = \lim_{(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d \rightarrow (+\infty, \dots, +\infty))} F_X(x_1, \dots, x_d)$$

- ou de plusieurs composantes X_{i_1}, \dots, X_{i_k} du vecteur aléatoire X

$$\begin{aligned} F_{X_{i_1}, \dots, X_{i_k}}(x_{i_1}, \dots, x_{i_k}) &= \mathbb{P}(X_{i_1} \leq x_{i_1}, \dots, X_{i_k} \leq x_{i_k}) \\ &= \lim_{x_{i_j} \rightarrow +\infty \text{ pour } j \notin \llbracket 1, k \rrbracket} F_X(x_1, \dots, x_d) \end{aligned}$$

Proposition

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $X_i : \Omega \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ d variables aléatoires.

Alors, les X_i sont indépendantes ssi :

$$\forall (x_1, \dots, x_d) \in \mathbb{R}^d, F_X(x_1, \dots, x_d) = F_{X_1}(x_1) \times \dots \times F_{X_d}(x_d)$$

Si les X_i admettent une densité f_{X_i} , alors elles sont indépendantes ssi :

$$\forall (x_1, \dots, x_d) \in \mathbb{R}^d, f_X(x_1, \dots, x_d) = f_{X_1}(x_1) \times \dots \times f_{X_d}(x_d)$$

Définition

On appelle **copule** de dimension 2 toute fonction $C : \mathbb{R}^2 \rightarrow [0, 1]$ tq :

- $C(x, y) = 0$ si $x \leq 0$ ou $y \leq 0$.
- $C(x, y) = x$ si $y \geq 1$.
- $C(x, y) = y$ si $x \geq 1$.
- $C(x, y) = 1$ si $x \geq 1$ et $y \geq 1$.
- $0 \leq a \leq b \leq 1$ et $0 \leq c \leq d \leq 1$ entraîne :

$$C(b, d) - C(b, c) - C(a, d) + C(a, c) \geq 0$$

Remarque : Il suffit de définir C sur $[0, 1]^2$.

Exemples : $C(x, y) = xy$ pour $(x, y) \in [0, 1]^2$ est une copule. On l'appelle la **copule d'indépendance**.

$C(x, y) = \min(x, y)$ pour $(x, y) \in [0, 1]^2$ est une copule. On l'appelle la **copule de comonotonie**.

$C(x, y) = e^{-((- \ln x)^\theta + (- \ln y)^\theta)^{\frac{1}{\theta}}}$ pour $(x, y) \in [0, 1]^2$ est une copule. On l'appelle la **copule de Gumbel** de paramètre $\theta \in [1, +\infty[$.

Théorème (Sklar)

- Soit $Z = (X, Y) : \Omega \rightarrow \mathbb{R}^2$ un vecteur aléatoire. On note F_Z la fonction de répartition (bi-variée) de Z , F_X et F_Y les fonctions de répartition de X et Y .

Alors, il existe une copule C de dimension 2 telle que $F_Z(x, y) = C(F_X(x), F_Y(y))$. Elle est unique si F_X et F_Y sont continues.

- Soit $X : \Omega \rightarrow \mathbb{R}$ et $Y : \Omega \rightarrow \mathbb{R}$ deux variables aléatoires, de fonctions de répartition F_X et F_Y . Soit C une copule de dimension 2.

Alors, on peut construire une variable aléatoire $Z : \Omega \rightarrow \mathbb{R}^2$ sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ dont la fonction de répartition est $F_Z(x, y) = C(F_X(x), F_Y(y))$.

Définition

On appelle **copule** de dimension d toute fonction $C : \mathbb{R}^d \rightarrow [0, 1]$ tq :

- $C(x_1, \dots, x_d) = 0$ si l'un des x_i est nul.
- $C(x_1, \dots, x_d) = x_j$ si $\forall i \in \llbracket 1, d \rrbracket \setminus \{j\}, x_i = 1$.
- $C(x_1, \dots, x_d) = 1$ si $\forall i \in \llbracket 1, d \rrbracket, x_i = 1$.
- $\forall i \in \llbracket 1, d \rrbracket, 0 \leq x_1^i \leq x_2^i \leq 1$ entraîne :

$$\sum_{(i_1, \dots, i_d) \in \{1, 2\}^d} (-1)^{(\sum_{j=1}^d i_j)} C(x_{i_1}^1, \dots, x_{i_d}^d) \geq 0$$

Remarque : Le théorème de Sklar se généralise aux copules de dimension d .

Section VIII.2 - Moments, Covariance

Définition

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et $X : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire tel que $\forall i \in \llbracket 1, d \rrbracket, X_i \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.
On appelle **espérance** de X le vecteur

$$\mathbb{E}(X) = \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_d) \end{pmatrix}$$

Définition

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, $X : \Omega \rightarrow \mathbb{R}$ et $Y : \Omega \rightarrow \mathbb{R}$ deux variables aléatoires dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$.
On appelle **covariance** de X et Y le réel

$$\text{Cov}(x, y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$$

Proposition

Soit $X : \Omega \rightarrow \mathbb{R}$ et $Y : \Omega \rightarrow \mathbb{R}$ deux variables aléatoires dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

- $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
- $\text{Cov}(X, X) = \text{Var}(X)$.
- $\text{Cov} : L^2(\Omega, \mathcal{F}, \mathbb{P}) \times L^2(\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ est bilinéaire.
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$

Proposition (inégalité de Cauchy-Schwarz)

Soit X et Y deux variables aléatoires réelles dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Alors :

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}$$

Démonstration : $|\text{Cov}(X, Y)| = |\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))| = |\langle X - \mathbb{E}(X), Y - \mathbb{E}(Y) \rangle_{L^2(\Omega, \mathcal{F}, \mathbb{P})}| \leq \|X - \mathbb{E}(X)\|_{L^2(\Omega, \mathcal{F}, \mathbb{P})} \|Y - \mathbb{E}(Y)\|_{L^2(\Omega, \mathcal{F}, \mathbb{P})} = \sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}$ par l'inégalité de Cauchy-Schwarz classique.

Définition

Soit X et Y deux variables aléatoires réelles dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$ de variance non nulle.
On appelle **coefficient de corrélation linéaire** le réel de $[-1, 1]$

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Proposition

Soit X et Y deux variables aléatoires réelles dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$ de variance non nulle. Alors :

$$\exists(a, b) \in \mathbb{R}^2, Y = aX + b \Leftrightarrow |\rho_{X,Y}| = 1$$

Proposition

Soit X et Y deux variables aléatoires réelles dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Si X et Y sont indépendantes, alors $\text{Cov}(X, Y) = 0$. Si de plus elles sont de variance non nulle, alors $\rho_{X,Y} = 0$.

Démonstration : $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$.

Définition

Soit X et Y deux variables aléatoires réelles dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

X et Y sont dites **linéairement indépendantes** si $\text{Cov}(X, Y) = 0$.

Remarques : Deux variables aléatoires indépendantes sont linéairement indépendantes, mais la réciproque est fausse. Par exemple, $X \sim \mathcal{U}([-1, 1])$ et $Y = X^2$ ne sont pas indépendantes, mais $\text{Cov}(X, Y) = \mathbb{E}(X^3) - \mathbb{E}(X^2)\mathbb{E}(X) = 0 - 1 \times 0 = 0$.

$\text{Cov} : L^2(\Omega, \mathcal{F}, \mathbb{P}) \times L^2(\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ est bilinéaire, symétrique et positive, mais pas définie : en effet $\text{Cov}(X, X) = \text{Var}(X) = 0$ n'implique pas que $X = 0$ (seulement X constante). On peut remédier à cela en considérant la relation d'équivalence \equiv définie par $X \equiv Y$ ssi X et Y diffèrent d'une constante ($\exists a \in \mathbb{R}, Y = X + a$). Cov est alors un produit scalaire sur $L^2(\Omega, \mathcal{F}, \mathbb{P}) / \equiv$. On a par ailleurs la complétude de $L^2(\Omega, \mathcal{F}, \mathbb{P}) / \equiv$, ce qui nous permet d'affirmer que $(L^2(\Omega, \mathcal{F}, \mathbb{P}) / \equiv, \text{Cov})$ est un espace de Hilbert pour lequel la norme induite est l'écart-type, et l'orthogonalité est l'indépendance linéaire.

Définition

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, $X : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire tel que $\forall i \in \llbracket 1, d \rrbracket, X_i \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. On appelle **matrice de covariances** de X la matrice

$$\Sigma = \begin{pmatrix} \text{Cov}(X_1, X_1) & \dots & \text{Cov}(X_1, X_d) \\ \vdots & \ddots & \vdots \\ \text{Cov}(X_d, X_1) & \dots & \text{Cov}(X_d, X_d) \end{pmatrix}$$

Proposition

Σ est la matrice de la forme quadratique q définie sur $L = \mathbb{R}^d$ par

$$\forall V \in \mathbb{R}^d, q(V) = \text{Var}(\langle X, V \rangle)$$

Démonstration : Soit q la forme quadratique associée à la matrice Σ . Alors :

$$\begin{aligned} q(V) &= {}^t V \Sigma V = \sum_{i=1}^d \sum_{j=1}^d \text{Cov}(X_i, X_j) V_i V_j = \sum_{i=1}^d \sum_{j=1}^d \text{Cov}(V_i X_i, V_j X_j) \\ &= \text{Cov}\left(\sum_{i=1}^d V_i X_i, \sum_{j=1}^d V_j X_j\right) = \text{Cov}(\langle X, V \rangle, \langle X, V \rangle) = \text{Var}(\langle X, V \rangle) \end{aligned}$$

Proposition

La matrice de covariances est symétrique et positive.

Démonstration : $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$ et $\forall V \in \mathbb{R}^d, {}^t V \Sigma V = \text{Var}(\langle X, V \rangle) \geq 0$.

Proposition

Soit $X : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire admettant une densité f_X dont le support est A . Soit $\phi : A \rightarrow B$ un difféomorphisme \mathcal{C}^1 et $Y = \phi(X)$.

Alors Y admet une densité f_Y définie par

$$f_Y = f_X \circ \phi^{-1} \frac{1}{|J_\phi \circ \phi^{-1}|} 1_B$$

où $J_\phi = \det D\phi$ est la jacobienne de ϕ .

Démonstration : Soit $V \in \mathcal{B}(\mathbb{R}^d)$ et $U = \phi^{-1}(V \cap B)$.

$$\mathbb{P}(X \in U) = \int_U f_X(x) \lambda^{(d)}(dx) = \int_{\phi^{-1}(V \cap B)} f_X(x) \lambda^{(d)}(dx)$$

$$\Rightarrow \mathbb{P}(Y \in B \cap D) = \int_{V \cap B} f_X(\phi^{-1}(y)) |J_{\phi^{-1}}(y)| \lambda^{(d)}(dy)$$

Chapitre IX. Transformée de Fourier, Fonction caractéristique

Section IX.1 - Transformée de Fourier d'une mesure

Définition

Soit μ une mesure finie sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. On appelle **transformée de Fourier** de μ la fonction $\hat{\mu} : \mathbb{R}^d \rightarrow \mathbb{C}$ définie par

$$\hat{\mu}(y) = \int_{\mathbb{R}^d} e^{i\langle x, y \rangle} \mu(dx)$$

Remarque : Le fait que l'on ait choisi une mesure finie rend $e^{i\langle x, y \rangle}$ intégrable. On ne peut pas définir la transformée de Fourier de $\lambda^{(d)}$.

Exemples : Soit $a \in \mathbb{R}^d$ et $\mu = \delta_a$, alors $\hat{\mu}(y) = \int_{\mathbb{R}^d} e^{i\langle x, y \rangle} \delta_a(dx) = e^{i\langle a, y \rangle}$.

Soit $\mu = 1_{[-1,1]} \frac{1}{2\pi} \lambda$, alors $\hat{\mu}(y) = \int_{\mathbb{R}^d} e^{ixy} 1_{[-1,1]} \frac{1}{2\pi} \lambda(dx)$.

Proposition

La fonction $\hat{\mu}$ est continue et bornée (par $\hat{\mu}(0) = \mu(\mathbb{R}^d)$)

Démonstration : Soit $y \in \mathbb{R}^d$. Alors $|\hat{\mu}(y)| = |\int_{\mathbb{R}^d} e^{i\langle x, y \rangle} \mu(dx)| \leq \int_{\mathbb{R}^d} |e^{i\langle x, y \rangle}| \mu(dx) = \int_{\mathbb{R}^d} \mu(dx) = \mu(\mathbb{R}^d)$. Ainsi $\hat{\mu}$ est bornée. Pour la continuité, on peut appliquer le théorème de continuité sous le signe somme avec domination de $|e^{i\langle x, y \rangle}|$ par 1.

Théorème

Soit μ et ν deux mesures finies sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, alors

$$\widehat{\mu * \nu} = \hat{\mu} \hat{\nu}$$

Démonstration : Soit $x \in \mathbb{R}^d$.

$$\widehat{\mu * \nu}(x) = \int_{\mathbb{R}^d} e^{i\langle x, y \rangle} (\mu * \nu)(dy)$$

$\mu * \nu$ est la mesure image de la somme pour la mesure produit donc :

$$\begin{aligned} \widehat{\mu * \nu}(x) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} e^{i\langle x, u+v \rangle} (\mu \otimes \nu)(du, dv) \\ &= \int_{\mathbb{R}^d} e^{i\langle x, u \rangle} \mu(du) \int_{\mathbb{R}^d} e^{i\langle x, v \rangle} \nu(dv) = \hat{\mu}(x) \hat{\nu}(x) \end{aligned}$$

D'où le résultat.

Théorème

Soit μ et ν deux mesures finies sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, alors

$$\hat{\mu} = \hat{\nu} \Leftrightarrow \mu = \nu$$

Section IX.2 - Transformée de Fourier d'une fonction

Définition

Soit $f \in L^1(\mathbb{R})$. On appelle **transformée de Fourier** de f la fonction $\hat{f} : \mathbb{R} \rightarrow \mathbb{C}$ définie par

$$\hat{f}(y) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-ixy} f(x) \lambda(dx)$$

On la note aussi $\mathcal{F}f$.

Remarque : La fonction $\mathcal{F}f$ est bien définie puisque $|e^{-ixy}f(x)| = |f(x)|$ et $f \in L^1(\mathbb{R})$.

Proposition

Lorsque $f \in L^1(\mathbb{R})$, la fonction $\mathcal{F}f$ est continue et bornée sur \mathbb{R} .
De plus, $\|\mathcal{F}f\|_\infty \leq \frac{1}{\sqrt{2\pi}}\|f\|_1$ et $\lim_{x \rightarrow -\infty} \mathcal{F}f(x) = \lim_{x \rightarrow +\infty} \mathcal{F}f(x) = 0$

Démonstration : Si f est à valeurs positives, on pose $\lambda_f = \frac{1}{2\pi}f\lambda$. Alors $\hat{\lambda}_f \leq \lambda_f(\mathbb{R})$, donc $\hat{\lambda}_f$ est borné par $\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f d\lambda = \frac{1}{\sqrt{2\pi}}\|f\|_1$. On a aussi la continuité par le résultat de la section précédente. Dans le cas général, on peut refaire un raisonnement analogue sur la fonction et non la mesure.

Les limites en $+\infty$ et $-\infty$ s'obtiennent en établissant le résultat sur les fonctions f en escalier puis en raisonnant par densité des fonctions en escalier dans $L^1(\mathbb{R})$.

Proposition

Soit $f, g \in L^1(\mathbb{R})$ et $a, b \in \mathbb{C}$. Alors :

- $\mathcal{F}(af + bg) = a\mathcal{F}f + b\mathcal{F}g$.
- $\forall c \in \mathbb{R}^*, \mathcal{F}(x \mapsto f(cx)) = y \mapsto \frac{1}{c}\mathcal{F}f(\frac{y}{c})$.
- $\forall x_0 \in \mathbb{R}^*, \mathcal{F}(x \mapsto f(x - x_0)) = e^{-ix_0 y}\mathcal{F}f$.
- $\mathcal{F}(f * g) = \sqrt{2\pi}\mathcal{F}f\mathcal{F}g$.

Démonstration : La première proposition découle de la linéarité de l'intégrale, les deux suivantes des changements de variables $x \mapsto cx$ et $x \mapsto x - x_0$ et la dernière en utilisant les mesures de densité f et g .

Proposition

Si f et $x \mapsto xf(x)$ sont dans $L^1(\mathbb{R})$, alors :

- $\mathcal{F}f \in C^1(\mathbb{R})$.
- $(\mathcal{F}f)' = \mathcal{F}(x \mapsto -ixf(x))$.

Démonstration : $\forall y \in \mathbb{R}, (x \mapsto f(x)e^{-ixy}) \in L^1(\mathbb{R}), \forall x \in \mathbb{R}$, l'application $y \mapsto f(x)e^{-ixy}$ est dérivable et $(y \mapsto f(x)e^{-ixy})' \leq |xf(x)|$ qui est intégrable. Le théorème de dérivation des intégrales à paramètre donne alors le résultat.

Proposition

Soit $f \in L^1(\mathbb{R}) \cap C^1(\mathbb{R})$ tel que $f' \in L^1(\mathbb{R})$.

Alors $\mathcal{F}(f') = y \mapsto iy(\mathcal{F}f)(y)$.

Démonstration : $\mathcal{F}(f')$ est bien définie puisque f' existe et $f' \in L^1(\mathbb{R})$. Soit $A > 0$, alors :

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{[-A, A]} f'(x)e^{-ixy} \lambda(dx) &= \frac{1}{\sqrt{2\pi}} \int_{-A}^A f'(x)e^{-ixy} dx \\ &= \frac{iy}{\sqrt{2\pi}} \int_{-A}^A f(x)e^{-ixy} dx + \frac{1}{\sqrt{2\pi}} [f(x)e^{-ixy}]_{-A}^A \end{aligned}$$

En faisant tendre A vers $+\infty$, on obtient le résultat.

Définition

Soit $F \in L^1(\mathbb{R})$. On appelle **transformée de Fourier inverse** de F la fonction $\overline{\mathcal{F}} : \mathbb{R} \rightarrow \mathbb{C}$ définie par :

$$\overline{\mathcal{F}}F = y \mapsto \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{ixy} F(x) \lambda(dx)$$

On la note aussi parfois \mathcal{F}^{-1} .

Remarque : La fonction $\mathcal{F}f$ est bien définie puisque $|e^{ixy}F(x)| = |F(x)|$ et $F \in L^1(\mathbb{R})$.

Proposition

Soit $f \in L^1(\mathbb{R})$ telle que $\mathcal{F}f \in L^1(\mathbb{R})$, alors

$$\overline{\mathcal{F}\mathcal{F}}f = f \text{ p.p.}$$

Démonstration :

$$(\overline{\mathcal{F}\mathcal{F}}f)(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{ixu} \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-iuy} f(y) \lambda(dy) \right) \lambda(du)$$

A ce stade, on pourrait être tenté d'appliquer Fubini, mais ce n'est pas possible ici car $(u, y) \mapsto e^{iu(x-y)} f(y) \notin L^1(\mathbb{R}^2)$. Cependant, pour $n \in \mathbb{N}^*$, posons $a_n(u) = \frac{1}{2\pi} e^{-\frac{|u|}{n}}$ et notons $k_n = \mathcal{F}a_n$.

$$\begin{aligned} k_n(x) &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ixu - \frac{|u|}{n}} \lambda(du) \\ &= \frac{1}{2\pi} \int_{\mathbb{R}^-} e^{-ixu + \frac{u}{n}} \lambda(du) + \frac{1}{2\pi} \int_{\mathbb{R}^+} e^{-ixu - \frac{u}{n}} \lambda(du) \\ &= \frac{1}{2\pi} \left(\frac{1}{-ix + \frac{1}{n}} + \frac{1}{ix + \frac{1}{n}} \right) = \frac{n}{\pi} \frac{1}{1 + (nx)^2} \end{aligned}$$

On remarque que $\int_{\mathbb{R}} k_n d\lambda = 1$, $\sup_{n \in \mathbb{N}} \int_{\mathbb{R}} |k_n| d\lambda < +\infty$ et pour tout $F \subset \mathbb{R}^*$ fermé, $\lim_{n \rightarrow +\infty} \int_F k_n d\lambda = 0$. Ainsi, k_n est un noyau de sommabilité. Ainsi $k_n * f \rightarrow f$ lorsque n tend vers $+\infty$ dans L^p . Or :

$$(k_n * f)(x) = \int_{\mathbb{R}} \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} a_n(u) e^{-u(x-y)} \lambda(du) \right) f(y) \lambda(dy)$$

et $(u, y) \mapsto e^{i(y-u)x} a_n(u) f(y) \in L^1(\mathbb{R}^2)$, ce qui nous permet d'utiliser le théorème de Fubini :

$$\begin{aligned} (k_n * f)(x) &= \int_{\mathbb{R}} a_n(u) e^{-iux} \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iuy} f(y) \lambda(dy) \right) \lambda(du) \\ &= a_n(-u) e^{iux} \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-iuy} f(y) \lambda(dy) \right) \lambda(du) = \int_{\mathbb{R}} a_n(u) e^{iux} (\mathcal{F}f)(u) \lambda(du) \end{aligned}$$

Puisque $|a_n(u) e^{iux} (\mathcal{F}f)(u)| \leq |(\mathcal{F}f)(u)|$ avec $\mathcal{F}f \in L^1(\mathbb{R})$, le théorème de convergence dominée s'applique. Puisque l'on a la convergence L^p du membre de gauche, on peut trouver une extractrice ϕ telle que la sous-suite $(k_{\phi(n)} * f)$ converge simplement vers f . Ainsi en passant à la limite lorsque $n \rightarrow +\infty$:

$$\begin{aligned} (k_{\phi(n)} * f)(x) &= \int_{\mathbb{R}} a_{\phi(n)}(u) e^{iux} (\mathcal{F}f)(u) \lambda(du) \\ &\Rightarrow f(x) = \overline{\mathcal{F}\mathcal{F}}f(x) \end{aligned}$$

Ce qui nous donne le résultat attendu.

Remarque : Si $f \in L^1(\mathbb{R})$, on a pas forcément $\mathcal{F}f \in L^1(\mathbb{R})$. Par exemple, avec $f = 1_{[-1,1]} \in L^1(\mathbb{R})$, le calcul fournit $\mathcal{F}f = \sqrt{\frac{2}{\pi}} \text{sinc} \notin L^1(\mathbb{R})$.

Définition

On appelle **espace de Schwartz** l'ensemble des fonctions $\phi \in C^\infty(\mathbb{R})$ à décroissance rapide, c'est-à-dire vérifiant

$$\forall (p, q) \in \mathbb{N}^2, \exists M > 0, \forall x \in \mathbb{R}, (1 + x^2)^p |\phi^{(q)}(x)| \leq M$$

On le note $\mathcal{S}(\mathbb{R})$.

Remarque : La décroissance rapide est équivalente à

$$\forall p \in \mathbb{N}, \exists C > 0, \sup_{\alpha \leq p, \beta \leq p} \|x^\alpha \phi^{(\beta)}\|_\infty \leq C$$

Proposition

Soit $\phi, \psi \in \mathcal{S}(\mathbb{R})$, $P \in \mathbb{R}[X]$ et $\lambda \in \mathbb{C}$.

Alors, ϕ' , ϕP , $\phi + \psi$, $\lambda \phi$ et $\phi \psi$ sont dans $\mathcal{S}(\mathbb{R})$.

Exemple : La fonction ϕ définie par $\phi(x) = e^{-x^2}$ est dans l'espace de Schwartz.

Définition

On dit que $\phi \in C^\infty(\mathbb{R})$ est dans $\mathcal{C}_C^{+\infty}(\mathbb{R}) = \mathcal{C}_0^{+\infty}(\mathbb{R}) = \mathcal{D}(\mathbb{R})$ si elle est à support compact, i.e. $\overline{\{x \in \mathbb{R}, \phi(x) \neq 0\}}$ compact.

Exemple : La fonction ϕ définie par

$$\phi(x) = \begin{cases} \exp\left(\frac{-1}{1-x^2}\right) & \text{si } |x| < 1 \\ 0 & \text{sinon} \end{cases}$$

appartient à $\mathcal{C}_C^{+\infty}(\mathbb{R})$.

Proposition

$$\mathcal{C}_0^{+\infty}(\mathbb{R}) \subset \mathcal{S}(\mathbb{R}) \subset L^p(\mathbb{R})$$

Démonstration : La première inclusion se déduit du fait qu'une fonction continue sur un compact est bornée. La seconde se déduit du fait que pour tout $p \in [1, +\infty[$, $x \mapsto (\frac{M}{1+x^2})^p$ est intégrable.

Définition

Soit $\phi \in \mathcal{S}(\mathbb{R})$. Pour α et β dans \mathbb{N} , on note $|\phi|_{\alpha,\beta} = \|x^{(\alpha)}\phi^{(\beta)}\|_\infty$ et on considère la topologie initiale associée aux fonctions $\phi \mapsto |\phi|_{\alpha,\beta}$, c'est-à-dire la topologie la plus fine rendant ces fonctions continues. On l'appelle **la topologie de $\mathcal{S}(\mathbb{R})$** .

Remarque : Soit $(\phi_n)_{n \in \mathbb{N}}$ une suite de $\mathcal{S}(\mathbb{R})$.

$\phi_n \rightarrow \phi$ lorsque $n \rightarrow +\infty$ signifie $\forall p \in \mathbb{N}$, $\lim_{n \rightarrow +\infty} \mathcal{N}_p(\phi_n - \phi) = 0$ où $\mathcal{N}_p(\cdot) = \sum_{0 \leq \alpha, \beta \leq p} |\cdot|_{\alpha,\beta}$.

Proposition

$\mathcal{C}_0^\infty(\mathbb{R})$ est dense dans \mathcal{S} .

Théorème

La transformée de Fourier \mathcal{F} est un automorphisme de $\mathcal{S}(\mathbb{R})$ et $\mathcal{F}^{-1} = \overline{\mathcal{F}}$.

Démonstration : Soit $\phi \in \mathcal{S}(\mathbb{R}) \subset L^1(\mathbb{R})$. On a aussi $x \mapsto x\phi(x) \in \mathcal{S}(\mathbb{R}) \subset L^1(\mathbb{R})$. Donc $\mathcal{F}\phi \in \mathcal{C}^1(\mathbb{R})$ et $(\mathcal{F}\phi)' = \mathcal{F}(x \mapsto -ix\phi(x))$. Par récurrence, on vérifie que $\forall \beta \in \mathbb{N}^*$, $(\mathcal{F}\phi)^{(\beta)} = (-1)^\beta \mathcal{F}(x \mapsto x^\beta \phi(x))$. Par ailleurs, $\phi \in \mathcal{S}(\mathbb{R})$ donc $\phi' \in L^1(\mathbb{R})$. Ainsi, $\mathcal{F}(\phi') = y \mapsto iy(\mathcal{F}\phi)(y)$. Par récurrence, $\forall \alpha \in \mathbb{N}^*$, $\mathcal{F}(\phi^{(\alpha)}) = y \mapsto (iy)^\alpha (\mathcal{F}\phi)(y)$. On a alors :

$$\begin{aligned} y^\alpha (\mathcal{F}\phi)^{(\beta)}(y) &= (-i)^{\alpha+\beta} (iy)^\alpha \mathcal{F}(x \mapsto x^\beta \phi(x))(y) \\ &= (-i)^{\alpha+\beta} \mathcal{F}((x \mapsto x^\beta \phi(x))^{(\alpha)})(y) \end{aligned}$$

On en déduit que $y^\alpha (\mathcal{F}\phi)^\beta$ est borné, et donc que $\mathcal{F}\phi \in \mathcal{S}(\mathbb{R})$. Puisque $\mathcal{F}\phi \in L^1(\mathbb{R})$, on a l'égalité $\overline{\mathcal{F}\mathcal{F}\phi} = \phi$ presque partout, ce qui achève la démonstration.

Théorème (Formule de Plancherel)

Pour tout ϕ et ψ dans $\mathcal{S}(\mathbb{R})$:

$$\langle \mathcal{F}\phi, \mathcal{F}\psi \rangle_{L^2(\mathbb{R})} = \langle \phi, \psi \rangle_{L^2(\mathbb{R})}$$

Démonstration : Soit ϕ et ψ dans $\mathcal{S}(\mathbb{R})$.

$$\begin{aligned} \langle \phi, \psi \rangle_{L^2(\mathbb{R})} &= \int_{\mathbb{R}} \overline{\phi(x)} \psi(x) \lambda(dx) \\ &= \int_{\mathbb{R}} \frac{1}{2\pi} \int_{\mathbb{R}} (\mathcal{F}\phi)(y) e^{ixy} \lambda(dy) \overline{\psi(x)} \lambda(dx) \end{aligned}$$

$(x, y) \mapsto \overline{\mathcal{F}\phi}(y)\psi(x) \in L^1(\mathbb{R})$ donc Fubini s'applique.

$$\begin{aligned}\langle \phi, \psi \rangle_{L^2(\mathbb{R})} &= \int_{\mathbb{R}} \frac{1}{2\pi} \overline{(\mathcal{F}\phi)(y)} \int_{\mathbb{R}} \psi(x) e^{-ixy} \lambda(dx) \lambda(dy) \\ &= \int_{\mathbb{R}} \overline{\mathcal{F}\phi}(y) (\mathcal{F}\psi)(y) \lambda(dy) = \langle \mathcal{F}\phi, \mathcal{F}\psi \rangle_{L^2(\mathbb{R})}\end{aligned}$$

Remarque : $\langle \mathcal{F}\phi, \psi \rangle_{L^2(\mathbb{R})} = \langle \phi, \overline{\mathcal{F}\psi} \rangle_{L^2(\mathbb{R})}$.

On dit que $\overline{\mathcal{F}}$ est l'opérateur adjoint de \mathcal{F} .

Définition

On définit \mathcal{F} de $L^2(\mathbb{R})$ dans $L^2(\mathbb{R})$ par densité. Si $f \in L^2(\mathbb{R})$, on peut construire une suite f_n d'éléments de $\mathcal{S}(\mathbb{R})$ qui converge vers f . \mathcal{F} étant une isométrie de $\mathcal{S}(\mathbb{R})$ et par complétude de $L^2(\mathbb{R})$, $\mathcal{F}f_n$ admet une limite dans $L^2(\mathbb{R})$, qu'on note $\mathcal{F}f$. Si $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, \mathcal{F} coïncide bien avec la définition donnée sur $L^1(\mathbb{R})$.

Proposition

Soit $f \in L^2(\mathbb{R})$. Alors :

$$\mathcal{F}f = \frac{1}{\sqrt{2\pi}} \lim_{n \rightarrow +\infty} \left(y \mapsto \int_{[-n, n]} f(x) e^{-ixy} \lambda(dx) \right) \text{ dans } L^2(\mathbb{R})$$

Démonstration : On pose $\phi_n = f1_{[-n, n]}$. $\forall n \in \mathbb{N}, \phi_n \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ car $f \in L^2(\mathbb{R})$ et $1_{[-n, n]} \in L^2(\mathbb{R})$. On a $\lim_{n \rightarrow +\infty} \|\phi_n - f\|_2 = 0$ donc $\lim_{n \rightarrow +\infty} \|\mathcal{F}\phi_n - \mathcal{F}f\|_2 = 0$. Or $\mathcal{F}\phi_n = \frac{1}{\sqrt{2\pi}} \int_{[-n, n]} f(x) e^{-ixy} \lambda(dx)$, ce qui donne le résultat attendu.

Proposition

Soit $f \in L^2(\mathbb{R})$. Alors :

$$\mathcal{F}f = \frac{1}{\sqrt{2\pi}} \frac{d}{dy} \int_{\mathbb{R}} f(x) \frac{1 - e^{-ixy}}{ix} \lambda(dx)$$

Démonstration : On pose $\phi_n = f1_{[-n, n]} \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Soit $y \in \mathbb{R}^+$. $\lim_{n \rightarrow +\infty} \langle 1_{[0, y]}, \mathcal{F}\phi_n \rangle = \langle 1_{[0, y]}, \mathcal{F}f \rangle$. En appliquant Fubini à $(x, t) \mapsto f(x) e^{-ixt} \in L^1([-n, n] \times [0, y])$:

$$\begin{aligned}\lim_{n \rightarrow +\infty} \int_{[0, y]} \frac{1}{\sqrt{2\pi}} \int_{[-n, n]} f(x) e^{-ixt} \lambda(dx) \lambda(dt) &= \int_{[0, y]} \mathcal{F}f(x) \lambda(dx) \\ \Rightarrow \frac{1}{\sqrt{2\pi}} \lim_{n \rightarrow +\infty} \int_{[-n, n]} \int_{[0, y]} f(x) e^{-ixt} \lambda(dt) \lambda(dx) &= \int_{[0, y]} \mathcal{F}f(x) \lambda(dx) \\ \Rightarrow \frac{1}{\sqrt{2\pi}} \lim_{n \rightarrow +\infty} \int_{[-n, n]} f(x) \frac{1 - e^{-ixy}}{ix} \lambda(dx) &= \int_{[0, y]} \mathcal{F}f(x) \lambda(dx)\end{aligned}$$

Comme $f \in L^2(\mathbb{R})$ et $x \mapsto \frac{1 - e^{-ixy}}{ix} \in L^2(\mathbb{R})$, on a $y \mapsto \int_{\mathbb{R}} f(x) \frac{1 - e^{-ixy}}{ix} \lambda(dx) \in L^2(\mathbb{R})$ et on peut alors appliquer le théorème de convergence dominée :

$$\int_{\mathbb{R}} f(x) \frac{1 - e^{-ixy}}{ix} \lambda(dx) = \int_{[0, y]} \mathcal{F}f(x) \lambda(dx)$$

D'où le résultat après dérivation.

Théorème (Plancherel)

\mathcal{F} est un automorphisme isométrique de $L^2(\mathbb{R})$.

Démonstration : C'est une conséquence directe du fait que \mathcal{F} est un automorphisme isométrique dans $\mathcal{S}(\mathbb{R})$. (cf. formule de Plancherel)

Proposition

Soit $f \in L^2(\mathbb{R}) \cap C^1(\mathbb{R})$, tel que $f' \in L^2(\mathbb{R})$. Alors :

$$\mathcal{F}(f') = (y \mapsto iy)\mathcal{F}f$$

Démonstration : La proposition s'établit dans $\mathcal{S}(\mathbb{R})$, puis en passant à la limite.

Section IX.3 - Fonction caractéristique

Définition

Soit X une variable aléatoire et P_X sa loi.

P_X s'appelle la **fonction caractéristique** de X , et se note Φ_X :

$$\Phi_X(t) = \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} P_X(dx) = \mathbb{E}(e^{i\langle t, X \rangle})$$

Remarque : Lorsque P_X a une densité f par rapport à la mesure de Lebesgue, alors

$$\Phi_X(t) = \int_{\mathbb{R}^d} f(x) e^{i\langle t, x \rangle} \lambda(dx)$$

Exemples : Pour $P_X = \sum_{k=1}^n \frac{1}{n} \delta_{a_k}$ (loi uniforme discrète) :

$$\Phi_X(t) = \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} \left(\sum_{k=1}^n \frac{1}{n} \delta_{a_k} \right) (dx) = \frac{1}{n} \sum_{k=1}^n e^{ia_k t}$$

Proposition

Soit X une variable aléatoire. Alors :

- $\Phi_X(0) = 1$
- $\forall t \in \mathbb{R}^d, |\Phi_X(t)| \leq 1$
- $\forall a \in \mathbb{R}, \forall b \in \mathbb{R}^d, \Phi_{aX+b} = e^{ibt} \Phi_X(at)$
- Φ_X est continue sur \mathbb{R}^d .

Proposition

Soit X une variable aléatoire dont la loi a une densité f_X par rapport à la mesure de Lebesgue. Alors :

- $\lim_{t \rightarrow -\infty} \Phi_X(t) = \lim_{t \rightarrow +\infty} \Phi_X(t) = 0$.
- $f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle t, x \rangle} \Phi_X(t) \lambda^{(d)}(dt)$

Proposition

Soit X une variable aléatoire. Φ_X satisfait :

$$\forall N \in \mathbb{N}^*, (t_1, \dots, t_N) \in \mathbb{R}^N, (x_1, \dots, x_N) \in \mathbb{R}^N, \sum_{1 \leq j, k \leq N} x_j \Phi_X(t_j - t_k) \overline{x_k} \geq 0$$

Démonstration : Cela provient de l'égalité :

$$\sum_{1 \leq j, k \leq N} x_j \Phi_X(t_j - t_k) \overline{x_k} = \mathbb{E} \left(\left| \sum_{i=1}^N x_i e^{i\langle t_i, X \rangle} \right|^2 \right) \geq 0$$

Théorème (Théorème d'unicité)

Deux variables aléatoires X et Y ont la même loi ssi $\Phi_X = \Phi_Y$.

Démonstration : Deux mesures ayant la même transformée de Fourier sont égales.

Théorème

Les variables aléatoires réelles X_1, \dots, X_n sont indépendantes ssi :

$$\forall (t_1, \dots, t_n) \in \mathbb{R}^n, \Phi_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = \prod_{k=1}^n \Phi_{X_k}(t_k)$$

Démonstration : Par définition de la mesure produit :

$$\int_{\mathbb{R}^n} e^{i\langle t, x \rangle} (P_{X_1} \otimes \dots \otimes P_{X_n})(dx_1, \dots, dx_n) = \prod_{k=1}^n \int_{\mathbb{R}} e^{it_k x_k} P_{X_k}(dx_k)$$

Le résultat équivaut donc à $P_{(X_1, \dots, X_n)} = P_{X_1} \otimes \dots \otimes P_{X_n}$, c'est-à-dire à l'indépendance des variables aléatoires.

Proposition

Soit X_1, \dots, X_n des variables aléatoires indépendantes. Alors :

$$\Phi_{X_1 + \dots + X_n} = \prod_{k=1}^n \Phi_{X_k}$$

Démonstration : On sait que $P_{X_1 + \dots + X_n} = P_{X_1} * \dots * P_{X_n}$. On a alors $\widehat{P_{X_1 + \dots + X_n}} = \widehat{P_{X_1}} \dots \widehat{P_{X_n}}$, d'où le résultat.

Proposition

Soit X une variable aléatoire dans $L^n(\Omega, \mathcal{F}, \mathbb{P})$ avec $n \in \mathbb{N}^*$.

Alors $\Phi_X \in C^n(\mathbb{R})$ et

$$\forall k \leq n, \forall t \in \mathbb{R}, \Phi_X^{(k)}(t) = i^k \mathbb{E}(X^k e^{itX})$$

Proposition

Soit X une variable aléatoire dans $L^n(\Omega, \mathcal{F}, \mathbb{P})$ avec $n \in \mathbb{N}^*$. Alors :

$$\mathbb{E}(X^k) = (-i)^n \Phi_X^{(k)}(0)$$

Démonstration : C'est un corollaire immédiat de la proposition précédente.

Chapitre X. Vecteurs Gaussiens

Section X.1 - Définition d'un vecteur gaussien

Définition

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et X_1, \dots, X_d des variables aléatoires sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

On dit que le vecteur $X = (X_1, \dots, X_d)$ est **gaussien** si $\forall (a_1, \dots, a_d) \in \mathbb{R}^d$, $a_1 X_1 + \dots + a_d X_d$ suit une loi normale.

Exemples : Soit $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(m_2, \sigma_2^2)$ deux variables aléatoires indépendantes. Alors $X = (X_1, X_2)$ est un vecteur aléatoire gaussien. En effet, $\forall (a_1, a_2) \in \mathbb{R}^2$, $a_1 X_1 + a_2 X_2 \sim \mathcal{N}(a_1 m_1 + a_2 m_2, a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2)$. (pour le montrer, utiliser le fait que la fonction caractéristique d'une somme de deux variables aléatoires indépendantes est le produit des fonctions caractéristiques de chaque variable)

Soit $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$, ϵ suivant la loi de Bernoulli $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ indépendante de X_1 et $X_2 = \epsilon X_1$. On a $\Phi_{X_2}(t) = \int_{\mathbb{R}^2} e^{itux} (\mathbb{P}_{X_1} \otimes P_\epsilon)(dx, du)$ par indépendance des variables aléatoires, ce qui se simplifie par application du théorème de Fubini en $\Phi_{X_2}(t) = \int_{\mathbb{R}} \cos(tx) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \lambda(dx) = e^{-\frac{1}{2}t^2}$. Ainsi $X_2 \sim \mathcal{N}(0, 1)$. Or, $X_1 + X_2 = (1 + \epsilon)X_1$ donc $\mathbb{P}(X_1 + X_2 = 0) = \frac{1}{2}$: $X_1 + X_2$ ne peut pas suivre de loi normale, et donc (X_1, X_2) n'est pas gaussien puisque l'on a trouvé une combinaison linéaire de X_1 et X_2 qui ne suit pas une loi normale.

On retiendra que si $X = (X_1, \dots, X_n)$ est gaussien, alors les X_i suivent une loi normale, mais que la réciproque est fausse.

Section X.2 - Caractérisation d'un vecteur gaussien

Proposition

Soit $X = (X_1, \dots, X_d)$ un vecteur gaussien.

Sa fonction caractéristique $\Phi_X : \mathbb{R}^d \rightarrow \mathbb{C}$ est donnée par

$$\Phi_X(t) = \exp \left(i \langle t, m \rangle - \frac{1}{2} \langle t, Dt \rangle \right)$$

où $m = (m_j)_{1 \leq j \leq d}$ est le vecteur d'espérance de X et $D = (D_{j,k})_{1 \leq j,k \leq d}$ est la matrice de covariances de X .

Démonstration : Soit $t = (t_1, \dots, t_d) \in \mathbb{R}^d$ et $Y = \langle t, X \rangle = t_1 X_1 + \dots + t_d X_d$. X étant gaussien, Y suit une loi normale. $\mathbb{E}(Y) = \sum_{k=1}^d t_k m_k = \langle t, m \rangle$ et $\text{Var}(Y) = \text{Cov}(Y, Y) = \sum_{1 \leq k, j \leq d} t_j D_{j,k} t_k = \langle t, Dt \rangle$. On en déduit que $\Phi_Y(u) = \exp(i \langle t, m \rangle u - \frac{1}{2} \langle t, Dt \rangle u^2)$. Or $\Phi_X(t) = \mathbb{E}(\exp(i \langle t, X \rangle)) = \mathbb{E}(\exp(iY)) = \Phi_Y(1)$, d'où le résultat.

Proposition

La loi d'un vecteur gaussien est entièrement caractérisée par son vecteur d'espérance $m \in \mathbb{R}^d$ et sa matrice de covariances $D \in \mathcal{M}_d(\mathbb{R})$.

On notera alors $\mathcal{N}(m, D)$ cette loi.

Démonstration : Φ_X caractérise la loi de X .

Théorème

Soit $X = (X_1, \dots, X_d)$ un vecteur gaussien. Les X_i sont indépendants si et seulement si la matrice D de covariance de X est diagonale.

Démonstration : Pour le sens direct, cela vient simplement du fait que l'indépendance entraîne la non-corrélation. Pour le sens indirect, si D est diagonale alors on a l'égalité $\Phi_X(t_1, \dots, t_d) = \prod_{k=1}^d \Phi_{X_k}(t_k)$.

Proposition

Soit $m \in \mathbb{R}^d$ et $D \in \mathcal{M}_d(\mathbb{R})$ symétrique et positive. Alors, il existe un vecteur gaussien à valeurs dans \mathbb{R}^d d'espérance m et de matrice de covariance D .

Démonstration : D étant symétrique et positive, elle admet une décomposition de Cholesky : $D = C^t C$. Soit d variables aléatoires $Y_i \sim \mathcal{N}(0, 1)$ indépendantes. Le vecteur $Y = (Y_1, \dots, Y_d)$ est gaussien, tout comme $X = CY + m$. Alors, $\mathbb{E}(X) = \mathbb{E}(CY) + m = m$, et $\text{Cov}(X_i, X_j) = \mathbb{E}((CY)_i(CY)_j) = \sum_{1 \leq k, l \leq d} C_{ik} C_{jl} \mathbb{E}(Y_k Y_l) = \sum_{1 \leq k, l \leq d} C_{ik} C_{jl} \delta_{kl} = \sum_{k=1}^d C_{ik} C_{jk} = (D)_{ij}$. On a donc construit le vecteur gaussien recherché.

Section X.3 - Loi d'un vecteur gaussien

Proposition

Soit $m \in \mathbb{R}^d$ et $D \in \mathcal{M}_d(\mathbb{R})$ symétrique positive.

D est inversible si et seulement si la loi $\mathcal{N}(m, D)$ est absolument continue par rapport à la mesure de Lebesgue.

La densité est alors la fonction de \mathbb{R}^d dans \mathbb{R} :

$$x \mapsto \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{\det D}} \exp\left(-\frac{1}{2} \langle x - m, D^{-1}(x - m) \rangle\right)$$

Démonstration : Supposons D inversible. On réécrit comme précédemment $D = C^t C$ et $X = CY + m$, de sorte que $Y = C^{-1}(X - m) = \phi(X)$ où, pour $x = (x_1, \dots, x_d)$, on a défini $\phi(x) = C^{-1}(x - m)$. Chaque Y_i suit une loi normale centrée réduite, et les Y_i sont indépendantes. La densité de la loi de Y est $(y_1, \dots, y_d) \mapsto \prod_{i=1}^d \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} y_i^2) = \frac{1}{(2\pi)^{\frac{d}{2}}} \exp(-\frac{1}{2} \langle y, y \rangle)$. Soit $A \in \mathcal{B}(\mathbb{R}^d)$ et $B = \phi(A)$, alors $\mathbb{P}(X \in A) = \mathbb{P}(Y \in B)$. Or,

$$\begin{aligned} \mathbb{P}(Y \in B) &= \int_B \frac{1}{(2\pi)^{\frac{d}{2}}} \exp(-\frac{1}{2} \langle y, y \rangle) \lambda^{(d)}(dy). \text{ On a donc } \mathbb{P}(X \in A) = \int_{\phi(A)} \frac{1}{(2\pi)^{\frac{d}{2}}} \exp(-\frac{1}{2} \langle y, y \rangle) \lambda^{(d)}(dy) \\ &= \int_A \frac{1}{(2\pi)^{\frac{d}{2}}} \exp(-\frac{1}{2} \langle C^{-1}(x - m), C^{-1}(x - m) \rangle) |\det C^{-1}| \lambda^{(d)}(dx) = \int_A \frac{1}{(2\pi)^{\frac{d}{2}} \det C} \exp(-\frac{1}{2} \langle x - m, D^{-1}(x - m) \rangle) \lambda^{(d)}(dx), \end{aligned}$$

et on en déduit, puisque $\det C = \sqrt{\det D}$, la densité attendue.

Réciproquement, on suppose que D est singulière, et $X \sim \mathcal{N}(m, D)$. Soit $v \in (\text{Ker } D) \setminus \{0\}$. On pose $Z = \langle v, X \rangle$. Alors, $\mathbb{E}(Z) = \mathbb{E}(\langle v, X \rangle) = \mathbb{E}(\sum_{i=1}^d v_i X_i) = \sum_{i=1}^d v_i \mathbb{E}(X_i) = \langle v, m \rangle$, et $\text{Var}(Z) = \text{Var}(\langle v, X \rangle) = {}^t v D v = 0$. On en déduit que Z est égale à son espérance presque partout, soit $\mathbb{P}(Z = \langle v, m \rangle) = 1$. Ainsi, $\mathbb{P}(\langle v, X \rangle = \langle v, m \rangle) = \mathbb{P}(\langle v, X - m \rangle = 0) = 1$. En notant H l'hyperplan de vecteur normal v , cela signifie que $\mathbb{P}(X - m \in H) = \mathbb{P}(X \in m + H) = P_X(m + H) = 1$. Or un hyperplan est de mesure nulle pour la mesure de Lebesgue $\lambda^{(d)}$; si P_X était absolument continue par rapport à la mesure de Lebesgue, on devrait donc avoir $P_X(m + H) = 0$. On conclut donc que si D est singulière alors la loi $\mathcal{N}(m, D)$ ne peut pas être absolument continue par rapport à la mesure de Lebesgue, ce qui achève la preuve.

Chapitre XI. Convergence de variables aléatoires

Section XI.1 - Les différents modes de convergence d'une v.a.

Définition

La suite de v.a. $(X_n)_{n \in \mathbb{N}}$ **converge en probabilité** vers la v.a. X ssi :

$$\forall \epsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$$

On note alors $X_n \xrightarrow{P} X$.

Exemple : Pour $n \in \mathbb{N}^*$, on considère $X_n : \mathbb{R} \rightarrow \mathbb{R}$ définie par :

$$X_n = \omega \mapsto \begin{cases} 0 & \text{si } \omega < 0 \\ 1 - n\omega & \text{si } \omega \in [0, \frac{1}{n}] \\ 0 & \text{si } \omega > \frac{1}{n} \end{cases}$$

Soit $\epsilon > 0$, et $X = \omega \mapsto 0$. Alors, $\mathbb{P}(|X_n - X| > \epsilon) = \mathbb{P}([0, \frac{1-\epsilon}{n}])$. Ainsi, si \mathbb{P} est une mesure à densité par rapport à la mesure de Lebesgue, $\lim_{n \rightarrow +\infty} \mathbb{P}([0, \frac{1-\epsilon}{n}]) = \lim_{n \rightarrow +\infty} \int_0^{\frac{1-\epsilon}{n}} f(x)dx = 0$, c'est-à-dire $X_n \xrightarrow{P} X$.

Définition

La suite de v.a. $(X_n)_{n \in \mathbb{N}}$ **converge presque sûrement** vers la v.a. X ssi :

$$\mathbb{P}(\{\omega \in \Omega; \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)\}) = 1$$

On note alors $X_n \xrightarrow{p.s.} X$.

Exemple : En reprenant X_n définie $\forall n \in \mathbb{N}$ comme précédemment et $X = \omega \mapsto 0$, on remarque que la suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ converge simplement vers X presque partout (il n'y a qu'en 0 qu'on a pas la convergence simple). Ainsi, $X_n \xrightarrow{p.s.} X$.

Proposition

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires convergeant presque sûrement vers X . Alors, $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X .

Démonstration : Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires convergeant presque sûrement vers X . Alors, $\Omega^* = \{\omega \in \Omega; \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)\}$ a pour mesure 1. Pour $\epsilon > 0$, on pose $\Omega^\epsilon = \{\omega \in \Omega; \exists N \in \mathbb{N}, \forall n \geq N, |X_n(\omega) - X(\omega)| < \epsilon\}$. On remarque que $\Omega^\epsilon = \bigcup_{N \in \mathbb{N}^*} \bigcap_{n \geq N} \{\omega \in \Omega, |X_n(\omega) - X(\omega)| < \epsilon\}$ est une union d'intersections d'ensembles mesurables, donc est mesurable, et que $\Omega^* \subset \Omega^\epsilon$. Ainsi, $\mathbb{P}(\Omega^\epsilon) = 1$. Posons $A_N^\epsilon = \bigcap_{n \geq N} \{\omega \in \Omega, |X_n(\omega) - X(\omega)| < \epsilon\}$. Alors, $(A_N^\epsilon)_{N \in \mathbb{N}^*}$ est croissante et $\bigcup_{N \in \mathbb{N}^*} A_N^\epsilon = \Omega^\epsilon$. Donc, $\lim_{N \rightarrow +\infty} \mathbb{P}(A_N^\epsilon) = 1$. Dit autrement, $\forall \delta > 0, \exists N \in \mathbb{N}^*, \mathbb{P}(A_N^\epsilon) > 1 - \delta$, avec pour $n \geq N$, $A_N^\epsilon \subset \{\omega \in \Omega; |X_n(\omega) - X(\omega)| < \epsilon\}$. Donc $\mathbb{P}(|X_n - X| < \epsilon) > 1 - \delta$. Ainsi, $\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| < \epsilon) = 1$, d'où $\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$ et donc $X_n \xrightarrow{p.s.} X$.

Proposition

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires convergeant en probabilité vers X . Alors, on peut extraire une sous-suite $(X_{\phi(n)})_{n \in \mathbb{N}}$ qui converge presque sûrement vers X .

Remarque : Généralement, la convergence en probabilité n'entraîne pas la convergence presque sûrement.

Par ailleurs, elle n'entraîne pas non plus la convergence des moments : en modifiant la définition de la suite de variables aléatoires définies dans le premier exemple par

$$X_n = \omega \mapsto \begin{cases} 0 & \text{si } \omega < 0 \\ n - n^2\omega & \text{si } \omega \in [0, \frac{1}{n}] \\ 0 & \text{si } \omega > \frac{1}{n} \end{cases}$$

alors on a $\forall n \in \mathbb{N}, \mathbb{E}(X_n) = \frac{1}{2}$ mais $\mathbb{E}(X) = 0$, donc $\lim_{n \rightarrow +\infty} \mathbb{E}(X_n) \neq \mathbb{E}(X)$.

Définition

Soit $p \geq 1$. La suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ **converge dans L^p** vers la v.a. X ssi toutes les variables aléatoires X_n et X sont dans L^p et :

$$\lim_{n \rightarrow +\infty} \mathbb{E}(|X_n - X|^p) = 0$$

On note alors $X_n \xrightarrow{L^p} X$.

Exemple : Soit $p \in [1, +\infty[$. On reprend la définition de $(X_n)_{n \in \mathbb{N}}$ du premier exemple :

$$X_n = \omega \mapsto \begin{cases} 0 & \text{si } \omega < 0 \\ 1 - n\omega & \text{si } \omega \in [0, \frac{1}{n}] \\ 0 & \text{si } \omega > \frac{1}{n} \end{cases}$$

et $X = \omega \mapsto 0$. Alors :

$$\mathbb{E}(|X_n - X|^p) = \int_0^{\frac{1}{n}} (1 - n\omega)^p d\omega = \left[\frac{1}{p+1} \frac{-1}{n} (1 - n\omega)^{p+1} \right]_0^{\frac{1}{n}} = \frac{1}{n(p+1)} \xrightarrow{n \rightarrow +\infty} 0$$

et donc $(X_n)_{n \in \mathbb{N}}$ converge vers X dans L^p .

Proposition

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires convergeant dans L^p vers X . Alors, $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X .

Démonstration : Cela résulte de l'inégalité de Markov : $\mathbb{P}(|X_n - X| > \epsilon) < \frac{1}{\epsilon^p} \mathbb{E}(|X_n - X|^p) \xrightarrow{n \rightarrow +\infty} 0$.

Théorème

Soit $p \in [1, +\infty[$ et $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires vérifiant $X_n \xrightarrow{L^p} X$ et $\exists Y \in L^p, \forall n \in \mathbb{N}, |X_n| \leq Y$. Alors, $X \in L^p$ et $X_n \xrightarrow{L^p} X$.

Proposition

Les limites ainsi définies par les convergences en probabilité, presque sûre et dans L^p vérifient l'unicité de la limite, la linéarité et le passage à la limite dans les inégalités. De plus, pour toute fonction f continue, on a $X_n \rightarrow X \Rightarrow f(X_n) \rightarrow f(X)$.

Section XI.2 - Loïs des grands nombres

Théorème (Loi faible des grands nombres)

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires de $L^2(\Omega, \mathcal{F}, \mathbb{P})$ indépendantes et identiquement distribuées.

On note $m = \mathbb{E}(X_n)$ et $M_N = \frac{1}{N} \sum_{n=1}^N X_n$. Alors, $M_N \xrightarrow{P} m$, c'est-à-dire $\forall \epsilon > 0, \lim_{N \rightarrow +\infty} \mathbb{P}(|M_N - m| > \epsilon) = 0$.

Démonstration : On note $m = \mathbb{E}(X_n)$ et $\sigma^2 = \text{Var}(X_n)$. Alors, $\mathbb{E}(M_N) = \frac{1}{N} \sum_{n=1}^N \mathbb{E}(X_n) = \frac{Nm}{N} = m$ et $\text{Var}(M_N) = \frac{1}{N^2} \sum_{n=1}^N \text{Var}(X_n) = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N}$. Pour tout $\epsilon > 0$, on applique l'inégalité de Chebyshev :

$$\mathbb{P}(|M_N - m| > \epsilon) \leq \frac{\sigma^2}{N\epsilon^2} \xrightarrow{N \rightarrow +\infty} 0$$

d'où $M_N \xrightarrow{P} m$.

Théorème (Loi forte des grands nombres)

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires de $L^2(\Omega, \mathcal{F}, \mathbb{P})$ indépendantes et identiquement distribuées. On note $m = \mathbb{E}(X_n)$ et $M_N = \frac{1}{N} \sum_{n=1}^N X_n$. Alors, $M_N \xrightarrow{p.s.} m$ et $M_N \xrightarrow{L^p} m$.

Remarque : Cela nous permet d'effectuer des approximations numériques, par exemple la méthode de Monte Carlo. On prend $X_n \sim \mathcal{U}([0, 1])$ une suite de variables aléatoires indépendantes, et alors on a :

$$\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=1}^N f(X_n) = \mathbb{E}(f(X_n)) = \int_{[0,1]} f(x) \lambda(dx)$$

Cela permet d'approcher la valeur d'intégrales par l'utilisation de variables aléatoires, et on peut par exemple en déduire une approximation de la valeur de π avec l'intégrale sur $[0, 1]$ de $f(x) = \sqrt{1 - x^2}$.

Section XI.3 - Convergence en loi

Définition

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles, $(F_{X_n})_{n \in \mathbb{N}}$ leurs fonctions de répartition respectives, et soit X une variable aléatoire de fonction de répartition F_X .

On dit que la suite des variables aléatoires $(X_n)_{n \in \mathbb{N}}$ **converge en loi** vers la variable aléatoire X ssi $(F_{X_n})_{n \in \mathbb{N}}$ converge simplement vers F_X , sauf éventuellement aux points de discontinuité de F_X . On note $X_n \xrightarrow{L} X$.

Exemple : Considérons $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{U}([0, 1]))$ et :

$$X_n = \omega \mapsto \begin{cases} 0 & \text{si } \omega < 0 \\ 1 - n\omega & \text{si } \omega \in [0, \frac{1}{n}] \\ 0 & \text{si } \omega > \frac{1}{n} \end{cases}$$

de fonctions de répartition respectives :

$$F_{X_n} = x \mapsto \begin{cases} 0 & \text{si } x < 0 \\ \frac{1}{n} + (1 - \frac{1}{n})x & \text{si } x \in [0, \frac{1}{n}] \\ 1 & \text{si } x > \frac{1}{n} \end{cases}$$

Alors, $(F_{X_n})_{n \in \mathbb{N}}$ converge simplement vers $F_X = 1_{[0, +\infty[}$. Ainsi $X_n \xrightarrow{L} 0$.

Définition

Soit $(\mu_n)_{n \in \mathbb{N}}$ une suite de mesures de probabilité sur E . On dit que $(\mu_n)_{n \in \mathbb{N}}$ **converge faiblement** (ou **étroitement**) vers μ ssi

$$\forall f \in \mathcal{C}_b(E), \lim_{n \rightarrow +\infty} \int_E f d\mu_n = \int_E f d\mu$$

où $\mathcal{C}_b(E)$ est l'ensemble des fonctions continues et bornées de E .

Proposition

Lorsque $E = \mathbb{R}$, la suite de variables $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X ssi la suite des lois de X_n converge vers la loi de X .

Définition

Lorsque $E \neq \mathbb{R}$, on dit que la suite de variables $(X_n)_{n \in \mathbb{N}}$ **converge en loi** vers X ssi la suite des lois de X_n converge vers la loi de X . On note $X_n \xrightarrow{L} X$.

Théorème (Portmanteau pour les mesures)

Soit $(\mu_n)_{n \in \mathbb{N}}$ une suite de mesures de probabilité sur E . Toutes les propositions suivantes sont équivalentes :

- $(\mu_n)_{n \in \mathbb{N}}$ converge faiblement vers μ .
- Pour toute fonction f de E uniformément continue et bornée, $\lim_{n \rightarrow +\infty} \int_E f d\mu_n = \int_E f d\mu$.
- Pour toute fonction f de E continue et à support compact, $\lim_{n \rightarrow +\infty} \int_E f d\mu_n = \int_E f d\mu$.
- Pour tout $A \subset E$ fermé, $\limsup_{n \rightarrow +\infty} \mu_n(A) \leq \mu(A)$.
- Pour tout $A \subset E$ ouvert, $\liminf_{n \rightarrow +\infty} \mu_n(A) \geq \mu(A)$.
- Pour tout $A \in \mathcal{B}(E)$ tel que $\mu(\partial A) = 0$, $\lim_{n \rightarrow +\infty} \mu_n(A) = \mu(A)$.

Théorème (Portmanteau pour les variables aléatoires)

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires sur E . Toutes les propositions suivantes sont équivalentes :

- $X_n \xrightarrow{\mathcal{L}} X$.
- Pour toute fonction f de E uniformément continue et bornée, $\lim_{n \rightarrow +\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$.
- Pour toute fonction f de E continue et à support compact, $\lim_{n \rightarrow +\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$.
- Pour tout $A \subset E$ fermé, $\limsup_{n \rightarrow +\infty} \mathbb{P}(X_n \in A) \leq \mathbb{P}(X \in A)$.
- Pour tout $A \subset E$ ouvert, $\liminf_{n \rightarrow +\infty} \mathbb{P}(X_n \in A) \geq \mathbb{P}(X \in A)$.
- Pour tout $A \in \mathcal{B}(E)$ tel que $\mathbb{P}(X \in \partial A) = 0$, $\lim_{n \rightarrow +\infty} \mathbb{P}(X_n \in A) = \mathbb{P}(X \in A)$.

Proposition

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires à valeurs dans un ensemble discret. Alors :

$$X_n \xrightarrow{\mathcal{L}} X \Leftrightarrow \forall k \in E, \lim_{n \rightarrow +\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k)$$

Proposition

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires de fonction caractéristiques respectives $\Phi_n = \Phi_{X_n}$ et X une variable aléatoire de fonction caractéristique $\Phi = \Phi_X$. Alors :

$$X_n \xrightarrow{\mathcal{L}} X \Leftrightarrow \Phi_n \rightarrow \Phi \text{ simplement}$$

Exemple : Soit $\lambda > 0$, $X_n \sim \mathcal{B}(n, \frac{\lambda}{n})$, et $X \sim \text{Pois}(\lambda)$. La fonction caractéristique de X_n est $\Phi_n = t \mapsto (1 - \frac{\lambda}{n} + \frac{\lambda}{n} e^{it})^n$, et $(\Phi(n))_{n \in \mathbb{N}}$ converge simplement vers $\phi = t \mapsto \exp(\lambda(e^{it} - 1))$ qui est précisément la fonction caractéristique de X . On en déduit que $X_n \xrightarrow{\mathcal{L}} X$.

Proposition

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires convergeant en probabilité vers X . Alors, $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X .

Démonstration : Supposons que $X_n \xrightarrow{P} X$. Soit f une fonction continue et bornée. Alors, $f(X_n) \xrightarrow{P} f(X)$ et puisque f est bornée, $|f(X_n)| \leq C \in L^1$ donc $f(X_n) \xrightarrow{L^1} f(X)$. Dit autrement, $\lim_{n \rightarrow +\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$, ce qui donne la convergence en loi par le théorème Portmanteau.

Proposition

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $E = \mathbb{R}^d$. On suppose que $X_n \xrightarrow{\mathcal{L}} X$ et $X = c$ presque sûrement, où $c \in E$ est une constante. Alors, $X_n \xrightarrow{P} X$.

Section XI.4 - Théorème Central Limite (TCL)

Théorème (Central Limite)

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires de $L^2(\Omega, \mathcal{F}, \mathbb{P})$ indépendantes et identiquement distribuées. On note $S_N = \sum_{n=1}^N X_n$, $m = \mathbb{E}(X_n)$ et $\sigma^2 = \text{Var}(X_n)$. On suppose que $\sigma \neq 0$. Alors :

$$\frac{S_N - Nm}{\sigma\sqrt{N}} \xrightarrow[N \rightarrow +\infty]{\mathcal{L}} Y$$

où $Y \sim \mathcal{N}(0, 1)$.

Démonstration : Quitte à remplacer X_n par $\frac{X_n - m}{\sigma}$, on suppose que $m = 0$ et $\sigma = 1$. Soit $Y_N = \frac{1}{\sqrt{N}} S_N = \frac{1}{\sqrt{N}} \sum_{n=1}^N X_n$. Alors, puisque les X_n sont indépendants et identiquement distribués :

$$\Phi_{Y_N}(t) = \Phi_{\frac{1}{\sqrt{N}} \sum_{n=1}^N X_n}(t) = \prod_{i=1}^n \Phi_{X_n} \left(\frac{t}{\sqrt{N}} \right) = \left(\Phi_X \left(\frac{t}{\sqrt{N}} \right) \right)^N$$

Les X_n sont dans L^2 , donc $\Phi_X \in \mathcal{C}^2$. On a alors $\Phi_X(0) = 1$, $\Phi'_X(t) = i\mathbb{E}(Xe^{itX})$ donc $\Phi'_X(0) = im = 0$ et $\Phi''_X(t) = -\mathbb{E}(X^2 e^{itX})$ donc $\Phi''_X(0) = -\sigma = -1$. On en déduit que :

$$\begin{aligned} \Phi_X(t) &= 1 - t^2 + o(t^2) \Rightarrow \Phi_X \left(\frac{t}{\sqrt{N}} \right) = 1 - \frac{t^2}{2N} + o \left(\frac{t^2}{\sqrt{N}} \right) \\ \Rightarrow \ln \Phi_X \left(\frac{t}{\sqrt{N}} \right) &= -\frac{t^2}{2N} + o \left(\frac{t^2}{\sqrt{N}} \right) \Rightarrow N \ln \Phi_X \left(\frac{t}{\sqrt{N}} \right) = -\frac{t^2}{2} + o(t^2) \end{aligned}$$

On obtient ainsi un équivalent à t fixé lorsque $N \rightarrow +\infty$. On en déduit que

$$\forall t \in \mathbb{R}, \Phi_{Y_N}(t) = \left(\Phi_X \left(\frac{t}{\sqrt{N}} \right) \right)^N \xrightarrow[N \rightarrow +\infty]{} \exp \left(-\frac{t^2}{2} \right)$$

On a établi que $(\Phi_{Y_N})_{N \in \mathbb{N}}$ converge simplement vers la fonction caractéristique de $Y \sim \mathcal{N}(0, 1)$, d'où $Y_N \xrightarrow{\mathcal{L}} Y$.

Chapitre XII. Introduction aux processus stochastiques

Section XII.1 - Espérance conditionnelle

Proposition

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire. Soit $\mathcal{G} \subset \mathcal{F}$ une sous-tribu. Alors, il existe une unique variable aléatoire $Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})$ vérifiant $\forall U \in L^2(\Omega, \mathcal{G}, \mathbb{P}), \mathbb{E}(XU) = \mathbb{E}(YU)$.

Démonstration : $H = L^2(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de Hilbert, de produit scalaire $\langle X, Y \rangle = \int_{\Omega} XY d\mathbb{P} = \mathbb{E}(XY)$. $A = L^2(\Omega, \mathcal{G}, \mathbb{P})$ est un sous-espace vectoriel fermé de H , on peut donc définir la projection orthogonale sur A . Ainsi, il existe un unique $Y \in A$ tel que $\forall U \in A, \langle X - Y, U \rangle = 0 \Rightarrow \forall U \in A, \mathbb{E}(XU) = \mathbb{E}(YU)$.

Définition

La variable aléatoire Y définie précédemment est appelée **espérance conditionnelle de X sachant \mathcal{G}** . Elle est notée $\mathbb{E}(X|\mathcal{G})$.

Proposition

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $\mathcal{G} \subset \mathcal{F}$ une sous-tribu. Alors :

- L'application $X \mapsto \mathbb{E}(X|\mathcal{G})$ est linéaire dans $L^1(\Omega, \mathcal{F}, \mathbb{P})$.
- $X \geq 0$ p.s. $\Rightarrow \mathbb{E}(X|\mathcal{G}) \geq 0$ p.s.
- $\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X)$.

Remarque : Cette proposition et un argument de densité permettent d'étendre la définition de $\mathbb{E}(X|\mathcal{G})$ à L^1 .

Définition

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire. Soit $\mathcal{G} \subset \mathcal{F}$ une sous-tribu. La variable aléatoire $Y \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ vérifiant pour toute variable aléatoire U \mathcal{G} -mesurable et bornée, $\mathbb{E}(XU) = \mathbb{E}(YU)$ est appelée **espérance conditionnelle de X sachant \mathcal{G}** . Elle est notée $\mathbb{E}(X|\mathcal{G})$.

Remarque : Cela équivaut à vérifier $\forall A \in \mathcal{G}, \int_A X d\mathbb{P} = \int_A Y d\mathbb{P}$.

Proposition

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $\mathcal{G} \subset \mathcal{F}$ une sous-tribu. Alors :

- L'application $X \mapsto \mathbb{E}(X|\mathcal{G})$ est linéaire dans $L^1(\Omega, \mathcal{F}, \mathbb{P})$.
- $X \geq 0$ p.s. $\Rightarrow \mathbb{E}(X|\mathcal{G}) \geq 0$ p.s.
- $\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X)$.
- $\mathcal{J} \subset \mathcal{G} \subset \mathcal{F} \Rightarrow \mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{J}) = \mathbb{E}(X|\mathcal{J})$.

Proposition

Soit X et Y des variables aléatoires réelles sur $(\Omega, \mathcal{F}, \mathbb{P})$, et $\mathcal{G} \subset \mathcal{F}$ une sous-tribu. On suppose que X est \mathcal{G} -mesurable. Si X, Y et XY sont intégrables (ou positives), alors $\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G})$.

Proposition (Inégalité de Jensen)

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, $\mathcal{G} \subset \mathcal{F}$ une sous-tribu et $\phi : \mathbb{R} \rightarrow \mathbb{R}$ convexe. Si X et $\phi(X)$ sont intégrables, alors $\phi(\mathbb{E}(X|\mathcal{G})) \leq \mathbb{E}(\phi(X)|\mathcal{G})$.

Exemple : On considère $(\Omega, \mathcal{F}, \mathbb{P}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{U}([0, 1]))$ et $X : \mathbb{R} \rightarrow \mathbb{R}$ définie par :

$$X = \omega \mapsto \begin{cases} 0 & \text{si } \omega < 0 \\ 1 - \omega & \text{si } \omega \in [0, 1] \\ 0 & \text{si } \omega > 1 \end{cases}$$

On pose $\mathcal{G} = \sigma(\{[\frac{i}{2}, \frac{i+1}{2}], i \in \mathbb{Z}\})$. On remarque que X n'est pas \mathcal{G} -mesurable. On cherche à déterminer $\mathbb{E}(X|\mathcal{G})$, qui doit être $L^1(\Omega, \mathcal{G}, \mathbb{P})$ et vérifier $\forall A \in \mathcal{G}, \int_A X d\mathbb{P} = \int_A \mathbb{E}(X|\mathcal{G}) d\mathbb{P}$. Puisque $\mathbb{E}(X|\mathcal{G})$ doit être \mathcal{G} -mesurable, elle doit être constante sur les intervalles de la forme $[\frac{i}{2}, \frac{i+1}{2}]$. En calculant l'intégrale de X sur chacun de ces intervalles, on trouve alors que :

$$\mathbb{E}(X|\mathcal{G}) = \omega \mapsto \begin{cases} 0 & \text{si } \omega < 0 \\ \frac{3}{4} & \text{si } \omega \in [0, \frac{1}{2}] \\ \frac{1}{4} & \text{si } \omega \in [\frac{1}{2}, 1] \\ 0 & \text{si } \omega > 1 \end{cases}$$

Pour $\mathcal{J} = \{\emptyset, \Omega\}$, on a $\mathbb{E}(X|\mathcal{J}) = \omega \mapsto \frac{1}{2}$. Puisque $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, on a $\mathbb{E}(X|\mathcal{F}) = X$.

Théorème

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire intégrable (ou positive). Alors :

$$\mathbb{E}(X|\mathcal{G}) = X \Leftrightarrow X \text{ est } \mathcal{G}\text{-mesurable}$$

Proposition

Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{E})$ une variable aléatoire et $B \in \mathcal{F}$ tel que $\mathbb{P}(B) > 0$ et $\mathbb{P}(\Omega \setminus B) > 0$. Alors, $\mathbb{E}(X|\sigma(B))$ est la variable aléatoire

$$\frac{\mathbb{E}(X1_B)}{\mathbb{P}(B)}1_B + \frac{\mathbb{E}(X1_{\Omega \setminus B})}{1 - \mathbb{P}(B)}1_{\Omega \setminus B}$$

Démonstration : $\sigma(B) = \{\emptyset, B, \Omega \setminus B, \Omega\}$. Soit $\omega_1 \in B$ et $x_1 = \mathbb{E}(X|\mathcal{G})(\omega_1)$. Puisque $\mathbb{E}(X|\mathcal{G})$ est mesurable, alors $\mathbb{E}(X|\mathcal{G})^{-1}(x_1) \in \sigma(B)$ soit, puisque cet ensemble est non vide, différent de Ω (la mesure de $\Omega \setminus B$ est non nulle) et contient $\omega_1 \in B$, $\mathbb{E}(X|\mathcal{G})^{-1}(x_1) = B$. De même, si $\omega_2 \in \Omega \setminus B$ et $x_2 = \mathbb{E}(X|\mathcal{G})(\omega_2)$, alors $\mathbb{E}(X|\mathcal{G})^{-1}(x_2) = \Omega \setminus B$. Donc, $\mathbb{E}(X|\mathcal{G}) = x_11_B + x_21_{\Omega \setminus B}$. Or, $\mathbb{E}(X1_B) = \mathbb{E}((x_11_B + x_21_{\Omega \setminus B})1_B) = x_1\mathbb{E}(1_B) = x_1\mathbb{P}(B)$ soit $x_1 = \frac{\mathbb{E}(X1_B)}{\mathbb{P}(B)}$ et de même, $x_2 = \frac{\mathbb{E}(X1_{\Omega \setminus B})}{\mathbb{P}(\Omega \setminus B)}$, ce qui donne le résultat attendu.

Définition

Soit $B \in \mathcal{F}$ tel que $\mathbb{P}(B) > 0$ et $\mathbb{P}(\Omega \setminus B) > 0$.

On appelle **espérance conditionnelle de X sachant B** , et on note $\mathbb{E}(X|B)$, le réel :

$$\mathbb{E}(X|B) = \frac{\mathbb{E}(X1_B)}{\mathbb{P}(B)}$$

Remarque : $\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|\sigma(B))) = \mathbb{E}(\mathbb{E}(X|B)1_B + \mathbb{E}(X|\Omega \setminus B)1_{\Omega \setminus B}) = \mathbb{E}(X|B)\mathbb{P}(B) + \mathbb{E}(X|\Omega \setminus B)\mathbb{P}(\Omega \setminus B)$.

Définition

Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ et $Y : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{E})$ deux variables aléatoires.

On appelle **espérance conditionnelle de X sachant Y** la variable aléatoire $\mathbb{E}(X|\sigma(Y))$. On la note $\mathbb{E}(X|Y)$.

De manière analogue, on notera $\mathbb{E}(X|Y_1, \dots, Y_n) = \mathbb{E}(X|\sigma(Y_1, \dots, Y_n))$.

Théorème

Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ et $Y : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{E})$ deux variables aléatoires.

Il existe une application borélienne $h : E \rightarrow \mathbb{R}$ telle que $\mathbb{E}(X|Y) = h(Y)$.

Proposition

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $(X, Y) : \Omega \rightarrow \mathbb{R}^2$ un vecteur aléatoire admettant une densité $f_{(X,Y)}$. On suppose que $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ et $\forall y \in \mathbb{R}, f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) \lambda(dx) > 0$. On pose :

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \text{ et } h(y) = \int_{\mathbb{R}} x f_{X|Y=y}(x) \lambda(dx)$$

Alors, $\mathbb{E}(X|Y) = h(Y)$.

Proposition

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires. Alors :

$$X_n \xrightarrow{\text{p.s.}} X \text{ et } \forall n \in \mathbb{N}, X_n \geq 0 \Rightarrow \mathbb{E}(X_n | \mathcal{G}) \xrightarrow{\text{p.s.}} \mathbb{E}(X | \mathcal{G}) \text{ (convergence monotone)}$$

$$X_n \xrightarrow{\text{p.s.}} X \Rightarrow \mathbb{E}(\liminf X_n | \mathcal{G}) \leq \liminf \mathbb{E}(X_n | \mathcal{G}) \text{ p.s. (lemme de Fatou)}$$

$$X_n \xrightarrow{\text{p.s.}} X \text{ et } \exists Z \in L^1(\Omega, \mathcal{F}, \mathbb{P}), \forall n \in \mathbb{N}, |X_n| \leq Z \Rightarrow \mathbb{E}(X_n | \mathcal{G}) \xrightarrow{\text{p.s.}} \mathbb{E}(X | \mathcal{G}) \text{ (convergence dominée)}$$

Définition

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $A \in \mathcal{F}$. Soit $\mathcal{G} \subset \mathcal{F}$ une sous-tribu.

On appelle **probabilité conditionnelle de A sachant G**, et on note $\mathbb{P}(A | \mathcal{G})$, la variable aléatoire :

$$\mathbb{P}(A | \mathcal{G}) = \mathbb{E}(1_A | \mathcal{G})$$

Remarque : Pour $B \in \mathcal{F}$, alors :

$$\mathbb{P}(A | \sigma(B)) = \frac{\mathbb{E}(1_A 1_B)}{\mathbb{P}(B)} 1_B + \frac{\mathbb{E}(1_A 1_{\Omega \setminus B})}{\mathbb{P}(\Omega \setminus B)} 1_{\Omega \setminus B} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} 1_B + \frac{\mathbb{P}(A \cap (\Omega \setminus B))}{\mathbb{P}(\Omega \setminus B)} 1_{\Omega \setminus B} = \mathbb{P}(A | B) 1_B + \mathbb{P}(A | (\Omega \setminus B)) 1_{\Omega \setminus B}.$$

Section XII.2 - Processus stochastiques

Définition

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et (E, \mathcal{E}) un espace mesuré. On appelle **processus stochastique** (ou **processus aléatoire**) toute collection de variables aléatoires $(X_t)_{t \in \mathcal{T}}$ sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans E .

On le note $X = \{X_t, t \in \mathcal{T}\}$. Lorsque $\mathcal{T} = \mathbb{N}$, le processus est dit **discret**.

Définition

Un processus stochastique discret $(S_n)_{n \in \mathbb{N}}$ est appelé **marche aléatoire** à un paramètre si ses accroissements $X_n = S_n - S_{n-1}$ pour $n \geq 1$ sont indépendants et identiquement distribués.

Définition

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. On appelle **filtration** toute suite croissante $(\mathcal{F}_n)_{n \in \mathbb{N}}$ de sous-tribus de \mathcal{F} .

Définition

On dit qu'un processus stochastique discret $X = \{X_n, n \in \mathbb{N}\}$ est **adapté à la filtration** \mathcal{F} si pour tout $n \in \mathbb{N}$, X_n est \mathcal{F}_n -mesurable.

Exemple : Soit $X = \{X_n, n \in \mathbb{N}\}$. La filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ définie par $\mathcal{F}_n = \sigma(X_k, k \in \llbracket 1, n \rrbracket)$ est adaptée au processus X . On l'appelle **filtration naturelle** de X .

Définition

Un processus discret X est appelé une **martingale** par rapport à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ ssi le processus est adapté à la filtration, pour tout $n \in \mathbb{N}$, $X_n \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ et pour tout $n \in \mathbb{N}$, $X_n = \mathbb{E}(X_{n+1} | \mathcal{F}_n)$ p.s. (*)

En remplaçant (*) par $X_n \leq \mathbb{E}(X_{n+1} | \mathcal{F}_n)$ p.s., on l'appelle une **sous-martingale**.

En remplaçant (*) par $X_n \geq \mathbb{E}(X_{n+1} | \mathcal{F}_n)$ p.s., on l'appelle une **sur-martingale**.

Proposition

Si X est une martingale, alors $\forall n \in \mathbb{N}, \mathbb{E}(X_n) = \mathbb{E}(X_0)$.

Démonstration : Pour $n \in \mathbb{N}$, $\mathbb{E}(X_n) = \mathbb{E}(\mathbb{E}(X_{n+1} | \mathcal{F}_n)) = \mathbb{E}(X_{n+1})$, et on conclut par récurrence.

Définition

Un processus X adapté à une filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ est **prévisible** si $\forall n \in \mathbb{N}, X_n$ est \mathcal{F}_{n+1} -mesurable.

Proposition

Soit \mathcal{S} une martingale et \mathcal{C} un processus prévisible et borné. Alors, le processus stochastique $((\mathcal{C} \cdot \mathcal{S})_n)_{n \in \mathbb{N}}$ défini par :

$$\begin{cases} (\mathcal{C} \cdot \mathcal{S})_0 = 0 \\ \forall n \in \mathbb{N}^*, (\mathcal{C} \cdot \mathcal{S})_n = \sum_{k=1}^n C_k(S_k - S_{k-1}) \end{cases}$$

est une martingale.

Définition

On appelle $\mathcal{C} \cdot \mathcal{S}$ la **transformée de la martingale \mathcal{S} par le processus \mathcal{C}** .