Polycopié d'Intégration de Lebesgue et Probabilités
L3 MIDO

Cours de M. Trashorras

Retranscrit par Samuel Lelouch

Compilé par Samuel Lelouch et Arris Bouzouane avec Gemini

3 décembre 2025

Table des matières

Chapitre 2 : Intégration par rapport à une mesure

Compléments et Rappels (Exercices)

Exemple 1: Exercice : Unicité de la mesure

Soit $(E, \mathcal{E})$ un espace mesurable et $\mathcal{C}$ une famille de parties de E, stable par intersection finie, tq $\sigma(\mathcal{C}) = \mathcal{E}$.

- Soient μ et ν deux mesures sur $(E, \mathcal{E})$ tq $\forall A \in \mathcal{C},\ \mu(A) = \nu(A)$.
- On suppose que $\forall n \geq 0$, il existe une suite **croissante** $(E_n)_{n\geq 0}$ d'éléments de $\mathcal{C}$ tq $\forall n, \mu(E_n) = \nu(E_n) < +\infty$ et $E = \bigcup_{n\in\mathbb{N}} E_n$.

Montrer que μ et ν sont σ-finies sur $(E, \mathcal{E})$ et égales.

Réponse : En effet, en posant $\forall n \geq 0$, et tout $A \in \mathcal{E}$:

$$\mu_n(A) = \mu(A \cap E_n)$$

$$\nu_n(A) = \nu(A \cap E_n)$$

On définit 2 mesures finies sur $(E, \mathcal{E})$ tq $\forall A \in \mathcal{C}$, on a :

$$\mu_n(A) = \mu(A \cap E_n) = \mu(E_n) - \mu(A^c \cap E_n) \quad \text{(car } \mathcal{C} \text{ stable par intersection)}$$

$$= \nu(E_n) - \nu(A^c \cap E_n) = \nu_n(A)$$

Or on a 2 cas :

- **1er cas :** $\mu_n(E) = \nu_n(E) = 0$. Alors $\mu_n = \nu_n$ sur $(E, \mathcal{E})$.
- **2ème cas :** $\mu_n(E) = \nu_n(E) \neq 0$. Et alors les probabilités $\tilde{\mu}_n(A) = \frac{\mu_n(A)}{\mu_n(E)}$ et $\tilde{\nu}_n(A) = \frac{\nu_n(A)}{\nu_n(E)}$ coïncident sur $\mathcal{C}$. Donc d'après l'exemple déjà vu sur $(E, \mathcal{E})$, on a $\nu_n = \mu_n$ sur $(E, \mathcal{E})$.

Ainsi, $\forall n \geq 0$ et tout $A \in \mathcal{E}$, on a :

$$\nu_n(A) = \nu(A \cap E_n) = \mu(A \cap E_n) = \mu_n(A)$$

De plus :

$$\nu(A) = \lim_{n\to\infty} \nu(A \cap E_n) = \lim_{n\to\infty} \mu(A \cap E_n) = \mu(A)$$

d'après la propriété de continuité croissante des mesures.

Remarque

Conséquence de l'exo précédent, $\mu = \nu$ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Puisque en notant $\mathcal{C} = \{]a, b],\ a, b \in \mathbb{Q}, a < b\}$, on a :

- $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$.
- $\mathcal{C}$ est stable par intersection finie.
- $\forall n \in \mathbb{N}^*, E_n =]-n, n]$ nous donne une suite croissante d'éléments de $\mathcal{C}$ tq $\mathbb{R} = \bigcup E_n$ et $\forall n \in \mathbb{N}^*, \nu(E_n) = \mu(E_n) = 2n$.

Exemple 2: Réponse question (dérivabilité/mesurabilité)

$\forall x \in \mathbb{R}$:

$$f'(x) = \lim_{n\to+\infty} \frac{f(x+\frac{1}{n}) - f(x)}{\frac{1}{n}}$$
$$= \limsup_{n} n\left(f(x+\frac{1}{n}) - f(x)\right)$$

où $\forall n \in \mathbb{N}^*$ et tout entier $n \geq 1$, $f_n(x) = \frac{f(x+\frac{1}{n})-f(x)}{\frac{1}{n}}$ est une fonction continue sur $\mathbb{R}$ donc mesurable. Donc f' est mesurable en tant que limsup d'une suite de fonctions mesurables.

I. Intégration des fonctions étagées positives

Définition 1: Fonction étagée

On dit qu'une fonction $f : E \to \mathbb{R}$ est étagée si et seulement si :

- f est mesurable ($\mathcal{E} - \mathcal{B}(\mathbb{R})$ mesurable).
- f prend un nombre fini de valeurs.

Ainsi $f(E) = \{f(x), x \in E\} \subset \mathbb{R}$ est un ensemble fini et :

$$f = \sum_{\alpha\in f(E)} \alpha \mathbb{1}_{f^{-1}(\{\alpha\})}$$

Cette représentation de f est unique, on dit que c'est la **représentation canonique** de f.

Remarque

Toutes les fonctions en escalier sont des fonctions étagées, mais la réciproque est fausse car par exemple $\mathbb{1}_{\mathbb{Q}}$ n'est pas une fonction en escalier.

Définition 2: Intégrale d'une fonction étagée

Soit $f : E \to F$ une fonction étagée positive de représentation canonique $f = \sum_{\alpha\in f(E)} \alpha \mathbb{1}_{f^{-1}(\{\alpha\})}$. On appelle intégrale de f sur E par rapport à la mesure μ :

$$\int_E f d\mu = \sum_{\alpha\in f(E)} \alpha\mu(f^{-1}(\{\alpha\}))$$

Remarque

1. f est à valeurs réelles.
2. Il est possible que $E = \mathbb{R}$, $\mathcal{E} = \mathcal{B}(\mathbb{R})$.
3. Ici, si $\alpha \in \mathbb{R}^+$, on pose :

$$\alpha\times+\infty = +\infty, \quad \infty\times\infty = \infty, \quad \alpha \geq 0 \implies \alpha\times\infty = +\infty, \quad \alpha = 0 \implies \alpha\times(+\infty) = 0$$

4. Si $f = \mathbb{1}_{\mathbb{Q}}$ définie sur $E = \mathbb{R}, \mathcal{E} = \mathcal{B}(\mathbb{R}), \mu = \lambda$ (la mesure de Lebesgue) :

$$\int_{\mathbb{R}} f d\lambda = 1 \times \lambda(\mathbb{Q}) + 0 \times \lambda(\mathbb{Q}^c)$$

puisque $\mathbb{1}_{\mathbb{Q}} = 1 \times \mathbb{1}_{\mathbb{Q}} + 0 \times \mathbb{1}_{\mathbb{Q}^c}$.

$$= 1 \times 0 + 0 \times (+\infty) = 0$$

5. Si f est étagée positive, nécessairement $\int_E f d\mu \in [0, +\infty]$. Si $f = \mathbb{1}_{\mathbb{R}+}$:

$$\int_{\mathbb{R}} g d\lambda = 1 \times \lambda(\mathbb{R}^+) + 0 \times \lambda(\mathbb{R}^*_-)$$

$$= 1 \times +\infty + 0 \times (+\infty) = +\infty + 0 = +\infty$$

Définition 3: Exemple (suite)

Soit $(E, \mathcal{E}, \mu)$ un espace mesuré. Si $F : E \to \mathbb{R}^+$ est une fonction étagée alors :

$$F = \sum_{\alpha \in f(E)} \alpha \mathbb{1}_{f^{-1}(\{\alpha\})}$$

$$= \sum_{\alpha \in f(E)} \alpha \mathbb{1}_{\{f = \alpha\}} \quad (\leftarrow \text{somme finie})$$

On définit :

$$\int_E F d\mu = \sum_{\alpha \in f(E)} \alpha \mu(f^{-1}(\{\alpha\})) \in [0, +\infty]$$

Exemple 3: Mesure de Dirac

Soit E un ensemble non vide, $a \in E$ et δ_a la mesure de Dirac en a sur $(E, \mathcal{P}(E))$. $\forall A \subset E$:

$$\delta_a(A) = \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{sinon} \end{cases}$$

Soit $f : E \to \mathbb{R}^+$ une fonction étagée positive $f = \sum \alpha \mathbb{1}_{f^{-1}(\{\alpha\})}$. Il existe un **unique** $\alpha \in f(E)$ tq $a \in \{f = \alpha\}$.

$$\int_E f d\delta_a = \sum_{\alpha \in f(E)} \alpha \delta_a(\{f = \alpha\})$$

$$= \sum_{\alpha \in f(E)} \alpha \delta_a(\{f = \alpha\}) \quad (\text{avec } \delta_a(\{f = \alpha\}) = 0 \text{ sauf pour } \alpha \text{ tq } f(a) = \alpha)$$

$$= f(a)$$

Proposition 1: Propriété de la représentation

Soit $f : E \to \mathbb{R}^+$ une fonction étagée positive. Pour toute représentation de f de la forme :

$$f = \sum_{i \in I} a_i \mathbb{1}_{A_i} \quad \text{où } I \text{ est un ensemble fini}$$

et les $(A_i)_{i \in I}$ constituent une partition de E en parties mesurables, on a :

$$\int_E f d\mu = \sum_{i \in I} a_i \mu(A_i)$$

Remarque

$E = [0, 2]$.

$$f(x) = \begin{cases} 1 & \text{si } x \in [0, 1] \\ 3 & \text{si } x \in]1, 2] \end{cases} \qquad F = 1 \times \mathbb{1}_{[0,1]} + 3\mathbb{1}_{]1,2]}$$

On peut aussi écrire :

$$\begin{aligned} F &= 1 \times \mathbb{1}_{[0,1]} + 1 \times \mathbb{1}_{]1,1.1]} + 3 \times \mathbb{1}_{]1.1,2]} \\ &= 1 \times \mathbb{1}_{[0,1.1]} + 2 \times \mathbb{1}_{]1,2]} \quad \text{(Pas une partition)} \end{aligned}$$

Preuve : Soit $f = \sum_{i \in I} a_i \mathbb{1}_{A_i}$ où I est un ensemble fini et les $(A_i)_{i \in I}$ constituent une partition de E en éléments de $\mathcal{E}$. En particulier, $\forall i \in I$, $a_i \in f(E)$.

$$\begin{aligned} \sum_{i \in I} a_i \mu(A_i) &= \sum_{\alpha \in f(E)} \left(\sum_{\substack{i \in I \\ a_i = \alpha}} a_i \mu(A_i) \right) = \sum_{\alpha \in f(E)} \left(\sum_{\substack{i \in I \\ a_i = \alpha}} \alpha \mu(A_i) \right) \\ &= \sum_{\alpha \in f(E)} \alpha \left(\sum_{\substack{i \in I \\ a_i = \alpha}} \mu(A_i) \right) = \sum_{\alpha \in f(E)} \alpha \mu \left(\bigcup_{\substack{i \in I \\ a_i = \alpha}} A_i \right) \\ &= \sum_{\alpha \in f(E)} \alpha \mu(f^{-1}(\{\alpha\})) \end{aligned}$$

Proposition 2: Propriétés de l'intégrale (Fonctions étagées)

Soient $f, g : E \to \mathbb{R}^+$ des fonctions étagées positives et $\lambda \in \mathbb{R}_*^+$.

1. La fonction λf est étagée positive et :

$$\int_E (\lambda f) d\mu = \lambda \int_E f d\mu$$

2. La fonction $f + g$ est étagée positive et :

$$\int_E (f + g) d\mu = \int_E f d\mu + \int_E g d\mu$$

3. Si $\forall x \in E, f(x) \geq g(x)$ alors :

$$\int_E f d\mu \geq \int_E g d\mu$$

Preuve : **1 -** Soit $f : E \to \mathbb{R}^+$ est étagée positive et $\lambda \in \mathbb{R}^+$. Comme $f = \sum_{\alpha \in f(E)} \alpha \mathbb{1}_{f^{-1}(\{\alpha\})}$, on a $\lambda f = \sum_{\alpha \in f(E)} \lambda\alpha \mathbb{1}_{f^{-1}(\{\alpha\})}$. Donc :

$$\int_E (\lambda f) d\mu = \sum_{\alpha \in f(E)} \lambda\alpha\mu(\{f = \alpha\}) = \lambda \sum_{\alpha \in f(E)} \alpha\mu(\{f = \alpha\}) = \lambda \int_E f d\mu$$

2 - Soient $f, g : E \to \mathbb{R}^+$ deux fonctions étagées positives.

$$f = \sum_{\alpha \in f(E)} \alpha \mathbb{1}_{\{f=\alpha\}} = \sum_{i \in I} a_i \mathbb{1}_{A_i}$$

$$g = \sum_{\beta \in g(E)} \beta \mathbb{1}_{\{g=\beta\}} = \sum_{j \in J} b_j \mathbb{1}_{B_j}$$

I est un ensemble fini, $(A_i)_{i\in I}$ est une partition mesurable de E. J est un ensemble fini, $(B_j)_{j\in J}$ est une partition mesurable de E. Les ensembles $(A_i \cap B_j)_{i\in I, j\in J}$ constituent une partition mesurable de E.

$$f + g = \sum_{(i,j)\in I\times J} (a_i + b_j)\mathbb{1}_{A_i \cap B_j}$$

$$\int_E (f+g) d\mu = \sum_{(i,j)\in I\times J} (a_i + b_j)\mu(A_i \cap B_j)$$

$$= \sum_{(i,j)\in I\times J} a_i\mu(A_i \cap B_j) + \sum_{(i,j)\in I\times J} b_j\mu(A_i \cap B_j)$$

$$= \sum_{i\in I} \left| \sum_{j\in J} a_i\mu(A_i \cap B_j) \right| + \sum_{j\in J} \left| \sum_{i\in I} b_j\mu(A_i \cap B_j) \right|$$

$$= \sum_{i\in I} a_i \left(\sum_{j\in J} \mu(A_i \cap B_j) \right) + \sum_{j\in J} b_j \left(\sum_{i\in I} \mu(A_i \cap B_j) \right)$$

$$= \sum_{i\in I} a_i\mu \left(\bigcup_{j\in J} A_i \cap B_j \right) + \sum_{j\in J} b_j\mu \left(\bigcup_{i\in I} A_i \cap B_j \right)$$

$$= \sum_{i\in I} a_i\mu(A_i) + \sum_{j\in J} b_j\mu(B_j)$$

$$= \int_E f d\mu + \int_E g d\mu$$

3 - Soient $f, g : E \to \mathbb{R}^+$ étagées positives, tq $\forall x \in E, f(x) \geq g(x)$. Alors :

$$\int_E f d\mu = \int_E (f - g + g) d\mu$$

$$= \int_E (f-g) d\mu + \int_E g d\mu$$

Or $f - g \geq 0$, donc $\int_E (f-g) d\mu \geq 0$.

$$\geq \int_E g d\mu$$

Proposition 3: Calcul via partition quelconque

Soit $N \in \mathbb{N}^*$, $a_1, \dots, a_N \in \mathbb{R}^+$ et $A_1, \dots, A_N \in \mathcal{E}$. Si $f = \sum_{i=1}^N a_i \mathbb{1}_{A_i}$ alors f est une fonction étagée positive et $\int_E f d\mu = \sum_{i=1}^N a_i \mu(A_i)$.

<u>Preuve :</u> D'après la proposition précédente, si $f = \sum_{i=1}^N a_i \mathbb{1}_{A_i}$ avec $N \in \mathbb{N}^*$, $a_1, \dots, a_N \geq 0$ et $A_1, \dots, A_N \in \mathcal{E}$, alors :

$$\begin{aligned}\int_E f d\mu = \int_E \left(\sum_{i=1}^N a_i \mathbb{1}_{A_i} \right) d\mu &= \sum_{i=1}^N \int_E a_i \mathbb{1}_{A_i} d\mu \\ &= \sum_{i=1}^N a_i \int_E \mathbb{1}_{A_i} d\mu \\ &= \sum_{i=1}^N a_i \mu(A_i)\end{aligned}$$

Remarque

Soient $a, b \in \mathbb{R}, a < b$ et $f : [a, b] \to \mathbb{R}^+$ une fonction en escalier. Il existe donc une subdivision $a = \sigma_0 < \sigma_1 < \dots < \sigma_N = b$ de $[a, b]$ tq f est constante sur les intervalles $]\sigma_i, \sigma_{i+1}[$ $\forall i \in [\![0, N-1]\!]$. On peut donc écrire :

$$f = \sum_{i=0}^{N-1} f\left(\frac{\sigma_i + \sigma_{i+1}}{2}\right) \mathbb{1}_{]\sigma_i, \sigma_{i+1}[} + \sum_{j=0}^{N} f(\sigma_j) \mathbb{1}_{\{\sigma_j\}}$$

et on a vu que $\int_a^b f(t) dt = \sum_{i=0}^{N-1} f\left(\frac{\sigma_i + \sigma_{i+1}}{2}\right)(\sigma_{i+1} - \sigma_i)$.

Or, en notant λ la mesure de Lebesgue :

$$\begin{aligned}\int_{[a,b]} f d\lambda = \int_{[a,b]} \left(\sum_{i=0}^{N-1} f\left(\frac{\sigma_i + \sigma_{i+1}}{2}\right) \mathbb{1}_{]\sigma_i, \sigma_{i+1}[} \right) + \sum_{j=0}^{N} f(\sigma_j) \mathbb{1}_{\{\sigma_j\}} d\lambda \\ = \sum_{i=0}^{N-1} \int_{[a,b]} f(\dots) \mathbb{1}_{]\sigma_i, \sigma_{i+1}[} d\lambda + \sum_{j=0}^{N} f(\sigma_j) \int_{[a,b]} \mathbb{1}_{\{\sigma_j\}} d\lambda \\ = \sum_{i=0}^{N-1} f\left(\frac{\sigma_i + \sigma_{i+1}}{2}\right) \lambda(]\sigma_i, \sigma_{i+1}[) + \sum_{j=0}^{N} f(\sigma_j) \lambda(\{\sigma_j\}) \\ = \sum_{i=0}^{N-1} f\left(\frac{\sigma_i + \sigma_{i+1}}{2}\right)(\sigma_{i+1} - \sigma_i) + \sum_{j=0}^{N} f(\sigma_j) \times 0 \\ = \int_a^b f(t) dt\end{aligned}$$

II. Intégrale des fonctions mesurables positives

Définition 4: Intégrale d'une fonction mesurable positive

Soit $(E, \mathcal{E}, \mu)$ un espace mesuré et $f : E \to [0, +\infty]$ une fonction mesurable positive. On appelle intégrale de f par rapport à la mesure μ la quantité notée :

$$\int_E f d\mu = \sup \left\{ \int_E h d\mu : \begin{array}{l} h : E \to \mathbb{R} \text{ étagée positive} \\ h \le f \end{array} \right\}$$

Remarque

1. Dès que f est mesurable et positive, $\int_E f d\mu$ est correctement définie.
2. $\int_E f d\mu \in [0, +\infty]$.
3. f est à valeurs dans $[0, +\infty]$.
4. Si $f : E \to \mathbb{R}$ est une fonction étagée positive, on a toujours avec cette nouvelle définition :

$$\int_E f d\mu = \sum_{\alpha \in f(E)} \alpha \mu(\{f = \alpha\})$$

Proposition 4: Propriétés basiques

Soit $(E, \mathcal{E}, \mu)$ un espace mesuré et $f, g : E \to [0, +\infty]$ deux fonctions mesurables.

1. Si $f \ge g$ alors $\int_E f d\mu \ge \int_E g d\mu$.
2. Si $\mu(\{f > 0\}) = \mu(\{x \in E : f(x) > 0\}) = 0$ alors $\int_E f d\mu = 0$.

Preuve : **1** - Puisque $f \ge g$.

$$\{h : E \to \mathbb{R}^+ \text{ étagée}, h \le g\} \subset \{h : E \to \mathbb{R}^+ \text{ étagée}, h \le f\}$$

donc $\{\int_E h d\mu, \dots, h \le g\} \subset \{\int_E h d\mu, \dots, h \le f\}$. donc $\sup \{\int_E h d\mu, h \le g\} \le \sup \{\int_E h d\mu, h \le f\}$.

$$\int_E g d\mu \le \int_E f d\mu$$

2 - Soit h une fonction étagée positive telle que $0 \le h \le f$. On a $h = \sum_{\alpha \in h(E)} \alpha \mathbb{1}_{\{h=\alpha\}}$. Et nécessairement pour tout $\alpha \in h(E)$, $\alpha = 0$ et $\mu(h = \alpha) = 0$. Car si on avait pour un $\alpha \in h(E)$, $\alpha > 0$ et $\mu(h = \alpha) \ne 0$, on aurait $\mu(f > 0) \ge \mu(h = \alpha) > 0$. Ainsi pour toute fonction étagée $h : E \to \mathbb{R}^+$ tq $h \le f$, on a :

$$\int_E h d\mu = \sum_{\alpha \in h(E)} \alpha \mu(h = \alpha) = 0$$

Donc $\int_E f d\mu = \sup_h \{\int_E h d\mu\} = \sup\{0\} = 0$.

Théorème 1: Théorème de Convergence Monotone (TCM)

Soit $(E, \mathcal{E}, \mu)$ un espace mesuré et $(f_n)_{n \ge 0}$ une suite **croissante** de fonctions **mesurables positives** définies sur E.

- $\forall x \in E$, la suite $(f_n(x))_{n \ge 0}$ est une suite croissante d'éléments de $[0, +\infty]$.

- $\forall n \in \mathbb{N}$, $f_n : E \to [0, +\infty]$ est mesurable.

Alors, la fonction $F : E \to [0, +\infty]$ définie par :

$$\forall x \in E, \quad F(x) = \lim_{n \to +\infty} f_n(x) = \sup_{n \geq 0} f_n(x)$$

est mesurable, à valeurs dans $[0, +\infty]$, et la suite $(\int_E f_n d\mu)_{n \geq 0}$ est une suite croissante qui tend vers $\int_E F d\mu$.

Preuve :

La fonction F est mesurable positive, donc $\int_E F d\mu$ est correctement définie. Comme la suite $(f_n)_{n \geq 0}$ est croissante, $\forall n \in \mathbb{N}$, on a $f_n \leq f_{n+1} \leq F$. D'après la proposition précédente (croissance de l'intégrale), on a :

$$\int_E f_n d\mu \leq \int_E f_{n+1} d\mu \leq \int_E F d\mu$$

La suite $(\int_E f_n d\mu)_{n \geq 0}$ est croissante, donc $\lim_{n \to +\infty} \int_E f_n d\mu$ est correctement définie et :

$$\lim_{n \to +\infty} \int_E f_n d\mu \leq \int_E F d\mu$$

Il reste à prouver que $L = \int_E F d\mu \leq \lim_{n \to +\infty} \int_E f_n d\mu$.

Des fonctions mesurables aux parties mesurables Soit $h : E \to \mathbb{R}^+$ une fonction étagée telle que $h \leq F$. Il existe un entier N, des réels $b_1, \dots, b_N \geq 0$ et des ensembles disjoints $B_1, \dots, B_N \in \mathcal{E}$ tels que $h = \sum_{i=1}^N b_i \mathbb{1}_{B_i}$.
Pour tout $n \geq 0$ et tout $\epsilon > 0$, on pose :

$$A_{n,\epsilon} = \{x \in E : f_n(x) \geq (1 - \epsilon) h(x)\}$$

Limite croissante de parties mesurables : $\forall \epsilon > 0$, on a :

- La suite $(A_{n,\epsilon})_{n \geq 0}$ est croissante puisque la suite $(f_n)_{n \geq 0}$ est croissante.
- Prouvons que $\bigcup_{n \geq 0} A_{n,\epsilon} = E$.

En effet, clairement $\bigcup_{n \geq 0} A_{n,\epsilon} \subset E$. Et $\forall x \in E$, puisque $f_n(x) \to F(x)$ et que $F(x) \geq h(x)$, on a $x \in \bigcup_{n \geq 0} A_{n,\epsilon}$ puisque :

- 1^{er} cas : $h(x) = 0$. Alors $x \in A_{0,\epsilon}$.
- $2^{\text{ème}}$ cas : $h(x) > 0$. Mais alors $F(x) \geq h(x) > (1 - \epsilon) h(x)$, donc il existe un $n_0 \in \mathbb{N}$ tel que $f_{n_0}(x) > (1 - \epsilon) h(x)$.

On a donc $\forall i \in [\![1, N]\!]$ et $\forall \epsilon > 0$, $\lim_{n \to +\infty} \mu(B_i \cap A_{n,\epsilon}) \underset{\substack{\text{Par continuité} \\ \text{croissante des mesures}}}{=} \mu(B_i \cap E) = \mu(B_i)$.

Donc, pour tout $n \geq 0$ et tout $\epsilon > 0$:

$$\begin{aligned}
\int_E f_n d\mu &\geq \int_E f_n \mathbb{1}_{A_{n,\epsilon}} d\mu \quad (\text{car } f_n \geq 0) \\
&\geq \int_E (1 - \epsilon) h \mathbb{1}_{A_{n,\epsilon}} d\mu \quad (\text{définition de } A_{n,\epsilon}) \\
&\geq \int_E (1 - \epsilon) \left(\sum_{i=1}^N b_i \mathbb{1}_{B_i \cap A_{n,\epsilon}} \right) d\mu \\
&= (1 - \epsilon) \sum_{i=1}^N b_i \mu(B_i \cap A_{n,\epsilon})
\end{aligned}$$

En passant à la limite quand $n \to +\infty$:

$$\lim_{n\to+\infty} \int_E f_n d\mu \geq (1-\epsilon) \sum_{i=1}^{N} b_i \lim_{n\to+\infty} \mu(B_i \cap A_{n,\epsilon}) = (1-\epsilon) \sum_{i=1}^{N} b_i \mu(B_i) = (1-\epsilon) \int_E h d\mu$$

Par conséquent,

$$\lim_{n\to+\infty} \int_E f_n d\mu \geq (1-\epsilon) \int_E h d\mu$$

Et ceci étant vrai $\forall \epsilon > 0$, on a :

$$\lim_{n\to+\infty} \int_E f_n d\mu \geq \int_E h d\mu$$

Ceci étant vrai pour toutes les fonctions étagées positives h telles que $h \leq F$, on a :

$$\lim_{n\to+\infty} \int_E f_n d\mu \geq \sup_{h \leq F, h \text{ étagée}} \int_E h d\mu$$

Or, par définition de l'intégrale de Lebesgue pour une fonction mesurable positive :

$$\int_E F d\mu = \sup\{\int_E h d\mu : h \text{ étagée}, 0 \leq h \leq F\}$$

Donc :

$$\lim_{n\to+\infty} \int_E f_n d\mu \geq \int_E F d\mu$$

Ce qui conclut la preuve. □

Corollaire 1: Séries de fonctions mesurables positives (Beppo-Levi)

Soit $(E, \mathcal{E}, \mu)$ un espace mesuré et $(f_n)_{n\geq 0}$ une suite de fonctions mesurables positives définies sur E. Alors la suite $(g_n)_{n\geq 0}$ définie par $g_n = \sum_{k=0}^{n} f_k$ est une suite croissante de fonctions mesurables positives. Donc d'après le Théorème de Convergence Monotone :

$$\lim_{n\to\infty} \int_E g_n d\mu = \int_E \lim_{n\to\infty} g_n d\mu$$

Donc :

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

Proposition 5: Approximation par des fonctions étagées

Soit $(E, \mathcal{E})$ un espace mesurable et $f : E \to [0, +\infty]$ une fonction mesurable. En posant $\forall n \in \mathbb{N}$ et $\forall i \in \{0, \ldots, n2^n - 1\}$:

$$A_n = \{x \in E : f(x) \geq n\}$$

$$B_{n,i} = \left\{ x \in E : \frac{i}{2^n} \leq f(x) < \frac{i+1}{2^n} \right\}$$

et

$$f_n = \sum_{i=0}^{n2^n-1} \frac{i}{2^n} \mathbb{1}_{B_{n,i}} + n \mathbb{1}_{A_n}$$

On se donne une suite croissante $(f_n)_{n\geq 0}$ de fonctions étagées positives telle que $\forall x \in E$: $f_n(x) \to f(x)$.

Remarque

Cette construction est canonique et permet d'approcher n'importe quelle fonction mesurable positive par une suite croissante de fonctions étagées.

Exercice

Donner un exemple de suite $(f_n)_{n\geq 0}$ de fonctions mesurables (mais non croissante) telle que $\int_E f_n d\mu$ soit définie et :

$$\lim_{n\to\infty} \int_E f_n d\mu \neq \int_E \lim_{n\to\infty} f_n d\mu$$

Réponse : $E = \mathbb{R}$, $\mathcal{E} = \mathcal{B}(\mathbb{R})$ et $\mu = \lambda$ la mesure de Lebesgue. Soit $f_n = \mathbb{1}_{[n,n+1]}$. C'est une suite de fonctions mesurables qui converge (simplement) vers $F = 0$. Alors que $\forall n \geq 0$, $\int_{\mathbb{R}} f_n d\mu = 1$. Donc :

$$\lim_{n\to\infty} \int_{\mathbb{R}} f_n d\mu = 1 \neq 0 = \int_{\mathbb{R}} 0 d\mu$$

Remarque

Résultat d'approximation : Si $f : E \to [0, +\infty]$ est mesurable, il existe une suite croissante $(f_n)_{n\geq 0}$ de fonctions étagées positives telle que $(f_n)_{n\geq 0}$ converge simplement vers f.

Proposition 6: Linéarité de l'intégrale pour les fonctions positives

Soient $f, g : E \to [0, +\infty]$ deux fonctions mesurables et $\lambda > 0$. On a :

1. $\int_E (f+g) d\mu = \int_E f d\mu + \int_E g d\mu$
2. $\int_E (\lambda f) d\mu = \lambda \int_E f d\mu$

Preuve :

1. Pour la somme : Soient $f, g : E \to [0, +\infty]$ mesurables. D'après le résultat d'approximation, il existe 2 suites croissantes de fonctions étagées positives $(f_n)_{n\geq 0}$ et $(g_n)_{n\geq 0}$ telles que :

$$f_n \xrightarrow{C.S} f \quad \text{et} \quad g_n \xrightarrow{C.S} g$$

La suite $(f_n + g_n)_{n\geq 0}$ est une suite croissante de fonctions étagées positives telle que $f_n + g_n \xrightarrow{C.S} f + g$.

$$\int_E (f+g) d\mu \underset{\text{TCM}}{=} \int_E \lim_{n\to\infty} (f_n + g_n) d\mu \underset{\text{TCM}}{=} \lim_{n\to\infty} \int_E (f_n + g_n) d\mu$$

Or, pour les fonctions étagées positives (propriété déjà vue) :

$$= \lim_{n\to\infty} \left(\int_E f_n d\mu + \int_E g_n d\mu \right)$$

Par linéarité des limites :

$$= \lim_{n\to\infty} \int_E f_n d\mu + \lim_{n\to\infty} \int_E g_n d\mu \underset{\text{TCM}}{=} \int_E f d\mu + \int_E g d\mu$$

2. Pour le produit par un scalaire : Soit $\lambda > 0$. Si f est une fonction étagée positive, alors λf est une fonction étagée positive, donc $\int_E (\lambda f) d\mu$ est correctement définie et vaut $\lambda \int_E f d\mu$ (par définition pour les étagées). Puisque f est mesurable positive, il existe une suite $(f_n)_{n\geq 0}$ croissante de fonctions étagées positives qui converge simplement vers f. Clairement, les fonctions (λf_n) sont des fonctions étagées positives et la suite $(\lambda f_n)_{n\geq 0}$ est une suite croissante qui converge simplement vers λf. Donc d'après le TCM :

$$\int_E (\lambda f) d\mu \underset{\text{TCM}}{=} \lim_{n\to\infty} \int_E \lambda f_n d\mu \underset{\text{Prop étagées}}{=} \lim_{n\to\infty} \lambda \int_E f_n d\mu \underset{\text{limite dans } \overline{\mathbb{R}}}{=} \lambda \lim_{n\to\infty} \int_E f_n d\mu$$

$$\underset{\text{TCM}}{=} \lambda \int_E f d\mu$$

□

Exemple 4: Mesure de Dirac

Soit E un ensemble non vide, $a \in E$ et δ_a la mesure de Dirac en a sur $(E, \mathcal{P}(E))$. Soit $f : E \to \mathbb{R}$ une fonction mesurable positive.

- 1^{er} cas : f est une fonction étagée positive. Alors $\int_E f d\delta_a = f(a)$.
- $2^{\text{ème}}$ cas : f est une fonction mesurable positive. D'après la proposition précédente, il existe une suite croissante $(f_n)_{n\geq 0}$ de fonctions étagées positives ayant f comme limite (simple). $\forall n \geq 0, \quad \int_E f_n d\delta_a = f_n(a)$. Or, d'après le TCM :

$$\int_E f d\delta_a = \lim_{n\to\infty} \int_E f_n d\delta_a = \lim_{n\to\infty} f_n(a) = f(a)$$

Proposition 7: Inégalité de Markov

Soit $f : E \to [0, +\infty]$ une fonction mesurable. $\forall a > 0$:

$$\mu(\{f \geq a\}) = \mu(\{x : f(x) \geq a\}) \leq \frac{1}{a} \int_E f d\mu$$

Preuve :

Soit $f : E \to [0, +\infty]$ une fonction mesurable. $\forall a > 0$,

$$f \geq a \mathbb{1}_{\{f \geq a\}}$$

Donc, par croissance de l'intégrale :

$$\int_E f d\mu \geq \int_E a \mathbb{1}_{\{f\geq a\}} d\mu = a\mu(\{f \geq a\})$$

D'où :

$$\mu(\{f \geq a\}) \leq \frac{1}{a} \int_E f d\mu$$

□

Proposition 8: Nullité de l'intégrale

Soit $f : E \to [0, +\infty]$ une fonction mesurable.

1. $\int_E f d\mu = 0 \iff \mu(\{f > 0\}) = 0$

2. $\int_E f d\mu < +\infty \implies \mu(\{f = +\infty\}) = 0$

Preuve :

1. On a déjà montré que $\mu(\{f > 0\}) = 0 \implies \int_E f d\mu = 0$. Supposons que $\int_E f d\mu = 0$. On a $\{f > 0\} = \{x \in E : f(x) > 0\} = \bigcup_{m \geq 1}\{f \geq \frac{1}{m}\}$. Or $\forall m \geq 1$, d'après l'inégalité de Markov :

$$\mu(\{f \geq \frac{1}{m}\}) \leq m \int_E f d\mu = 0$$

Donc $\mu(\{f > 0\}) = \mu(\bigcup_{m \geq 1}\{f \geq \frac{1}{m}\}) \leq \sum_{m=1}^{\infty} \mu(\{f \geq \frac{1}{m}\}) = 0$.
2. On a $\{f = +\infty\} = \bigcap_{n \geq 1}\{f \geq n\}$. C'est la limite d'une suite décroissante d'ensembles de E. Or $\forall n \geq 1$, d'après l'inégalité de Markov :

$$\mu(\{f \geq n\}) \leq \frac{1}{n} \int_E f d\mu$$

Puisque $\int_E f d\mu < +\infty$, on a :

$$\frac{1}{n} \int_E f d\mu \xrightarrow{n \to \infty} 0$$

En particulier $\exists n_0 \in \mathbb{N}$ tel que $\mu(\{f \geq n_0\}) < +\infty$. Donc par continuité décroissante des mesures :

$$\mu(\{f = +\infty\}) = \mu\left(\bigcap_{n \in \mathbb{N}}\{f \geq n\}\right) = \lim_{n \to \infty} \mu(\{f \geq n\}) = 0$$

□

Lemme 1: Lemme de Fatou

Soit $(f_n)_{n \geq 0}$ une suite de fonctions mesurables positives. On a :

$$0 \leq \int_E (\liminf_n f_n) d\mu \leq \liminf_n \left(\int_E f_n d\mu\right)$$

Preuve :

Soit $(f_n)_{n \geq 0}$ une suite de fonctions mesurables positives définies sur E. En posant $\forall m \in \mathbb{N}$, $h_m = \inf_{n \geq m} f_n$. On a $h_m(x) = \inf_{n \geq m} f_n(x)$. On se donne une suite croissante de fonctions mesurables positives telle que $\forall n \geq 0$ et tout entier $n \geq m$, on a :

$$h_m \leq f_n$$

Donc $\forall m \geq 0$ et $\forall n \geq m$, on a :

$$\int_E h_m d\mu \leq \int_E f_n d\mu$$

Donc $\forall m \geq 0$, on a :

$$\int_E h_m d\mu \leq \inf_{n \geq m} \int_E f_n d\mu$$

Donc :

$$\lim_{m \to \infty} \int_E h_m d\mu \leq \lim_{m \to \infty} \inf_{n \geq m} \int_E f_n d\mu = \liminf_n \int_E f_n d\mu$$

Mais d'après le TCM, comme la suite $(h_m)_m$ est croissante positive :

$$\lim_{m \to \infty} \int_E h_m d\mu = \int_E \lim_{m \to \infty} h_m d\mu = \int_E \lim_{m \to \infty} (\inf_{n \geq m} f_n) d\mu = \int_E \liminf_n f_n d\mu$$

D'où :

$$0 \leq \int_E \liminf f_n d\mu \leq \liminf \int_E f_n d\mu$$

$\square$

Exercice

Soit $(f_n)_{n\geq 0}$ une suite de fonctions mesurables positives définies sur E qui converge simplement vers une fonction $f : E \to [0, +\infty]$ mesurable. Montrer que si $\sup_{n\geq 0} \int_E f_n d\mu < +\infty$, alors $\int_E f d\mu < +\infty$.

III. Égalité de fonctions μ-pp

Définition 5: Égalité presque partout

On dit que deux fonctions f, g définies sur E, mesurables et à valeurs dans $(\bar{\mathbb{R}}, \bar{\mathcal{B}})$, sont égales μ-presque partout (noté μ-pp) quand :

$$\mu(\{f \neq g\}) = \mu(\{x \in E : f(x) \neq g(x)\}) = 0$$

Proposition 9: Intégrale et égalité presque partout

Soient $f, g : E \to [0, +\infty]$ des fonctions mesurables. Si $f = g$ μ-pp, alors :

$$\int_E f d\mu = \int_E g d\mu$$

Preuve :

En notant $f \wedge g = \min(f, g)$ et $f \vee g = \max(f, g)$. On a que $f = g$ μ-pp $\implies f \wedge g = f \vee g$ μ-pp. Car $\{f \wedge g \neq f \vee g\} \subset \{f \neq g\}$, donc $0 \leq \mu(\{f \wedge g \neq f \vee g\}) \leq \mu(\{f \neq g\}) = 0$.
Or :

$$\begin{aligned}\int_E f \vee g d\mu &= \int_E (f \vee g - f \wedge g + f \wedge g) d\mu \\ &= \int_E (f \vee g - f \wedge g) d\mu + \int_E f \wedge g d\mu\end{aligned}$$

Car :

- $f \vee g - f \wedge g \geq 0$
- $\mu(\{f \vee g - f \wedge g > 0\}) = 0$

Ceci implique que $\int_E (f \vee g - f \wedge g) d\mu = 0$. Donc :

$$\int_E f \vee g d\mu = \int_E f \wedge g d\mu$$

Or $f \wedge g \leq f \leq f \vee g$ et $f \wedge g \leq g \leq f \vee g$. Donc par croissance de l'intégrale :

$$\begin{aligned}\int_E f \wedge g d\mu \leq \int_E f d\mu \leq \int_E f \vee g d\mu \\ \implies \int_E f d\mu = \int_E f \wedge g d\mu\end{aligned}$$

De même pour g, donc $\int_E f d\mu = \int_E g d\mu$. $\square$

Remarque

Pour décider si $f = g$ μ-pp, il n'est pas nécessaire de connaître f et g sur E tout entier. En effet, si il existe $A \in \mathcal{E}, B \in \mathcal{E}$ tels que $f\mathbb{1}_A = g\mathbb{1}_B$ avec $\mu(A^c) = 0 = \mu(B^c)$, alors $f = g$ μ-pp. En particulier, dès que $\mu(A^c) = 0$, on a $f = f\mathbb{1}_A$ μ-pp donc :

$$\int_E f d\mu = \int_E f\mathbb{1}_A d\mu$$

Donc pour calculer $\int_E f d\mu$, il n'est pas nécessaire de connaître f sur E mais il suffit de savoir que f est mesurable positive et connaître $f\mathbb{1}_A$.

Exercice

Soit $\mathcal{L}^0 = \{f : E \to \overline{\mathbb{R}} \text{ mesurables}\}$. Montrer que la relation binaire $\sim$ sur $\mathcal{L}^0$ définie par :

$$f \sim g \iff f = g \quad \mu\text{-pp}$$

est une relation d'équivalence.

Exercice

Soient $f, g : E \to \mathbb{R}$ mesurables telles que $f = 0$ μ-pp et $g = 0$ μ-pp. Prouver que $fg = 0$ μ-pp.
Réponse : $\{fg \neq 0\} = \{f \neq 0, f = g\} \cup \{f \neq 0, f \neq g\} \subset \{f \neq 0\} \cup \{g \neq 0\}$ (Note : la transcription de la réponse manuscrite est ici adaptée pour être logique mathématiquement, le manuscrit contient des ratures et inclusions). On a :

$$0 \leq \mu(\{fg \neq 0\}) \leq \mu(\{f \neq 0\}) + \mu(\{g \neq 0\}) = 0$$

Réponse (Relation d'équivalence)

- $\sim$ est réflexive (évident).
- $\sim$ est symétrique (évident).
- $\sim$ est transitive : Si $f \sim g$ et $g \sim h$, on a $\{f \neq h\} \subset \{f \neq g\} \cup \{g \neq h\}$. Donc $0 \leq \mu(\{f \neq h\}) \leq \mu(\{f \neq g\}) + \mu(\{g \neq h\}) = 0$.

IV. Fonctions intégrables

Définition 6: Partie positive et négative

Pour tout $x \in \mathbb{R}$, on appelle partie positive de x, $x^+ = \max(x, 0) = \begin{cases} x & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$. Pour tout $x \in \mathbb{R}$, on appelle partie négative de x, $x^- = \max(-x, 0) = \begin{cases} 0 & \text{si } x \geq 0 \\ -x & \text{sinon} \end{cases}$.

On a donc $\forall x \in \mathbb{R}$:

$$x = x^+ - x^-$$

$$|x| = x^+ + x^-$$

Pour toute fonction $f : E \to \mathbb{R}$, on note $f^+ : x \mapsto (f(x))^+ = \max(f(x), 0)$. Pour toute fonction $f : E \to \mathbb{R}$, on note $f^- : x \mapsto (f(x))^- = \max(-f(x), 0)$.
Si $f : (E, \mathcal{E}) \to \mathbb{R}$ est mesurable, f^+ et f^- sont mesurables et :

$$f = f^+ - f^-$$

$$|f| = f^+ + f^-$$

Définition 7: Fonction intégrable

Soit $(E, \mathcal{E}, \mu)$ un espace mesuré. Soit $f : E \to \overline{\mathbb{R}}$ une fonction mesurable. On dit que f est **intégrable par rapport à** μ (ou μ-intégrable) si :

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

Comme $0 \leq f^+ \leq |f|$ et $0 \leq f^- \leq |f|$, on a alors :

$$\int_E f^+ d\mu < +\infty \quad \text{et} \quad \int_E f^- d\mu < +\infty$$

Et on pose :

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

que l'on appelle intégrale de f par rapport à μ.

Remarque

1. Si $\int_E f^+ d\mu < +\infty$ et $\int_E f^- d\mu < +\infty$, alors $\int_E |f| d\mu < +\infty$.
2. Quand est-ce que $\int_E f d\mu$ est définie ?
   - Si $f : E \to \mathbb{R}$ est mesurable et de signe constant :
     - Si $f \geq 0$, c'est l'intégrale d'une fonction mesurable positive.
     - Si $f \leq 0$, alors en posant $\int_E f d\mu = -\int_E (-f) d\mu$.
   - Si $f : E \to \mathbb{R}$ n'est pas de signe constant, $\int_E f d\mu$ est définie uniquement si $\int_E |f| d\mu < +\infty$.

Exemple 5: Mesure de Dirac

Soit E un ensemble non vide et $a \in E$. Une fonction $f : E \to \mathbb{R}$ est intégrable par rapport à δ_a si $\int_E |f| d\delta_a = |f(a)| < +\infty$, donc si $f(a) \in \mathbb{R}$. Dans ce cas, i.e f est intégrable à δ_a, comme :

$$\int_E f^+ d\delta_a = f(a)^+ \quad \text{et} \quad \int_E f^- d\delta_a = f(a)^-$$

On a :

$$\int_E f d\delta_a = \int_E f^+ d\delta_a - \int_E f^- d\delta_a = f(a)^+ - f(a)^- = f(a)$$

Exercice

Soit $(E, \mathcal{E}, \mu)$ un espace mesuré. $f, g : E \to \overline{\mathbb{R}}$ mesurables telles que $|f| \leq |g|$ et g est μ-intégrable. Montrer que f est μ-intégrable.

Proposition 10: Propriétés de l'intégrale

Soit $(E, \mathcal{E}, \mu)$ un espace mesuré et $f, g : E \to \mathbb{R}$ des fonctions mesurables.

1. Si f est intégrable et $\lambda \in \mathbb{R}$, alors λf est intégrable et
$$\int_E \lambda f d\mu = \lambda \int_E f d\mu$$
2. Si f et g sont intégrables alors $f + g$ est intégrable et
$$\int_E (f+g) d\mu = \int_E f d\mu + \int_E g d\mu$$
3. Si f est intégrable, $\left|\int_E f d\mu\right| \leq \int_E |f| d\mu$.
4. Si f est intégrable et $f = g$ μ-pp, alors g est intégrable et $\int_E f d\mu = \int_E g d\mu$.
5. Si f est intégrable et $f \geq 0$ μ-pp, alors $\int_E f d\mu \geq 0$.
6. Si f et g sont intégrables et $f \geq g$ μ-pp, alors $\int_E f d\mu \geq \int_E g d\mu$.

Preuve :

1. On a $\int_E |\lambda f| d\mu = \int_E |\lambda||f| d\mu = |\lambda| \int_E |f| d\mu < +\infty$ dès que f est intégrable. Si $\lambda \geq 0$, $(\lambda f)^+ = \lambda f^+$ et $(\lambda f)^- = \lambda f^-$. Donc :
$$\int_E \lambda f d\mu = \int_E (\lambda f)^+ d\mu - \int_E (\lambda f)^- d\mu = \int_E \lambda f^+ d\mu - \int_E \lambda f^- d\mu$$
$$= \lambda \left(\int_E f^+ d\mu - \lambda \int_E f^- d\mu \right) = \lambda \left(\int_E f^+ d\mu - \int_E f^- d\mu \right) = \lambda \int_E f d\mu$$
Si $\lambda < 0$, $(\lambda f)^+ = -\lambda f^-$ et $(\lambda f)^- = -\lambda f^+$. Donc :
$$\int_E \lambda f d\mu = \int_E (\lambda f)^+ d\mu - \int_E (\lambda f)^- d\mu$$
$$= \int_E (-\lambda f^-) d\mu - \int_E (-\lambda f^+) d\mu$$
$$= \int_E -\lambda f^- d\mu - \int_E -\lambda f^+ d\mu$$
$$= -\lambda \int_E f^- d\mu + \lambda \int_E f^+ d\mu = \lambda \left(\int_E f^+ d\mu - \int_E f^- d\mu \right) = \lambda \int_E f d\mu$$
2. Soient $f, g : E \to \mathbb{R}$ mesurables et intégrables. Puisque $|f+g| \leq |f| + |g|$, on a $\int_E |f+g| d\mu \leq \int_E |f| d\mu + \int_E |g| d\mu < +\infty$. Donc $f + g$ est intégrable. On n'a pas $(f+g)^+ = f^+ + g^+$ ni $(f+g)^- = f^- + g^-$. Mais $f + g = (f+g)^+ - (f+g)^-$ et $f + g = f^+ - f^- + g^+ - g^-$. Donc :
$$(f+g)^+ - (f+g)^- = f^+ - f^- + g^+ - g^-$$
$$\iff (f+g)^+ + f^- + g^- = f^+ + g^+ + (f+g)^-$$

Donc :

$$\int_E ((f+g)^+ + f^- + g^-)d\mu = \int_E (f^+ + g^+ + (f+g)^-)d\mu$$

Donc :

$$\int_E (f+g)^+ d\mu + \int_E f^- d\mu + \int_E g^- d\mu = \int_E f^+ d\mu + \int_E g^+ d\mu + \int_E (f+g)^- d\mu$$

f, g et $f+g$ étant intégrables, ces quantités sont des nombres réels. Donc :

$$\int_E (f+g)^+ d\mu - \int_E (f+g)^- d\mu = \left(\int_E f^+ d\mu - \int_E f^- d\mu\right) + \left(\int_E g^+ d\mu - \int_E g^- d\mu\right)$$

D'où :

$$\int_E (f+g)d\mu = \int_E f d\mu + \int_E g d\mu$$

3. Si f est intégrable :

$$\begin{aligned} \left|\int_E f d\mu\right| = \left|\int_E f^+ d\mu - \int_E f^- d\mu\right| &\leq \left|\int_E f^+ d\mu\right| + \left|\int_E f^- d\mu\right| \\ &= \int_E f^+ d\mu + \int_E f^- d\mu = \int_E (f^+ + f^-)d\mu \\ &= \int_E |f| d\mu \end{aligned}$$

4. Soient f, g tels que f est intégrable et $f = g$ μ-pp. Alors $f^+ = g^+$ μ-pp et $f^- = g^-$ μ-pp. (car pour toute fonction $\psi : \mathbb{R} \to \mathbb{R}$ mesurable, si $f = g$ pp alors $\psi(f) = \psi(g)$ pp). Donc si f est intégrable :

$$\int_E g^+ d\mu = \int_E f^+ d\mu < +\infty \quad \text{et} \quad \int_E g^- d\mu = \int_E f^- d\mu < +\infty$$

Donc :

$$\int_E |g| d\mu < +\infty \quad \text{et} \quad \int_E f d\mu = \int_E g d\mu$$

5. Si f est intégrable et $f \geq 0$ μ-pp, on a $f = f^+$ μ-pp. Donc $\int_E f d\mu = \int_E f^+ d\mu \geq 0$.
6. Si f et g vérifient $f \geq g$ μ-pp alors :

$$\int_E (f-g)d\mu = \int_E ((f-g)+g)d\mu$$

(La transcription du manuscrit semble contenir une petite confusion dans la preuve 6, qui reprend la linéarité. Le chemin direct est d'utiliser la linéarité et la positivité)

$$\int_E f d\mu - \int_E g d\mu = \int_E (f-g)d\mu \geq 0 \quad (\text{car } f - g \geq 0 \text{ pp})$$

Donc $\int_E f d\mu \geq \int_E g d\mu$. □

Définition 8: Fonctions à valeurs complexes

Soit $(E, \mathcal{E}, \mu)$ un espace mesuré et $f : E \to \mathbb{C}$. On dit que f est mesurable si $\mathrm{Re}(f) : E \to \mathbb{R}$ et $\mathrm{Im}(f) : E \to \mathbb{R}$ le sont. On dit que f est intégrable par rapport à μ si les fonctions $\mathrm{Re}(f) : E \to \mathbb{R}$

et $\mathrm{Im}(f) : E \to \mathbb{R}$ sont intégrables par rapport à μ.

$$\int_E f d\mu = \int_E \mathrm{Re}(f) d\mu + i \int_E \mathrm{Im}(f) d\mu$$

Exercice (Propriété de transfert / Mesure image)

Soit $(E, \mathcal{E}, \mu)$ un espace mesuré et $(F, \mathcal{F})$ un espace mesurable. Et $f : E \to F$ une fonction mesurable. On a vu que :

$$\mu^f : B \in \mathcal{F} \to \mu(f^{-1}(B)) \quad \text{est une mesure sur } (F, \mathcal{F})$$

Montrer que $\forall g : F \to \mathbb{R}$ mesurable, g est μ^f-intégrable ssi $g \circ f$ est μ-intégrable et si l'une des conditions est vérifiée :

$$\int_F g d\mu^f = \int_E g \circ f d\mu$$

Réponse : Soient $f, g : E \to F$ mesurables et $\varphi : F \to G$ mesurable. Si $f = g$ μ-pp alors $\varphi(f) = \varphi(g)$ μ-pp. Puisque $\{\varphi(f) \neq \varphi(g)\} \subset \{f \neq g\}$. Donc $0 \leq \mu(|\varphi(f) \neq \varphi(g)|) \leq \mu(f \neq g) = 0$. En particulier $f = g$ μ-pp $\implies f^+ = g^+$ μ-pp et $f^- = g^-$ μ-pp.
Propriété de transfert (Réponse) : Soit $f : E \to F$ une fonction mesurable. On définit une mesure sur $(F, \mathcal{F})$ en posant sur $\mathcal{F}$:

$$\mu^f : B \in \mathcal{F} \to \mu^f(B) = \mu(f^{-1}(B))$$

Pour toute fonction $g : F \to \mathbb{R}$ mesurable : g est μ^f intégrable ssi $g \circ f$ est μ-intégrable. Et, si l'une ou l'autre des conditions est vérifiée :

$$\int_F g \circ f d\mu = \int_F g d\mu^f$$

Chapitre 3 : Espaces L^p

I. Définitions et inégalité de Hölder

Définition 9: Espaces $\mathcal{L}^p$

Soit $(E, \mathcal{E}, \mu)$ un espace mesuré et p un nombre réel, $p \geq 1$. On pose :

$$\mathcal{L}^p(E, \mathcal{E}, \mu) = \left\{ f : E \to \mathbb{R}, \text{ mesurable, tq } \int_E |f|^p d\mu < +\infty \right\}$$

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

Remarque

On a vu que la relation binaire $\sim$ définie sur l'ensemble des fonctions mesurables sur $(E, \mathcal{E}, \mu)$ à valeurs réelles :

$$f \sim g \iff f = g \ \mu\text{-p.p.}$$

est une relation d'équivalence.

Définition 10: Espace L^p

$\forall p \in [1, +\infty]$, on pose $L^p(E, \mathcal{E}, \mu) = \mathcal{L}^p(E, \mathcal{E}, \mu)/ \sim$.

Définition 11: Normes

$\forall p \in [1, +\infty]$ et toute fonction mesurable on pose :

- Si $p \in [1, +\infty[$, $\|f\|_p = \left(\int_E |f|^p d\mu\right)^{\frac{1}{p}}$
- $\|f\|_\infty = \inf\{A \in [0, +\infty] : |f| \leq A \ \mu\text{-p.p.}\}$

Exemple 6

Soit $f : \mathbb{R} \to \mathbb{R}$ définie par $f = \mathbb{1}_\mathbb{Q}$. On a que $\|f\|_\infty = 0$ puisque $|f| \leq 0$ λ-p.p.

Proposition 11: Supremum essentiel

Soit $f : E \to \mathbb{R}$ mesurable. On a que $\|f\|_\infty$ est le plus petit élément de $\overline{\mathbb{R}}^+$ qui majore $|f|$ μ-p.p. Autrement dit, si $A \in \overline{\mathbb{R}}^+$ vérifie $|f| \leq A$ μ-p.p. alors $A \geq \|f\|_\infty$.

Preuve : Soit $B = \{A \in [0, +\infty] \text{ tq } |f| \leq A \ \mu\text{-p.p.}\}$.

- **1er cas :** $B = \emptyset$. Alors $\|f\|_\infty = +\infty$ et il n'existe aucun réel A tel que $|f| \leq A$ p.p., donc $+\infty$ est bien le plus petit élément de $[0, +\infty]$ qui majore $|f|$ μ-p.p.
- **2ème cas :** $B \neq \emptyset$. On pose $D = \bigcap_{m \in B \cap \mathbb{Q}} \{|f| \leq m\}$. On a :

$$\mu(D^c) \leq \sum_{m \in B \cap \mathbb{Q}} \mu(\{|f| > m\}) = 0$$

Comme B est un intervalle de la forme $[z, +\infty[$ ou $]z, +\infty[$, on a que $B \cap \mathbb{Q}$ contient une suite (m_n) qui décroît vers $\inf B = \|f\|_\infty$. Or $\forall n \geq 0$ et $\forall x \in D$, on a $|f(x)| \leq m_n$, donc $\forall x \in D$, $|f(x)| \leq \|f\|_\infty$. Donc $|f| \leq \|f\|_\infty$ μ-p.p.
Par définition de $\|f\|_\infty$, il ne peut y avoir de réel M strictement plus petit que $\|f\|_\infty$ tel que $|f| \leq M$ μ-p.p.

Définition 12: Exposants conjugués

Soient $p, q \in [1, +\infty]$. On dit que p et q sont conjugués si $\frac{1}{p} + \frac{1}{q} = 1$.

Remarque

1 et $+\infty$ sont conjugués.

Proposition 12: Inégalité de Hölder

Soit $(E, \mathcal{E}, \mu)$ un espace mesuré. Soient $p, q \in [1, +\infty]$ conjugués et $f, g : E \to \mathbb{R}$ mesurables :

$$\int_E |fg| d\mu \leq \|f\|_p \|g\|_q$$

<u>Preuve :</u>

- **1er cas :** $\|f\|_p = 0$ ou $\|g\|_q = 0$. Supposons que $\|f\|_p = 0$, alors $f = 0$ μ-p.p., donc $fg = 0$ μ-p.p., donc $\int_E |fg| d\mu = 0 \leq \|f\|_p \|g\|_q$.
- **2ème cas :** $\|f\|_p > 0$ et $\|g\|_q > 0$. On peut supposer que $f \in L^p$ et $g \in L^q$.
  - **1er sous-cas :** Si $p = 1$ et $q = \infty$. On a $|g| \leq \|g\|_\infty$ μ-p.p. Donc $|fg| \leq |f| \|g\|_\infty$ μ-p.p. Donc $\int_E |fg| d\mu \leq \|g\|_\infty \int_E |f| d\mu \leq \|g\|_\infty \|f\|_1$.
  - **2ème sous-cas :** Si $p, q \in]1, +\infty[$. $\forall u, v \in \mathbb{R}^+$ et $\forall \alpha \in]0, 1[$, $u^\alpha v^{1-\alpha} \leq \alpha u + (1-\alpha) v$. Car :
    - $u = 0$ ou $v = 0$ (trivial).
    - $u > 0$ et $v > 0$, c'est alors une conséquence de la convexité de $x \mapsto e^x$.

    $\forall x \in E$, en prenant $\alpha = \frac{1}{p}$, $1 - \alpha = 1 - \frac{1}{p} = \frac{1}{q}$.

    $$u = \frac{|f(x)|^p}{\|f\|_p^p}, \quad v = \frac{|g(x)|^q}{\|g\|_q^q}$$

    Donc on a :
    $$u^\alpha v^{1-\alpha} = \frac{|f(x)|}{\|f\|_p} \frac{|g(x)|}{\|g\|_q} \leq \frac{1}{p} \frac{|f(x)|^p}{\|f\|_p^p} + \frac{1}{q} \frac{|g(x)|^q}{\|g\|_q^q}$$

    Donc :
    $$\int_E \frac{|f(x)||g(x)|}{\|f\|_p \|g\|_q} d\mu \leq \frac{1}{p} \int_E \frac{|f|^p}{\|f\|_p^p} d\mu + \frac{1}{q} \int_E \frac{|g|^q}{\|g\|_q^q} d\mu = \frac{1}{p} + \frac{1}{q} = 1$$

    Donc $\int_E |fg| d\mu \leq \|f\|_p \|g\|_q$.

Remarque

1. Si $f \in L^p$ et $g \in L^q$ avec p et q conjugués alors $fg \in L^1$.
2. Si $p = q = 2$, c'est l'inégalité de Cauchy-Schwarz.

3. Si μ est une mesure finie, en prenant $g = 1, \forall p \geq 1$, on a :

$$\int_E |f| d\mu = \int_E |f \cdot 1| d\mu \leq \left(\int_E |f|^p d\mu \right)^{\frac{1}{p}} \times (\mu(E))^{\frac{1}{q}}$$

Donc $\forall p \geq 1,\ L^p \subset L^1$. En particulier, $\forall r \geq 1, \forall p \geq 1$, en posant $r' = pr \geq r$. On a :

$$\int_E |f|^r d\mu \leq \left(\int_E |f|^{rp} d\mu \right)^{\frac{1}{p}} (\mu(E))^{\frac{1}{q}}$$

donc :

$$\left(\int_E |f|^r d\mu \right)^{\frac{1}{r}} = \|f\|_r \leq \|f\|_{r'} (\mu(E))^{\frac{1}{qr}}$$

donc si $r' \geq r$, on a $L^{r'} \subset L^r$. Si μ est une probabilité, $r \mapsto \|f\|_r$ est croissante.

II. Inégalité de Jensen

Proposition 13: Inégalité de Jensen

Soit $(E, \mathcal{E}, \mu)$ un espace de probabilité, $I \subset \mathbb{R}$ un intervalle ouvert, $\varphi : I \to \mathbb{R}$ une fonction convexe, et $F : E \to I$ une fonction mesurable. Si les fonctions F et $\varphi(F)$ sont μ-intégrables alors :

$$\varphi\left(\int_E F d\mu \right) \leq \int_E \varphi(F) d\mu$$

Preuve : On pose $E_\varphi = \{(a,b) \in \mathbb{R}^2 : \forall t \in I,\ \varphi(t) \geq at + b\}$. Et $\forall t \in I$, on a $\varphi(t) = \sup_{(a,b) \in E_\varphi}(at + b)$. Donc $\forall x \in E, a, b \in E_\varphi$, on a $\varphi(F(x)) \geq aF(x) + b$. Donc $\forall (a,b) \in E_\varphi$, on a :

$$\int_E \varphi(F) d\mu \geq a \int_E F d\mu + b$$

donc :

$$\int_E \varphi(F) d\mu \geq \sup_{(a,b) \in E_\varphi} \left(a \int_E F d\mu + b \right) = \varphi\left(\int_E F d\mu \right)$$

III. L'espace de Banach $L^p(E, \mathcal{E}, \mu)$

Proposition 14: Inégalité de Minkowski

Soit $p \in [1, +\infty]$. Pour toutes les fonctions $f, g : E \to \mathbb{R}$ mesurables :

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

Preuve :

- **1er cas :** $p = 1$. $\forall x \in E,\ |f(x) + g(x)| \leq |f(x)| + |g(x)|$ d'où $\|f + g\|_1 \leq \|f\|_1 + \|g\|_1$.
- **2ème cas :** $p = +\infty$. On peut supposer que $\|f\|_\infty < +\infty$ et $\|g\|_\infty < +\infty$. Alors $|f| \leq \|f\|_\infty$ et $|g| \leq \|g\|_\infty$ p.p. Donc comme $\forall x \in E,\ |f(x) + g(x)| \leq |f(x)| + |g(x)|$, on a $|f + g| \leq \|f\|_\infty + \|g\|_\infty$ μ-p.p. Donc $\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$.

- **3ème cas :** $1 < p < +\infty$. Si $\|f+g\|_p = 0$, l'inégalité est évidemment vraie. On va supposer que $\|f+g\|_p \neq 0$. On a :

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

Donc :

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

On applique Hölder avec $u = |f|$ ou $|g|$ et $v = |f+g|^{p-1}$:

$$\leq \left(\int_E |f|^p d\mu\right)^{\frac{1}{p}} \left(\int_E |f+g|^{(p-1)q} d\mu\right)^{\frac{1}{q}} + \left(\int_E |g|^p d\mu\right)^{\frac{1}{p}} \left(\int_E |f+g|^{(p-1)q} d\mu\right)^{\frac{1}{q}}$$

Or $(p-1)q = (p-1)\frac{p}{p-1} = p$. Donc :

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

Donc :

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

Or $p - \frac{p}{q} = p(1-\frac{1}{q}) = p\frac{1}{p} = 1$. D'où $\|f+g\|_p \leq \|f\|_p + \|g\|_p$.

Remarque

$\forall p \in [1, +\infty]$, tous $f, g \in L^p$ et $\lambda \in \mathbb{R}$, on a $f + \lambda g \in L^p$. Donc L^p est un SEV de L^0.

$$\mathcal{L}^0 = \{f : E \to \mathbb{R} \text{ mesurables}\} \text{ est un espace vectoriel}$$

$$L^0 = \mathcal{L}^0 / \sim$$

Proposition 15: Espace de Banach

$\forall p \in [1, +\infty]$, l'espace L^p est un espace vectoriel normé complet (un espace de Banach).

Preuve :

- **Espace vectoriel :** ✓(Déjà vu)
- **Norme :**
  - Inégalité triangulaire : $\iff$ Inégalité de Minkowski.
  - Homogénéité : Si $p \in [1, +\infty[$, $f \in L^p$ et $\lambda \in \mathbb{R}$:

$$\|\lambda f\|_p = \left(\int |\lambda f|^p d\mu\right)^{\frac{1}{p}} = |\lambda| \|f\|_p$$

    Si $p = +\infty$, exercice.
  - Séparation : Soit $f \in L^p$. $\|f\|_p = 0 \implies f = 0$ μ-p.p.
- **Complet :**
  1er cas : $p \in [1, +\infty[$. Soit $(f_n)_{n\geq 0}$ une suite de Cauchy d'éléments de L^p. Pour montrer que $(f_n)_{n\geq 0}$ converge, il suffit de montrer que $(f_n)_{n\geq 0}$ admet une sous-suite convergente. On peut extraire de $(f_n)_{n\geq 0}$ une sous-suite $(g_n)_{n\geq 0}$ telle que :

$$\forall n \geq 0 \text{ et } \forall m \geq n, \quad \|g_n - g_m\|_p \leq \frac{1}{2^n}$$

En particulier, $\forall n \geq 0$, on a $\|g_n - g_{n+1}\|_p \leq \frac{1}{2^n}$.
On considère la série de fonctions positives $\sum |g_{n+1} - g_n|$.

continuité de $x \mapsto x^{\frac{1}{p}}$

$$\begin{aligned}
\left(\int_E \left(\sum_{n=1}^{\infty} |g_{n+1} - g_n|\right)^p d\mu\right)^{\frac{1}{p}} &\overset{\text{C.S}}{=} \left(\int_E \left(\lim_{N\to\infty} \sum_{n=1}^{N} |g_{n+1} - g_n|\right)^p d\mu\right)^{\frac{1}{p}} \\
&= \left(\int_E \lim_{N\to\infty} \left(\sum_{n=1}^{N} |g_{n+1} - g_n|\right)^p d\mu\right)^{\frac{1}{p}} \\
&\overset{\text{TCM}}{=} \left(\lim_{N\to\infty} \int_E \left(\sum_{n=1}^{N} |g_{n+1} - g_n|\right)^p d\mu\right)^{\frac{1}{p}} \\
&= \lim_{N\to\infty} \boxed{\left(\int_E \left(\sum_{n=1}^{N} |g_{n+1} - g_n|\right)^p d\mu\right)^{\frac{1}{p}}}
\end{aligned}$$

$$= \left\|\sum_{n=1}^{N} |g_{n+1} - g_n|\right\|_p$$

$$\begin{aligned}
&= \lim_{N\to\infty} \left(\left\|\sum_{n=1}^{N} |g_{n+1} - g_n|\right\|_p\right) \\
&\leq \lim_{N\to\infty} \left(\sum_{n=1}^{N} \|g_{n+1} - g_n\|_p\right) \\
&< +\infty
\end{aligned}$$

Inégalité de Minkowski

Donc la série de terme général $|g_{n+1} - g_n|$ est absolument convergente μ-p.p. Donc en posant $h = g_1 + \sum_{n=1}^{\infty}(g_{n+1} - g_n)$ (quitte à prendre $h = 0$ sur la partie mesurable de E où la série de terme général $g_{n+1} - g_n$ n'est pas absolument convergente), on définit une fonction mesurable h telle que $h = \lim_{n\to\infty} g_n$ μ-p.p.

1. **Est-ce que** $h \in L^p$ **?** On a :

$$\int_E |h|^p d\mu = \int_E \liminf |g_n|^p d\mu \leq \liminf \int_E |g_n|^p d\mu < +\infty$$

(car (g_n) est une suite de Cauchy dans $L^p \implies (\|g_n\|_p)_n$ est bornée). Donc $h \in L^p$.

2. **Est-ce que** $g_n \xrightarrow{L^p} h$ **?** Autrement dit, $\|g_n - h\|_p \xrightarrow{n\to\infty} 0$.

$$\int_E |g_n - h|^p d\mu = \int_E \liminf_{N\to\infty} |g_n - g_N|^p d\mu \leq \liminf_{N\to\infty} \int_E |g_n - g_N|^p d\mu$$

$$\leq \left(\frac{1}{2^{n_p}}\right) \xrightarrow{n\to\infty} 0$$

<u>**2ème cas :**</u> $p = +\infty$. Soit $(f_n)_{n\geq 0}$ une suite de Cauchy de L^∞. On a :

$$\forall \epsilon > 0, \exists N_\epsilon,\ \forall n, p \geq N_\epsilon, \quad \|f_n - f_p\|_\infty < \epsilon$$

Donc $\forall \epsilon > 0, \exists N_\epsilon,\ \forall n, p \geq N_\epsilon,\ \exists A_{n,p,\epsilon} \in \mathcal{E}$ telle que $\forall x \notin A_{n,p,\epsilon}$, $|f_n(x) - f_p(x)| < \epsilon$ et $\mu(A_{n,p,\epsilon}) = 0$.

En prenant :

$$A = \bigcap_{\epsilon \in \mathbb{Q}_+^*} \bigcap_{n \geq N_\epsilon} \bigcap_{p \geq N_\epsilon} A^c_{n,p,\epsilon} \in \mathcal{E}$$

On a $\mu(A^c) = 0$ car A^c est une réunion dénombrable de parties mesurables de mesure nulle.
On peut modifier la suite $(f_n)_{n\geq 0}$ en posant $\tilde{f}_n(x) = 0, \forall x \in A^c, \forall n \geq 0$. $\forall x \in A$, on a : $\forall \epsilon > 0, \exists N_\epsilon, \forall n, p \geq N_\epsilon,\ |\tilde{f}_n(x) - \tilde{f}_p(x)| < \epsilon$. Autrement dit, $\forall x \in A$, $(\tilde{f}_n(x))_{n\geq 0}$ est une suite de Cauchy d'éléments de $\mathbb{R}$, donc convergente, et on note $f(x) = \lim_{n\to\infty} \tilde{f}_n(x)$.
1 - Est-ce que $\|f_n - f\|_\infty \xrightarrow{n\to\infty} 0$ **?** $\forall n, p \in \mathbb{N}$.

$$|\tilde{f}_n(x) - \tilde{f}_p(x)| \leq \|\tilde{f}_n - \tilde{f}_p\|_\infty \quad \mu\text{-p.p.}$$

Donc :

$$|\tilde{f}_n(x) - f(x)| = \lim_{p\to\infty} |\tilde{f}_n(x) - \tilde{f}_p(x)| \leq \sup_{m\geq n} \|f_n - f_m\|_\infty \quad \mu\text{-p.p.}$$

Donc $\|f_n - f\|_\infty \leq \sup_{m\geq n} \|f_n - f_m\|_\infty \xrightarrow{n\to\infty} 0$.
2 - Est-ce que $f \in L^\infty$ **?** $\forall n \in \mathbb{N}, |f(x)| \leq |f(x) - f_n(x)| + |f_n(x)| \leq \|f - f_n\|_\infty + \|f_n\|_\infty$ μ-p.p.
Donc comme $\|f - f_n\|_\infty \to 0$ et que $(f_n)_{n\geq 0}$ est de Cauchy dans L^p (ici L^∞), alors $(\|f_n\|_\infty)_{n\geq 0}$ est bornée et donc f est bien bornée. Donc $f \in L^\infty$, ce qui conclut la preuve.

Proposition 16: Extraction de sous-suite

$\forall p \in [1, +\infty]$, si $(f_n)_{n\geq 0}$ est une suite de Cauchy de L^p, alors $(f_n)_{n\geq 0}$ admet une sous-suite qui converge μ-p.p.

Définition 13: Espaces de suites ℓ^p

Dans le cas particulier où $E = \mathbb{N}$ et $\mu =$ la mesure de comptage, on note $\ell^p = L^p(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu) = \{u \in \mathbb{R}^{\mathbb{N}} : \sum_{n=0}^{\infty} |u_n|^p < +\infty\}$ et :

$$\ell^\infty = L^\infty(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu) = \{u \in \mathbb{R}^{\mathbb{N}} : u \text{ est une suite bornée}\}$$

Proposition 17: Espace de Hilbert L^2

L'espace $L^2(E, \mathcal{E}, \mu)$ muni du produit scalaire :

$$\langle f, g\rangle = \int_E fg d\mu$$

est un espace de Hilbert.

Théorème 2: Représentation de Riesz

Si H est un espace de Hilbert réel, alors pour toute forme linéaire continue ϕ définie sur H, il existe $v \in H$ tq $\forall u \in H,\ \phi(u) = \langle u, v\rangle$.

IV. Le Théorème de Radon-Nikodym

Définition 14: Absolue continuité et singularité

Soit $(E, \mathcal{E})$ un espace mesurable et μ et ν 2 mesures définies sur $\mathcal{E}$.

1. On dit que ν est **absolument continue** par rapport à μ (noté $\nu \ll \mu$) si :
$$\forall A \in \mathcal{E}, \quad \mu(A) = 0 \implies \nu(A) = 0$$
2. On dit que ν et μ sont **étrangères** (noté $\nu \perp \mu$) s'il existe $N \in \mathcal{E}$ tel que $\mu(N) = 0$ et $\nu(N^c) = 0$.

Remarque

Si $(E, \mathcal{E}, \mu)$ est un espace mesuré et $f : E \to [0, +\infty]$ mesurable, alors on définit une mesure ν sur $\mathcal{E}$ en posant :
$$\forall A \in \mathcal{E}, \quad \nu(A) = \int_E f \mathbb{1}_A d\mu$$
et ν est absolument continue par rapport à μ.

Théorème 3: de Radon-Nikodym

Soient $(E, \mathcal{E})$ un espace mesurable et μ et ν deux mesures σ-finies sur $(E, \mathcal{E})$. Il existe un unique couple (ν_a, ν_s) de mesures σ-finies sur $(E, \mathcal{E})$ tq :

1. $\nu = \nu_a + \nu_s$
2. ν_a est absolument continue par rapport à μ ($\nu_a \ll \mu$) et ν_s et μ sont étrangères ($\nu_s \perp \mu$).

De plus, il existe une fonction mesurable positive g définie sur E tq :
$$\forall A \in \mathcal{E}, \quad \nu_a(A) = \int_E g \mathbb{1}_A d\mu$$
et g est unique à un ensemble de μ-mesure près.

Preuve : **1er cas :** On commence par supposer que μ et ν sont des mesures finies.

1er sous-cas : On suppose que pour toute fonction $f : E \to [0, +\infty]$ mesurable, on a $\int_E f d\nu \leq \int_E f d\mu$ (c'est-à-dire $\nu \leq \mu$). On définit sur $L^2(E, \mathcal{E}, \mu)$ la fonction ϕ suivante :
$$f \mapsto \int_E f d\nu$$
ϕ est correctement définie car μ étant une mesure finie, $L^2(\mu) \subset L^1(\mu)$ et $\forall f \in L^1(\mu)$, $\int_E |f| d\nu \leq \int_E |f| d\mu < +\infty$. L'application ϕ est linéaire, donc ϕ est une forme linéaire et pour tout $f \in L^2(E, \mathcal{E}, \mu)$, d'après l'inégalité de Hölder :
$$|\phi(f)| = \left| \int_E f d\nu \right| \leq \int_E |f| d\nu \leq \left(\int_E f^2 d\nu \right)^{\frac{1}{2}} (\nu(E))^{\frac{1}{2}}$$
$$\leq \left(\int_E f^2 d\mu \right)^{\frac{1}{2}} (\nu(E))^{\frac{1}{2}} \leq \|f\|_2 (\nu(E))^{\frac{1}{2}}$$

Donc ϕ est une forme linéaire continue définie sur l'espace de Hilbert $L^2(E, \mathcal{E}, \mu)$. Donc d'après le théorème de Riesz, il existe $h \in L^2(E, \mathcal{E}, \mu)$ tel que :

$$\forall f \in L^2(\mu), \quad \phi(f) = \langle f, h \rangle = \int_E f h d\mu$$

En particulier, $\forall A \in \mathcal{E}$, en prenant $f = \mathbb{1}_A$, on a :

$$\nu(A) = \int_E \mathbb{1}_A d\nu = \phi(f) = \int_E \mathbb{1}_A h d\mu$$

donc $\nu \ll \mu$. On va maintenant montrer que $0 \leq h \leq 1$ μ-p.p. En effet, $\forall \varepsilon > 0$:

$$\mu(\{h > 1 + \varepsilon\}) \geq \nu(\{h > 1 + \varepsilon\}) = \int_E \mathbb{1}_{\{h > 1+\varepsilon\}} h d\mu \geq (1 + \varepsilon)\mu(\{h > 1 + \varepsilon\})$$

donc $\forall \varepsilon > 0$, $\mu(\{h > 1 + \varepsilon\}) = 0$. De même pour $\mu(\{h < -\varepsilon\}) = 0$. Donc $0 \leq h \leq 1$ μ-p.p.

2ème sous-cas : On ne suppose plus que $\nu \leq \mu$. Cependant, on a encore que $\nu \leq \nu + \mu$ puisque si $f : E \to [0, +\infty]$ est mesurable :

$$\int_E f d(\mu + \nu) = \int_E f d\nu + \int_E f d\mu \geq \int_E f d\nu$$

Donc d'après ce qui précède, il existe une fonction mesurable h telle que $\forall x \in E, h(x) \in [0, 1]$ et pour toute fonction $f \in L^2(\mu + \nu)$:

$$\int_E f d\nu = \int_E f h d(\mu + \nu) = \int_E f h d\mu + \int_E f h d\nu$$

Donc :

$$\int_E f(1 - h) d\nu = \int_E f h d\mu \quad (*)$$

En particulier, l'égalité précédente est vraie pour toutes les fonctions étagées positives f et, comme $\forall x \in E, h(x) \in [0, 1]$, elle est vraie pour toutes les fonctions mesurables positives comme conséquence du TCM.

On pose $N = \{x \in E : h(x) = 1\} \in \mathcal{E}$. L'égalité $(*)$ avec $f = \mathbb{1}_N$ donne :

$$\int_E \mathbb{1}_N (1 - h) d\nu = \int_E \mathbb{1}_N 0 d\nu = 0$$

$$\int_E f h d\mu = \int_E \mathbb{1}_N h d\mu = \int_E \mathbb{1}_N 1 d\mu = \mu(N)$$

Donc $\mu(N) = 0$. La mesure ν_s définie sur $\mathcal{E}$ par $\nu_s(A) = \nu(A \cap N)$ est donc étrangère à μ. Pour toute fonction $f : E \to [0, +\infty[$ mesurable, en appliquant $(*)$ à $\frac{f}{1-h}\mathbb{1}_{N^c}$:

$$\int_E \frac{f}{1 - h} \mathbb{1}_{N^c} (1 - h) d\nu = \int_E f \mathbb{1}_{N^c} d\nu = \int_E \frac{f}{1 - h} \mathbb{1}_{N^c} h d\mu$$

$$= \int_E f g d\mu \quad \text{avec } g = \frac{h}{1 - h} \mathbb{1}_{N^c}$$

On définit une mesure sur $\mathcal{E}$ en posant $\nu_a(A) = \nu(A \cap N^c)$. On vient d'établir que pour toute fonction $f : E \to [0, +\infty[$ mesurable :

$$\int_E f \mathbb{1}_{N^c} d\nu = \int_E f g d\mu$$

En particulier pour tout $A \in \mathcal{E}$, en prenant $f = \mathbb{1}_A$:

$$\int_E \mathbb{1}_A \mathbb{1}_{N^c} d\nu = \nu(A \cap N^c) = \nu_a(A) = \int_E \mathbb{1}_A g d\mu$$

Donc ν_a est bien absolument continue par rapport à μ et $\nu = \nu_a + \nu_s$.

Unicité : Supposons qu'il existe une autre décomposition $(\widetilde{\nu_a}, \widetilde{\nu_s})$ de ν avec les mêmes propriétés.

$$\nu_a(A) + \nu_s(A) = \widetilde{\nu_a}(A) + \widetilde{\nu_s}(A)$$

$$\nu_a(A) - \widetilde{\nu_a}(A) = \widetilde{\nu_s}(A) - \nu_s(A)$$

Or $\nu_a \ll \mu$ et $\widetilde{\nu_a} \ll \mu \implies \nu_a - \widetilde{\nu_a} \ll \mu$. Et $\nu_s \perp \mu$ et $\widetilde{\nu_s} \perp \mu \implies \widetilde{\nu_s} - \nu_s \perp \mu$. Le seul moyen d'être à la fois absolument continu et étranger à μ est d'être la mesure nulle. Donc $\nu_a = \widetilde{\nu_a}$ et $\nu_s = \widetilde{\nu_s}$.

Finalement, $g = \tilde{g}$ μ-p.p.

$$\int_E g \mathbb{1}_{\{g > \tilde{g}\}} d\mu = \nu_a(\{g > \tilde{g}\})$$

$$\int_E \tilde{g} \mathbb{1}_{\{\tilde{g} > g\}} d\mu = \widetilde{\nu_a}(\{\tilde{g} > g\})$$

Mais $\nu_a(\{g > \tilde{g}\}) = \widetilde{\nu_a}(\{g > \tilde{g}\})$. Donc $\int_E (g - \tilde{g}) \mathbb{1}_{\{g > \tilde{g}\}} d\mu = 0$. Donc $\mu(\{g > \tilde{g}\}) = 0$.

Chapitre 4 : Mesures Produits

I. Généralités sur les espaces produits et les tribus produits

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

Définition 15: Tribu produit

On appelle tribu produit la tribu sur $E_1 \times E_2$, notée $\mathcal{E}_1 \otimes \mathcal{E}_2$, définie par :

$$\mathcal{E}_1 \otimes \mathcal{E}_2 = \sigma\left(\{B_1 \times B_2,\ B_1 \in \mathcal{E}_1,\ B_2 \in \mathcal{E}_2\}\right)$$

On appelle **pavé** toute partie $B_1 \times B_2$ de $E_1 \times E_2$ où $B_1 \in \mathcal{E}_1$ et $B_2 \in \mathcal{E}_2$.

Remarque

L'intersection de 2 pavés est un pavé.
Soient $A_1, B_1 \in \mathcal{E}_1$ et $A_2, B_2 \in \mathcal{E}_2$.

$$\begin{aligned}(A_1 \times A_2) \cap (B_1 \times B_2) &= \{(x_1, x_2) : x_1 \in A_1 \text{ et } x_1 \in B_1,\ x_2 \in A_2 \text{ et } x_2 \in B_2\} \\ &= (A_1 \cap B_1) \times (A_2 \cap B_2)\end{aligned}$$

Proposition 18: Propriété des projections

La tribu $\mathcal{E}_1 \otimes \mathcal{E}_2$ est la plus petite tribu sur $E_1 \times E_2$ pour laquelle les projections :

$$\begin{aligned} p_1 : E_1 \times E_2 &\to E_1 \\ (x_1, x_2) &\mapsto x_1 \end{aligned} \quad \text{et} \quad \begin{aligned} p_2 : E_1 \times E_2 &\to E_2 \\ (x_1, x_2) &\mapsto x_2 \end{aligned}$$

sont mesurables.

Preuve :

- p_1 est $\mathcal{E}_1 \otimes \mathcal{E}_2 - \mathcal{E}_1$ mesurable car $\forall A_1 \in \mathcal{E}_1$, on a :

$$p_1^{-1}(A_1) = A_1 \times E_2 \in \mathcal{E}_1 \otimes \mathcal{E}_2$$

- Soit $\mathcal{F}$ une tribu sur $E_1 \times E_2$ telle que p_1 et p_2 sont $\mathcal{F} - \mathcal{E}_1$ mesurables et $\mathcal{F} - \mathcal{E}_2$ mesurables.
On a alors $\forall A_1 \in \mathcal{E}_1$, que $p_1^{-1}(A_1) = A_1 \times E_2 \in \mathcal{F}$ et $\forall A_2 \in \mathcal{E}_2$, $p_2^{-1}(A_2) = E_1 \times A_2 \in \mathcal{F}$.
Et comme $\mathcal{F}$ est une tribu :

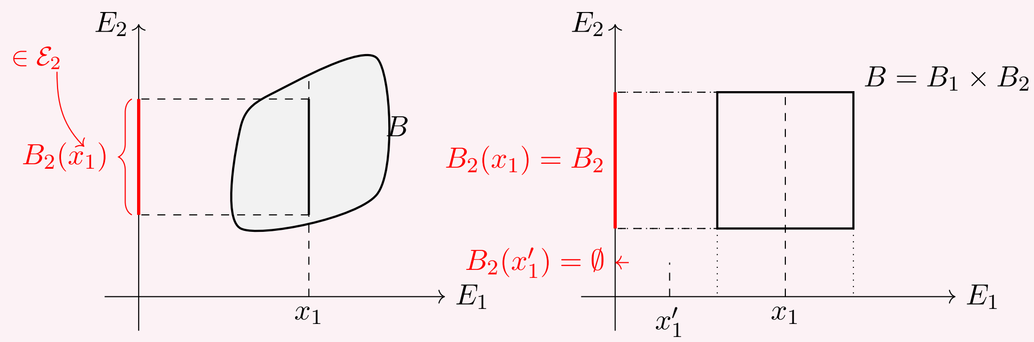
$$(A_1 \times E_2) \cap (E_1 \times A_2) = A_1 \times A_2 \in \mathcal{F}$$

Donc $\mathcal{F}$ contient tous les pavés, donc $\mathcal{E}_1 \otimes \mathcal{E}_2 \subset \mathcal{F}$.

Proposition 19: Propriété de section

Tout $B \in \mathcal{E}_1 \otimes \mathcal{E}_2$ vérifie la propriété suivante :

- $\forall x_1 \in E_1, \quad B_2(x_1) = \{x_2 \in E_2 : (x_1, x_2) \in B\} \in \mathcal{E}_2$
- $\forall x_2 \in E_2, \quad B_1(x_2) = \{x_1 \in E_1 : (x_1, x_2) \in B\} \in \mathcal{E}_1$



Preuve : On note $\mathcal{D}$ la classe des parties de $E_1 \times E_2$ ayant la propriété suivante :

$$B \in \mathcal{D} \iff \forall x_1 \in E_1,\ B_2(x_1) = \{x_2 \in E_2 : (x_1, x_2) \in B\} \in \mathcal{E}_2$$

On montre que $\mathcal{D}$ est une tribu sur $E_1 \times E_2$.

- $E_1 \times E_2 \in \mathcal{D}$. En effet $\forall x_1 \in E_1$, $(E_1 \times E_2)_2(x_1) = E_2 \in \mathcal{E}_2$.
- Soit $A \in \mathcal{D}$.
$$\forall x_1 \in E_1,\ (A^c)_2(x_1) = \{x_2 \in E_2 : (x_1, x_2) \in A^c\}$$
$$= \{x_2 \in E_2 : (x_1, x_2) \notin A\} = (A_2(x_1))^c \in \mathcal{E}_2 \quad \text{puisque } \mathcal{E}_2 \text{ est une tribu.}$$
Donc $\mathcal{D}$ est stable par passage au complémentaire.
- Soit $(A_n)_{n\geq 0}$ une suite d'éléments de $\mathcal{D}$.
$$\forall x_1 \in E_1,\ (\bigcup_{n\geq 0} A_n)_2(x_1) = \{x_2 \in E_2 : (x_1, x_2) \in \bigcup_{n\geq 0} A_n\}$$
$$= \bigcup_{n\geq 0} \{x_2 \in E_2 : (x_1, x_2) \in A_n\} = \bigcup_{n\geq 0} (A_n)_2(x_1) \in \mathcal{E}_2$$
Donc $\mathcal{D}$ est stable par union dénombrable, donc $\mathcal{D}$ est une tribu.
- $\mathcal{D}$ contient les pavés.
$$\forall (B_1, B_2) \in E_1 \times E_2,$$
$$\forall x_1 \in E_1, \quad (B_1 \times B_2)_2(x_1) = \begin{cases} \emptyset & \text{si } x_1 \notin B_1 \\ B_2 & \text{si } x_1 \in B_1 \end{cases} \in \mathcal{E}_2$$

Donc $\mathcal{D}$ est une tribu qui contient tous les pavés, donc elle contient $\mathcal{E}_1 \otimes \mathcal{E}_2$.

$$\forall B \in \mathcal{E}_1 \otimes \mathcal{E}_2,\ \forall x_1 \in E_1,\ B_2(x_1) \in \mathcal{E}_2.$$

II. Construction de la mesure produit

Proposition 20: Rappel

$\forall C \in \mathcal{C}\ \pi(\mathcal{C}) = \mathcal{P}(\mathcal{C})$ et si deux mesures σ-finies coïncident sur un π-système alors elles sont égales.

Proposition 21: Construction des mesures produits

Soient $(E_1, \mathcal{E}_1, \mu_1)$ et $(E_2, \mathcal{E}_2, \mu_2)$ deux espaces mesurés où μ_1 et μ_2 sont des mesures σ-finies.

1. $\forall A \in \mathcal{E}_1 \otimes \mathcal{E}_2$, les applications :
$$\begin{aligned} g_1 : E_1 &\to [0, +\infty] \\ x_1 &\mapsto \mu_2(A_2(x_1)) \end{aligned} \qquad \text{et} \qquad \begin{aligned} g_2 : E_2 &\to [0, +\infty] \\ x_2 &\mapsto \mu_1(A_1(x_2)) \end{aligned}$$
sont mesurables.
2. Les fonctions m et m' définies sur $\mathcal{E}_1 \otimes \mathcal{E}_2$ par :
$$m(A) = \int_{E_1} g_1 d\mu_1 \quad \text{et} \quad m'(A) = \int_{E_2} g_2 d\mu_2$$
sont des mesures.
3. Pour tout $A \in \mathcal{E}_1, B \in \mathcal{E}_2$, on a :
$$m(A \times B) = \mu_1(A)\mu_2(B) \quad \text{et} \quad m'(A \times B) = \mu_1(A)\mu_2(B)$$
et m et m' sont des mesures σ-finies.
4. Les mesures m et m' sont égales sur $\mathcal{E}_1 \otimes \mathcal{E}_2$. Donc $\forall A \in \mathcal{E}_1 \otimes \mathcal{E}_2$:
$$\int_{E_1} g_1 d\mu_1 = \int_{E_2} g_2 d\mu_2$$
et on note : $\mu_1 \otimes \mu_2 = m = m'$.

Preuve :

1. On commence par supposer que μ_2 est une mesure finie. On pose $\mathcal{D} = \{A \in \mathcal{E}_1 \otimes \mathcal{E}_2 : x_1 \mapsto \mu_2(A_2(x_1)) \text{ est mesurable}\}$. Et on va montrer que $\mathcal{D}$ est une classe monotone sur $E_1 \times E_2$ qui contient les pavés.
   i) $E_1 \times E_2 \in \mathcal{D}$ car $\forall x_1 \in E_1$, $(E_1 \times E_2)_2(x_1) = E_2$. Donc $x_1 \mapsto \mu_2((E_1 \times E_2)_2(x_1)) = \mu_2(E_2)$ est constante donc mesurable.
   ii) Soient $A, B \in \mathcal{D}$ tq $A \subset B$. Pour tout $x_1 \in E_1$:
$$\begin{aligned} (B \backslash A)_2(x_1) = \{y \in E_2 : (x_1, y) \in B \backslash A\} &= \{y \in E_2 : (x_1, y) \in B\} \backslash \{y \in E_2 : (x_1, y) \in A\} \\ &= B_2(x_1) \setminus A_2(x_1) \end{aligned}$$
   Donc comme μ_2 est une mesure finie :
$$\forall x_1 \in E_1, \ \mu_2((B \setminus A)_2(x_1)) = \mu_2(B_2(x_1)) - \mu_2(A_2(x_1))$$
   donc $x_1 \mapsto \mu_2((B \setminus A)_2(x_1))$ est mesurable (car différence de fonctions mesurables).
   iii) Soit $(A_n)_{n \geq 0}$ une suite croissante d'éléments de $\mathcal{D}$. Comme $\forall x_1 \in E_1$, $(\bigcup_{n \geq 0} A_n)_2(x_1) = \bigcup_{n \geq 0} (A_n)_2(x_1)$, on a :
$$\mu_2((\bigcup_{n \geq 0} A_n)_2(x_1)) = \mu_2(\bigcup_{n \geq 0} (A_n)_2(x_1)) = \lim_{n \to +\infty} \mu_2((A_n)_2(x_1))$$
   donc $x_1 \mapsto \mu_2((\bigcup_{n \geq 0} A_n)_2(x_1))$ est mesurable comme limite simple d'une suite de fonctions mesurables. Donc $\bigcup_{n \geq 0} A_n \in \mathcal{D}$.

iv) $\forall (B_1 \times B_2) \in E_1 \times E_2$ et $\forall x_1 \in E_1$, on a $(B_1 \times B_2)_2(x_1) = \begin{cases} \emptyset & \text{si } x_1 \notin B_1 \\ B_2 & \text{si } x_1 \in B_1 \end{cases}$. Donc

$$\mu_2((B_1 \times B_2)_2(x_1)) = \begin{cases} 0 & \text{si } x_1 \notin B_1 \\ \mu_2(B_2) & \text{si } x_1 \in B_1 \end{cases} = \mu_2(B_2)\mathbb{1}_{B_1}(x_1) \text{ qui est mesurable.}$$

Donc $\mathcal{D}$ est une classe monotone qui contient les pavés. La classe des pavés étant stable par intersection finie et engendrant $\mathcal{E}_1 \otimes \mathcal{E}_2$, on a $\mathcal{E}_1 \otimes \mathcal{E}_2 \subset \mathcal{D}$. Donc $\forall A \in \mathcal{E}_1 \otimes \mathcal{E}_2$, $x_1 \mapsto \mu_2(A_2(x_1))$ est mesurable.

À présent, considérons μ_2 σ-finie plutôt que finie. Il existe une suite croissante $(F_n)_{n\geq 0}$ d'éléments de $\mathcal{E}_2$ tq $\forall n \geq 0$, $\mu_2(F_n) < +\infty$ et $\bigcup_{n\geq 0} F_n = E_2$. Définissons la mesure sur $\mathcal{E}_2$ par $\mu_2^n(A) = \mu_2(A \cap F_n)$. Elle est une mesure finie $\forall n \geq 0$. Donc $x_1 \mapsto \mu_2^n(A_2(x_1))$ est mesurable $\forall A \in \mathcal{E}_1 \otimes \mathcal{E}_2$. Mais $\forall x_1 \in E_1$:

$$\begin{aligned}\mu_2(A_2(x_1)) = \mu_2(A \cap (\bigcup_{n\geq 0} F_n)_2(x_1)) &= \lim_{n\to+\infty} \mu_2((A \cap F_n)_2(x_1)) \\ &= \lim_{n\to+\infty} \mu_2^n(A_2(x_1))\end{aligned}$$

Donc $x_1 \mapsto \mu_2(A_2(x_1))$ est limite simple d'une suite de fonctions mesurables, donc mesurable.

2. On va montrer que l'application définie sur $\mathcal{E}_1 \otimes \mathcal{E}_2$ par $m : A \to \int_{E_1} \mu_2(A_2(x_1))\mu_1(dx_1)$ est une mesure sur $\mathcal{E}_1 \otimes \mathcal{E}_2$.
   - $m(\emptyset) = 0$.
   - Soit $(A^n)_{n\geq 0}$ une suite d'éléments de $\mathcal{E}_1 \otimes \mathcal{E}_2$ 2-à-2 disjoints. Alors $\forall x_1 \in E_1$, $((A^n)_2(x_1))_{n\geq 0}$ est une suite d'éléments de $\mathcal{E}_2$ 2-à-2 disjoints. Donc :

$$m(\bigcup_{n\geq 0} A^n) = \int_{E_1} \mu_2((\bigcup_{n\geq 0} A^n)_2(x_1))\mu_1(dx_1) = \int_{E_1} \mu_2(\bigcup_{n\geq 0} (A^n)_2(x_1))\mu_1(dx_1)$$

$$= \int_{E_1} \sum_{n\geq 0} \mu_2((A^n)_2(x_1))\mu_1(dx_1) = \sum_{n\geq 0} \int_{E_1} \mu_2((A^n)_2(x_1))\mu_1(dx_1) = \sum_{n\geq 0} m(A^n)$$

3. $\forall A \in \mathcal{E}_1$ et $B \in \mathcal{E}_2$.

$$m(A \times B) = \int_{E_1} \mu_2((A \times B)_2(x_1))\mu_1(dx_1) = \int_{E_1} \mu_2(B)\mathbb{1}_A(x_1)\mu_1(dx_1) = \mu_2(B)\mu_1(A)$$

$$m'(A \times B) = \int_{E_2} \mu_1((A \times B)_1(y))\mu_2(dy) = \int_{E_2} \mu_1(A)\mathbb{1}_B(y)\mu_2(dy) = \mu_1(A)\mu_2(B)$$

4. Puisque μ_1 est σ-finie, il existe une suite croissante $(A_n)_{n\geq 0}$ d'éléments de $\mathcal{E}_1$ tq $E_1 = \bigcup A_n$ et $\forall n \geq 0, \mu_1(A_n) < \infty$. Idem pour μ_2 avec une suite $(B_n)_{n\geq 0}$ d'éléments de $\mathcal{E}_2$. On a ainsi une suite croissante $(A_n \times B_n)_{n\geq 0}$ d'éléments de $\mathcal{E}_1 \otimes \mathcal{E}_2$ tq :

$$E_1 \times E_2 = \bigcup_{n\geq 0} A_n \times B_n \quad \text{et} \quad \forall n \geq 0,\ m(A_n \times B_n) = \mu_1(A_n)\mu_2(B_n) = m'(A_n \times B_n) < +\infty$$

Donc m et m' sont σ-finies et comme de plus m et m' sont égales sur les pavés, elles sont égales sur $\mathcal{E}_1 \otimes \mathcal{E}_2$ (d'après le résultat rappelé au début du paragraphe).

Remarque

Sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on considère $\mu_1 = \lambda$ (Lebesgue), $\mu_2 =$ mesure de comptage. μ_2 n'est pas σ-finie.

Avec $C=\{(x,x),\ x\in\mathbb{R}\}\in\mathcal{B}(\mathbb{R}^2)=\mathcal{B}(\mathbb{R})\otimes\mathcal{B}(\mathbb{R})$.

$$\int_{\mathbb{R}}\mu_2(C_2(x_1))\mu_1(dx_1)=\int_{\mathbb{R}}1d\lambda=+\infty$$

$$\int_{\mathbb{R}}\mu_1(C_1(x_2))\mu_2(dx_2)=\int_{\mathbb{R}}0d\mu_2=0$$

III. Théorèmes de Fubini

Proposition 22: Théorème de Fubini-Tonelli ou Fubini-positif

Soient $(E_1,\mathcal{E}_1,\mu_1)$ et $(E_2,\mathcal{E}_2,\mu_2)$ deux espaces mesurés avec μ_1 et μ_2 des mesures σ-finies et $F:E_1\times E_2\to\bar{\mathbb{R}}$ une fonction $\mathcal{E}_1\otimes\mathcal{E}_2-\mathcal{B}(\mathbb{R})$ mesurable **positive**.

1. Les fonctions

$$x_1\in E_1\mapsto\int_{E_2}F(x_1,x_2)\mu_2(dx_2)$$
$$x_2\in E_2\mapsto\int_{E_1}F(x_1,x_2)\mu_1(dx_1)$$

sont mesurables.

2. On a :

$$\begin{aligned}\int_{E_1\times E_2}Fd(\mu_1\otimes\mu_2)&=\int_{E_1}\left(\int_{E_2}F(x_1,x_2)\mu_2(dx_2)\right)\mu_1(dx_1)\\&=\int_{E_2}\left(\int_{E_1}F(x_1,x_2)\mu_1(dx_1)\right)\mu_2(dx_2)\end{aligned}$$

Preuve :

1. Soit $C\in\mathcal{E}_1\otimes\mathcal{E}_2$ et $F=\mathbb{1}_C$. On a :

$$\begin{aligned}\mathbb{1}_C(x_1,x_2)=\begin{cases}1&\text{si }(x_1,x_2)\in C\\0&\text{sinon}\end{cases}&=\begin{cases}1&\text{si }x_2\in C_2(x_1)\\0&\text{sinon}\end{cases}\\&=\mathbb{1}_{C_2(x_1)}(x_2)=\mathbb{1}_{C_1(x_2)}(x_1)\end{aligned}$$

Dans ce cas particulier :

$$x_1\mapsto\int_{E_2}\mathbb{1}_C(x_1,x_2)\mu_2(dx_2)=\int_{E_2}\mathbb{1}_{C_2(x_1)}(x_2)\mu_2(dx_2)=\mu_2(C_2(x_1))$$

est bien mesurable (d'après la construction de la mesure produit). Donc les affirmations sont vraies pour les fonctions indicatrices. Par linéarité de l'intégrale, elles sont vraies pour les fonctions étagées positives.

2. On commence par supposer que F est étagée positive. On note $(a_\ell,A^\ell)_{\ell=1,\dots,n}$ les éléments de la représentation canonique de F. Pour tout $(x_1,x_2)\in E_1\times E_2$, on a :

$$F(x_1,x_2)=\sum_{\ell=1}^n a_\ell\mathbb{1}_{A^\ell}(x_1,x_2)=\sum_{\ell=1}^n a_\ell\mathbb{1}_{A_2^\ell(x_1)}(x_2)=\sum_{\ell=1}^n a_\ell\mathbb{1}_{A_1^\ell(x_2)}(x_1)$$

$$\int_{E_1\times E_2} F d(\mu_1\otimes\mu_2) = \sum_{\ell=1}^n a_\ell (\mu_1\otimes\mu_2)(A^\ell) \quad \text{(def de } \mu_1\otimes\mu_2)$$
$$= \sum_{\ell=1}^n a_\ell \int_{E_1} \mu_2(A_2^\ell(x_1))\mu_1(dx_1)$$
$$= \int_{E_1} \left(\sum_{\ell=1}^n a_\ell \mu_2(A_2^\ell(x_1))\right)\mu_1(dx_1)$$

Or $\int_{E_2}\sum_{\ell=1}^n a_\ell \mathbb{1}_{A^\ell}(x_1,x_2)\mu_2(dx_2) = \int_{E_2} F(x_1,x_2)\mu_2(dx_2)$.

$$= \int_{E_1}\left(\int_{E_2} F(x_1,x_2)\mu_2(dx_2)\right)\mu_1(dx_1)$$

Les deux égalités sont donc vraies pour les fonctions étagées positives.

Maintenant, soit F une fonction mesurable positive et $(f_n)_{n\geq 0}$ une suite croissante de fonctions étagées positives qui converge simplement vers F. D'après le TCM (Théorème de Convergence Monotone), on a :

$$\int_{E_2} F(x_1,x_2)\mu_2(dx_2) = \int_{E_2} \lim_{n\to\infty} f_n(x_1,x_2)\mu_2(dx_2) = \lim_{n\to\infty}\int_{E_2} f_n(x_1,x_2)\mu_2(dx_2)$$

donc $x_1 \mapsto \int_{E_2} F(x_1,x_2)\mu_2(dx_2)$ est une fonction limite simple d'une suite de fonctions mesurables donc elle est mesurable.

De plus :

$$\int_{E_1\times E_2} F d(\mu_1\otimes\mu_2) = \lim_{n\to\infty}\int_{E_1\times E_2} f_n d(\mu_1\otimes\mu_2)$$
$$= \lim_{n\to\infty}\int_{E_1}\left(\int_{E_2} f_n(x_1,x_2)\mu_2(dx_2)\right)\mu_1(dx_1)$$
$$= \int_{E_1}\lim_{n\to\infty}\left(\int_{E_2} f_n(x_1,x_2)\mu_2(dx_2)\right)\mu_1(dx_1) \quad \text{(par le TCM sur } E_1)$$
$$= \int_{E_1}\left(\int_{E_2} F(x_1,x_2)\mu_2(dx_2)\right)\mu_1(dx_1)$$

Proposition 23: Théorème de Fubini-Lebesgue

Soient $(E_1,\mathcal{E}_1,\mu_1)$ et $(E_2,\mathcal{E}_2,\mu_2)$ deux espaces mesurés où μ_1 et μ_2 sont σ-finies et $F\in L^1(E_1\times E_2,\mathcal{E}_1\otimes\mathcal{E}_2,\mu_1\otimes\mu_2)$.

1. On a :
   - $x_2\mapsto F(x_1,x_2)$ est dans $L^1(E_2,\mathcal{E}_2,\mu_2)$ pour μ_1-presque tout x_1.
   - $x_1\mapsto F(x_1,x_2)$ est dans $L^1(E_1,\mathcal{E}_1,\mu_1)$ pour μ_2-presque tout x_2.
2. Les fonctions :
   - $x_1\mapsto \int_{E_2} F(x_1,x_2)\mu_2(dx_2)$ sont définies μ_1-p.p. et appartiennent à $L^1(E_1,\mathcal{E}_1,\mu_1)$.
   - $x_2\mapsto \int_{E_1} F(x_1,x_2)\mu_1(dx_1)$ sont définies μ_2-p.p. et appartiennent à $L^1(E_2,\mathcal{E}_2,\mu_2)$.
3. On a :
$$\int_{E_1\times E_2} F d(\mu_1\otimes\mu_2) = \int_{E_1}\left(\int_{E_2} F(x_1,x_2)\mu_2(dx_2)\right)\mu_1(dx_1)$$
$$= \int_{E_2}\left(\int_{E_1} F(x_1,x_2)\mu_1(dx_1)\right)\mu_2(dx_2)$$

Preuve :

1. Puisque $F \in L^1(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2, \mu_1 \otimes \mu_2)$, on a :
$$\int_{E_1 \times E_2} |F| d(\mu_1 \otimes \mu_2) = \int_{E_1} \left(\int_{E_2} |F(x_1, x_2)| \mu_2(dx_2) \right) \mu_1(dx_1) < +\infty$$
Donc $\int_{E_2} |F(x_1, x_2)| \mu_2(dx_2) < +\infty$ μ_1-p.p. Donc $x_2 \mapsto F(x_1, x_2)$ est dans $L^1(E_2, \mathcal{E}_2, \mu_2)$ μ_1-p.p.
2. Donc pour μ_1 presque partout x_1 :
$$x_1 \mapsto \int_{E_2} F(x_1, x_2) \mu_2(dx_2) = \int_{E_2} F^+(x_1, x_2) \mu_2(dx_2) - \int_{E_2} F^-(x_1, x_2) \mu_2(dx_2)$$
est bien définie (différence finie). De plus :
$$\begin{aligned} \int_{E_1} \left| \int_{E_2} F(x_1, x_2) \mu_2(dx_2) \right| \mu_1(dx_1) &\leq \int_{E_1} \left(\int_{E_2} |F(x_1, x_2)| \mu_2(dx_2) \right) \mu_1(dx_1) \\ &= \int_{E_1 \times E_2} |F| d(\mu_1 \otimes \mu_2) < +\infty \end{aligned}$$
Donc $x_1 \mapsto \int_{E_2} F(x_1, x_2) \mu_2(dx_2)$ est un élément de $L^1(E_1, \mathcal{E}_1, \mu_1)$.
3. Puisque $F \in L^1(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2, \mu_1 \otimes \mu_2)$, on a :
$$\int_{E_1 \times E_2} F^+ d(\mu_1 \otimes \mu_2) = \int_{E_1} \left(\int_{E_2} F^+(x_1, x_2) \mu_2(dx_2) \right) \mu_1(dx_1) < +\infty$$
$$\int_{E_1 \times E_2} F^- d(\mu_1 \otimes \mu_2) = \int_{E_1} \left(\int_{E_2} F^-(x_1, x_2) \mu_2(dx_2) \right) \mu_1(dx_1) < +\infty$$
D'où l'égalité énoncée par soustraction.

IV. La mesure de Lebesgue sur $\mathbb{R}^d$

Définition 16: Mesure de Lebesgue sur $\mathbb{R}^d$

Sur la base de $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, où λ est la mesure de Lebesgue, on définit par récurrence sur $d \geq 2$, la mesure de Lebesgue sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ en utilisant que :
$$\begin{aligned} \mathbb{R}^d &= \mathbb{R}^{d-1} \times \mathbb{R} \\ \mathcal{B}(\mathbb{R}^d) &= \mathcal{B}(\mathbb{R}^{d-1}) \otimes \mathcal{B}(\mathbb{R}) \\ \lambda^{\otimes d} &= \lambda^{\otimes d-1} \otimes \lambda \end{aligned}$$

Exemple 7: Hyperplan

Soient $a_1, \ldots, a_d$ des réels tous non nuls et $\alpha \in \mathbb{R}$. On pose $D = \{(x_1, \ldots, x_d) \in \mathbb{R}^d : a_1 x_1 + \cdots + a_d x_d = \alpha\}$. Et on suppose que $a_d \neq 0$. On a que $\lambda^{\otimes d}(D) = 0$. Car :
$$\lambda^{\otimes d}(D) = \int_{\mathbb{R}^{d-1}} \left(\int_{\mathbb{R}} \mathbb{1}_D(x_1, \ldots, x_d) \lambda(dx_d) \right) \lambda^{\otimes d-1}(dx_1 \ldots dx_{d-1})$$
Or pour $x_1, \ldots, x_{d-1}$ fixés, l'ensemble $\{x_d \in \mathbb{R} : (x_1, \ldots, x_d) \in D\}$ est un singleton (car $a_d \neq 0$) ou vide. Sa mesure de Lebesgue est donc nulle.
$$= 0$$

Proposition 24: Propriétés de la mesure de Lebesgue

Soit d un entier naturel, $d \geq 1$.

1. La mesure de Lebesgue sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est invariante par translation : $\forall x \in \mathbb{R}^d$ et $\forall B \in \mathcal{B}(\mathbb{R}^d)$, on a :
$$\lambda(x+B) = \lambda(B)$$
2. **Réciproquement**, si μ est une mesure sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ finie sur les parties mesurables bornées de $\mathbb{R}^d$ et invariante par translation, alors il existe une constante $c \geq 0$ telle que $\mu = c\lambda$.

Preuve :

1. Soit $x \in \mathbb{R}^d$, on note $\tau_x : \mathbb{R}^d \to \mathbb{R}^d$ définie par $u \mapsto u - x$. On a $\forall B \in \mathcal{B}(\mathbb{R}^d)$, $\tau_x^{-1}(B) = x + B$. Donc $\lambda(\tau_x^{-1}(B)) = \lambda(x+B)$. Donc $B \in \mathcal{B}(\mathbb{R}^d) \mapsto \lambda(x+B)$ est une mesure sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.
   - **Supposons que** $d = 1$ **:** Il suffit de montrer que λ et λ^{τ_x} coïncident sur les boréliens de la forme $]a,b[$ pour pouvoir conclure. (Ceci est une conséquence du corollaire du théorème de la classe monotone déjà utilisé précédemment). Clairement :
   $$\lambda(x+]a,b[) = \lambda(]a+x, b+x[) = (b+x) - (a+x) = b - a = \lambda(]a,b[)$$
   - **Si** $d \geq 2$ **:** $\forall A_1, \ldots, A_d \in \mathcal{B}(\mathbb{R})$, on a que $A_1 \times A_2 \times \cdots \times A_d \in \mathcal{B}(\mathbb{R}^d)$. Et on a que $x + A_1 \times \cdots \times A_d = (x_1 + A_1) \times (x_2 + A_2) \times \cdots \times (x_d + A_d) \in \mathcal{B}(\mathbb{R}^d)$ avec $x = (x_1, \ldots, x_d)$. Donc :
   $$\lambda^{\otimes d}(A_1 \times \cdots \times A_d) = \lambda_1(A_1) \times \cdots \times \lambda_1(A_d)$$
   Et :
   $$\begin{aligned}\lambda^{\otimes d}(x + A_1 \times \cdots \times A_d) &= \lambda^{\otimes d}((x_1 + A_1) \times \cdots \times (x_d + A_d)) \\ &= \lambda_1(x_1 + A_1) \ldots \lambda_1(x_d + A_d) \\ &= \lambda_1(A_1) \ldots \lambda_1(A_d) = \lambda^{\otimes d}(A_1 \times \cdots \times A_d)\end{aligned}$$
   Donc $\lambda^{\otimes d}$ et $\lambda^{\otimes d, \tau_x}$ coïncident sur la classe des pavés. Comme cette classe est stable par intersection finie et qu'elle engendre $\mathcal{B}(\mathbb{R}^d)$, à nouveau, d'après le corollaire du théorème de la classe monotone déjà utilisé, $\lambda^{\otimes d} = \lambda^{\otimes d, \tau_x}$.
2. Soit μ une mesure sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ finie sur les parties bornées mesurables de $\mathbb{R}^d$ et invariante par translation. On pose $C = \mu([0,1[^d)$. $\forall m \geq 1$, on découpe $[0,1[^d$ en translatés de $[0, 1/m[^d$. La propriété d'additivité de μ entraîne que :
$$C = \mu([0,1[^d) = m^d \mu([0,1/m[^d)$$
donc $\mu([0,1/m[^d) = \frac{C}{m^d}$. Soient $a_1, \ldots, a_d \geq 0$, $\forall m \geq 1$.
$$\prod_{i=1}^d [0, \frac{\lfloor m a_i \rfloor}{m}[\subset \prod_{i=1}^d [0, a_i[\subset \prod_{i=1}^d [0, \frac{\lfloor m a_i \rfloor + 1}{m}[$$
$$\mu\left(\prod_{i=1}^d [0, \frac{\lfloor m a_i \rfloor}{m}[\right) \leq \mu\left(\prod_{i=1}^d [0, a_i[\right) \leq \mu\left(\prod_{i=1}^d [0, \frac{\lfloor m a_i \rfloor + 1}{m}[\right)$$
$$\frac{C}{m^d} \prod_{i=1}^d \lfloor m a_i \rfloor \leq \mu\left(\prod_{i=1}^d [0, a_i[\right) \leq \frac{C}{m^d} \prod_{i=1}^d (\lfloor m a_i \rfloor + 1)$$
Donc en prenant $m \to \infty$, $\mu(\prod_{i=1}^d [0, a_i[) = C \prod_{i=1}^d a_i = C\lambda(\prod_{i=1}^d [0, a_i[)$. Donc μ et $c\lambda$ sont égales sur tous les rectangles $\prod_{i=1}^d [0, a_i[$ (avec $a_i > 0$). Donc sur tous les rectangles $\prod_{i=1}^d [a_i, b_i[$ puisqu'elles sont invariantes par translation. Comme la classe des pavés rectangles engendre $\mathcal{B}(\mathbb{R}^d)$ et est stable par intersection finie, d'après le corollaire habituel du théorème des classes monotones, $\mu = c\lambda$.

V. Changement de variable

Proposition 25: Formule du changement de variable affine

Soit M une matrice $d \times d$ inversible, $a \in \mathbb{R}^d$ et $\phi : \mathbb{R}^d \to \mathbb{R}^d$ définie par $\phi(x) = Mx + a$. La mesure image λ^d par ϕ^{-1} vérifie :

$$\forall A \in \mathcal{B}(\mathbb{R}^d), \quad \lambda^d(\phi^{-1}(A)) = \lambda^d(\phi(A)) = |\det M|\lambda^d(A)$$

Preuve : Comme ϕ est bijective et ϕ^{-1} continue, on a bien que $\phi(A) = (\phi^{-1})^{-1}(A) \in \mathcal{B}(\mathbb{R}^d)$ pour $A \in \mathcal{B}(\mathbb{R}^d)$ et même $A \in \mathcal{B}(\mathbb{R}^d) \mapsto \lambda(\phi(A)) = \lambda^{\phi^{-1}}(A)$ est une mesure.

$$\lambda^d(\phi(A)) = \lambda^d(MA + a) = \lambda^d(MA)$$

puisque λ^d est invariante par translation. Et la mesure $A \in \mathcal{B}(\mathbb{R}^d) \mapsto \lambda^d(MA)$ est invariante par translation elle aussi car :

$$\forall b \in \mathbb{R}^d, \ \lambda^d(M(A+b)) = \lambda^d(MA + Mb) = \lambda^d(MA)$$

car λ^d est invariante par translation. On a que l'image MA de toute partie A bornée de $\mathbb{R}^d$ par $x \mapsto Mx$ est bornée, donc $\forall A \subset \mathbb{R}^d$ mesurable bornée, $\lambda^d(MA) < \infty$. Donc d'après le résultat précédent, il existe un réel $c > 0$ tel que $\forall A \in \mathcal{B}(\mathbb{R}^d)$, $\lambda^d(MA) = c\lambda^d(A)$.

- **1er cas :** On suppose que M est orthogonale. En prenant B la boule de rayon 1 de $\mathbb{R}^d$, on a $MB = B$ donc :
$$\lambda^d(MB) = \lambda^d(B)$$
et $\lambda^d(MB) = c\lambda^d(B)$ donc $c = 1 = |\det(M)|$.
- **2ème cas :** M est symétrique et définie positive. Comme M est inversible, M est définie positive donc il existe une matrice orthogonale P et des réels $\alpha_1, \dots, \alpha_d > 0$ tels que :
$$P^{-1}MP = \begin{pmatrix} \alpha_1 & & 0 \\ & \ddots & \\ 0 & & \alpha_d \end{pmatrix} = D = \lambda_d\left(\prod_{i=1}^d [0, \alpha_i[\right) = \prod_{i=1}^d \alpha_i = \det(M)$$
En prenant $A = P[0,1]^d$, on a :
$$\lambda^d(MA) = \lambda^d(MP[0,1]^d) = \lambda^d(PP^{-1}MP[0,1]^d)$$
$$= \lambda^d(PD[0,1]^d) = \lambda^d(P\prod_{i=1}^d [0, \alpha_i])$$
Comme P est orthogonale, $\lambda^d(P \dots) = \lambda^d(\dots)$, donc :
$$= \lambda^d(\prod_{i=1}^d [0, \alpha_i]) = \prod_{i=1}^d \alpha_i = \det(M) = \det(M)\lambda^d([0,1]^d)$$
Donc $c = \det(M)$.
- **3ème cas :** M est inversible quelconque. Alors $M = PS$ où $S = (M^TM)^{1/2}$ est symétrique définie positive et $P = MS^{-1}$ est orthogonale (Décomposition polaire).
$$\forall A \in \mathcal{B}(\mathbb{R}^d), \ \lambda^d(MA) = \lambda^d(PSA) \overset{1er\_cas}{=} \lambda^d(SA) = \det S \cdot \lambda^d(A)$$
$$= |\det M|\lambda^d(A)$$
Mais $|\det S|^2 = \det(M)^2$ et $\det S > 0$ donc $\det(S) = |\det(M)|$.

Théorème 4: Formule du changement de variables

Soient $U, V \subset \mathbb{R}^d$ des ouverts et $\phi : U \to V$ avec ϕ un C^1-difféomorphisme.

1. Pour toute fonction $f : V \to \mathbb{R}^+$ borélienne :

$$\int_V f(y)dy = \int_U f(\phi(x))|J_\phi(x)|dx$$

2. $\forall f : V \to \mathbb{R}$ borélienne, on a que :

$$f \in L^1(V, \mathcal{B}(V), \lambda_d) \iff (f \circ \phi)|J_\phi| \in L^1(U, \mathcal{B}(U), \lambda_d)$$

   et si c'est le cas :

$$\int_V f(y)dy = \int_U f(\phi(x))|J_\phi(x)|dx$$

Chapitre 5 : Fondements de la théorie des probabilités

I. Espace de probabilité

Expérience aléatoire : Une expérience dont on ne connaît pas le résultat à l'avance mais dont on connaît tous les résultats possibles.

Définition 17: Espace de probabilité

On appelle espace de probabilité un triplet $(\Omega, \mathcal{A}, \mathbb{P})$ où :

- Ω est un ensemble (Ensemble de tous les résultats possibles de l'expérience aléatoire).
- $\mathcal{A}$ est une tribu sur Ω (Ensemble de questions sur le résultat de l'expérience qui a pour réponse oui ou non).
- $\mathbb{P}$ est une mesure de probabilité définie sur $\mathcal{A}$.

On appelle **événements** les éléments de $\mathcal{A}$ et $\mathbb{P}$ mesure la probabilité d'observer des événements considérés.

II. Variable aléatoire

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. $(E, \mathcal{E})$ un espace mesurable.

Définition 18: Variable aléatoire

Une variable aléatoire est une fonction mesurable définie sur Ω à valeurs dans E.

Définition 19: Loi d'une variable aléatoire

Soit $X : \Omega \to E$ une variable aléatoire. On appelle **loi de X** l'image par X de la probabilité $\mathbb{P}$. C'est la probabilité $\mathbb{P}^X$ sur $(E, \mathcal{E})$ définie par :

$$B \in \mathcal{E} \longmapsto \mathbb{P}^X(B) = \mathbb{P}(X \in B) = \mathbb{P}(X^{-1}(B))$$

Définition 20: Variable aléatoire discrète

Soit $X : \Omega \to E$ une variable aléatoire. On dit que X est une variable aléatoire **discrète** s'il existe $M \in \mathcal{E}$ dénombrable t.q. :

$$\mathbb{P}(X \in M) = 1$$

Remarque

On peut supposer sans perte de généralité que $M \subset X(\Omega)$.

Proposition 26: Loi d'une v.a. discrète

Soit $X : \Omega \to E$ une variable aléatoire discrète. On a :

$$\mathbb{P}^X = \sum_{x \in M} p_x \delta_x$$

où $M \in \mathcal{E}$, $M \subset X(\Omega)$ est dénombrable et $\forall x \in M$, $p_x = \mathbb{P}(X = x)$.

Preuve :

Soit $B \in \mathcal{E}$.

$$\begin{aligned}
\mathbb{P}^X(B) &= \mathbb{P}(X \in B) \\
&= \mathbb{P}(X \in B \cap M) \\
&= \sum_{x \in B \cap M} \mathbb{P}(X = x) \\
&= \sum_{x \in M} p_x \delta_x(B) \\
&= \left(\sum_{x \in M} p_x \delta_x \right)(B)
\end{aligned}$$

$\square$

Exemple 8: Exemple de calcul

Soit X une v.a. à valeurs dans $\mathbb{N}$ de loi donnée par :

$$p_n = \begin{cases} 0 & \text{si } n = 0 \\ (1-p)^{n-1} p & \text{si } n \geq 1 \end{cases}$$

(Loi géométrique). Calculons $\mathbb{P}(X \text{ est un nombre pair})$.

$$\begin{aligned}
\mathbb{P}(X \in 2\mathbb{N}) &= \sum_{k \geq 1} \mathbb{P}(X = 2k) \\
&= \sum_{k \geq 1} (1-p)^{2k-1} p \\
&= p(1-p) \sum_{k \geq 1} ((1-p)^2)^{k-1} \\
&= p(1-p) \sum_{j \geq 0} ((1-p)^2)^j \\
&= \frac{p(1-p)}{1-(1-p)^2} = \frac{1-p}{2-p} \xrightarrow[p \to 1]{} 0 \quad \text{et} \quad \xrightarrow[p \to 0]{} \frac{1}{2}
\end{aligned}$$

Définition 21: Variable aléatoire à densité

On dit d'une variable aléatoire X à valeurs dans $\mathbb{R}^d$ qu'elle est une v.a. **à densité** quand sa loi est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. D'après le théorème de Radon-Nikodym, il existe une fonction $f : \mathbb{R}^d \to \mathbb{R}^+$ mesurable, positive, unique à un ensemble de mesure de Lebesgue nulle près telle que :

$$\forall B \in \mathcal{B}(\mathbb{R}^d), \quad \mathbb{P}(X \in B) = \int_B f(x) d\lambda(x)$$

On dit de f qu'elle est la **densité** de X.

Définition 22: Tribu engendrée

Soit $X : \Omega \to E$ une variable aléatoire. On appelle tribu engendrée par X la tribu :

$$\sigma(X) = \{X^{-1}(B) : B \in \mathcal{E}\}$$

C'est la plus petite tribu sur Ω pour laquelle X est mesurable.

Proposition 27: Factorisation (Lemme de Doob-Dynkin)

Soit $X : \Omega \to E$ une variable aléatoire et $Y : \Omega \to \mathbb{R}$ une variable aléatoire réelle. Il y a équivalence entre :

1. Y est $\sigma(X)$-mesurable.
2. Il existe une fonction $f : E \to \mathbb{R}$ mesurable telle que $Y = f(X)$.

Preuve :

$2 \implies 1$: $\forall B \in \mathcal{B}(\mathbb{R})$, $Y^{-1}(B) = X^{-1}(f^{-1}(B)) \in \sigma(X)$ car $f^{-1}(B) \in \mathcal{E}$. Donc Y est $\sigma(X)$-mesurable.
$1 \implies 2$: On commence par supposer que Y est étagée. Il existe $N \geq 1$, $\lambda_1, \ldots, \lambda_N \in \mathbb{R}$, $A_1, \ldots, A_N \in \sigma(X)$ t.q.

$$Y = \sum_{i=1}^{N} \lambda_i \mathbb{1}_{A_i}$$

$\forall i \in [\![1, N]\!]$, puisque $A_i \in \sigma(X)$, il existe $B_i \in \mathcal{E}$ t.q. $A_i = X^{-1}(B_i)$. Donc $Y = \sum_{i=1}^{N} \lambda_i \mathbb{1}_{X^{-1}(B_i)} = \sum_{i=1}^{N} \lambda_i \mathbb{1}_{B_i}(X) = f(X)$ où $f = \sum_{i=1}^{N} \lambda_i \mathbb{1}_{B_i}$.
Soit Y une v.a. $\sigma(X)$-mesurable quelconque. On sait qu'il existe une suite $(Y_n)_{n\geq 0}$ de v.a. étagées et $\sigma(X)$-mesurables t.q. pour $\omega \in \Omega$, $Y_n(\omega) \to Y(\omega)$. D'après le point précédent, il existe une suite $(f_n)_{n\geq 0}$ de fonctions mesurables définies sur E à valeurs dans $\mathbb{R}$ t.q. $\forall \omega \in \Omega$, $Y_n(\omega) = f_n(X(\omega))$. En posant $D = \{x \in E : \lim_{n\to\infty} f_n(x) \text{ existe}\}$. On se donne une fonction $f : E \to \mathbb{R}$ mesurable t.q. $Y = f(X)$ en posant la limite sur D et 0 sinon. $\square$

Remarque

Si $A \in \sigma(X)$ alors $\mathbb{1}_A$ est $\sigma(X)$-mesurable donc il existe une fonction f mesurable t.q. $\mathbb{1}_A = f(X)$. Donc on peut décider si A a lieu ou pas dès qu'on connait la valeur de X.

III. Espérance mathématique

Définition 23: Espérance

Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle. Si X est de signe constant ou si $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, on appelle espérance de X :

$$\mathbb{E}[X] = \int_\Omega X d\mathbb{P}$$

Si $X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} \in \mathbb{R}^d$ et que pour chaque $i = 1 \dots d$, $\mathbb{E}[X_i]$ est définie, on pose :

$$\mathbb{E}[X] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_d] \end{pmatrix}$$

Remarque

Quand elle est définie, $\mathbb{E}[X]$ est une intégrale de Lebesgue. On peut donc utiliser toutes les propriétés de l'intégrale de Lebesgue dans l'étape de $\mathbb{E}[X]$.

Proposition 28: Théorème du transfert

Soit $X : \Omega \to E$ une v.a. et $g : E \to \mathbb{R}$ une fonction mesurable t.q. $g(X)$ est de signe constant ou $g(X) \in L^1(\mathbb{P})$. On a :

$$\mathbb{E}[g(X)] = \int_\Omega g(X) d\mathbb{P} = \int_E g d\mathbb{P}^X$$

Proposition 29: Loi image

Soit $X : \Omega \to E$ une v.a. Si il existe une probabilité ν sur $(E, \mathcal{E})$ t.q. pour toute fonction $g : E \to \mathbb{R}$ mesurable bornée on a :

$$\mathbb{E}[g(X)] = \int_E g d\nu$$

Alors ν est la loi de X.

Preuve :

En utilisant $g = \mathbb{1}_B, \forall B \in \mathcal{E}$. $\mathbb{P}^X(B) = \mathbb{E}[\mathbb{1}_B(X)] = \int \mathbb{1}_B d\nu = \nu(B)$. $\square$

Remarque

Si $X : \Omega \to E$ et $Y : \Omega \to E$ des v.a. sont telles que pour toute fonction $g : E \to \mathbb{R}$ mesurable bornée, on a $\mathbb{E}[g(X)] = \mathbb{E}[g(Y)]$, alors X et Y ont la même loi.

Exemple 9: Densité uniforme

Si X admet pour densité $\mathbb{1}_{[0,1]}$ (X suit la loi uniforme sur $[0,1]$), alors X et $1-X$ ont la même loi. En effet, $\forall g : \mathbb{R} \to \mathbb{R}$ mesurable bornée :

$$\mathbb{E}[g(1-X)] = \int_{\mathbb{R}} g(1-x)\mathbb{1}_{[0,1]}(x)dx = \int_{\mathbb{R}} g(u)\mathbb{1}_{[0,1]}(u)du$$

(changement de variable $u = 1-x$). Donc X et $1-X$ ont la même loi. Cependant, $\mathbb{P}(X = 1-X) = \mathbb{P}(X = 1/2) = 0$.

Proposition 30: Espérance variable discrète

Soit $X : \Omega \to E$ une v.a. discrète et $g : E \to \mathbb{R}$ une fonction mesurable. Si g est de signe constant ou si $g(X) \in L^1(\mathbb{P})$, on a :

$$\mathbb{E}[g(X)] = \sum_{x \in M} g(x)\mathbb{P}(X = x)$$

où M est la partie dénombrable de X.

Preuve :

On suppose que $g \geq 0$.

$$\mathbb{E}[g(X)] = \int_E g d\mathbb{P}^X = \int_E g d\left(\sum_{x\in M} p_x \delta_x\right)$$

$$= \sum_{x\in M} p_x \left(\int_E g d\delta_x\right) = \sum_{x \in M} g(x)p_x = \sum_{x\in M} g(x)\mathbb{P}(X = x)$$

Si $g(X) \in L^1(\mathbb{P})$, alors $\sum_{x\in M} |g(x)|\mathbb{P}(X = x) < +\infty$. Et $\int_E g^+ d\mathbb{P}^X = \sum_{x\in M} g^+(x)\mathbb{P}(X = x)$, $\int_E g^- d\mathbb{P}^X = \sum_{x \in M} g^-(x)\mathbb{P}(X = x)$. Donc $\int_E g d\mathbb{P}^X = \sum_{x\in M}(g^+(x) - g^-(x))\mathbb{P}(X = x) = \sum_{x\in M} g(x)\mathbb{P}(X = x)$. □

Proposition 31: Espérance variable à densité

Soit $X : \Omega \to \mathbb{R}$ une v.a. de densité f, et $g : \mathbb{R} \to \mathbb{R}$ une fonction mesurable. Si $g(x)$ est de signe constant ou $g(X) \in L^1(\mathbb{P})$, alors :

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

Preuve :

$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g d\mathbb{P}^X$ et $\mathbb{P}^X$ admet f pour densité : $\forall B \in \mathcal{B}(\mathbb{R}), \mathbb{P}^X(B) = \int_{\mathbb{R}} f \mathbb{1}_B d\lambda$. (Suite classique intégrale de Lebesgue). □

IV. Variables aléatoires de carré intégrable

Puisque $\mathbb{P}(\Omega) = 1$, $\forall p, q$ t.q. $1 \leq p \leq q$, on a $L^q(\mathbb{P}) \subset L^p(\mathbb{P})$ et même $\forall X \in L^p$, on a $\|X\|_p \leq \|X\|_q$. En particulier, $L^2(\mathbb{P}) \subset L^1(\mathbb{P})$ et si $X, Y \in L^2(\mathbb{P})$ alors $X * Y \in L^1(\mathbb{P})$ puisque $\mathbb{E}[|XY|] \leq \sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}$ (Cauchy-Schwarz).

L'espace $L^2(\mathbb{P})$ muni de $< X, Y >= \mathbb{E}[XY]$ est un espace de Hilbert.

Définition 24: Variance

Soit $X \in L^2(\mathbb{R})$. On appelle variance de X :

$$V(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

Remarque

1. $V(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ (König-Huygens).
2. $\forall a \in \mathbb{R}$, $\mathbb{E}[(X-a)^2] = \mathbb{E}[(X - \mathbb{E}[X] + \mathbb{E}[X] - a)^2] = \text{Var}(X) + 2\mathbb{E}[X - \mathbb{E}[X]](\mathbb{E}[X] - a) + (\mathbb{E}[X] - a)^2$. Comme $\mathbb{E}[X - \mathbb{E}[X]] = 0$, on a $\mathbb{E}[(X-a)^2] = \text{Var}(X) + (\mathbb{E}[X] - a)^2$. Donc $\inf_{a\in\mathbb{R}} \mathbb{E}[(X-a)^2] = \text{Var}(X)$ et cet inf est atteint en $a = \mathbb{E}[X]$.

Proposition 32: Inégalité de Bienaymé-Tchebychev

Soit $X \in L^2(\mathbb{P})$. $\forall a > 0$ on a :

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

Preuve :

$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) = \mathbb{P}(|X - \mathbb{E}[X]|^2 \geq a^2) \leq \frac{1}{a^2}\mathbb{E}[|X - \mathbb{E}[X]|^2] = \frac{1}{a^2}\text{Var}(X)$ par Markov. □

Proposition 33: Propriétés de la variance

- $\forall (a,b) \in \mathbb{R}$, $V(aX+b) = a^2V(X)$.
- Si $\text{Var}(X) = 0$ alors $X = \mathbb{E}[X]$ p.s.

Preuve :

$\text{Var}(X) = \int_\Omega |X - \mathbb{E}[X]|^2 d\mathbb{P} = 0 \implies X - \mathbb{E}[X] = 0$ p.s. car c'est une intégrale de Lebesgue d'une fonction mesurable positive. □

Définition 25: Covariance et Dispersion

Soit $(X,Y) \in (L^2(\mathbb{P}))^2$ des v.a. réelles. La covariance de X et Y est définie par :

$$\text{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Si $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \in \mathbb{R}^n$ t.q. $\forall i \in [\![1,n]\!]$, $X_i \in L^2(\mathbb{P})$.

$$D_X = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T] = (\text{Cov}(X_i, X_j))_{1\leq i,j\leq n}$$

est appelée **matrice de dispersion** ou covariance de X.

Proposition 34: Propriétés de la Covariance

Soit $(X, Y) \in (L^2(\mathbb{P}))^2$.

1. Si X est une v.a. réelle : $\text{Cov}(X, X) = \text{Var}(X)$.
2. Si X et Y sont des v.a. réelles : $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
3. Sur l'ensemble des v.a. réelles de carré intégrable, $(X, Y) \mapsto \text{Cov}(X, Y)$ est bilinéaire.
4. Si $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \in \mathbb{R}^n$ avec chacune de ses composantes dans $L^2(\mathbb{P})$. $\forall A \in \mathcal{M}_{k,n}(\mathbb{R}), B \in \mathbb{R}^k$:
$$D_{AX+B} = AD_X A^T$$

Preuve :

$Y = AX + B$. $Y - \mathbb{E}[Y] = Y - \mathbb{E}[AX + B] = A(X - \mathbb{E}[X])$. Donc :

$$\begin{aligned} D_{AX+B} &= \mathbb{E}[(AX + B - \mathbb{E}[AX + B])(AX + B - \mathbb{E}[AX + B])^T] \\ &= \mathbb{E}[A(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T A^T] \\ &= A\mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T]A^T \\ &= AD_X A^T \end{aligned}$$

□

Remarque

Si $X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} \in L^2(\mathbb{P})$: $D_X = (\text{Cov}(X_i, X_j))_{1 \le i,j \le d}$ est une matrice symétrique associée à D_X et est positive. En effet, $\forall \lambda = (\lambda_1, \ldots, \lambda_d) \in \mathbb{R}^d$:

$$\text{Var}(\sum_{i=1}^{d} \lambda_i X_i) = \text{Cov}(\sum_{i=1}^{d} \lambda_i X_i, \sum_{j=1}^{d} \lambda_j X_j) = \sum_{i=1}^{d} \sum_{j=1}^{d} \lambda_i \lambda_j \text{Cov}(X_i, X_j) \ge 0$$

Si la forme bilinéaire associée à D_X n'est pas définie positive (i.e. D_X non inversible), alors $\exists (\lambda_1, \ldots, \lambda_d) \neq 0$ t.q. $\lambda_1 X_1 + \cdots + \lambda_d X_d = \lambda_0$ p.s.

Exemple 10: Problème de Bercy (Approximation affine)

Question : Soient $X, Y_1, \ldots, Y_d$ des v.a. de carrés intégrables. Quelle est la meilleure approximation de X par une fonction affine de $Y_1, \ldots, Y_d$? Autrement dit, on cherche à minimiser $(\beta_0, \ldots, \beta_d) \mapsto \mathbb{E}[|X - \beta_0 - \beta_1 Y_1 - \cdots - \beta_d Y_d|^2]$.

Proposition : On a $\inf_{\beta_0,\ldots,\beta_d} \mathbb{E}[|X - \beta_0 - \cdots - \beta_d Y_d|^2] = \mathbb{E}[|X - Z|^2]$ où $Z = \mathbb{E}[X] + \sum_{j=1}^{d} \alpha_j (Y_j - \mathbb{E}[Y_j])$ et les coefficients α étant n'importe quelle solution de :

$$\sum_{j=1}^{d} \alpha_j \text{Cov}(Y_j, Y_k) = \text{Cov}(X, Y_k), \forall k \in [\![1, d]\!]$$

Si D_Y est inversible, $\alpha = D_Y^{-1} \text{Cov}(X, Y)$.

Preuve : Soit H le sous-espace de $L^2(\mathbb{P})$. $H = \text{Vect}(1, Y_1 - \mathbb{E}[Y_1], \ldots, Y_d - \mathbb{E}[Y_d])$. C'est un s.e.v.

de dim finie de $L^2(\mathbb{P})$. La variable aléatoire $Z \in H$ qui minimise $U \mapsto \|X-U\|_2$ sur H est la projection orthogonale de X sur H. i.e. $X-Z \perp$ à tous les éléments de H. En notant $Z = \alpha_0 + \sum_{j=1}^d \alpha_j (Y_j - \mathbb{E}[Y_j])$. Nécessairement $\mathbb{E}[Z] = \alpha_0 = \mathbb{E}[X]$ (car orthogonal à 1). $\forall k \in [\![1,d]\!]$:

$$\begin{aligned}
0 &= \mathbb{E}[(X-Z)(Y_k - \mathbb{E}[Y_k])] \\
&= \mathbb{E}[(X - \mathbb{E}[X] + \mathbb{E}[X] - Z)(Y_k - \mathbb{E}[Y_k])] \\
&= \mathrm{Cov}(X, Y_k) - \mathbb{E}[(Z - \mathbb{E}[Z])(Y_k - \mathbb{E}[Y_k])] \\
&= \mathrm{Cov}(X, Y_k) - \mathbb{E}\left[\sum_{j=1}^d \alpha_j (Y_j - \mathbb{E}[Y_j])(Y_k - \mathbb{E}[Y_k])\right] \\
&= \mathrm{Cov}(X, Y_k) - \sum_{j=1}^d \alpha_j \mathrm{Cov}(Y_j, Y_k)
\end{aligned}$$

Réciproquement, si $\alpha_1, \dots, \alpha_d$ vérifient les conditions trouvées, alors $X - Z$ est orthogonal à H donc Z minimise bien la fonction considérée.

Exemple 11: Exercice

Si X et Y sont 2 v.a. réelles ayant la même fonction de répartition, alors elles ont la même loi.
Réponse : $\forall x \in \mathbb{R}, \mathbb{P}(X \leq x) = \mathbb{P}(Y \leq x)$. Alors $\mathbb{P}^X$ et $\mathbb{P}^Y$ coïncident sur les éléments de la classe $\mathcal{C} = \{]-\infty, x], x \in \mathbb{R}\}$. Mais $\mathcal{C}$ est stable par intersections finies et $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$. Et $\mathbb{P}^X(\mathbb{R}) = \mathbb{P}^Y(\mathbb{R}) = 1$. Donc $\mathbb{P}^X = \mathbb{P}^Y$.

V. Fonction caractéristique

Définition 26: Transformée de Fourier d'une mesure

Soit μ une mesure de probabilité sur $\mathbb{R}^d$. On appelle transformée de Fourier la fonction définie sur $\mathbb{R}^d$ par :

$$\hat{\mu}(\xi) = \int_{\mathbb{R}^d} e^{i\xi \cdot x} \mu(dx)$$

où $\xi \cdot x = \xi_1 x_1 + \cdots + \xi_d x_d$.

Proposition 35: Injectivité

La fonction définie sur l'ensemble des probabilités sur $\mathbb{R}^d$ par $F[\mu] : \mu \to \hat{\mu}$ est injective.

Définition 27: Fonction caractéristique

Soit X une v.a. à valeurs dans $\mathbb{R}^d$. On appelle fonction caractéristique de X la fonction définie sur $\mathbb{R}^d$ par :

$$\varphi_X(\xi) = \mathbb{E}[e^{i\xi \cdot X}]$$

où $\xi \cdot X = \xi_1 X_1 + \cdots + \xi_d x_d$.

Remarque

$F(\mathbb{P}^X)(\xi) = \varphi_X(\xi)$. $\varphi_X = \varphi_Y \implies X$ et Y ont la même loi.

Chapitre 6 : Indépendance

I. Événements indépendants

Définition 28: Indépendance de deux événements

Deux événements $A, B \in \mathcal{A}$ sont indépendants si et seulement si :

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

Remarque

1. Deux événements indépendants sous $\mathbb{P}$ peuvent ne plus l'être sous une autre probabilité.
2. Tous les événements de probabilité 0 ou 1 sont indépendants de tous les événements.

Exemple 12: Exercice

Si A et B sont indépendants, alors :

- A^c et B sont indépendants.
- A et B^c sont indépendants.
- A^c et B^c sont indépendants.

(Preuve laissée en exercice)

Définition 29: Probabilité conditionnelle

Soit $A \in \mathcal{A}$ t.q. $\mathbb{P}(A) > 0$. On appelle probabilité « sachant A » la probabilité définie par :

$$\mathbb{P}(\cdot|A) : B \in \mathcal{A} \longmapsto \mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

Remarque

1. Si A et B sont indépendants (et $\mathbb{P}(A) > 0$), alors $\mathbb{P}(B|A) = \mathbb{P}(B)$.
2. Il ne faut pas confondre « indépendant » et « incompatible » (disjoints).

Définition 30: Indépendance mutuelle

Soit un entier $n \geq 2$. On dit que des événements $A_1, \ldots, A_n$ sont mutuellement indépendants si $\forall p \in [\![2, n]\!]$ et $\forall \{i_1, \ldots, i_p\} \subset [\![1, n]\!]$ (2 à 2 distincts), on a :

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_p}) = \mathbb{P}(A_{i_1}) \times \cdots \times \mathbb{P}(A_{i_p})$$

Exemple 13: Indépendance 2 à 2 vs Mutuelle

A_1, A_2, A_3 sont mutuellement indépendants ssi :

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

- ET $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$
- ET $\mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_3)$
- ET $\mathbb{P}(A_2 \cap A_3) = \mathbb{P}(A_2)\mathbb{P}(A_3)$

Exemple 14: Lancer de dés

Lancer de dés distincts à 6 faces, équilibrés (Rouge et Vert).

- $E =$ Rouge tombe sur 4.
- $F =$ Somme égale à 6 $\leftarrow \{(1,5),(2,4),\dots,(5,1)\}$ (5 cas).
- $G =$ Somme égale à 7 $\leftarrow \{(1,6),\dots,(6,1)\}$ (6 cas).

E est indépendant de G mais E n'est pas indépendant de F.

- $\mathbb{P}(E) = \frac{1}{6}$.
- $\mathbb{P}(E \cap F) = \frac{1}{36} \neq \mathbb{P}(E)\mathbb{P}(F) = \frac{1}{6} \times \frac{5}{36}$.
- $\mathbb{P}(E \cap G) = \frac{1}{36} = \mathbb{P}(E)\mathbb{P}(G)$ (car $\mathbb{P}(G) = \frac{6}{36} = \frac{1}{6}$).

Définition 31: Système complet d'événements

Soit I un ensemble dénombrement (fini ou en bijection avec $\mathbb{N}$). On dit qu'une famille d'événements $(A_i)_{i\in I}$ est un système complet d'événements ssi :

- Les $(A_i)_{i\in I}$ sont 2 à 2 disjoints (incompatibles).
- $\sum_{i\in I} \mathbb{P}(A_i) = \mathbb{P}(\bigcup_{i\in I} A_i) = 1$.

Proposition 36: Formule des probabilités totales

Soit $(A_i)_{i\in I}$ un système complet d'événements (Causes). $\forall B \in \mathcal{A}$ (Conséquence) :

$$\mathbb{P}(B) = \sum_{i\in I} \mathbb{P}(B \cap A_i)$$

Si $\forall i \in I, \mathbb{P}(A_i) > 0$, alors :

$$\mathbb{P}(B) = \sum_{i\in I} \mathbb{P}(B|A_i)\mathbb{P}(A_i)$$

Proposition 37: Formule de Bayes (Probabilité des causes)

Soit $(A_i)_{i\in I}$ un système complet d'événements t.q. $\forall i \in I, \mathbb{P}(A_i) > 0$. Alors $\forall B \in \mathcal{A}$ et $\forall j \in I$ t.q. $\mathbb{P}(B) > 0$:

$$\mathbb{P}(A_j|B) = \frac{\mathbb{P}(A_j \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_j)\mathbb{P}(A_j)}{\sum_{i\in I} \mathbb{P}(B|A_i)\mathbb{P}(A_i)}$$

Exemple 15: Questionnaire étudiant

Étudiante interrogée. Connait la réponse avec proba $p \in]0,1[$. Quelle est la probabilité qu'elle connaisse la réponse sachant qu'elle a répondu correctement ? Système complet :

- C : Connait la réponse.

- C^c : Ne connait pas la réponse.

On cherche $\mathbb{P}(\text{Connait la réponse}|\text{réponse est correcte})$. On suppose que si elle ne connait pas, elle choisit au hasard parmi m réponses ($\mathbb{P}(\text{correct}|C^c) = 1/m$). Si elle connait, $\mathbb{P}(\text{correct}|C) = 1$.

$$= \frac{\mathbb{P}(\text{Réponse correcte}|\text{Connait})\mathbb{P}(\text{Connait})}{\mathbb{P}(\text{Correct}|\text{Connait})\mathbb{P}(\text{Connait}) + \mathbb{P}(\text{Correct}|\text{Connait pas})\mathbb{P}(\text{Connait pas})}$$

$$= \frac{1 \times p}{1 \times p + \frac{1}{m}(1-p)} \xrightarrow[m\to\infty]{} 1$$

II. Tribus indépendantes et variables aléatoires indépendantes

Soit $n \geq 2$.

Définition 32: Indépendance de sous-tribus

Soient $\mathcal{B}_1, \dots, \mathcal{B}_n$ des sous-tribus de $\mathcal{A}$. On dit que $\mathcal{B}_1, \dots, \mathcal{B}_n$ sont indépendantes ssi $\forall A_1 \in \mathcal{B}_1, \dots, \forall A_n \in \mathcal{B}_n$:

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \times \dots \times \mathbb{P}(A_n)$$

Remarque

1. $A, B \in \mathcal{A}$ sont indépendants ssi $\sigma(\{A\})$ et $\sigma(\{B\})$ sont indépendantes. Rappel : $\sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\}$.
2. Si $\mathcal{B}_1, \dots, \mathcal{B}_n$ sont indépendantes, $\forall p \in [\![2, n]\!]$ et $\forall (i_1, \dots, i_p) \in [\![1, n]\!]$ (2 à 2 distincts), $\mathcal{B}_{i_1}, \dots, \mathcal{B}_{i_p}$ sont indépendantes.

On suppose donnés des espaces mesurables $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$ et $(F_1, \mathcal{F}_1), \dots, (F_n, \mathcal{F}_n)$.

Définition 33: Variables aléatoires indépendantes

Soient $X_1, \dots, X_n$ des v.a. à valeurs dans $E_1, \dots, E_n$. On dit que $X_1, \dots, X_n$ sont indépendantes ssi $\sigma(X_1), \dots, \sigma(X_n)$ sont indépendantes.

Remarque

1. $X_1, \dots, X_n$ sont indépendantes ssi $\forall A_1 \in \mathcal{E}_1, \dots, \forall A_n \in \mathcal{E}_n$:
$$\mathbb{P}(X_1 \in A_1 \cap \dots \cap X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \times \dots \times \mathbb{P}(X_n \in A_n)$$
Rappel : $\sigma(X_i) = \{X_i^{-1}(A) : A \in \mathcal{E}_i\}$.
2. S'il existe des sous-tribus sur Ω, $\mathcal{B}_1, \dots, \mathcal{B}_n$ t.q. :
   - $\forall i \in [\![1, n]\!]$, X_i est $\mathcal{B}_i - \mathcal{E}_i$ mesurable.
   - $\mathcal{B}_1, \dots, \mathcal{B}_n$ sont indépendantes.

   Alors $X_1, \dots, X_n$ sont indépendantes.
3. Si $X_1, \dots, X_n$ sont indépendantes et $\forall i \in [\![1, n]\!]$, $f_i : E_i \to F_i$ mesurable, alors : $f_1(X_1), \dots, f_n(X_n)$ sont indépendantes.

Lemme 2: Indépendance et π-systèmes

Soient $\mathcal{B}_1, \dots, \mathcal{B}_n$ des sous-tribus de $\mathcal{A}$. On suppose que $\forall i \in [\![1, n]\!]$, $\exists \mathcal{C}_i \subset \mathcal{B}_i$ t.q. :

- $\sigma(\mathcal{C}_i) = \mathcal{B}_i$
- $\mathcal{C}_i$ est stable par intersection finie (π-système).
- $\Omega \in \mathcal{C}_i$

et $\forall C_1 \in \mathcal{C}_1, \dots, C_n \in \mathcal{C}_n$:

$$\mathbb{P}(C_1 \cap \dots \cap C_n) = \mathbb{P}(C_1) \times \dots \times \mathbb{P}(C_n)$$

Alors $\mathcal{B}_1, \dots, \mathcal{B}_n$ sont indépendantes.

Preuve :

La preuve se fait en n étapes. **1-** Soient $C_2 \in \mathcal{C}_2, \dots, C_n \in \mathcal{C}_n$ quelconques mais fixés. Soit $\mathcal{M}_1 = \{B_1 \in \mathcal{B}_1 : \mathbb{P}(B_1 \cap C_2 \cap \dots \cap C_n) = \mathbb{P}(B_1)\mathbb{P}(C_2)\dots\mathbb{P}(C_n)\}$. $\mathcal{M}_1$ est une classe monotone :

- $\Omega \in \mathcal{M}_1$ car $\mathbb{P}(\Omega \cap C_2 \dots) = \mathbb{P}(C_2 \dots) = 1 \times \mathbb{P}(C_2)\dots$
- Soient $A, B \in \mathcal{M}_1$ t.q. $A \subset B$. $\mathbb{P}((B \setminus A) \cap \dots) = \mathbb{P}(B \cap \dots) - \mathbb{P}(A \cap \dots)$ par linéarité... $= (\mathbb{P}(B) - \mathbb{P}(A))\mathbb{P}(C_2) \dots = \mathbb{P}(B \setminus A)\mathbb{P}(C_2)\dots$ Donc $B \setminus A \in \mathcal{M}_1$.
- Soit $(B_m)_{m \geq 0}$ une suite croissante d'éléments de $\mathcal{M}_1$. Par continuité croissante de $\mathbb{P}$: $\mathbb{P}((\bigcup B_m) \cap C_2 \dots) = \lim \mathbb{P}(B_m \cap \dots) = \lim \mathbb{P}(B_m)\mathbb{P}(C_2) \dots = \mathbb{P}(\bigcup B_m)\mathbb{P}(C_2)\dots$ Donc $\bigcup B_m \in \mathcal{M}_1$.

Comme $\mathcal{C}_1 \subset \mathcal{M}_1$, on a $\mathcal{M}(\mathcal{C}_1) \subset \mathcal{M}_1$. Mais $\mathcal{C}_1$ est stable par intersection finie, donc $\mathcal{B}_1 = \sigma(\mathcal{C}_1) = \mathcal{M}(\mathcal{C}_1) \subset \mathcal{M}_1$. Donc $\forall B_1 \in \mathcal{B}_1, \forall C_2 \in \mathcal{C}_2, \dots, \forall C_n \in \mathcal{C}_n$, l'égalité produit est vraie.
2- Soient $B_1 \in \mathcal{B}_1, C_3 \in \mathcal{C}_3, \dots, C_n \in \mathcal{C}_n$ quelconques mais fixés. $\mathcal{M}_2 = \{B_2 \in \mathcal{B}_2 : \mathbb{P}(B_1 \cap B_2 \cap C_3 \dots) = \mathbb{P}(B_1)\mathbb{P}(B_2)\dots\}$. Est une classe monotone qui contient $\mathcal{C}_2$, donc $\mathcal{B}_2 = \sigma(\mathcal{C}_2) \subset \mathcal{M}_2$ pour les mêmes raisons qu'en (1). Donc $\forall B_1 \in \mathcal{B}_1, \forall B_2 \in \mathcal{B}_2, \forall C_3 \in \mathcal{C}_3 \dots$ l'égalité tient.
3- et plus : Et ainsi de suite jusqu'à n. □

Proposition 38: Propriété d'indépendance par paquets

Soient $\mathcal{B}_1, \dots, \mathcal{B}_n$ des sous-tribus sur Ω de $\mathcal{A}$, indépendantes. $\forall p \in [\![2, n]\!]$ et tous entiers $0 = n_0 < n_1 < \dots < n_p = n$. Les tribus :

$$\begin{aligned} \mathcal{D}_1 &= \sigma(\mathcal{B}_1, \dots, \mathcal{B}_{n_1}) \\ \mathcal{D}_j &= \sigma(\mathcal{B}_{n_{j-1}+1}, \dots, \mathcal{B}_{n_j}) \\ &\vdots \\ \mathcal{D}_p &= \sigma(\mathcal{B}_{n_{p-1}+1}, \dots, \mathcal{B}_{n_p}) \end{aligned}$$

sont indépendantes.

Preuve :

$\forall j \in [\![1, p]\!]$, on pose :

$$\mathcal{G}_j = \{B_{n_{j-1}+1} \cap \dots \cap B_{n_j} \mid B_k \in \mathcal{B}_k \text{ pour } k \in [\![n_{j-1}+1, n_j]\!]\} \subset \mathcal{D}_j$$

Les $\mathcal{G}_j$ sont des π-systèmes (stables par intersection) car l'intersection de deux éléments de $\mathcal{G}_j$ reste une intersection d'éléments des $\mathcal{B}_k$ (qui sont des tribus, donc stables par intersection). $\Omega \in \mathcal{G}_j$.
De plus, les $\mathcal{G}_j$ vérifient la condition d'indépendance produit car les $\mathcal{B}_k$ sont toutes indépendantes entre elles. D'après le Lemme précédent, il suffit de montrer que $\sigma(\mathcal{G}_j) = \mathcal{D}_j$.

* Puisque $\mathcal{G}_j \subset \mathcal{D}_j$, on a $\sigma(\mathcal{G}_j) \subset \mathcal{D}_j$.
* Réciproquement : $\forall i \in \{n_{j-1}+1, \dots, n_j\}$, $\forall B \in \mathcal{B}_i$, on a :

$$B = \Omega \cap \dots \cap \Omega \cap B \cap \Omega \cap \dots \cap \Omega \in \mathcal{G}_j$$

Donc $\forall i \in \{n_{j-1}+1, \dots, n_j\}$, $\mathcal{B}_i \subset \mathcal{G}_j \subset \sigma(\mathcal{G}_j)$. Donc :

$$\sigma(\mathcal{B}_{n_{j-1}+1}, \dots, \mathcal{B}_{n_j}) = \sigma\left(\bigcup_{i=n_{j-1}+1}^{n_j} \mathcal{B}_i\right) \subset \sigma(\mathcal{G}_j)$$

Donc $\mathcal{D}_j \subset \sigma(\mathcal{G}_j)$.
Conclusion : $\mathcal{D}_j = \sigma(\mathcal{G}_j)$ et les $\mathcal{D}_j$ sont indépendantes. $\square$

Remarque

Si $X_1, \dots, X_n$ sont des V.A. indépendantes $X_i : \Omega \to E_i$. Et pour $0 = n_0 < n_1 < \dots < n_p = n$, on a des fonctions :

$$\begin{aligned} f_1 &: E_1 \times \dots \times E_{n_1} \to F_1 \\ f_j &: E_{n_{j-1}+1} \times \dots \times E_{n_j} \to F_j \\ &\vdots \end{aligned}$$

mesurables, alors : $f_1(X_1, \dots, X_{n_1}), \dots, f_j(X_{n_{j-1}+1}, \dots, X_{n_j}), \dots$ sont indépendantes.

Proposition 39

Les v.a. $X_1, \dots, X_n$ sont indépendantes si et seulement si :

$$\mathbb{P}^{(X_1, \dots, X_n)} = \mathbb{P}^{X_1} \otimes \dots \otimes \mathbb{P}^{X_n}$$

Ceci est une égalité sur $E_1 \otimes \dots \otimes E_n$.
Si $X_1, \dots, X_n$ sont indépendantes, pour toutes fonctions mesurables positives $f_i : E_i \to \mathbb{R}^+$ $(1 \leq i \leq n)$, alors :

$$\mathbb{E}\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}[f_i(X_i)]$$

Preuve :

$X_1, \dots, X_n$ sont indépendantes ssi $\forall F_1 \in \mathcal{E}_1, \dots, F_n \in \mathcal{E}_n$:

$$\begin{aligned} \mathbb{P}^{(X_1, \dots, X_n)}(F_1 \times \dots \times F_n) &= \mathbb{P}((X_1, \dots, X_n) \in F_1 \times \dots \times F_n) \\ &= \mathbb{P}(X_1 \in F_1, \dots, X_n \in F_n) \\ &= \mathbb{P}(X_1 \in F_1) \times \dots \times \mathbb{P}(X_n \in F_n) \\ &= \mathbb{P}^{X_1}(F_1) \times \dots \times \mathbb{P}^{X_n}(F_n) \\ &= (\mathbb{P}^{X_1} \otimes \dots \otimes \mathbb{P}^{X_n})(F_1 \times \dots \times F_n) \end{aligned}$$

Donc $X_1, \dots, X_n$ sont indépendantes ssi $\mathbb{P}^{(X_1, \dots, X_n)}$ et $\mathbb{P}^{X_1} \otimes \dots \otimes \mathbb{P}^{X_n}$ coïncident sur la classe des pavés. Cette classe :

- est stable par intersection finie,
- engendre $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$,

- contient $E_1 \times \cdots \times E_n$ (qui vérifie l'égalité).

Donc $X_1, \ldots, X_n$ sont indépendantes ssi $\mathbb{P}^{(X_1,\ldots,X_n)} = \mathbb{P}^{X_1} \otimes \cdots \otimes \mathbb{P}^{X_n}$.
Finalement, si $X_1, \ldots, X_n$ sont indépendantes et $\forall i \in [\![1, n]\!]$, $f_i : E_i \to \mathbb{R}^+$ sont mesurables positives :

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^{n} f_i(X_i)\right] &= \int_{E_1 \times \cdots \times E_n} f_1(x_1) \ldots f_n(x_n) d\mathbb{P}^{(X_1,\ldots,X_n)} \\
&= \int (f_1(x_1) \ldots f_n(x_n)) d(\mathbb{P}^{X_1} \otimes \cdots \otimes \mathbb{P}^{X_n}) \\
&= \int_{E_1 \times \cdots \times E_{n-1}} (f_1 \ldots f_{n-1}) \left(\int_{E_n} f_n(x_n) d\mathbb{P}^{X_n}(x_n) \right) d\mathbb{P}^{X_1} \otimes \cdots \otimes d\mathbb{P}^{X_{n-1}} \\
&\quad \text{(Fubini-Tonelli car positif)} \\
&= \mathbb{E}[f_n(X_n)] \times \cdots \times \mathbb{E}[f_1(X_1)] \\
&= \prod_{i=1}^{n} \mathbb{E}[f_i(X_i)]
\end{aligned}$$

□

Proposition 40

Si $X_1, \ldots, X_n$ sont indépendantes ($X_i : \Omega \to E_i$) et on se donne $g_1, \ldots, g_n$ mesurables $g_i : E_i \to \mathbb{R}$.
Alors si $\forall i \in [\![1, n]\!]$, $g_i(X_i)$ est intégrable, on a que $\prod_{i=1}^{n} g_i(X_i)$ est intégrable et :

$$\mathbb{E}\left[\prod_{i=1}^{n} g_i(X_i)\right] = \prod_{i=1}^{n} \mathbb{E}[g_i(X_i)]$$

Remarque

Si X_1 et X_2 sont des v.a.r. indépendantes et intégrables, alors $X_1 X_2$ est intégrable.
Sans l'hypothèse d'indépendance, "X_1 et X_2 sont intégrables" ne suffit pas pour garantir que $X_1 X_2$ est intégrable.
Cependant, si $X_1, X_2 \in L^2(\mathbb{P})$ alors $X_1 X_2 \in L^1(\mathbb{P})$ (Cauchy-Schwarz).

Remarque

X_1 et X_n sont indépendantes ssi pour toutes les fonctions $g_1, \ldots, g_n$ mesurables **bornées** $g_i : E_i \to \mathbb{R}$, on a :

$$\mathbb{E}\left[\prod_{i=1}^{n} g_i(X_i)\right] = \prod_{i=1}^{n} \mathbb{E}[g_i(X_i)]$$

Remarque

Si X_1 et X_2 sont indépendantes et intégrables :

$$\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2] \implies \mathrm{Cov}(X_1, X_2) = 0$$

La réciproque est fausse : il existe des v.a. t.q. $\mathrm{Cov}(X_1, X_2) = 0$ mais X_1 et X_2 ne sont pas indépendantes.

Preuve de la proposition :

Par récurrence :
$\underline{\mathcal{P}(2)}$: On pose $f(x_1, x_2) = g_1(x_1)g_2(x_2)$.
On a :

$$\begin{aligned} f^+(x_1, x_2) &= g_1^+(x_1)g_2^+(x_2) + g_1^-(x_1)g_2^-(x_2) \\ f^-(x_1, x_2) &= g_1^+(x_1)g_2^-(x_2) + g_1^-(x_1)g_2^+(x_2) \end{aligned}$$

Donc :

$$\begin{aligned} \mathbb{E}[f^+(X_1, X_2)] &= \mathbb{E}[g_1^+(X_1)g_2^+(X_2) + g_1^-(X_1)g_2^-(X_2)] \\ &= \mathbb{E}[g_1^+(X_1)]\mathbb{E}[g_2^+(X_2)] + \mathbb{E}[g_1^-(X_1)]\mathbb{E}[g_2^-(X_2)] < +\infty \end{aligned}$$

(car les espérances sont finies par hypothèse d'intégrabilité).
De même, $\mathbb{E}[f^-(X_1, X_2)] < +\infty$.
Donc $f(X_1, X_2) = g_1(X_1)g_2(X_2)$ est intégrable.
De plus :

$$\begin{aligned} \mathbb{E}[g_1(X_1)g_2(X_2)] &= \mathbb{E}\left[(g_1^+(X_1) - g_1^-(X_1))(g_2^+(X_2) - g_2^-(X_2))\right] \\ &= \mathbb{E}\left[g_1^+(X_1)g_2^+(X_2) - g_1^+(X_1)g_2^-(X_2) - g_1^-(X_1)g_2^+(X_2) + g_1^-(X_1)g_2^-(X_2)\right] \\ &= \mathbb{E}[g_1^+(X_1)g_2^+(X_2)] - \mathbb{E}[g_1^+(X_1)g_2^-(X_2)] - \mathbb{E}[g_1^-(X_1)g_2^+(X_2)] + \mathbb{E}[g_1^-(X_1)g_2^-(X_2)] \\ &= \mathbb{E}[g_1^+(X_1)]\mathbb{E}[g_2^+(X_2)] - \mathbb{E}[g_1^+(X_1)]\mathbb{E}[g_2^-(X_2)] - \mathbb{E}[g_1^-(X_1)]\mathbb{E}[g_2^+(X_2)] + \mathbb{E}[g_1^-(X_1)]\mathbb{E}[g_2^-(X_2)] \\ &= \left(\mathbb{E}[g_1^+(X_1)] - \mathbb{E}[g_1^-(X_1)]\right)\left(\mathbb{E}[g_2^+(X_2)] - \mathbb{E}[g_2^-(X_2)]\right) \\ &= \mathbb{E}[g_1(X_1)]\mathbb{E}[g_2(X_2)] \end{aligned}$$

Soit n quelconque mais fixé. On suppose que $\mathcal{P}(n)$ est vraie.
Soient $g_1, \dots, g_{n+1}$ des fonctions mesurables $g_i : E_i \to \mathbb{R}$ telles que $g_1(X_1), \dots, g_{n+1}(X_{n+1})$ sont intégrables.
On a donc $g_1(X_1) \dots g_n(X_n)g_{n+1}(X_{n+1}) = h(X_1, \dots, X_n)g_{n+1}(X_{n+1})$.
Avec $h(X_1, \dots, X_n)$ intégrable d'après $\mathcal{P}(n)$.
C'est le produit de 2 v.a. intégrables et indépendantes (indépendance par paquets). Donc $g_1(X_1) \dots g_{n+1}(X_{n+1})$ est intégrable d'après $\mathcal{P}(2)$ et :

$$\begin{aligned} \mathbb{E}[g_1(X_1) \dots g_{n+1}(X_{n+1})] &= \mathbb{E}[h(X_1 \dots X_n)g_{n+1}(X_{n+1})] \\ &= \mathbb{E}[g_1(X_1) \dots g_n(X_n)]\mathbb{E}[g_{n+1}(X_{n+1})] \quad (\text{d'après } \mathcal{P}(2)) \\ &= \prod_{i=1}^{n} \mathbb{E}[g_i(X_i)] \times \mathbb{E}[g_{n+1}(X_{n+1})] \quad (\text{d'après } \mathcal{P}(n)) \end{aligned}$$

La propriété est donc vraie $\forall n \in \mathbb{N}$ ($n \geq 2$). $\square$

Proposition 41

Soient $X_1, X_2, \dots, X_n$ des v.a. réelles.

1. Si $(X_1, \dots, X_n)$ admet $p(x_1, \dots, x_n)$ comme densité, alors chaque X_i admet $x_i \mapsto \int_{\mathbb{R}^{n-1}} p(x_1, \dots, x_n)dx_1 \dots dx_{i-1}dx_{i+1} \dots dx_n$ comme densité.
2. Si $X_1, \dots, X_n$ sont indépendantes et chaque X_i admet une densité p_i, alors $(X_1, \dots, X_n)$ admet $\prod_{i=1}^n p_i$ pour densité.
3. S'il existe des fonctions mesurables positives $g_1, \dots, g_n$ telles que $(X_1, \dots, X_n)$ admet $\prod_{i=1}^n g_i(x_i)$ pour densité, alors $X_1, \dots, X_n$ sont indépendantes. (Et pour tout i, $\exists c_i > 0$ t.q. X_i admet $c_i g_i$ comme densité).

Preuve :

1) Soit $A \in \mathcal{B}(\mathbb{R})$. On a que $\forall i \in [\![1, n]\!]$:

$$\begin{aligned}
\mathbb{P}^{X_i}(A) &= \mathbb{P}(X_i \in A) \\
&= \mathbb{P}((X_1, \ldots, X_i, \ldots, X_n) \in \mathbb{R} \times \cdots \times A \times \cdots \times \mathbb{R}) \\
&= \int_{\mathbb{R} \times \cdots \times A \times \cdots \times \mathbb{R}} p(x_1, \ldots, x_n) dx_1 \ldots dx_n \\
&= \int_A \left(\int_{\mathbb{R}^{n-1}} p(x_1, \ldots, x_n) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n \right) dx_i \quad \text{(Fubini positif)}
\end{aligned}$$

D'où $x_i \mapsto \int_{\mathbb{R}^{n-1}} p(x_1, \ldots, x_n) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n$ est une densité pour X_i.

2) Pour toute fonction $g : \mathbb{R}^n \to \mathbb{R}$ mesurable bornée :

$$\begin{aligned}
\mathbb{E}[g(X_1, \ldots, X_n)] &= \int_{\mathbb{R}^n} g(x_1, \ldots, x_n) d\mathbb{P}^{(X_1, \ldots, X_n)}(x_1, \ldots, x_n) \\
&= \int_{\mathbb{R}^n} g(x_1, \ldots, x_n) d(\mathbb{P}^{X_1} \otimes \cdots \otimes \mathbb{P}^{X_n})(dx_1 \ldots dx_n) \\
&\quad \text{(Fubini-Lebesgue car g est bornée)} \\
&= \int_{\mathbb{R}^{n-1}} \left(\int_{\mathbb{R}} g(x_1 \ldots x_n) p_n(x_n) dx_n \right) d\mathbb{P}^{(X_1 \ldots X_{n-1})}(dx_1 \ldots dx_{n-1}) \\
&= \ldots \text{ (Récurrence / Fubini successifs)} \\
&= \int_{\mathbb{R}^n} g(x_1, \ldots, x_n) p_1(x_1) \ldots p_n(x_n) dx_1 \ldots dx_n
\end{aligned}$$

3) Puisque $\prod_{i=1}^n g_i$ est une densité, on a :

$$1 = \int_{\mathbb{R}^n} \left(\prod_{i=1}^n g_i(x_i) \right) dx_1 \ldots dx_n = \prod_{i=1}^n \left(\int_{\mathbb{R}} g_i(x_i) dx_i \right)$$

Donc en posant $K_i = \int_{\mathbb{R}} g_i(x_i) dx_i$, on a $K_i \in]0, +\infty[$ et $\prod K_i = 1$.
D'après le premier point de la proposition, chaque X_i admet une densité qui se déduit de $\prod g_i$ en intégrant :

$$\begin{aligned}
p_i(x_i) &= \int_{\mathbb{R}^{n-1}} g_1(x_1) \ldots g_i(x_i) \ldots g_n(x_n) dx_1 \ldots \hat{dx_i} \ldots dx_n \\
&= g_i(x_i) \int_{\mathbb{R}^{n-1}} \prod_{j \neq i} g_j(x_j) d(\ldots) \\
&= g_i(x_i) \prod_{j \neq i} \left(\int_{\mathbb{R}} g_j(x_j) dx_j \right) \\
&= \frac{g_i(x_i)}{K_i} \quad \left(\text{car } \prod K_j = 1\right)
\end{aligned}$$

Reste à montrer que $X_1 \ldots X_n$ sont indépendantes.
Pour tout choix de fonctions boréliennes bornées $h_1 \ldots h_n$, on a :

$$\begin{aligned}
\mathbb{E}[h_1(X_1) \ldots h_n(X_n)] &= \int_{\mathbb{R}^n} h_1(x_1) \ldots h_n(x_n) g_1(x_1) \ldots g_n(x_n) dx_1 \ldots dx_n \\
&= \int_{\mathbb{R}^n} h_1(x_1) \ldots h_n(x_n) \frac{g_1(x_1)}{K_1} \cdots \frac{g_n(x_n)}{K_n} dx_1 \ldots dx_n \quad \left(\text{car } \prod K_i = 1\right) \\
&= \prod_{i=1}^n \left(\int_{\mathbb{R}} h_i(x_i) \frac{g_i(x_i)}{K_i} dx_i \right) \\
&= \prod_{i=1}^n \mathbb{E}[h_i(X_i)]
\end{aligned}$$

Donc $X_1 \dots X_n$ sont indépendantes. $\square$

Proposition 42

Les composantes d'un vecteur aléatoire $(X_1, \dots, X_n) \in \mathbb{R}^n$ sont indépendantes ssi $\forall \xi \in \mathbb{R}^n$ on a :

$$\varphi_{(X_1,\dots,X_n)}(\xi) = \varphi_{X_1}(\xi_1) \dots \varphi_{X_n}(\xi_n)$$

(Où $\xi = (\xi_1, \dots, \xi_n)$).

Preuve :

Sens direct ($\implies$) : Supposons que $X_1 \dots X_n$ sont indépendantes. Alors $\forall \xi = (\xi_1, \dots, \xi_n)$:

$$\begin{aligned}\varphi_{(X_1,\dots,X_n)}(\xi) &= \mathbb{E}\left[e^{i\sum_{j=1}^n \xi_j X_j}\right] \\ &= \mathbb{E}\left[\prod_{j=1}^n e^{i\xi_j X_j}\right]\end{aligned}$$

Or les fonctions $u \mapsto e^{i\xi_j u}$ sont mesurables bornées.
Donc, par indépendance :

$$= \prod_{j=1}^n \mathbb{E}[e^{i\xi_j X_j}] = \varphi_{X_1}(\xi_1) \dots \varphi_{X_n}(\xi_n)$$

Réciproquement ($\impliedby$) : Soit $\xi = (\xi_1, \dots, \xi_n)$.

$$\begin{aligned}\varphi_{(X_1,\dots,X_n)}(\xi) &= \mathbb{E}\left[e^{i\sum_{j=1}^n \xi_j X_j}\right] = \int_{\mathbb{R}^n} e^{i\sum_{j=1}^n \xi_j x_j} d\mathbb{P}^{(X_1 \dots X_n)}(x_1 \dots x_n) \\ &= \varphi_{\mathbb{P}^{(X_1 \dots X_n)}}(\xi_1, \dots, \xi_n)\end{aligned}$$

Et :

$$\begin{aligned}\varphi_{X_1}(\xi_1) \dots \varphi_{X_n}(\xi_n) &= \prod_{j=1}^n \mathbb{E}[e^{i\xi_j X_j}] = \prod_{j=1}^n \int_{\mathbb{R}} e^{i\xi_j x_j} d\mathbb{P}^{X_j}(x_j) \\ &= \int_{\mathbb{R}^n} \prod_{j=1}^n e^{i\xi_j x_j} d(\mathbb{P}^{X_1} \otimes \dots \otimes \mathbb{P}^{X_n}) \\ &= \varphi_{\mathbb{P}^{X_1} \otimes \dots \otimes \mathbb{P}^{X_n}}(\xi_1, \dots, \xi_n)\end{aligned}$$

Comme la transformation de Fourier est injective :

$$\mathbb{P}^{(X_1 \dots X_n)} = \mathbb{P}^{X_1} \otimes \dots \otimes \mathbb{P}^{X_n}$$

Donc $X_1, \dots, X_n$ sont indépendantes. $\square$

III. Sommes de variables aléatoires

Proposition 43

On considère l'espace mesuré $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$ et $f, g \in L^1(\lambda)$. On a :

$$\int_{\mathbb{R}^d} |f(x-y)g(y)|\lambda(dy) < +\infty \quad \text{pour } \lambda\text{-presque tout } x \in \mathbb{R}^d$$

et même :

$$\int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} f(x-y)g(y)\lambda(dy) \right| \lambda(dx) \leq \|f\|_1 \|g\|_1$$

Preuve :

D'après le théorème de Fubini :

$$\begin{aligned}
\int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f(x-y)||g(y)|\lambda(dy) \right) \lambda(dx) &= \int_{\mathbb{R}^d} |g(y)| \left(\int_{\mathbb{R}^d} |f(x-y)|\lambda(dx) \right) \lambda(dy) \\
&\quad \text{(changement de variable } u = x - y\text{)} \\
&= \int_{\mathbb{R}^d} |g(y)| \left(\int_{\mathbb{R}^d} |f(u)|\lambda(du) \right) \lambda(dy) \\
&= \left(\int_{\mathbb{R}^d} |g(y)|\lambda(dy) \right) \left(\int_{\mathbb{R}^d} |f(u)|\lambda(du) \right) \\
&= \|g\|_1 \|f\|_1 < +\infty
\end{aligned}$$

On a donc que pour λ-presque tout $x \in \mathbb{R}^d$:

$$\int_{\mathbb{R}^d} |f(x-y)||g(y)|dy < +\infty \text{ donc } \int_{\mathbb{R}^d} f(x-y)g(y)dy < +\infty$$

De plus :

$$\begin{aligned}
\int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} f(x-y)g(y)dy \right| \lambda(dx) &\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(x-y)||g(y)|\lambda(dy)\lambda(dx) \\
&\leq \|f\|_1 \|g\|_1
\end{aligned}$$

□

Définition 34: Produit de convolution

Soient $f, g \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$. On appelle produit de convolution de f et g la fonction définie sur $\mathbb{R}^d$ par :

$$(f * g)(x) = \begin{cases} \int_{\mathbb{R}^d} f(x-y)g(y)dy & \text{si } \int |f(x-y)g(y)|dy < +\infty \\ 0 & \text{sinon} \end{cases}$$

Proposition 44: Propriétés de la convolution

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

1. $f * g \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$
2. $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$
3. $f * g = g * f$

Preuve :

1 et 2 sont des conséquences de la proposition précédente.
3- On a $\forall x \in \mathbb{R}^d$ t.q. $\int_{\mathbb{R}^d} |f(x-y)||g(y)|dy < +\infty$:

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

(en posant $u = x - y$).
Si $\int_{\mathbb{R}^d} |f(x-y)||g(y)|dy = +\infty$, alors $(f * g) = 0$ mais aussi $\int |f(u)||g(x-u)|du = +\infty$ donc $(g * f)(x) = 0$. $\square$

Définition 35: Convolution de mesures

Soient μ et ν deux probabilités sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. On appelle produit de convolution de μ et ν la mesure de probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ qui est la mesure image de $\mu \otimes \nu$ par l'application $s : (x, y) \mapsto x + y$. Notée $\mu * \nu$.

Remarque

$$\mathbb{R}^d \times \mathbb{R}^d, \mu \otimes \nu \xrightarrow{s:(x,y)\mapsto x+y} \mathbb{R}^d, \mu * \nu$$
$$\forall A \in \mathcal{B}(\mathbb{R}^d), \quad (\mu * \nu)(A) = (\mu \otimes \nu)(s^{-1}(A))$$

Proposition 45

Soient μ et ν des mesures de probabilités sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Pour toute fonction $\varphi : \mathbb{R}^d \to \mathbb{R}$ mesurable positive ou mesurable bornée :

$$\int_{\mathbb{R}^d} \varphi(z)(\mu * \nu)(dz) = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \varphi(x+y)\mu(dx) \right) \nu(dy)$$

Preuve :

$0^{\text{ème}}$ cas : Il existe $A \in \mathcal{B}(\mathbb{R}^d)$ t.q. $\varphi = \mathbb{1}_A$. Alors $\int_{\mathbb{R}^d} \varphi(z)(\mu * \nu)(dz) = (\mu * \nu)(A) = (\mu \otimes \nu)(s^{-1}(A))$.

$$\begin{aligned}
\int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \mathbb{1}_{x+y \in A} \mu(dx) \right) \nu(dy) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{1}_A(x+y)(\mu \otimes \nu)(dx, dy) \\
&= \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{1}_{s^{-1}(A)}(x, y)(\mu \otimes \nu)(dx, dy) \\
&= (\mu \otimes \nu)(s^{-1}(A))
\end{aligned}$$

D'où l'égalité dans ce cas.
1^{er} cas : φ est une fonction étagée positive. Alors l'égalité est une conséquence du $0^{\text{ème}}$ cas et de la propriété de linéarité de l'intégrale. Etc... (Standard measure theory extension). $\square$

Proposition 46: Somme de v.a. indépendantes

Soient X et Y des v.a. indépendantes à valeurs dans $\mathbb{R}^d$.
1- La v.a. $X + Y$ est de la loi $\mathbb{P}^X * \mathbb{P}^Y$. Si de plus X admet une densité p_X et Y admet p_Y, alors $X + Y$ admet $p_X * p_Y$ comme densité.
2- On a $\varphi_{X+Y} = \varphi_X \varphi_Y$.

3- Si X et Y sont de carré intégrable :

$$D_{X+Y} = D_X + D_Y$$

Preuve :

Pour toute fonction $\varphi : \mathbb{R}^d \to \mathbb{R}$ mesurable bornée :

$$\begin{aligned}
\mathbb{E}[\varphi(X+Y)] &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \varphi(x+y) d\mathbb{P}^{(X,Y)}(dx,dy) \\
&= \int_{\mathbb{R}^d \times \mathbb{R}^d} \varphi(x+y) d(\mathbb{P}^X \otimes \mathbb{P}^Y)(dx,dy) \quad \text{(car indépendance)} \\
&= \int_{\mathbb{R}^d} \varphi(u) d(\mathbb{P}^X * \mathbb{P}^Y)(du) \quad \text{(d'après la prop précédente)}
\end{aligned}$$

Donc $X+Y$ a pour loi $\mathbb{P}^X * \mathbb{P}^Y$.
Si de plus X et Y admettent des densités alors (d'après Fubini, puisque tout est positif) :

$$\begin{aligned}
\mathbb{E}[\varphi(X+Y)] &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(x+y) p_X(x) dx \, p_Y(y) dy \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(x+y) p_X(x) dx \, p_Y(y) dy \\
&= \int_{\mathbb{R}^d \times \mathbb{R}^d} \varphi(u) p_X(u-v) p_Y(v) du dv \quad \text{(changement } u = x+y, v = y\text{)} \\
&= \int_{\mathbb{R}^d} \varphi(u) \left(\int_{\mathbb{R}^d} p_X(u-v) p_Y(v) dv \right) du \\
&= \int_{\mathbb{R}^d} \varphi(u) (p_X * p_Y)(u) du
\end{aligned}$$

Donc $X+Y$ admet $p_X * p_Y$ comme densité.
2- $\forall \xi = (\xi_1, \dots, \xi_n) \in \mathbb{R}^d$, on a :

$$\begin{aligned}
\varphi_{X+Y}(\xi) &= \mathbb{E}\left[e^{i \sum_{j=1}^d \xi_j (X_j + Y_j)}\right] \\
&= \mathbb{E}\left[e^{i \sum \xi_j X_j} \times e^{i \sum \xi_j Y_j}\right] \\
&= \mathbb{E}\left[e^{i \sum \xi_j X_j}\right] \mathbb{E}\left[e^{i \sum \xi_j Y_j}\right] \quad \text{(par indépendance)} \\
&= \varphi_X(\xi) \varphi_Y(\xi)
\end{aligned}$$

3- X est carré intégrable ssi $\forall i \in [\![1, d]\!]$, X_i est de carré intégrable. Donc si X et Y sont de carré intégrable, $X+Y$ est de carré intégrable. $\forall (i,j) \in [\![1, d]\!]^2$:

$$\begin{aligned}
[D_{X+Y}]_{i,j} &= \mathrm{Cov}(X_i + Y_i, X_j + Y_j) \\
&= \mathrm{Cov}(X_i, X_j) + \mathrm{Cov}(Y_i, Y_j) + \underbrace{\mathrm{Cov}(X_i, Y_j)}_{=0} + \underbrace{\mathrm{Cov}(Y_i, X_j)}_{=0} \\
&= [D_X]_{i,j} + [D_Y]_{i,j}
\end{aligned}$$

□

IV. Variables gaussiennes, vecteurs gaussiens

Définition 36: Variable gaussienne réelle

On dit qu'une v.a. réelle X est de loi gaussienne $\mathcal{N}(m, \sigma^2)$ ($m \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+$) si :

- $X = m$ p.s. (auquel cas $\sigma^2 = 0$)
- OU $\text{Var}[X] \neq 0$ et X admet $x \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$ comme densité.

Proposition 47: Propriétés $\mathcal{N}(m, \sigma^2)$

Si $X \sim \mathcal{N}(m, \sigma^2)$ alors $\mathbb{E}[X] = m$, $\text{Var}[X] = \sigma^2$. Et $\forall t \in \mathbb{R}$:

- $\mathbb{E}[e^{tX}] = e^{tm + \frac{t^2\sigma^2}{2}}$
- $\mathbb{E}[e^{itX}] = e^{itm - \frac{t^2\sigma^2}{2}}$

Preuve :

(Voir TD, exo 6) □

Définition 37: Vecteur gaussien

Soit $X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} \in \mathbb{R}^d$ un vecteur aléatoire. On dit que X est un vecteur gaussien ssi $\forall (t_1, \ldots, t_d) \in \mathbb{R}^d, \sum_{i=1}^d t_i X_i$ est une v.a. réelle gaussienne.

Remarque

Si $X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$ est un vecteur gaussien, alors $\forall i \in [\![1, d]\!]$, X_i est une v.a. réelle gaussienne. **Attention :** La réciproque est fausse.

Proposition 48: Transformation affine

Soit $X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$ un vecteur gaussien et $B \in \mathcal{M}_{n,d}(\mathbb{R})$, $R \in \mathbb{R}^n$. Le vecteur $Y = BX + R \in \mathbb{R}^n$ est un vecteur gaussien avec :

$$\mathbb{E}[Y] = B\mathbb{E}[X] + R \quad \text{et} \quad D_Y = BD_XB^T$$

Preuve :

Toute combinaison linéaire de $Y_1 \ldots Y_n$ est une combinaison linéaire de composantes de X plus une constante, donc une v.a. réelle gaussienne. Donc Y est un vecteur gaussien. □

Proposition 49: Fonction caractéristique d'un vecteur gaussien

Soit $X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$ un vecteur gaussien. Pour tout $\xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d$, on a :

$$\varphi_X(\xi) = e^{i\xi \cdot m - \frac{1}{2}\xi^T D_X \xi}$$

où $m = \mathbb{E}[X]$ et D_X est la matrice de dispersion de X.

Preuve :

Soit $\forall \xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d$. $Y = \xi \cdot X = \xi_1 X_1 + \dots + \xi_d X_d = \xi^T X$ est une v.a. réelle gaussienne. $\mathbb{E}[\xi^T X] = \xi^T \mathbb{E}[X] = \xi^T m$. Var$[\xi^T X] = D_{\xi^T X} = \xi^T D_X \xi$. Donc $\xi \cdot X \sim \mathcal{N}(\xi^T m, \xi^T D_X \xi)$. Donc :

$$\begin{aligned} \varphi_X(\xi) = \mathbb{E}[e^{i\xi^T X}] &= \varphi_{\xi^T X}(1) \\ &= e^{i(1)\xi^T m - \frac{1}{2}(1)^2 \xi^T D_X \xi} \\ &= e^{i\xi^T m - \frac{1}{2}\xi^T D_X \xi} \end{aligned}$$

□

Proposition 50: Composantes indépendantes et vecteur gaussien

Si $X_1, \dots, X_n$ sont des v.a. réelles **indépendantes** de lois $\mathcal{N}(m_1, \sigma_1^2), \dots, \mathcal{N}(m_d, \sigma_d^2)$ respectivement. Alors $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ est un vecteur gaussien.

Preuve :

$\forall \xi \in \mathbb{R}^d$:

$$\begin{aligned} \varphi_X(\xi) &= \varphi_{X_1}(\xi_1) \dots \varphi_{X_d}(\xi_d) = \prod_{j=1}^{d} e^{im_j\xi_j - \frac{1}{2}\sigma_j^2 \xi_j^2} \\ &= e^{i\xi^T m - \frac{1}{2}\xi^T D_X \xi} \end{aligned}$$

Où $m = \begin{pmatrix} m_1 \\ \vdots \\ m_d \end{pmatrix}$ et $D_X = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_d^2 \end{pmatrix}$. Donc X est bien un vecteur gaussien de moyenne m et de matrice de dispersion D_X. □

Proposition 51: Existence de vecteurs gaussiens

À toute matrice D symétrique et positive et à tout vecteur $m \in \mathbb{R}^d$, on peut associer un vecteur gaussien de loi $\mathcal{N}(m, D)$.

Preuve :

Comme D est symétrique et positive, il existe une matrice C t.q. $D = CC^T$. En partant de variables aléatoires $X_1, \ldots, X_d$ indépendantes toutes de loi $\mathcal{N}(0,1)$, on définit $X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} \sim \mathcal{N}(0, I_d)$.
Et $Y = CX + m$ est un vecteur gaussien avec : $\mathbb{E}[Y] = C\mathbb{E}[X] + m = m$ Var$[Y] = CD_X C^T = CI_d C^T = CC^T = D$. Donc $Y \sim \mathcal{N}(m, D)$. $\square$

Remarque

$$\begin{tikzcd} X_i \sim \mathcal{N} \text{ indép.} \arrow[r, Rightarrow] & X = (X_i) \text{ vecteur gaussien} \arrow[r, Rightarrow, "\text{Recip. Fausse}"'] & X_i \text{ v.a. gaussiennes} \arrow[l, dashed, bend left, "\text{Pas forcément indépendantes}"] \end{tikzcd}$$

Que $X_1, \ldots, X_d$ soient des v.a. gaussiennes ne garantit pas que X est un vecteur gaussien.

Proposition 52: Densité d'un vecteur gaussien

Soit $X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} \in \mathbb{R}^d$ un vecteur de loi $\mathcal{N}_d(m, D)$. Alors X admet une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ ssi D est inversible, et dans ce cas :

$$f(x_1, \ldots, x_d) = \frac{1}{(2\pi)^{d/2}\sqrt{\det D}} e^{-\frac{1}{2}(x-m)^T D^{-1}(x-m)}$$

est la densité de la loi de X.

Remarque

Si $X \sim \mathcal{N}(0,1)$ dans $\mathbb{R}$, $Y = \begin{pmatrix} X \\ -X \end{pmatrix}$ est un vecteur gaussien. $Y = \begin{pmatrix} 1 \\ -1 \end{pmatrix} X$. $D_Y = \begin{pmatrix} 1 \\ -1 \end{pmatrix} (1) \begin{pmatrix} 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ (non inversible).

Proposition 53: Caractérisation de l'indépendance

Les composantes d'un vecteur gaussien $X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} \in \mathbb{R}^d$ sont indépendantes ssi D_X est diagonale.

Preuve :

($\Longrightarrow$) Si $X_1, \ldots, X_d$ sont indépendantes alors $\forall i, j \in [\![1, d]\!], i \neq j$: Cov$(X_i, X_j) = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] = \mathbb{E}[\ldots]\mathbb{E}[\ldots] = 0$. Donc D_X est diagonale.

($\Longleftarrow$) Réciproquement : $\forall \xi \in \mathbb{R}^d$,

$$
\begin{aligned}
\varphi_X(\xi) &= e^{im^T \xi - \frac{1}{2}\xi^T D_X \xi} \\
&\quad \text{(car } D_X \text{ est diagonale)} \\
&= e^{i \sum_{j=1}^d m_j \xi_j - \frac{1}{2} \sum_{j=1}^d \sigma_j^2 \xi_j^2} \\
&= \prod_{j=1}^d e^{i m_j \xi_j - \frac{1}{2}\sigma_j^2 \xi_j^2} \\
&= \prod_{j=1}^d \varphi_{X_j}(\xi_j)
\end{aligned}
$$

Donc les composantes de X sont indépendantes. $\square$

Chapitre 7 : Convergence de variables aléatoires

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité.

I. Plusieurs notions de convergence

Définition 38: Modes de convergence

Soient $(X_n)_{n\geq 1}$ et X des v.a. à valeurs dans un espace métrique (E, d).

1. On dit que (X_n) converge vers X **presque sûrement** (p.s.) si :
$$\mathbb{P}(\{\omega \in \Omega : X_n(\omega) \xrightarrow[n\to\infty]{} X(\omega)\}) = 1$$
On note : $X_n \xrightarrow{p.s.} X$.
2. On dit que (X_n) converge **en probabilité** ($\mathbb{P}$) vers X si :
$$\forall \epsilon > 0, \mathbb{P}(\|X_n - X\| > \epsilon) \xrightarrow[n\to\infty]{} 0$$
On note : $X_n \xrightarrow{\mathbb{P}} X$.
3. Si (X_n) et X sont à valeurs dans $\mathbb{R}^d$, on dit que X_n converge vers X **dans** L^p ($p \in [1, +\infty[$) si :
$$\mathbb{E}[\|X_n - X\|^p] \xrightarrow[n\to\infty]{} 0$$
On note : $X_n \xrightarrow{L^p} X$.
4. On dit que (X_n) converge **en loi** vers X si :
$$\forall \psi : E \to \mathbb{R} \text{ continue bornée} : \mathbb{E}[\psi(X_n)] \xrightarrow[n\to\infty]{} \mathbb{E}[\psi(X)]$$

Remarque

$\forall \epsilon, \epsilon' > 0$
$$\epsilon \leq \epsilon' \implies \{\|X_n - X\| > \epsilon'\} \subseteq \{\|X_n - X\| > \epsilon\}$$
donc si $\mathbb{P}(\|X_n - X\| > \epsilon) \to 0 \implies \mathbb{P}(\|X_n - X\| > \epsilon') \to 0$.

On note $L^0(\Omega, \mathcal{A}, \mathbb{P})$ (ou L^0) l'ensemble des v.a. à valeurs dans $\mathbb{R}^d$.

Proposition 54: Métrique de la convergence en probabilité

On se donne $d : L^0 \times L^0 \to \mathbb{R}_+$ définie par :
$$d(X, Y) = \mathbb{E}[\|X - Y\| \wedge 1]$$
Alors :

1. d est une distance (sur l'espace L^0 où l'on identifie les v.a. égales p.s.).
2. d est compatible avec la convergence en probabilité :
$$X_n \xrightarrow{\mathbb{P}} X \iff d(X_n, X) \xrightarrow[n\to\infty]{} 0$$
3. (L^0, d) est un espace métrique complet.

Preuve :

1) d est une distance

- Symétrie : $d(X,Y) = \mathbb{E}[\|X-Y\| \wedge 1] = \mathbb{E}[\|Y-X\| \wedge 1] = d(Y,X)$.
- Séparation : $d(X,Y) = 0 \iff \mathbb{E}[\|X-Y\| \wedge 1] = 0$. Comme $\|X-Y\| \wedge 1 \geq 0$, ceci est équivalent à $\|X-Y\| \wedge 1 = 0$ p.s. $\implies \|X-Y\| = 0$ p.s. $\implies X = Y$ p.s.
- Inégalité triangulaire :

Exemple 16: Exercice

Montrer que $\forall (x,y,z) \in (\mathbb{R}^d)^3$, on a $\|x-y\| \wedge 1 \leq \|x-z\| \wedge 1 + \|z-y\| \wedge 1$. (En intégrant cette inégalité, on obtient $d(X,Y) \leq d(X,Z) + d(Z,Y)$).

2) Compatibilité avec la convergence $\mathbb{P}$

($\implies$) Supposons $X_n \xrightarrow{\mathbb{P}} X$. Soit $\epsilon \in]0,1[$.

$$\begin{aligned} d(X_n,X) &= \mathbb{E}[\|X_n - X\| \wedge 1] \\ &= \mathbb{E}[(\|X_n - X\| \wedge 1) \cdot \mathbb{1}_{\|X_n - X\| > \epsilon}] + \mathbb{E}[(\|X_n - X\| \wedge 1) \cdot \mathbb{1}_{\|X_n - X\| \leq \epsilon}] \\ &\leq \mathbb{E}[1 \cdot \mathbb{1}_{\|X_n - X\| > \epsilon}] + \mathbb{E}[\epsilon \cdot \mathbb{1}_{\|X_n - X\| \leq \epsilon}] \\ &\leq \mathbb{P}(\|X_n - X\| > \epsilon) + \epsilon \end{aligned}$$

Comme $X_n \xrightarrow{\mathbb{P}} X$, on a $\mathbb{P}(\|X_n - X\| > \epsilon) \xrightarrow[n\to\infty]{} 0$. Donc, $\forall \epsilon \in]0,1[$, il existe N t.q. $\forall n \geq N$, $\mathbb{P}(\|X_n - X\| > \epsilon) \leq \epsilon$. Pour $n \geq N$, on a $d(X_n,X) \leq \epsilon + \epsilon = 2\epsilon$. Ceci montre que $d(X_n,X) \xrightarrow[n\to\infty]{} 0$.

($\impliedby$) Réciproquement, supposons $d(X_n,X) \to 0$. Soit $\epsilon \in]0,1[$ (comme sur l'image).

$$\begin{aligned} d(X_n,X) &= \mathbb{E}[\|X_n - X\| \wedge 1] \\ &= \mathbb{E}[(\|X_n - X\| \wedge 1)\mathbb{1}_{\|X_n - X\| > \epsilon}] + \mathbb{E}[(\|X_n - X\| \wedge 1)\mathbb{1}_{\|X_n - X\| \leq \epsilon}] \\ &\geq \mathbb{E}[(\|X_n - X\| \wedge 1)\mathbb{1}_{\|X_n - X\| > \epsilon}] \end{aligned}$$

Sur l'événement $\{\|X_n - X\| > \epsilon\}$, comme $\epsilon < 1$, on a $\|X_n - X\| \wedge 1 \geq \epsilon \wedge 1 = \epsilon$.

$$\begin{aligned} d(X_n,X) &\geq \mathbb{E}[\epsilon \cdot \mathbb{1}_{\|X_n - X\| > \epsilon}] \\ d(X_n,X) &\geq \epsilon \cdot \mathbb{P}(\|X_n - X\| > \epsilon) \end{aligned}$$

On a donc (l'inégalité de Markov sur la v.a. $\|X_n - X\| \wedge 1$) :

$$\mathbb{P}(\|X_n - X\| > \epsilon) \leq \frac{1}{\epsilon} d(X_n,X)$$

Puisque $d(X_n,X) \to 0$, on a $\mathbb{P}(\|X_n - X\| > \epsilon) \to 0$.

3) (L^0, d) **est complet** (La preuve de ce point est l'objet de la Proposition 55 qui suit). $\square$

Proposition 55: Complétude de l'espace $L^1(d)$

Soit $(X_n)_{n\geq 1}$ une suite de v.a. à valeur dans $\mathbb{R}^d$ qui est une suite de Cauchy pour $d(X,Y) = \mathbb{E}[\|X-Y\| \wedge 1]$. Alors il existe une v.a. X à valeurs dans $\mathbb{R}^d$ telle que $d(X_n,X) \xrightarrow[n\to\infty]{} 0$.

Preuve :

- Il suffit de montrer que $(X_n)_{n\geq 1}$ admet une sous-suite convergente.

- $(X_n)_{n\geq 1}$ admet une sous-suite $(Y_n)_{n\geq 1}$ t.q. $\forall n \geq 1$:

$$\mathbb{E}[\|Y_{n+1} - Y_n\| \wedge 1] = d(Y_n, Y_{n+1}) \leq \frac{1}{2^n}$$

On a, d'après le Théorème de Convergence Monotone (ou Fubini-Tonelli) :

$$\mathbb{E}\left[\sum_{k=1}^{+\infty} \|Y_{k+1} - Y_k\| \wedge 1\right] = \sum_{k=1}^{+\infty} \mathbb{E}[\|Y_{k+1} - Y_k\| \wedge 1]$$

$$\leq \sum_{k=1}^{+\infty} \frac{1}{2^k} = 1 < +\infty$$

Donc $\sum_{k=1}^{+\infty} \|Y_{k+1} - Y_k\| \wedge 1 < +\infty$ p.s.
Or, $\forall \omega \in \Omega$, si $\sum_{k=1}^{+\infty}(\|Y_{k+1}(\omega) - Y_k(\omega)\| \wedge 1) < +\infty$, nécessairement il n'y a qu'un nombre fini d'entiers k t.q. $\|Y_{k+1}(\omega) - Y_k(\omega)\| > 1$. Par conséquent $\sum_{k=1}^{+\infty} \|Y_{k+1} - Y_k\| < +\infty$ p.s.
Par conséquent, la série $\sum_{k=1}^{+\infty}(Y_{k+1} - Y_k)$ converge presque partout (car $\mathbb{R}^d$ est complet). Et on pose $X = Y_1 + \sum_{k=1}^{+\infty}(Y_{k+1} - Y_k)$. On a alors $\forall n \geq 1$:

$$\begin{aligned}
\|X - Y_n\| &= \left\| Y_1 + \sum_{k=1}^{+\infty}(Y_{k+1} - Y_k) - \left(Y_1 + \sum_{k=1}^{n-1}(Y_{k+1} - Y_k)\right)\right\| \\
&= \left\|\sum_{k=n}^{+\infty}(Y_{k+1} - Y_k)\right\| \\
&\leq \sum_{k=n}^{+\infty} \|Y_{k+1} - Y_k\| \xrightarrow[n\to\infty]{} 0 \quad \text{(p.s.)}
\end{aligned}$$

On a donc $Y_n \xrightarrow{p.s.} X$.
Donc $d(Y_n, X) = \mathbb{E}[\|Y_n - X\| \wedge 1] \xrightarrow[n\to\infty]{} 0$. d'après le TCD (Théorème de Convergence Dominée) car :

- $\|Y_n - X\| \wedge 1 \xrightarrow[n\to\infty]{p.s.} 0$
- $\forall n \geq 1, \|Y_n - X\| \wedge 1 \leq 1$ et $\mathbb{E}[1] = 1 < +\infty$.

La sous-suite (Y_n) converge vers X pour la métrique d. Puisque (X_n) est de Cauchy, cela implique que la suite entière (X_n) converge vers X pour d. □

Proposition 56: 0

Si $(X_n)_{n\geq 1}$ et X sont des v.a. à valeurs dans $\mathbb{R}^d$ et $X_n \xrightarrow{\mathbb{P}} X$ alors $(X_n)_{n\geq 1}$ admet une sous-suite $(Y_n)_{n\geq 1}$ t.q. $Y_n \xrightarrow{p.s.} X$.

Exemple 17: Unicité de la limite

Soient $(X_n)_{n\geq 1}$, X et Y des v.a. à valeurs dans $\mathbb{R}^d$. Montrer que :

1. Si $X_n \xrightarrow{p.s.} X$ et $X_n \xrightarrow{p.s.} Y$ alors $X = Y$ p.s.
2. Si $X_n \xrightarrow{\mathbb{P}} X$ et $X_n \xrightarrow{\mathbb{P}} Y$ alors $X = Y$ p.s.
3. Si $X_n \xrightarrow{\mathbb{P}} X$ et $X_n \xrightarrow{p.s.} Y$ alors $X = Y$ p.s.

Proposition 57: 1

Soit $(X_n)_{n\geq 1}$ une suite de v.a. à valeurs dans $\mathbb{R}^d$ t.q. il existe une v.a. X à valeurs dans $\mathbb{R}^d$ t.q. $X_n \xrightarrow{L^p} X$ et t.q. il existe $r \in]p, +\infty]$ t.q. la suite $(X_n)_{n\geq 1}$ est bornée dans L^r (i.e. $\sup_n \mathbb{E}[\|X_n\|^r] < +\infty$). Alors $\forall p' \in [p, r[$ on a $X_n \xrightarrow{L^{p'}} X$.

Proposition 58: 2 : Implications vers la convergence en probabilité

Soient $(X_n)_{n\geq 1}$ et X des v.a. à valeurs dans $\mathbb{R}^d$.

- Si $X_n \xrightarrow{p.s.} X$, alors $X_n \xrightarrow{\mathbb{P}} X$.
- Et s'il existe $p \in [1, +\infty]$ tel que $X_n \xrightarrow{L^p} X$, alors $X_n \xrightarrow{\mathbb{P}} X$.

Preuve de la proposition 2 :

Premier cas : Si $X_n \xrightarrow{p.s.} X$, alors

$$\mathbb{E}[|X_n - X| \wedge 1] = d(X_n, X) \xrightarrow[n\to\infty]{} 0$$

D'après le Théorème de Convergence Dominée (TCD) car :

- $X_n \xrightarrow{p.s.} X \implies |X_n - X| \wedge 1 \xrightarrow{p.s.} 0$
- $\forall n \geq 1, \quad |X_n - X| \wedge 1 \leq 1$ (et 1 est intégrable car la mesure est finie).

Donc $X_n \xrightarrow{\mathbb{P}} X$.

Deuxième cas : Si $X_n \xrightarrow{L^p} X$:

$$\mathbb{E}[|X_n - X| \wedge 1] \leq \mathbb{E}[|X_n - X|] = \|X_n - X\|_1 \underbrace{\leq}_{\text{Hölder}} \|X_n - X\|_p \xrightarrow[n\to\infty]{} 0$$

Donc $X_n \xrightarrow{\mathbb{P}} X$. $\square$

Preuve de la proposition (1) :

D'après la Proposition 56, $(X_n)_{n\geq 1}$ admet une sous-suite $(Y_n)_{n\geq 1}$ qui converge presque sûrement vers X. $\forall p' \in [p, r[$ on a que...

Remarque

Rappel : Dans les espaces de proba, $\forall p' \geq p \geq 1$, l'application $p \mapsto \|X\|_{L^p} = (\mathbb{E}[\|X\|^p])^{1/p}$ est croissante (par l'inégalité de Jensen).

Soit $p' \in [p, r[$. $\forall \epsilon > 0$,

$$\begin{aligned}\mathbb{E}[\|X_n - X\|^{p'}] &= \mathbb{E}[\|X_n - X\|^{p'} \mathbb{1}_{\|X_n - X\| \leq \epsilon}] + \mathbb{E}[\|X_n - X\|^{p'} \mathbb{1}_{\|X_n - X\| > \epsilon}] \\ &\leq \epsilon^{p'} + \mathbb{E}[\|X_n - X\|^{p'} \mathbb{1}_{\|X_n - X\| > \epsilon}]\end{aligned}$$

On utilise l'inégalité de Hölder avec $q = r/p'$ et r_H (indice Hölder) t.q. $1/q + 1/r_H = 1 \implies r_H = \frac{r}{r-p'}$.

$$\begin{aligned}\mathbb{E}[\|X_n - X\|^{p'} \mathbb{1}_{\dots}] &\leq \left(\mathbb{E}[(\|X_n - X\|^{p'})^{r/p'}]\right)^{p'/r} (\mathbb{E}[(\mathbb{1}_{\dots})^{r_H}])^{1/r_H} \\ &= (\mathbb{E}[\|X_n - X\|^r])^{p'/r} (\mathbb{P}(\|X_n - X\| > \epsilon))^{(r-p')/r}\end{aligned}$$

La suite (X_n) est bornée dans L^r, et $X \in L^r$ aussi (par Fatou sur la sous-suite p.s.), donc $\|X_n - X\|_r$ est borné par une constante C.

$$\mathbb{E}[\|X_n - X\|^{p'}] \leq \epsilon^{p'} + (C)^{p'} \left(\mathbb{P}(\|X_n - X\| > \epsilon)\right)^{(r-p')/r}$$

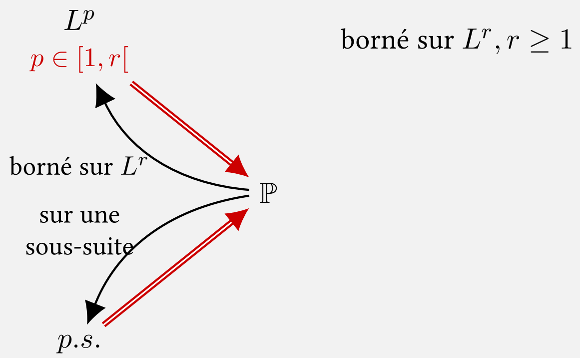
Puisque $X_n \xrightarrow{L^p} X$, on a $X_n \xrightarrow{\mathbb{P}} X$ (par Proposition 58). Donc $\mathbb{P}(\|X_n - X\| > \epsilon) \to 0$.

$$\limsup_{n \to \infty} \mathbb{E}[\|X_n - X\|^{p'}] \leq \epsilon^{p'}$$

Ceci est vrai $\forall \epsilon > 0$, donc $\lim_{n \to \infty} \mathbb{E}[\|X_n - X\|^{p'}] = 0$. □

Remarque

Schéma des convergences :



Exemple 18: Exercice : Contre-exemples

1. Trouver un exemple de suite $(X_n)_{n \geq 1}$ et X t.q. $X_n \xrightarrow{\mathbb{P}} X$ et on n'a pas $X_n \xrightarrow{p.s.} X$.
2. Trouver $(Y_n)_{n \geq 1}$ et Y t.q. $Y_n \xrightarrow{\mathbb{P}} Y$ et $Y_n \xrightarrow{L^p} Y$ et on n'a pas $Y_n \xrightarrow{L^q} Y$ (pour $q > p$).

II. Loi des grands nombres

Théorème 5: Loi faible des grands nombres (WLLN)

Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles, indépendantes, de même loi t.q. $\mathbb{E}[X_1^2] < +\infty$. On a :

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow[n \to \infty]{L^2} \mathbb{E}[X_1]$$

Preuve :

Calculons l'espérance et la variance de la moyenne empirique $\bar{X}_n$.

$$\mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\frac{1}{n} \sum_{k=1}^{n} X_k\right] = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}[X_k]$$

Comme les X_k sont de même loi, $\mathbb{E}[X_k] = \mathbb{E}[X_1]$ pour tout k.

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}[X_1] = \frac{1}{n}(n\mathbb{E}[X_1]) = \mathbb{E}[X_1]$$

Calculons la variance :

$$\mathbb{E}\left[|\bar{X}_n - \mathbb{E}[\bar{X}_n]|^2\right] = \text{Var}[\bar{X}_n] = \text{Var}\left[\frac{1}{n}\sum_{k=1}^{n} X_k\right]$$

Comme les X_k sont indépendantes, la variance de la somme est la somme des variances.

$$\text{Var}[\bar{X}_n] = \frac{1}{n^2}\sum_{k=1}^{n} \text{Var}[X_k]$$

Comme les X_k sont de même loi, $\text{Var}[X_k] = \text{Var}[X_1]$ (qui est finie car $\mathbb{E}[X_1^2] < \infty$).

$$\text{Var}[\bar{X}_n] = \frac{1}{n^2}(n\text{Var}[X_1]) = \frac{1}{n}\text{Var}[X_1] \xrightarrow[n\to\infty]{} 0$$

On a donc $\mathbb{E}[|\bar{X}_n - \mathbb{E}[X_1]|^2] \to 0$, ce qui est la définition de $\bar{X}_n \xrightarrow{L^2} \mathbb{E}[X_1]$. (Notons $S_n = \sum X_k$, on a $\frac{S_n}{n} \xrightarrow{L^2} \mathbb{E}[X_1]$). □

Exemple 19: Exercice (Variantes WLLN)

Montrer que le résultat (WLLN en $\mathbb{P}$) est encore vrai si :

- On remplace "indépendantes" par "$\forall n \neq m, \text{Cov}(X_n, X_m) = 0$" (variables décorrélées).
- On remplace "de même loi" par "$\forall n \geq 1, \mathbb{E}[X_n] = \mu$ et $\text{Var}[X_n] \leq C < +\infty$" (espérance constante et variances uniformément bornées).
- On remplace "de même loi de carré intégrable" par "il existe $C > 0, \forall n, \mathbb{E}[X_n^2] < C$".

Théorème 6: Loi forte des grands nombres (SLLN)

Soit $(X_n)_{n\geq 1}$ une suite de v.a. réelles indépendantes de même loi (i.i.d.) telle que X_n est intégrable (i.e. $\mathbb{E}[|X_1|] < +\infty$). On a :

$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to\infty]{p.s.} \mathbb{E}[X_1]$$

"Preuve" sous condition d'intégrabilité de X_1^2 ?

Preuve :

De même que dans la preuve de la loi faible, $\forall n \geq 1$:

$$\mathbb{E}[|\bar{X}_n - \mathbb{E}[X_1]|^2] = \frac{1}{n}\text{Var}[X_1]$$

En notant $\forall n \geq 1, S_n = \sum_{k=1}^{n} X_k$, on a $\bar{X}_n = S_n/n$. Regardons la sous-suite des carrés n^2 :

$$\mathbb{E}\left[\left(\frac{S_{n^2}}{n^2} - \mathbb{E}[X_1]\right)^2\right] = \frac{1}{n^2}\text{Var}[X_1]$$

Considérons la série de ces espérances :

$$\sum_{n=1}^{+\infty}\mathbb{E}\left[\left(\frac{S_{n^2}}{n^2} - \mathbb{E}[X_1]\right)^2\right] = \sum_{n=1}^{+\infty}\frac{1}{n^2}\text{Var}[X_1] = \text{Var}[X_1]\sum_{n=1}^{+\infty}\frac{1}{n^2} < +\infty$$

Par Fubini-Tonelli (ou TCM), on peut intervertir somme et espérance :

$$\mathbb{E}\left[\sum_{n=1}^{+\infty}\left(\frac{S_{n^2}}{n^2}-\mathbb{E}[X_1]\right)^2\right]<+\infty$$

On a donc $\sum_{n=1}^{+\infty}\left(\frac{S_{n2}}{n^2}-\mathbb{E}[X_1]\right)^2<+\infty$ p.s. Le terme général d'une série convergente tend vers 0, donc :

$$\frac{S_{n^2}}{n^2}\xrightarrow[n\to\infty]{p.s.}\mathbb{E}[X_1]$$

Ceci prouve la convergence p.s. pour la sous-suite n^2. Il faut maintenant l'étendre à n.
On suppose d'abord que les v.a. X_n sont positives ($X_n \geq 0$). $\forall n \geq 1$, il existe un unique entier $k \geq 1$ t.q. $k^2 \leq n < (k+1)^2$.
donc $\frac{S_n}{n}\xrightarrow{L^2}\mathbb{E}[X_1]$
donc :

$$\frac{S_{k^2}}{(k+1)^2}\leq\frac{S_n}{n}\leq\frac{S_{(k+1)^2}}{k^2}$$

$$\underbrace{\frac{S_{k^2}}{k^2}}_{\substack{\downarrow p.s.\\ \mathbb{E}[X_1]}}\cdot\underbrace{\frac{k^2}{(k+1)^2}}_{\to 1}\leq\frac{S_n}{n}\leq\underbrace{\frac{S_{(k+1)^2}}{(k+1)^2}}_{\substack{\downarrow p.s.\\ \mathbb{E}[X_1]}}\cdot\underbrace{\frac{(k+1)^2}{k^2}}_{\to 1}$$

donc $\frac{S_n}{n}\xrightarrow{\mathbb{P}}\mathbb{E}[X_1]$
donc $\frac{S_n}{n}\xrightarrow[n\to\infty]{p.s.}\mathbb{E}[X_1]$
Dans le cas général on a $X_k = X_k^+ - X_k^-$ et $\frac{1}{n}\sum_{k=1}^n X_k^+ \xrightarrow{p.s.} \mathbb{E}[X_1^+]$ et $\frac{1}{n}\sum_{k=1}^n X_k^- \xrightarrow{p.s.} \mathbb{E}[X_1^-]$
donc :

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^n X_k &= \frac{1}{n}\sum_{k=1}^n X_k^+ - \frac{1}{n}\sum_{k=1}^n X_k^- \\ &\xrightarrow[n\to\infty]{p.s.}\mathbb{E}[X_1^+]-\mathbb{E}[X_1^-]=\mathbb{E}[X_1^+-X_1^-]=\mathbb{E}[X_1]\end{aligned}$$

□

Définition 39

Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires. On note $\mathcal{B}_n = \sigma(X_k : k \geq n)$.
On appelle **tribu asymptotique** de $(X_n)_{n\geq 0}$ la tribu :

$$\mathcal{B}_\infty = \bigcap_{n\geq 0}\mathcal{B}_n$$

Exemple 20

Soit $(X_n)_{n\geq 1}$ une suite de v.a. réelles.

$$\left\{\limsup_{n\to+\infty}\frac{1}{n}(X_1+\cdots+X_n)\leq\alpha\right\}\in\mathcal{B}_\infty$$

$$\left\{\limsup_{n\to+\infty}(X_1+\cdots+X_n)\leq\alpha\right\}\notin\mathcal{B}_\infty$$

En effet, $\forall k \in \mathbb{N}, \forall \omega \in \Omega$:

$$\limsup_{n\to+\infty} \frac{1}{n}(X_1(\omega) + \cdots + X_n(\omega)) = \limsup_{n\to+\infty} \frac{1}{n}(X_k(\omega) + \cdots + X_n(\omega))$$

car $\frac{1}{n}(X_1(\omega) + \cdots + X_{k-1}(\omega)) \xrightarrow{n\to\infty} 0$.
Le terme de droite est mesurable par rapport à $\mathcal{B}_k$.
Ceci étant vrai pour tout k, la limite supérieure appartient à $\mathcal{B}_\infty$.
Autre exemple : $A_n = \{X_n \geq \alpha\}$.
On a que $\limsup A_n = \bigcap_{n\geq 1} \bigcup_{k\geq n} A_k \in \mathcal{B}_\infty$.
En effet, la suite d'ensembles $(\bigcup_{k\geq n} A_k)_{n\geq 0}$ est décroissante, donc $\forall m \geq 1$:

$$\limsup A_n = \bigcap_{n\geq m} \bigcup_{k\geq n} A_k \in \mathcal{B}_m$$

Donc $\limsup A_n \in \bigcap_{m\geq 1} \mathcal{B}_m = \mathcal{B}_\infty$.
De même : $\{\omega \in \Omega : X_n(\omega) \text{ converge}\} = \{\omega \in \Omega : \liminf X_n(\omega) = \limsup X_n(\omega)\} \in \mathcal{B}_\infty$.

Proposition 59: Loi du 0-1 de Kolmogorov

Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires indépendantes.
Alors $\forall B \in \mathcal{B}_\infty$, on a :

$$\mathbb{P}(B) \in \{0, 1\}$$

Preuve :

$\forall n > 0$, on pose $\mathcal{D}_n = \sigma(X_k, k \leq n)$ (tribu engendrée par le passé).
On a que $\forall n > 0$, $\mathcal{D}_n$ est indépendante de $\mathcal{B}_{n+1} = \sigma(X_k, k \geq n+1)$.
Comme $\mathcal{B}_\infty \subset \mathcal{B}_{n+1}$, on a que $\mathcal{D}_n$ est indépendante de $\mathcal{B}_\infty$.
Donc $\forall n > 0, \forall A \in \mathcal{D}_n, \forall B \in \mathcal{B}_\infty, \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
Or $\bigcup_{n\geq 1} \mathcal{D}_n$ est stable par intersection finie.
Donc $\sigma(\bigcup_{n\geq 1} \mathcal{D}_n) = \sigma(X_n, n \geq 0)$ est indépendante de $\mathcal{B}_\infty$.
En particulier, $\mathcal{B}_\infty$ est indépendante d'elle-même (car $\mathcal{B}_\infty \subset \sigma(X_n, n \geq 0)$).
$\forall B \in \mathcal{B}_\infty, \mathbb{P}(B \cap B) = \mathbb{P}(B)\mathbb{P}(B) \implies \mathbb{P}(B) = \mathbb{P}(B)^2$.
Donc $\mathbb{P}(B) \in \{0, 1\}$. □

Exemple 21: Application

Si $(X_n)_{n\geq 1}$ est une suite de v.a. réelles indépendantes, alors :

$$\left\{\limsup \frac{1}{n}(X_1 + \cdots + X_n) \leq \alpha\right\} \in \mathcal{B}_\infty$$

Donc la fonction de répartition de $\limsup \frac{1}{n}(X_1 + \cdots + X_n)$ ne prend que les valeurs 0 ou 1.
Par conséquent, $\limsup \frac{1}{n}(X_1 + \cdots + X_n)$ est une constante presque sûrement.

Loi du 0-1 de Borel

Théorème 7

Soit $(A_n)_{n\geq 0}$ une suite d'événements indépendants. On a :

$$\mathbb{P}(\limsup A_n) \in \{0,1\} \quad \text{et} \quad \sum_{n\geq 0} \mathbb{P}(A_n) < +\infty \iff \mathbb{P}(\limsup A_n) = 0$$

Preuve :

On a déjà vu que $\limsup A_n = \bigcap_{n\geq 0} \bigcup_{k\geq n} A_k \in \mathcal{B}_\infty$ où $\mathcal{B}_\infty$ est la tribu asymptotique de la suite de variables indépendantes $(\mathbb{1}_{A_n})_{n\geq 0}$.
Donc $\mathbb{P}(\limsup A_n) \in \{0,1\}$ d'après la loi de Kolmogorov.
Supposons que $\sum_{n\geq 0} \mathbb{P}(A_n) < +\infty$.
La suite $(\bigcup_{k\geq n} A_k)_{n\geq 0}$ est une suite décroissante, donc :

$$\mathbb{P}(\limsup A_n) = \mathbb{P}(\bigcap_{n\geq 1} \bigcup_{k\geq n} A_k) = \lim_{n\to\infty} \mathbb{P}(\bigcup_{k\geq n} A_k) \leq \lim_{n\to\infty} \sum_{k=n}^{+\infty} \mathbb{P}(A_k) = 0$$

(Reste d'une série convergente).
Réciproquement, on suppose que $\sum_{n\geq 0} \mathbb{P}(A_n) = +\infty$.
On a $(\limsup A_n)^c = (\bigcap_{n\geq 1} \bigcup_{k\geq n} A_k)^c = \bigcup_{n\geq 1} \bigcap_{k\geq n} A_k^c$.
Donc $\mathbb{P}((\limsup A_n)^c) \leq \sum_{n\geq 1} \mathbb{P}(\bigcap_{k\geq n} A_k^c)$.
$\forall \ell \geq 0$:

$$\bigcap_{k\geq n} A_k^c \subset \bigcap_{k=n}^{n+\ell} A_k^c$$

Donc par indépendance :

$$\mathbb{P}(\bigcap_{k\geq n} A_k^c) \leq \mathbb{P}(\bigcap_{k=n}^{n+\ell} A_k^c) = \prod_{k=n}^{n+\ell} \mathbb{P}(A_k^c) = \prod_{k=n}^{n+\ell} (1 - \mathbb{P}(A_k))$$

En utilisant l'inégalité $1 - x \leq e^{-x}$, on obtient :

$$\mathbb{P}(\bigcap_{k\geq n} A_k^c) \leq \prod_{k=n}^{n+\ell} e^{-\mathbb{P}(A_k)} = e^{-\sum_{k=n}^{n+\ell} \mathbb{P}(A_k)}$$

En passant à la limite en $\ell \to \infty$, comme la série diverge, le terme de droite tend vers 0.
Donc $\mathbb{P}((\limsup A_n)^c) = 0$, d'où $\mathbb{P}(\limsup A_n) = 1$. □

Remarque

Cette propriété (la première partie : si la somme converge alors probabilité 0) est vraie même si les événements ne sont pas indépendants (Lemme de Borel-Cantelli).

III. Convergence en Loi

Définition 40

Une suite $(\mu_n)_{n\geq 0}$ de mesures de probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ converge **étroitement** vers une probabilité μ sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ si :

Pour toute fonction $\varphi : \mathbb{R}^d \to \mathbb{R}$ continue et bornée, on a :

$$\int_{\mathbb{R}^d} \varphi d\mu_n \xrightarrow{n\to+\infty} \int_{\mathbb{R}^d} \varphi d\mu$$

On note alors $\mu_n \xrightarrow{e} \mu$ (ou $\xrightarrow{\mathcal{L}}$).

Remarque

$$X_n \xrightarrow{\mathcal{L}} X \iff \mathbb{P}_{X_n} \xrightarrow{e} \mathbb{P}_X$$

où $\mathbb{P}_{X_n}$ désigne la loi de X_n.

Proposition 60

Soient $(X_n)_{n\geq 0}$ des v.a. à valeurs dans $\mathbb{R}^d$ et X une v.a.
Alors $X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{\mathcal{L}} X$.

Preuve :

Par l'absurde.
On suppose qu'il existe $\varphi : \mathbb{R}^d \to \mathbb{R}$ continue bornée telle que $(\mathbb{E}[\varphi(X_n)])_{n\geq 0}$ ne converge pas vers $\mathbb{E}[\varphi(X)]$.
Ainsi, $\exists \varepsilon > 0$ et une sous-suite $(Y_n)_{n\geq 0}$ de $(X_n)_{n\geq 0}$ telle que $\forall n \geq 0$:

$$|\mathbb{E}[\varphi(Y_n)] - \mathbb{E}[\varphi(X)]| > \varepsilon$$

Or $Y_n \xrightarrow{\mathbb{P}} X$, donc on peut extraire de $(Y_n)_{n\geq 0}$ une sous-suite $(Z_n)_{n\geq 0}$ telle que $Z_n \xrightarrow{p.s.} X$.
Mais alors, d'après le TCD (Théorème de Convergence Dominée) :

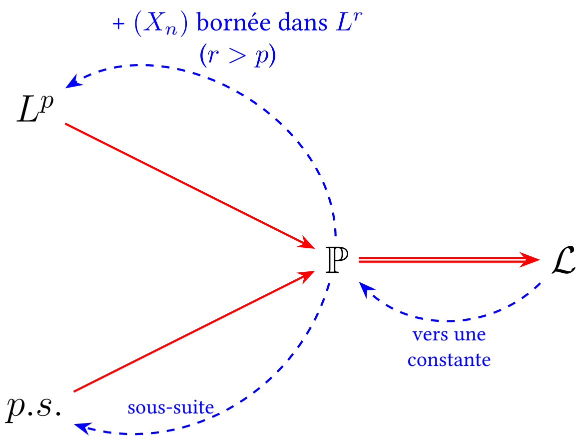
$$\mathbb{E}[\varphi(Z_n)] \to \mathbb{E}[\varphi(X)]$$

Puisque :

- φ étant continue, $Z_n \xrightarrow{p.s.} X \implies \varphi(Z_n) \xrightarrow{p.s.} \varphi(X)$.
- φ étant bornée, $\exists C \in \mathbb{R}$ tq $\forall n \geq 0, |\varphi(Z_n)| \leq C$ (domination par une constante intégrable car mesure de proba).

Ceci contredit le fait que $\forall n \geq 0, |\mathbb{E}[\varphi(Z_n)] - \mathbb{E}[\varphi(X)]| > \varepsilon$. $\square$

Schéma récapitulatif des convergences :



Exemple 22: Exercice

Trouver un exemple de suite $(X_n)_{n\geq 0}$ telle que $X_n \xrightarrow{\mathcal{L}} X$ mais $X_n \not\to^{\mathbb{P}} X$.

Théorème 8: Porte-manteau

Soient $(\mu_n)_{n\geq 0}$ et μ des probabilités sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.
Les affirmations suivantes sont équivalentes :

1. $\mu_n \xrightarrow{e} \mu$
2. $\forall G \subset \mathbb{R}^d$ ouvert, $\liminf \mu_n(G) \geq \mu(G)$
3. $\forall F \subset \mathbb{R}^d$ fermé, $\limsup \mu_n(F) \leq \mu(F)$
4. $\forall B \in \mathcal{B}(\mathbb{R}^d)$ tq $\mu(\partial B) = 0$, on a $\mu_n(B) \xrightarrow{n\to\infty} \mu(B)$.
   (Rappel : $\partial B = \bar{B} \setminus \mathring{B}$).

Exemple 23: Corrigé d'exercice

Question : Unicité de la limite.

1. **Si** $X_n \xrightarrow{p.s.} X$ **et** $X_n \xrightarrow{p.s.} Y$**, alors** $X = Y$ **p.s.**
   Soient les ensembles :
   $$A = \{\omega \in \Omega : X_n(\omega) \to X(\omega)\} \quad \text{et} \quad B = \{\omega \in \Omega : X_n(\omega) \to Y(\omega)\}$$
   On a $\mathbb{P}(A) = \mathbb{P}(B) = 1$.
   Pour tout $\omega \in A \cap B$, on a $X(\omega) = Y(\omega)$ par unicité de la limite réelle. Et $\mathbb{P}(A \cap B) = 1$ puisque :
   $$\mathbb{P}((A \cap B)^c) = \mathbb{P}(A^c \cup B^c) \leq \mathbb{P}(A^c) + \mathbb{P}(B^c) = 0$$
2. **Si** $X_n \xrightarrow{\mathbb{P}} X$ **et** $X_n \xrightarrow{p.s.} Y$**, alors** $X = Y$ **p.s.**
   Puisque $X_n \xrightarrow{\mathbb{P}} X$, il existe une sous-suite $(Z_n)_{n\geq 0}$ de $(X_n)_{n\geq 0}$ telle que $Z_n \xrightarrow{p.s.} X$. Or, comme sous-suite de (X_n), on a aussi $Z_n \xrightarrow{p.s.} Y$.
   D'après le point précédent, $X = Y$ p.s.
3. **Si** $X_n \xrightarrow{\mathbb{P}} X$ **et** $X_n \xrightarrow{\mathbb{P}} Y$, alors il existe une sous suite $(Z_n)_{n\geq 0}$ de $(X_n)_{n\geq 0}$ telle que $Z_n \xrightarrow{\mathbb{P}} X$, $Z_n \xrightarrow{p.s.} X$.
   Donc $X = Y$ p.s.

Exemple 24: Corrigé d'exercice

Soit X une variable de Bernoulli de paramètre $\frac{1}{2}$.

$$\mathbb{P}(X = 1) = \mathbb{P}(X = 0) = \frac{1}{2}$$

Alors $1 - X$ est aussi une variable aléatoire de loi de Bernoulli de paramètre $\frac{1}{2}$.

$$\mathbb{P}(1 - X = 1) = \mathbb{P}(X = 0) = \frac{1}{2}$$

On définit la suite (X_n) par :

$$X_n = \begin{cases} X & \text{si } n \text{ est pair} \\ 1 - X & \text{si } n \text{ est impair} \end{cases}$$

Convergence en loi : La suite des lois $\mathbb{P}_{X_n}$ est constante, donc elle converge vers cette "constante" qui est la loi de X.
Cependant, $(X_n)_{n\geq 0}$ ne converge pas en probabilité vers X. Car $\forall \varepsilon \in]0,1[$, $\forall n$ impair :

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|(1-X) - X| > \varepsilon) = \mathbb{P}(|1-2X| > \varepsilon)$$

Or $\mathbb{P}(|2X-1| = 1) = 1$. Donc la probabilité vaut 1, qui ne tend pas vers 0.

Exemple 25: Corrigé d'exercice

Sur $\Omega = [0,1]$, $\mathcal{A} = \mathcal{B}([0,1])$, $\mathbb{P} = \lambda$. Soit $p \in [1, +\infty[$.
Définissons :

$$X_n(\omega) = \begin{cases} n & \text{si } \omega \in]0, \frac{1}{n^p}] \\ 0 & \text{sinon} \end{cases}$$

- Pour tout $\omega \in \Omega$, $X_n(\omega) \xrightarrow[n\to\infty]{} 0$. Donc $X_n \xrightarrow{p.s.} 0$.
- Donc en particulier, $X_n \xrightarrow{\mathbb{P}} 0$.

Si la suite converge dans L^p, c'est nécessairement vers 0. Regardons la norme L^p pour tout $n > 0$:

$$\mathbb{E}[|X_n - 0|^p] = n^p \times \frac{1}{n^p} + 0 \times (1 - \frac{1}{n^p}) = 1$$

Donc $(X_n)_{n\geq 0}$ n'est pas une suite convergente de L^p.

Exemple 26: Corrigé d'exercice (Suite "Bossepied")

On considère $\Omega = [0,1]$, $\mathcal{A} = \mathcal{B}([0,1])$, $\mathbb{P} = \lambda$.

- $X_1 = \mathbb{1}_{[0,1[}$
- $X_2 = \mathbb{1}_{[0,1/2[}$, $X_3 = \mathbb{1}_{[1/2,1[}$
- $X_4 = \mathbb{1}_{[0,1/4[}, \dots$

On définit l'application $i \in \mathbb{N}^* \to n_i = \left[\frac{\ln(i)}{\ln(2)}\right]$. Soit le reste de la division euclidienne de i par 2^{n_i}.
On pose :

$$X_i(\omega) = \mathbb{1}_{[\frac{R_i}{2^{n_i}}, \frac{R_i+1}{2^{n_i}}[}(\omega)$$

Pour tout $\omega \in [0,1[$:

$$\limsup X_n(\omega) = 1 \quad \text{et} \quad \liminf X_n(\omega) = 0$$

(La suite converge en probabilité et dans L^p vers 0, mais ne converge pas presque sûrement).

Exemple 27: Corrigé d'exercice (Convergence des variances)

Soit $(X_n)_{n\geq 1}$ une suite de v.a. de carré intégrable telle que :

- $n \neq m \implies \text{Cov}(X_n, X_m) = 0$
- $\forall n \geq 1, \quad \mathbb{E}[X_n] = \mathbb{E}[X_1]$
- $\exists C > 0$ tq $\forall n \geq 1, \quad \mathbb{E}[X_n^2] < C$

Alors pour tout $n \in \mathbb{N}^*$:

$$\begin{aligned}
\mathbb{E}[(\overline{X}_n - \mathbb{E}[X_1])^2] &= \text{Var}[\overline{X}_n] \\
&= \text{Var}\left[\frac{1}{n}(X_1 + \cdots + X_n)\right] \\
&= \frac{1}{n^2}\text{Var}\left(\sum_{j=1}^{n} X_j\right) \\
&= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\text{Cov}(X_j, X_k) \\
&= \frac{1}{n^2}\sum_{j=1}^{n}\text{Var}(X_j) \quad \text{(car les covariances croisées sont nulles)} \\
&\leq \frac{1}{n^2}n(C - (\mathbb{E}[X_1])^2) \xrightarrow[n\to\infty]{} 0
\end{aligned}$$

Preuve du Théorème Porte-manteau :

i) $\implies$ ii) Soit $G \subset \mathbb{R}^d$, où G est un ouvert.

Pour tout $p \geq 1$, on pose :

$$\varphi_p(x) = (p \times d(x, G^c)) \wedge 1 = \min(p \times d(x, G^c), 1) \leq 1$$

où pour tout $A \subset \mathbb{R}^d$ et $\forall x \in \mathbb{R}^d$: $d(x, A) = \inf_{y\in A}|x - y|$.

La suite $(\varphi_p)_{p\geq 0}$ a les propriétés suivantes :

- $\forall p \geq 1$, on a $0 \leq \varphi_p \leq \mathbb{1}_G$.
  En effet, $\forall x \in G^c$, $d(x, G^c) = 0$ donc $\varphi_p(x) = 0$.
- $\forall p \geq 1$, φ_p est continue.

Pour justifier la continuité, il suffit de montrer que $\forall A \subset \mathbb{R}^d, x \mapsto d(x, A)$ est continue. Soient $(x, y) \in \mathbb{R}^d$ et $z \in A$:

$$|x - z| \leq |x - y| + |y - z|$$

En passant à l'infimum sur $z \in A$:

$$d(x, A) \leq |x - y| + d(y, A) \implies d(x, A) - d(y, A) \leq |x - y|$$

De même, $d(y, A) - d(x, A) \leq |x - y|$. Donc $|d(x, A) - d(y, A)| \leq |x - y|$. L'application $x \mapsto d(x, A)$ est donc 1-lipschitzienne, donc continue.

- La suite $(\varphi_p)_{p\geq 0}$ est une suite croissante qui converge simplement vers $\mathbb{1}_G$ car :
  - $\forall x \in G, \quad d(x, G^c) > 0$ (car G est ouvert).
  - $\forall x \in G^c, \quad d(x, G^c) = 0$.

Pour tout $n \geq 1$ et $p \geq 1$:

$$\mu^n(G) = \int \mathbb{1}_G d\mu^n \geq \int \varphi_p d\mu^n$$

En passant à la limite inférieure sur n (puisque φ_p est continue bornée et $\mu^n \xrightarrow{e} \mu$) :

$$\liminf_n \mu^n(G) \geq \liminf_n \int \varphi_p d\mu^n = \int \varphi_p d\mu$$

Par conséquent, en prenant le sup sur p et en appliquant le Théorème de Convergence Monotone (TCM) :

$$\liminf_n \mu^n(G) \geq \sup_{p\geq 1} \int \varphi_p d\mu \underset{\text{TCM}}{=} \int \mathbb{1}_G d\mu = \mu(G)$$

ii) $\iff$ iii)
Par passage au complémentaire, car les μ^n et μ sont des mesures de probabilité (donc de masse totale 1).
ii) + iii) $\implies$ iv)
Soit $B \in \mathcal{B}(\mathbb{R}^d)$ tel que $\mu(\partial B) = 0$.

- D'après ii), $\liminf \mu^n(\mathring{B}) \geq \mu(\mathring{B})$.
- D'après iii), $\limsup \mu^n(\overline{B}) \leq \mu(\overline{B})$.

Or, on a toujours $\mu^n(\mathring{B}) \leq \mu^n(B) \leq \mu^n(\overline{B})$. Donc :

$$\mu(\mathring{B}) \leq \liminf \mu^n(\mathring{B}) \leq \liminf \mu^n(B) \leq \limsup \mu^n(B) \leq \limsup \mu^n(\overline{B}) \leq \mu(\overline{B})$$

Or $\overline{B} = \partial B \cup \mathring{B}$ (union disjointe). Comme $\mu(\partial B) = 0$, on a $\mu(\overline{B}) = \mu(\mathring{B}) = \mu(B)$. On conclut que $\mu^n(B) \to \mu(B)$.
iv) $\implies$ i)
Soit $\varphi : \mathbb{R}^d \to \mathbb{R}$ une fonction continue bornée. Quitte à rajouter une constante à φ, on peut supposer qu'elle est positive. Soit K un majorant de φ. φ est à valeurs dans $[0, K]$.

$$\int_{\mathbb{R}^d} \varphi d\mu = \int_{\mathbb{R}^d} \varphi(x)\mu(dx)$$

Utilisons l'identité $\varphi(x) = \int_0^K \mathbb{1}_{\{t<\varphi(x)\}} dt$:

$$= \int_{\mathbb{R}^d} \left(\int_{[0,K]} \mathbb{1}_{\{t<\varphi(x)\}} dt \right) \mu(dx)$$

D'après Fubini-Tonelli (Fubini positif) :

$$= \underbrace{\int_{[0,K]} \left(\int_{\mathbb{R}^d} \mathbb{1}_{\{t<\varphi(x)\}} \mu(dx) \right) dt}_{\text{Fubini positif}} = \int_{[0,K]} \mu(\{x \in \mathbb{R}^d : \varphi(x) > t\}) dt$$

Posons $E_t^\varphi = \{x \in \mathbb{R}^d : \varphi(x) > t\}$. On a donc $\int_{\mathbb{R}^d} \varphi d\mu = \int_{[0,K]} \mu(E_t^\varphi) dt$.
Comme φ est continue, $\forall t \in [0, K]$, l'ensemble E_t^φ est ouvert. De plus, $\{x : \varphi(x) > t\} \subset \overline{E_t^\varphi}$, donc la frontière vérifie $\partial E_t^\varphi \subset \{x : \varphi(x) = t\}$.
Considérons la fonction $t \mapsto \mu(E_t^\varphi)$. Comme une fonction de répartition (ou de survie), elle est monotone.

- L'ensemble des points de discontinuité de $t \mapsto \mu(E_t^\varphi)$ est au plus dénombrable.
- Les points de discontinuité sont les $t \in [0, K]$ tels que $\mu(\{x : \varphi(x) = t\}) > 0$.

Il existe donc une partie dénombrable $D \subset [0, K]$ telle que $\forall t \in [0, K] \setminus D$, on a $\mu(\{x : \varphi(x) = t\}) = 0$, et donc $\mu(\partial E_t^\varphi) = 0$.
D'après le point iv), pour tout $t \in [0, K] \setminus D$:

$$\mu^n(E_t^\varphi) \to \mu(E_t^\varphi)$$

On applique le Théorème de Convergence Dominée (TCD) sur l'intégrale par rapport à t (sur $[0, K]$) :

- Convergence simple pour presque tout t (sur le complémentaire de D, qui est de mesure de Lebesgue nulle).
- Domination : $\forall t \in [0, K], \forall n \geq 1, \quad |\mu^n(E_t^\varphi)| \leq 1$ (intégrable sur $[0, K]$).

On obtient :

$$\int \varphi d\mu^n = \int_{[0,K]} \mu^n(E_t^\varphi)dt \longrightarrow \int_{[0,K]} \mu(E_t^\varphi)dt = \int \varphi d\mu$$

□

Proposition 61: Caractérisation par la fonction de répartition

Soient $(X_n)_{n\geq 0}$ et X des variables aléatoires réelles.
On a $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si pour tout $x \in \mathbb{R}$ point de continuité de F_X, on a :

$$F_{X_n}(x) = \mathbb{P}(X_n \leq x) \underset{n\to\infty}{\longrightarrow} \mathbb{P}(X \leq x) = F_X(x)$$

Preuve :

Sens $\implies$:
Supposons que $X_n \xrightarrow{\mathcal{L}} X$, c'est-à-dire $\mu^{X_n} \xrightarrow{e} \mu^X$. Soit $x \in \mathbb{R}$ un point de continuité de F_X. On a :

$$\mu^X(\partial] - \infty, x]) = \mu^X(\{x\}) = \mathbb{P}(X = x) = 0$$

D'après le théorème Porte-manteau (point iv), on a $\mu^{X_n}(]-\infty, x]) \to \mu^X(]-\infty, x])$, ce qui signifie exactement $F_{X_n}(x) \to F_X(x)$.

Sens $\impliedby$:
On suppose que $F_{X_n}(x) \to F_X(x)$ pour tout point de continuité x de F_X. Soit $A \subset \mathbb{R}$ un ouvert. Il existe une famille dénombrable d'intervalles ouverts deux à deux disjoints $(I_k)_{k\geq 0}$ tels que $A = \bigcup_{k\geq 0} I_k$, avec $I_k =]a_k, b_k[$.
Soit $\varepsilon > 0$ fixé. Pour tout $k \geq 0$, on cherche $J_k =]\alpha_k, \beta_k[\subset I_k$ tel que α_k, β_k soient des points de continuité de F_X et :

$$\mathbb{P}(X \in I_k) \leq \mathbb{P}(X \in J_k) + \varepsilon 2^{-(k+1)}$$

Or, pour tout $k \geq 0$:

$$\begin{aligned}
\mathbb{P}(X_n \in J_k) &= \mathbb{P}(X_n \in]\alpha_k, \beta_k]) \quad \text{(à la limite car points de continuité)} \\
&= F_{X_n}(\beta_k) - F_{X_n}(\alpha_k) \\
&\underset{n\to\infty}{\longrightarrow} F_X(\beta_k) - F_X(\alpha_k) = \mathbb{P}(X \in J_k)
\end{aligned}$$

On peut alors écrire :

$$\begin{aligned}
\liminf_n \mathbb{P}(X_n \in A) &\geq \sum_{k\geq 0} \liminf_n \mathbb{P}(X_n \in J_k) \quad \text{(par Fatou)} \\
&= \sum_{k\geq 0} \mathbb{P}(X \in J_k) \\
&\geq \sum_{k\geq 0} \left(\mathbb{P}(X \in I_k) - \varepsilon 2^{-k}\right) \\
&= \sum_{k\geq 0} \mathbb{P}(X \in I_k) - 2\varepsilon \\
&= \mathbb{P}(X \in A) - 2\varepsilon
\end{aligned}$$

Ceci étant vrai pour tout $\varepsilon > 0$, on obtient :

$$\liminf_n \mathbb{P}^{X_n}(A) \geq \mathbb{P}^X(A), \quad \forall A \subset \mathbb{R} \text{ ouvert.}$$

D'après le théorème Porte-manteau (point ii), on conclut que $\mathbb{P}^{X_n} \xrightarrow{e} \mathbb{P}^X$ (convergence en loi). $\square$

Proposition 62: Convergence en loi vers une constante

Soit $(X_n)_{n\geq 0}$ une suite de v.a. réelles qui converge en loi vers une v.a. constante $X = a \in \mathbb{R}$ p.s. Alors :

$$X_n \xrightarrow{\mathbb{P}} a$$

Preuve :

$X_n \xrightarrow{\mathcal{L}} X$ équivaut à $\mathbb{P}^{X_n} \xrightarrow{e} \mathbb{P}^X$. Ici $\mathbb{P}^X = \delta_a$ (masse de Dirac en a) puisque $X = a$ p.s.
Soit $\varepsilon > 0$. On considère l'ensemble :

$$\{|X_n - a| > \varepsilon\} = \{X_n \in]-\infty, a-\varepsilon[\cup]a+\varepsilon, +\infty[\}$$

Notons $A =]-\infty, a-\varepsilon[\cup]a+\varepsilon, +\infty[$. La frontière de cet ouvert est $\partial A = \{a-\varepsilon, a+\varepsilon\}$.
On a $\mathbb{P}^X(\partial A) = \delta_a(\{a-\varepsilon, a+\varepsilon\}) = 0$.
D'après le théorème Porte-manteau (point iv), puisque la frontière est de mesure nulle pour la loi limite :

$$\mathbb{P}(|X_n - a| > \varepsilon) = \mathbb{P}^{X_n}(A) \underset{n\to\infty}{\longrightarrow} \mathbb{P}^X(A) = \delta_a(]-\infty, a-\varepsilon[\cup]a+\varepsilon, +\infty[) = 0$$

Donc $X_n \xrightarrow{\mathbb{P}} a$. $\square$

Proposition 63: Régularité de la fonction caractéristique

Si X est une variable aléatoire réelle de carré intégrable ($X \in L^2$), alors $\varphi_X : t \mapsto \mathbb{E}[e^{itX}]$ est de classe C^2 sur $\mathbb{R}$. En particulier, pour tout $t \in \mathbb{R}$, on a le développement de Taylor-Young :

$$\varphi_X(t) = 1 + i\mathbb{E}[X]t - \frac{1}{2}\mathbb{E}[X^2]t^2 + o(t^2) \quad (t \to 0)$$

Preuve :

On peut appliquer le théorème de dérivation des intégrales à paramètre à la fonction $t \mapsto \mathbb{E}[e^{itX}]$. Car :

- $\forall t \in \mathbb{R}$, $\omega \mapsto e^{itX(\omega)}$ est intégrable (puisque bornée par 1).
- Pour tout $\omega \in \Omega$, $t \mapsto e^{itX(\omega)}$ est dérivable sur $\mathbb{R}$.
- $\forall t \in \mathbb{R}, \forall \omega \in \Omega$, $|\frac{d}{dt}e^{itX(\omega)}| = |iX(\omega)e^{itX(\omega)}| \leq |X(\omega)|$. Or X est intégrable car de carré intégrable ($L^2 \subset L^1$).

Donc φ_X est dérivable sur $\mathbb{R}$ et $\varphi_X'(t) = \mathbb{E}[iXe^{itX}]$. En particulier, $\varphi_X'(0) = i\mathbb{E}[X]$.
φ_X' est continue (conséquence du TCD) et dérivable sur $\mathbb{R}$ par la même conséquence du théorème de dérivation (en dominant la dérivée seconde $|(iX)^2e^{itX}| \leq X^2$ qui est intégrable).
On a pour tout $t \in \mathbb{R}$:

$$\varphi_X''(t) = \mathbb{E}[i^2X^2e^{itX}] = -\mathbb{E}[X^2e^{itX}]$$

En particulier, $\varphi_X''(0) = -\mathbb{E}[X^2]$.
D'après la formule de Taylor-Young appliquée à φ_X en 0 à l'ordre 2 :

$$\varphi_X(t) = \varphi_X(0) + t\varphi_X'(0) + \frac{t^2}{2}\varphi_X''(0) + o(t^2)$$

Ce qui donne le résultat annoncé. $\square$

Théorème 9: Théorème de Lévy

Soient $(X_n)_{n\geq 0}$ et X des v.a. à valeurs dans $\mathbb{R}^d$. La suite $(X_n)_{n\geq 0}$ converge en loi vers X si et seulement si la suite $(\varphi_{X_n})_{n\geq 0}$ des fonctions caractéristiques converge simplement vers φ_X.

Proposition 64: Théorème Central Limite (Préliminaire)

Soit $(X_n)_{n\geq 0}$ une suite de v.a. réelles indépendantes de même loi et de carré intégrable. En notant $\mu = \mathbb{E}[X_1]$ et $\sigma^2 = \mathrm{Var}(X_1)$, si $\sigma^2 > 0$:

$$\sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

Preuve :

On commence par supposer que $\mu = 0$ et $\sigma^2 = 1$. En notant $S_n = X_1 + \cdots + X_n$, il s'agit de montrer que :

$$\forall t \in \mathbb{R}, \quad \varphi_{\frac{S_n}{\sqrt{n}}}(t) \longrightarrow e^{-\frac{t^2}{2}}$$

(où $e^{-\frac{t^2}{2}}$ est la fonction caractéristique de la loi $\mathcal{N}(0,1)$).
Or, pour tout $t \in \mathbb{R}$:

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = \mathbb{E}\left[e^{it\frac{X_1+\cdots+X_n}{\sqrt{n}}}\right]$$

Par indépendance des X_i et car ils ont la même loi :

$$= \varphi_{\frac{X_1}{\sqrt{n}}}(t)\dots\varphi_{\frac{X_n}{\sqrt{n}}}(t) = \left(\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

On a que $\varphi_{X_1}(u) = 1 - \frac{u^2}{2} + o(u^2)$ (car $\mathbb{E}[X_1] = 0, \mathbb{E}[X_1^2] = 1$). Donc :

$$\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)$$

On utilise le lemme suivant sur les nombres complexes : $\forall (a,b) \in \mathbb{C}$ tels que $|a| \leq 1$ et $|b| \leq 1$, et $\forall n \geq 1$:

$$|a^n - b^n| = \left|(a-b)\sum_{k=0}^{n-1} a^{n-1-k}b^k\right| \leq |a-b|\sum_{k=0}^{n-1} 1 = n|a-b|$$

Pour n assez grand, on pose $a_n = \varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)$ et $b_n = 1 - \frac{t^2}{2n}$.

$$\left|\left(\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n - \left(1 - \frac{t^2}{2n}\right)^n\right| \leq n\left|\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right) - \left(1 - \frac{t^2}{2n}\right)\right|$$

$$= n\left|1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) - \left(1 - \frac{t^2}{2n}\right)\right| = n\left|o\left(\frac{t^2}{n}\right)\right| \underset{n\to\infty}{\longrightarrow} 0$$

Comme $\left(1 - \frac{t^2}{2n}\right)^n \to e^{-\frac{t^2}{2}}$, on conclut que $\varphi_{\frac{S_n}{\sqrt{n}}}(t) \to e^{-\frac{t^2}{2}}$.
Cas général : Si $\mu \neq 0$ ou $\sigma^2 \neq 1$ (mais $\sigma^2 > 0$).
On a :

$$\sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right) = \frac{\sqrt{n}}{n}\sum_{k=1}^{n}\left(\frac{X_k - \mu}{\sigma}\right)$$

On pose $Y_k = \frac{X_k - \mu}{\sigma}$. On vérifie que :

$$\mathbb{E}[Y_k] = 0 \quad \text{et} \quad \text{Var}(Y_k) = 1$$

Ainsi, en posant $V_n = \sum_{k=1}^{n} Y_k$:

$$\sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Y_k = \frac{V_n}{\sqrt{n}}$$

Où V_n est la somme de $Y_1, \ldots, Y_n$ qui sont des variables aléatoires indépendantes, de même loi, d'espérance nulle ($\mathbb{E}[Y_1] = 0$) et de variance 1 ($\text{Var}(Y_1) = 1$).
On se ramène donc exactement au premier cas démontré précédemment :

$$\sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right) = \frac{V_n}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

$\square$

Lemme 3: Lemme de Slutsky

Soient $(X_n)_{n \geq 0}$, $(Y_n)_{n \geq 0}$, X et Y des variables aléatoires réelles telles que :

- $X_n \xrightarrow{\mathcal{L}} X$
- $Y_n \xrightarrow{\mathcal{L}} Y$ où Y est une constante p.s.

Alors :

$$(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, Y)$$

En particulier :

$$X_n + Y_n \xrightarrow{\mathcal{L}} X + Y \quad \text{et} \quad X_n Y_n \xrightarrow{\mathcal{L}} XY$$

Preuve :

On commence par supposer que $Y = 0$ p.s. D'après le théorème de Lévy, il s'agit de montrer que pour tout $(s, t) \in \mathbb{R}^2$:

$$\varphi_{(X_n, Y_n)}(s, t) = \mathbb{E}[e^{i(sX_n + tY_n)}] \longrightarrow \mathbb{E}[e^{i(sX + t \cdot 0)}] = \varphi_{(X,0)}(s, t)$$

Or, pour tout $(s, t) \in \mathbb{R}^2$ et $n \geq 1$, on a par l'inégalité triangulaire :

$$|\varphi_{(X_n, Y_n)}(s, t) - \varphi_{(X,0)}(s, t)| \leq \underbrace{|\varphi_{(X_n, Y_n)}(s, t) - \varphi_{(X_n,0)}(s, t)|}_{(A)} + \underbrace{|\varphi_{(X_n,0)}(s, t) - \varphi_{(X,0)}(s, t)|}_{(B)}$$

Étude du terme (B) :

$$|\varphi_{(X_n,0)}(s, t) - \varphi_{(X,0)}(s, t)| = |\varphi_{X_n}(s) - \varphi_X(s)|$$

Puisque $X_n \xrightarrow{\mathcal{L}} X$, ce terme tend vers 0 quand $n \to \infty$.
Étude du terme (A) :

$$\begin{aligned}
(A) &= \left|\mathbb{E}[e^{i(sX_n + tY_n)}] - \mathbb{E}[e^{isX_n}]\right| \\
&= \left|\mathbb{E}\left[e^{isX_n}(e^{itY_n} - 1)\right]\right| \\
&\leq \mathbb{E}\left[|e^{isX_n}| \cdot |e^{itY_n} - 1|\right] \\
&= \mathbb{E}[|e^{itY_n} - 1|] \quad (\text{car } |e^{isX_n}| = 1)
\end{aligned}$$

Puisque la fonction $y \mapsto e^{iy}$ est continue et vaut 1 en 0, pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que si $|Y_n| \leq \delta$, alors $|e^{itY_n} - 1| < \varepsilon$. Soit $\varepsilon > 0$ quelconque mais fixé et $\delta > 0$ associé.

$$\mathbb{E}[|e^{itY_n} - 1|] \leq \varepsilon + 2\mathbb{P}(|Y_n| > \delta)$$

Or $Y_n \xrightarrow{\mathcal{L}} 0 \implies Y_n \xrightarrow{\mathbb{P}} 0$. Donc le terme de droite tend vers ε. Comme c'est vrai pour tout ε, le terme (A) tend vers 0.

Cas général : Si Y est constante p.s. mais différente de 0, alors $Y_n - Y \xrightarrow{\mathcal{L}} 0$. En appliquant le résultat précédent (avec le couple $(X_n, Y_n - Y)$), on a pour tout $(s,t) \in \mathbb{R}^2$:

$$\begin{aligned}
\varphi_{(X_n,Y_n)}(s,t) &= \mathbb{E}[e^{i(sX_n+tY_n)}] \\
&= \mathbb{E}[e^{i(sX_n+t(Y_n-Y)+tY)}] \\
&= \mathbb{E}[e^{i(sX_n+t(Y_n-Y))}e^{itY}] \\
&\xrightarrow[n\to\infty]{} \mathbb{E}[e^{isX}]e^{itY} \quad \text{(car } e^{itY} \text{ est constant)} \\
&= \mathbb{E}[e^{i(sX+tY)}]
\end{aligned}$$

Donc $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, Y)$.
Pour la somme $X_n + Y_n$, on considère une fonction $\psi : \mathbb{R} \to \mathbb{R}$ continue bornée.

$$\mathbb{E}[\psi(X_n + Y_n)] = \mathbb{E}[\psi(s(X_n, Y_n))]$$

où $s(x, y) = x + y$ est une fonction continue. Puisque $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, Y)$, on a par définition de la convergence en loi (via les fonctions continues bornées) :

$$\mathbb{E}[\psi(s(X_n, Y_n))] \xrightarrow[n\to\infty]{} \mathbb{E}[\psi(s(X, Y))] = \mathbb{E}[\psi(X + Y)]$$

Donc $X_n + Y_n \xrightarrow{\mathcal{L}} X + Y$. □

Chapitre 8 : Conditionnement

I. Cas discret

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Soient X et X' des v.a. réelles. Soient Y et Y' des v.a. à valeurs dans un ensemble E dénombable.

Définition 41: Probabilité conditionnelle

$\forall y \in E$ tel que $\mathbb{P}(Y = y) > 0$, on définit une probabilité sur $(\Omega, \mathcal{A})$ en posant :

$$\mathbb{P}_{Y=y}(B) = \frac{\mathbb{P}(B \cap \{Y = y\})}{\mathbb{P}(Y = y)}$$

qu'on appelle "probabilité sachant $Y = y$".

Définition 42: Loi conditionnelle

La mesure image de cette probabilité par X est une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

$$\mathbb{P}^X_{Y=y}(B) = \mathbb{P}_{Y=y}(X \in B) = \frac{\mathbb{P}(\{X \in B\} \cap \{Y = y\})}{\mathbb{P}(Y = y)}$$

est appelée loi de X sachant $Y = y$.

Proposition 65: Calcul de l'espérance conditionnelle

Soit $y \in E$ tel que $\mathbb{P}(Y = y) > 0$.

1. Si X est une v.a. positive :
$$\int_\Omega X d\mathbb{P}_{Y=y} = \frac{\mathbb{E}[X \mathbb{1}_{Y=y}]}{\mathbb{P}(Y = y)}$$
2. Si $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, alors $X \in L^1(\Omega, \mathcal{A}, \mathbb{P}_{Y=y})$ et
$$\int_\Omega X d\mathbb{P}_{Y=y} = \frac{\mathbb{E}[X \mathbb{1}_{Y=y}]}{\mathbb{P}(Y = y)}$$

Preuve :

1) On suppose que $X \geq 0$.
$0^{\text{ème}}$ cas : On suppose $\exists A \in \mathcal{A}$ t.q. $X = \mathbb{1}_A$.

$$\int_\Omega X d\mathbb{P}_{Y=y} = \int_\Omega \mathbb{1}_A d\mathbb{P}_{Y=y} = \mathbb{P}_{Y=y}(A) = \frac{\mathbb{P}(A \cap \{Y = y\})}{\mathbb{P}(Y = y)}$$

et

$$\frac{\mathbb{E}[X \mathbb{1}_{Y=y}]}{\mathbb{P}(Y = y)} = \frac{\mathbb{E}[\mathbb{1}_A \mathbb{1}_{Y=y}]}{\mathbb{P}(Y = y)} = \frac{\mathbb{P}(A \cap \{Y = y\})}{\mathbb{P}(Y = y)}$$

1^{er} cas : Si X est étagée positive, l'égalité découle du cas précédent et de la linéarité des intégrales de fonctions positives.
$2^{\text{ème}}$ cas : Si X est mesurable positive, l'égalité est une conséquence du cas précédent et du Théorème de Convergence Monotone (TCM).

2) Si $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, alors :

$$\int_\Omega |X| d\mathbb{P}_{Y=y} = \frac{\mathbb{E}[|X|\mathbb{1}_{Y=y}]}{\mathbb{P}(Y=y)} \leq \frac{\mathbb{E}[|X|]}{\mathbb{P}(Y=y)} < +\infty$$

donc $X \in L^1(\Omega, \mathcal{A}, \mathbb{P}_{Y=y})$.
En particulier, $X^+, X^- \in L^1(\Omega, \mathcal{A}, \mathbb{P}_{Y=y})$ et alors :

$$\begin{aligned}\int_\Omega X d\mathbb{P}_{Y=y} &= \int_\Omega X^+ d\mathbb{P}_{Y=y} - \int_\Omega X^- d\mathbb{P}_{Y=y} \\ &= \frac{\mathbb{E}[X^+\mathbb{1}_{Y=y}] - \mathbb{E}[X^-\mathbb{1}_{Y=y}]}{\mathbb{P}(Y=y)} \\ &= \frac{\mathbb{E}[X\mathbb{1}_{Y=y}]}{\mathbb{P}(Y=y)}\end{aligned}$$

□

Définition 43: Espérance conditionnelle (Nombre)

Soit $y \in E$ t.q. $\mathbb{P}(Y=y) > 0$. Si X est positive ou si $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, on appelle espérance conditionnelle de X sachant $Y = y$ la quantité notée :

$$\mathbb{E}[X|Y=y] = \int_\Omega X d\mathbb{P}_{Y=y}$$

Définition 44: Espérance conditionnelle (Variable Aléatoire)

Si X est positive ou si $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, on appelle espérance conditionnelle de X sachant Y la variable aléatoire $\mathbb{E}[X|Y] = \psi(Y)$ où :

$$\psi(y) = \begin{cases} \mathbb{E}[X|Y=y] & \text{si } \mathbb{P}(Y=y) > 0 \\ 0 & \text{si } \mathbb{P}(Y=y) = 0 \end{cases}$$

Exemple 28: Lancer de dé

- $\mathcal{E}$: lancer d'un dé à 6 faces équilibré.
- $\Omega = \{1, \dots, 6\}$.
- $\mathbb{P}$: probabilité uniforme.
- X : résultat du lancer.
- $Y(\omega) = \begin{cases} 1 & \text{si } \omega \text{ est impair} \\ 0 & \text{si } \omega \text{ est pair} \end{cases}$

Pour calculer $\mathbb{E}[X|Y]$, on commence par calculer $\mathbb{E}[X|Y=y]$ pour $y = 0, 1$.
Pour $y = 1$, $\forall n \in \{1, \dots, 6\}$:

$$\begin{aligned}\mathbb{P}(X=n|Y=1) = \frac{\mathbb{P}(X=n \cap Y=1)}{\mathbb{P}(Y=1)} &= \frac{\mathbb{P}(\{X=n\} \cap \{X \text{ est impair}\})}{1/2} \\ &= \begin{cases} \frac{1/6}{1/2} = \frac{1}{3} & \text{si } n \in \{1,3,5\} \\ 0 & \text{si } n \in \{2,4,6\} \end{cases}\end{aligned}$$

Alors $\mathbb{E}[X|Y=1] = 1 \times \frac{1}{3} + 2 \times 0 + 3 \times \frac{1}{3} + 4 \times 0 + 5 \times \frac{1}{3} + 6 \times 0 = 3$.

De même, $\mathbb{E}[X|Y=0]=4$. Donc $\mathbb{E}[X|Y]=3\mathbb{1}_{\{1,3,5\}}(X)+4\mathbb{1}_{\{2,4,6\}}(X)$ (ou plus simplement $3\mathbb{1}_{Y=1}+4\mathbb{1}_{Y=0}$).

Exemple 29: Boîtes et boules

Soit $n\in\mathbb{N}^*$. On dispose de n boîtes numérotées de 1 à n. $\forall k\in[\![1,n]\!]$, la boîte numérotée k contient k boules numérotées de 1 à k.

- On tire une boîte au hasard uniformément. On note Y le numéro de la boîte.
- Dans cette boîte, on tire une boule au hasard uniformément. On note X le numéro de la boule.

Pour calculer $\mathbb{E}[X|Y]$, on considère $\forall y\in\{1,\dots,n\}$. $\mathbb{P}$ sachant $Y=y$. $\forall x\in\{1,\dots,n\}$, on a :

$$\mathbb{P}(X=x|Y=y)=\begin{cases}\frac{1}{y} & \text{si } x\in\{1,\dots,y\}\\ 0 & \text{sinon}\end{cases}$$

Donc :

$$\mathbb{E}[X|Y=y]=\sum_{x=1}^{n}x\mathbb{P}(X=x|Y=y)$$

$$=\sum_{x=1}^{y}x\times\frac{1}{y}+0=\frac{1}{y}\sum_{x=1}^{y}x=\frac{1}{y}\frac{y(y+1)}{2}=\frac{y+1}{2}=\psi(y)$$

Donc $\mathbb{E}[X|Y]=\frac{Y+1}{2}$.

Proposition 66: Propriétés

1. Si X est positive, pour toute variable aléatoire Z positive et $\sigma(Y)$-mesurable, on a :

$$\mathbb{E}[ZX]=\mathbb{E}[Z\mathbb{E}[X|Y]]$$

2. Si $X\in L^1(\Omega,\mathcal{A},\mathbb{P})$, alors :

$$\mathbb{E}[|\mathbb{E}[X|Y]|]\leq\mathbb{E}[|X|]$$

donc $\mathbb{E}[X|Y]\in L^1(\Omega,\mathcal{A},\mathbb{P})$ et pour toute variable aléatoire Z bornée et $\sigma(Y)$-mesurable, on a :

$$\mathbb{E}[ZX]=\mathbb{E}[Z\mathbb{E}[X|Y]]$$

Exemple 30: "Boules et boîtes" (suite)

D'après la proposition précédente avec $Z=1$ (qui est nécessairement $\sigma(Y)$-mesurable), on a :

$$\mathbb{E}[X]=\mathbb{E}[\mathbb{E}[X|Y]]$$

Donc :

$$\mathbb{E}[X]=\mathbb{E}\left[\frac{Y+1}{2}\right]=\frac{1}{2}\mathbb{E}[Y]+\frac{1}{2}$$

Et comme Y est de loi uniforme sur $\{1,\dots,n\}$, on a $\mathbb{E}[Y]=\frac{n+1}{2}$. Donc $\mathbb{E}[X]=\frac{n+1}{4}+\frac{1}{2}$. (À comparer avec le calcul direct de $\mathbb{E}[X]$).

Preuve :

1) Si Z est positive et $\sigma(Y)$-mesurable, il existe une fonction mesurable positive g telle que $Z = g(Y)$.

$$\begin{aligned}
\mathbb{E}[Z\mathbb{E}[X|Y]] &= \mathbb{E}[g(Y)\psi(Y)] \\
&= \sum_{y\in E} g(y)\psi(y)\mathbb{P}(Y=y) \\
&= \sum_{y\in E, \mathbb{P}(Y=y)>0} g(y)\psi(y)\mathbb{P}(Y=y) \\
&= \sum_{y\in E, \mathbb{P}(Y=y)>0} g(y)\frac{\mathbb{E}[X\mathbb{1}_{Y=y}]}{\mathbb{P}(Y=y)}\mathbb{P}(Y=y) \\
&= \sum_{y\in E, \mathbb{P}(Y=y)>0} g(y)\mathbb{E}[X\mathbb{1}_{Y=y}] \\
&= \sum_{y\in E, \mathbb{P}(Y=y)>0} \mathbb{E}[Xg(y)\mathbb{1}_{Y=y}] \\
&= \sum_{y\in E, \mathbb{P}(Y=y)>0} \mathbb{E}[g(Y)X\mathbb{1}_{Y=y}] \\
&= \mathbb{E}\left[g(Y)X\sum_{y\in E, \mathbb{P}(Y=y)>0} \mathbb{1}_{Y=y}\right] \quad \text{(par linéarité et convergence monotone/somme)} \\
&= \mathbb{E}[g(Y)X] \quad \left(\text{car } \sum \mathbb{1}_{Y=y} = 1 \text{ p.s.}\right) \\
&= \mathbb{E}[XZ]
\end{aligned}$$

2) Si $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, alors $\forall y \in E$, avec $\mathbb{P}(Y=y) > 0$, on a $X \in L^1(\Omega, \mathcal{A}, \mathbb{P}_{Y=y})$, donc $\mathbb{E}[X|Y=y]$ est correctement définie et

$$\begin{aligned}
\mathbb{E}[|\mathbb{E}[X|Y]|] &= \mathbb{E}[|\psi(Y)|] \\
&= \sum_{y\in E, \mathbb{P}(Y=y)>0} |\mathbb{E}[X\mathbb{1}_{Y=y}]|\frac{\mathbb{P}(Y=y)}{\mathbb{P}(Y=y)} \\
&\leq \sum_{y\in E, \mathbb{P}(Y=y)>0} \mathbb{E}[|X|\mathbb{1}_{Y=y}] \\
&\leq \mathbb{E}\left[|X|\left(\sum_{y\in E, \mathbb{P}(Y=y)>0} \mathbb{1}_{Y=y}\right)\right] \leq \mathbb{E}[|X|]
\end{aligned}$$

Donc $X \in L^1(\Omega, \mathcal{A}, \mathbb{P}) \implies \mathbb{E}[X|Y] \in L^1(\Omega, \mathcal{A}, \mathbb{P})$.
Soit Z bornée et $\sigma(Y)$-mesurable. Alors :

- Z^+ et Z^- sont positives, bornées et $\sigma(Y)$-mesurables.
- X^+ et X^- sont positives et intégrables.

On a :

$$\begin{aligned}
\mathbb{E}[XZ] &= \mathbb{E}[(X^+ - X^-)(Z^+ - Z^-)] \\
&= \mathbb{E}[X^+Z^+ - X^-Z^+ - X^+Z^- + X^-Z^-] \\
&= \mathbb{E}[X^+Z^+] - \mathbb{E}[X^-Z^+] - \mathbb{E}[X^+Z^-] + \mathbb{E}[X^-Z^-] \\
&= \mathbb{E}[Z^+\mathbb{E}[X^+|Y]] - \mathbb{E}[Z^+\mathbb{E}[X^-|Y]] - \mathbb{E}[Z^-\mathbb{E}[X^+|Y]] + \mathbb{E}[Z^-\mathbb{E}[X^-|Y]] \\
&= \mathbb{E}[Z^+(\mathbb{E}[X^+|Y] - \mathbb{E}[X^-|Y])] - \mathbb{E}[Z^-(\mathbb{E}[X^+|Y] - \mathbb{E}[X^-|Y])] \\
&= \mathbb{E}[Z^+\mathbb{E}[X|Y]] - \mathbb{E}[Z^-\mathbb{E}[X|Y]] \\
&= \mathbb{E}[Z\mathbb{E}[X|Y]]
\end{aligned}$$

□

Proposition 67: Propriété caractéristique de l'espérance conditionnelle

Si $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ et V est une v.a. intégrable et $\sigma(Y)$-mesurable telle que pour toute v.a. Z $\sigma(Y)$-mesurable et bornée, on a :

$$\mathbb{E}[XZ] = \mathbb{E}[VZ]$$

Alors $V = \mathbb{E}[X|Y]$ p.s.

Preuve :

On sait déjà que pour toute v.a. Z $\sigma(Y)$-mesurable et bornée, on a $\mathbb{E}[XZ] = \mathbb{E}[Z\mathbb{E}[X|Y]]$.
Soit V une v.a. $\sigma(Y)$-mesurable et intégrable telle que $\forall Z$ $\sigma(Y)$-mesurable et bornée, $\mathbb{E}[XZ] = \mathbb{E}[ZV]$.
Alors $\forall Z$ $\sigma(Y)$-mesurable et bornée, on a :

$$\mathbb{E}[ZV] = \mathbb{E}[Z\mathbb{E}[X|Y]]$$

D'où :

$$\mathbb{E}[Z(V - \mathbb{E}[X|Y])] = 0$$

Soit $V - \mathbb{E}[X|Y]$ est $\sigma(Y)$-mesurable, donc l'égalité précédente avec les cas particuliers :

- $\mathbb{1}_{\{V-\mathbb{E}[X|Y]>0\}}$ qui est $\sigma(Y)$-mesurable bornée
- $\mathbb{1}_{\{V-\mathbb{E}[X|Y]<0\}}$ qui est $\sigma(Y)$-mesurable bornée

donne :

$$\mathbb{E}\left[(V - \mathbb{E}[X|Y])\mathbb{1}_{\{V-\mathbb{E}[X|Y]>0\}}\right] = 0$$

Donc $(V - \mathbb{E}[X|Y])^+ = 0$ p.s.
Et

$$\mathbb{E}\left[(\mathbb{E}[X|Y] - V)\mathbb{1}_{\{V-\mathbb{E}[X|Y]<0\}}\right] = 0$$

Donc $(V - \mathbb{E}[X|Y])^- = 0$ p.s.
Donc $V - \mathbb{E}[X|Y] = 0$ p.s. □

Proposition 68

Si $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ et Y et Y' sont à valeurs dans E et telles que $\sigma(Y) = \sigma(Y')$, alors $\mathbb{E}[X|Y] = \mathbb{E}[X|Y']$ p.s.

Preuve :

Pour tout v.a. Z bornée et $\sigma(Y) = \sigma(Y')$ mesurable :

$$\mathbb{E}[XZ] = \mathbb{E}[Z\mathbb{E}[X|Y]] = \mathbb{E}[Z\mathbb{E}[X|Y']]$$

Donc $\mathbb{E}[Z(\mathbb{E}[X|Y] - \mathbb{E}[X|Y'])] = 0$. Et en prenant $Z = \mathbb{1}_{\{\mathbb{E}[X|Y]-\mathbb{E}[X|Y']>0\}}$, on peut affirmer que $(\mathbb{E}[X|Y] - \mathbb{E}[X|Y'])^+ = 0$ p.s. Et $Z = \mathbb{1}_{\{\mathbb{E}[X|Y']-\mathbb{E}[X|Y]>0\}}$, on peut affirmer que $(\mathbb{E}[X|Y] - \mathbb{E}[X|Y'])^- = 0$ p.s.
Donc $\mathbb{E}[X|Y] = \mathbb{E}[X|Y']$ p.s. □

II. Espérance conditionnelle des v.a. intégrables

Proposition 69

Soit $\mathcal{B} \subset \mathcal{A}$ une sous-tribu de $\mathcal{A}$ et $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$.
Il existe un unique élément de $L^1(\Omega, \mathcal{B}, \mathbb{P})$ noté $\mathbb{E}[X|\mathcal{B}]$ tel que $\forall B \in \mathcal{B}$ on a :

$$\mathbb{E}[X\mathbb{1}_B] = \mathbb{E}[\mathbb{1}_B \mathbb{E}[X|\mathcal{B}]]$$

Preuve : Unicité : Soient X' et X'' des v.a. intégrables et $\mathcal{B}$-mesurables tq :

$$\forall B \in \mathcal{B}, \quad \mathbb{E}[X\mathbb{1}_B] = \mathbb{E}[X'\mathbb{1}_B]$$

$$\mathbb{E}[X\mathbb{1}_B] = \mathbb{E}[X''\mathbb{1}_B]$$

Alors $\{X' > X''\} \in \mathcal{B}$ et $\{X'' > X'\} \in \mathcal{B}$.
Et puisque $\forall B \in \mathcal{B}$, $\mathbb{E}[X'\mathbb{1}_B] = \mathbb{E}[X''\mathbb{1}_B]$, on a :

$$\mathbb{E}[X'\mathbb{1}_{\{X'>X''\}}] = \mathbb{E}[X''\mathbb{1}_{\{X'>X''\}}]$$

Donc :

$$\mathbb{E}[(X' - X'')\mathbb{1}_{\{X'>X''\}}] = 0 \implies (X' - X'')^+ = 0 \text{ p.s.}$$

Et :

$$\mathbb{E}[X''\mathbb{1}_{\{X''>X'\}}] = \mathbb{E}[X'\mathbb{1}_{\{X''>X'\}}]$$

Donc :

$$\mathbb{E}[(X'' - X')\mathbb{1}_{\{X''>X'\}}] = 0 \implies (X' - X'')^- = 0 \text{ p.s.}$$

Donc $X' = X''$ p.s.

Existence : On commence par supposer que $X \geqslant 0$ p.s.
En posant $Q(B) = \mathbb{E}[X\mathbb{1}_B]$, on définit une probabilité sur $(\Omega, \mathcal{B})$.
Si $\mathbb{P}(B) = 0$, on a nécessairement $Q(B) = 0$, donc Q est absolument continue par rapport à $\mathbb{P}$. D'après le théorème de Radon-Nikodym, il existe une fonction $\widetilde{X} \geqslant 0$ et $\mathcal{B}$-mesurable tq $\forall B \in \mathcal{B}$:

$$Q(B) = \mathbb{E}[\widetilde{X}\mathbb{1}_B] = \mathbb{E}[X\mathbb{1}_B]$$

De plus, en prenant $B = \Omega$, on a $\mathbb{E}[\widetilde{X}] = \mathbb{E}[X] < +\infty$, donc $\widetilde{X} \in L^1(\Omega, \mathcal{B}, \mathbb{P})$.

Si X n'est pas positive p.s., on pose :

$$\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X^+|\mathcal{B}] - \mathbb{E}[X^-|\mathcal{B}]$$

Remarque

- De façon équivalente, $\mathbb{E}[X|\mathcal{B}]$ est l'unique élément de $L^1(\Omega, \mathcal{B}, \mathbb{P})$ tq pour toute v.a. Z, $\mathcal{B}$-mesurable bornée :
$$\mathbb{E}[XZ] = \mathbb{E}[Z\mathbb{E}[X|\mathcal{B}]] \quad \text{(déf)}$$
- Si Y est une v.a., $\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)]$.
- $(X, Y) \overset{\mathcal{L}}{=} (X', Y)$ alors $\forall B \in \sigma(Y)$, $\exists \varphi$ mesurable tq $\mathbb{1}_B = \varphi(Y)$.
$$\mathbb{E}[X\mathbb{1}_B] = \mathbb{E}[\mathbb{1}_B \mathbb{E}[X|Y]]$$
$$\mathbb{E}[X'\mathbb{1}_B] = \mathbb{E}[\mathbb{1}_B \mathbb{E}[X'|Y]]$$
Donc d'après la propriété d'unicité $\mathbb{E}[X|Y] = \mathbb{E}[X'|Y]$ p.s.

Exemple 31: Cas Discret

$\Omega =]0,1]$.
$\mathcal{A} = \mathcal{B}(]0,1])$.
$\mathbb{P} = \lambda$ mesure de Lebesgue.
Soit $n \in \mathbb{N}^*$. $\mathcal{B} = \sigma\left(\left]\frac{i-1}{2^n}, \frac{i}{2^n}\right], i \in \{1, \dots, 2^n\}\right)$.
Soit $f \in L^1(\Omega, \mathcal{A}, \mathbb{P})$.
Comme $\mathbb{E}[f|\mathcal{B}]$ est $\mathcal{B}$-mesurable, il existe $\alpha_1, \dots, \alpha_{2^n} \in \mathbb{R}$ tq :

$$\mathbb{E}[f|\mathcal{B}] = \sum_{i=1}^{2^n} \alpha_i \mathbb{1}_{]\frac{i-1}{2^n}, \frac{i}{2^n}]}$$

C'est une conséquence de la propriété caractéristique que pour tout $j = 1, \dots, 2^n$:

$$\mathbb{E}\left[f\mathbb{1}_{]\frac{j-1}{2^n}, \frac{j}{2^n}]}\right] = \mathbb{E}\left[\mathbb{1}_{]\frac{j-1}{2^n}, \frac{j}{2^n}]}\mathbb{E}[f|\mathcal{B}]\right]$$

Or :

$$\mathbb{E}\left[f\mathbb{1}_{]\frac{j-1}{2^n}, \frac{j}{2^n}]}\right] = \int_{]0,1]} f\mathbb{1}_{]\frac{j-1}{2^n}, \frac{j}{2^n}]} d\lambda = \int_{]\frac{j-1}{2^n}, \frac{j}{2^n}]} f d\lambda$$

Et :

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}_{]\frac{j-1}{2^n}, \frac{j}{2^n}]}\mathbb{E}[f|\mathcal{B}]\right] &= \int_{]0,1]} \left(\sum_{i=1}^{2^n} \alpha_i \mathbb{1}_{]\frac{i-1}{2^n}, \frac{i}{2^n}]}\right) \mathbb{1}_{]\frac{j-1}{2^n}, \frac{j}{2^n}]} d\lambda \\
&= \alpha_j \int_{]\frac{j-1}{2^n}, \frac{j}{2^n}]} 1 d\lambda = \frac{\alpha_j}{2^n}
\end{aligned}$$

Donc :

$$\alpha_j = \frac{1}{\frac{1}{2^n}} \int_{]\frac{j-1}{2^n}, \frac{j}{2^n}]} f d\lambda$$

Exemple 32: Cas Continu (Densité)

$(X, Y) \in \mathbb{R}^2$ admet $p : \mathbb{R}^2 \to \mathbb{R}^+$ pour densité.
Alors Y admet $q(y) = \int_{\mathbb{R}} p(x,y) dx$ comme densité et si $q(y) = 0$ alors $p(x,y) = 0$ pour x λ-p.p.
On cherche $\mathbb{E}[h(X)|Y]$ où $h : \mathbb{R} \to \mathbb{R}$ est tq $h(X) \in L^1(\Omega, \mathcal{A}, \mathbb{P})$.
C'est-à-dire, on cherche $\varphi : \mathbb{R} \to \mathbb{R}$ mesurable tq pour toute fonction $g : \mathbb{R} \to \mathbb{R}$ mesurable bornée :

$$\mathbb{E}[h(X)g(Y)] = \mathbb{E}[g(Y)\varphi(Y)]$$

En effet, pour tout Z, $\sigma(Y)$ mesurable bornée, $\forall g(Y)$, on a :

$$\mathbb{E}[h(X)Z] = \mathbb{E}[Z\mathbb{E}[h(X)|\sigma(Y)]]$$

Ici $Z = g(Y)$ et $\mathbb{E}[h(X)|\sigma(Y)] = \varphi(Y)$.

Pour toute fonction $g : \mathbb{R} \to \mathbb{R}$ mesurable bornée :

$$\begin{aligned}
\mathbb{E}[h(X)g(Y)] &= \int_{\mathbb{R}^2} h(x)g(y)p(x,y)dxdy \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x)p(x,y)dx \right) g(y)dy \\
&= \int_{\mathbb{R}} \left(\frac{\int_{\mathbb{R}} h(x)p(x,y)dx}{q(y)} \right) g(y)q(y)\mathbb{1}_{\{q(y)>0\}}dy + 0 \\
&= \int_{\mathbb{R}} \left(\frac{\int_{\mathbb{R}} h(x)p(x,y)dx}{q(y)} \mathbb{1}_{\{q(y)>0\}} \right) g(y)q(y)dy \\
&= \mathbb{E}[\varphi(Y)g(Y)]
\end{aligned}$$

où :

$$\varphi(y) = \begin{cases} \frac{\int_{\mathbb{R}} h(x)p(x,y)dx}{q(y)} & \text{si } q(y) > 0 \\ h(0) & \text{si } q(y) = 0 \end{cases}$$

III. Propriétés de l'espérance conditionnelle

Proposition 70: Propriétés de l'espérance conditionnelle

Soit $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. Soit $\mathcal{B}$ une sous-tribu de $\mathcal{A}$.

1. Si X est $\mathcal{B}$-mesurable, alors $\mathbb{E}[X \mid \mathcal{B}] = X$ p.s.
2. **Linéarité :** Pour tout $\lambda \in \mathbb{R}$ et $X' \in L^1$,
$$\mathbb{E}[X + \lambda X' \mid \mathcal{B}] = \mathbb{E}[X \mid \mathcal{B}] + \lambda \mathbb{E}[X' \mid \mathcal{B}] \quad \text{p.s.}$$
3. **Espérance totale :** $\mathbb{E}[\mathbb{E}[X \mid \mathcal{B}]] = \mathbb{E}[X]$.
4. **Positivité :** Si $X \geq 0$, alors $\mathbb{E}[X \mid \mathcal{B}] \geq 0$ p.s.
5. **Inégalité de Jensen (version conditionnelle) :**
$$|\mathbb{E}[X \mid \mathcal{B}]| \leq \mathbb{E}[|X| \mid \mathcal{B}] \quad \text{p.s.}$$
En particulier, $\mathbb{E}[|\mathbb{E}[X \mid \mathcal{B}]|] \leq \mathbb{E}[|X|]$.
6. **Croissance :** Si $X \leq X'$ p.s., alors $\mathbb{E}[X \mid \mathcal{B}] \leq \mathbb{E}[X' \mid \mathcal{B}]$ p.s.

Preuve :

1) Pour toute v.a. Z bornée et $\mathcal{B}$-mesurable, on a trivialement :
$$\mathbb{E}[XZ] = \mathbb{E}[XZ]$$
Or, comme X est elle-même $\mathcal{B}$-mesurable (par hypothèse), elle vérifie la propriété caractéristique de l'espérance conditionnelle (avec le "candidat" étant X lui-même). Par unicité (p.s.), on a $\mathbb{E}[X \mid \mathcal{B}] = X$ p.s.

2) Soit $\lambda \in \mathbb{R}$. On sait que $X + \lambda X' \in L^1(\Omega, \mathcal{A}, \mathbb{P})$.
Pour toute v.a. Z bornée et $\mathcal{B}$-mesurable :
$$\mathbb{E}[Z(X + \lambda X')] = \mathbb{E}[Z\mathbb{E}[X + \lambda X' \mid \mathcal{B}]] \quad \text{(par définition)}$$

Mais par linéarité de l'espérance classique, on a aussi :

$$\begin{aligned}\mathbb{E}[Z(X+\lambda X')] &= \mathbb{E}[ZX]+\lambda\mathbb{E}[ZX'] \\ &= \mathbb{E}[Z\mathbb{E}[X \mid \mathcal{B}]]+\lambda\mathbb{E}[Z\mathbb{E}[X' \mid \mathcal{B}]] \\ &= \mathbb{E}[Z(\mathbb{E}[X \mid \mathcal{B}]+\lambda\mathbb{E}[X' \mid \mathcal{B}])] \quad \text{(par linéarité de l'intégrale)}\end{aligned}$$

D'après la propriété caractéristique (qui assure l'unicité p.s.), on identifie :

$$\mathbb{E}[X+\lambda X' \mid \mathcal{B}] = \mathbb{E}[X \mid \mathcal{B}]+\lambda\mathbb{E}[X' \mid \mathcal{B}] \quad \text{p.s.}$$

3) Cela découle directement de la définition en prenant la variable test $Z = 1$ (qui est bien bornée et $\mathcal{B}$-mesurable).

$$\mathbb{E}[1\cdot X] = \mathbb{E}[1\cdot\mathbb{E}[X \mid \mathcal{B}]] \implies \mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{B}]].$$

4) Soit $A = \{\mathbb{E}[X \mid \mathcal{B}] < 0\}$. Comme $\mathbb{E}[X \mid \mathcal{B}]$ est $\mathcal{B}$-mesurable, alors $A \in \mathcal{B}$. On considère la variable test $Z = \mathbb{1}_A$.

$$\mathbb{E}[\mathbb{1}_A X] = \mathbb{E}[\mathbb{1}_A\mathbb{E}[X \mid \mathcal{B}]]$$

Comme $X \geq 0$, le terme de gauche $\mathbb{E}[\mathbb{1}_A X]$ est ≥ 0. Le terme de droite est l'intégrale d'une fonction strictement négative sur A qui est donc inferieure ou egale a 0. On en deduit que $\mathbb{E}[\mathbb{1}_A\mathbb{E}[X \mid \mathcal{B}]] = 0$ p.s On doit donc avoir $\mathbb{1}_A\mathbb{E}[X \mid \mathcal{B}] = 0$ p.s., ce qui implique $\mathbb{1}_A = 0$ p.s. Ainsi $\mathbb{P}(A) = 0$, donc $\mathbb{E}[X \mid \mathcal{B}] \geq 0$ p.s.

5) On écrit $X = X^+ - X^-$.

$$|\mathbb{E}[X \mid \mathcal{B}]| = |\mathbb{E}[X^+ \mid \mathcal{B}] - \mathbb{E}[X^- \mid \mathcal{B}]| \leq \mathbb{E}[X^+ \mid \mathcal{B}] + \mathbb{E}[X^- \mid \mathcal{B}] = \mathbb{E}[X^+ + X^- \mid \mathcal{B}] = \mathbb{E}[|X| \mid \mathcal{B}].$$

5 bis) En intégrant cette inégalité : $\mathbb{E}[|\mathbb{E}[X \mid \mathcal{B}]|] \leq \mathbb{E}[\mathbb{E}[|X| \mid \mathcal{B}]] = \mathbb{E}[|X|]$.

6) Si $X \leq X'$ p.s., alors $X' - X \geq 0$ p.s. Par linéarité et positivité (point 4) :

$$\mathbb{E}[X' - X \mid \mathcal{B}] = \mathbb{E}[X' \mid \mathcal{B}] - \mathbb{E}[X \mid \mathcal{B}] \geq 0 \quad \text{p.s.}$$

D'où $\mathbb{E}[X \mid \mathcal{B}] \leq \mathbb{E}[X' \mid \mathcal{B}]$ p.s.

□

IV. Espérance conditionnelle des v.a. positives

Proposition 71: Définition et Extension

Soit X une v.a. positive (non nécessairement dans L^1). En posant :

$$\mathbb{E}[X \mid \mathcal{B}] = \lim_{n\to+\infty} \mathbb{E}[X \wedge n \mid \mathcal{B}] \quad \text{p.s.}$$

On définit une v.a. $\mathcal{B}$-mesurable positive telle que pour toute v.a. Z $\mathcal{B}$-mesurable positive :

$$\mathbb{E}[ZX] = \mathbb{E}[Z\mathbb{E}[X \mid \mathcal{B}]]$$

Preuve :

Si X est positive, pour tout $n \in \mathbb{N}$, $X \wedge n \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ (car bornée). Donc $\mathbb{E}[X \wedge n \mid \mathcal{B}]$ est correctement définie. La suite $(\mathbb{E}[X \wedge n \mid \mathcal{B}])_{n\geq 0}$ est une suite croissante de v.a. positives p.s. (par la propriété de croissance vue précédemment, car $X \wedge n \leq X \wedge (n+1)$). Donc la limite $\lim_{n\to\infty} \mathbb{E}[X \wedge n \mid \mathcal{B}]$ existe p.s. (dans $\bar{\mathbb{R}}_+$).

Pour toute v.a. Z $\mathcal{B}$-mesurable et positive (on peut supposer Z bornée dans un premier temps puis étendre par TCM) :

$$\begin{aligned}
\mathbb{E}[Z\mathbb{E}[X \mid \mathcal{B}]] &= \mathbb{E}\left[Z \lim_{n\to\infty} \mathbb{E}[X \wedge n \mid \mathcal{B}]\right] \\
&= \lim_{n\to\infty} \mathbb{E}[Z\mathbb{E}[X \wedge n \mid \mathcal{B}]] \quad \text{(TCM, car tout est positif)} \\
&= \lim_{n\to\infty} \mathbb{E}[Z(X \wedge n)] \quad \text{(Propriété caractéristique pour les v.a. intégrables)} \\
&= \mathbb{E}[Z \lim_{n\to\infty} (X \wedge n)] \quad \text{(TCM)} \\
&= \mathbb{E}[ZX]
\end{aligned}$$

Unicité ? Soit X une v.a. ≥ 0. Supposons qu'il existe X' et X'' deux v.a. positives, $\mathcal{B}$-mesurables telles que pour toute v.a. Z positive et $\mathcal{B}$-mesurable :

$$\mathbb{E}[ZX] = \mathbb{E}[ZX'] \quad \text{et} \quad \mathbb{E}[ZX] = \mathbb{E}[ZX'']$$

Alors $\mathbb{E}[ZX'] = \mathbb{E}[ZX'']$.
Soit $\mathcal{Q}^+ = \mathbb{Q} \cap [0, +\infty[$. Pour tous $a, b \in \mathcal{Q}^+$ tels que $a < b$, on considère l'ensemble

$$A = \{X' \leq a < b \leq X''\} \in \mathcal{B}$$

En prenant $Z = \mathbb{1}_A$ (qui est bien $\mathcal{B}$-mesurable), on a :

$$\mathbb{E}[X'\mathbb{1}_{\{X' \leq a < b \leq X''\}}] = \mathbb{E}[X''\mathbb{1}_{\{X' \leq a < b \leq X''\}}]$$

Or sur l'ensemble A, on a $X' \leq a$ et $X'' \geq b$. Donc :

$$\mathbb{E}[X'\mathbb{1}_A] \leq a\mathbb{P}(A) \quad \text{et} \quad \mathbb{E}[X''\mathbb{1}_A] \geq b\mathbb{P}(A)$$

L'égalité des espérances implique donc $b\mathbb{P}(A) \leq a\mathbb{P}(A)$. Comme $b > a$, cela force $\mathbb{P}(A) = 0$. Donc $\mathbb{P}(X' \leq a < b \leq X'') = 0$.
Or $\{X' < X''\} = \bigcup_{a,b\in\mathcal{Q}^+, a<b} \{X' \leq a < b \leq X''\}$. C'est une union dénombrable d'ensembles de mesure nulle, donc :

$$\mathbb{P}(X' < X'') = 0$$

On prouve de la même façon que $\mathbb{P}(X'' < X') = 0$. Donc $X' = X''$ p.s. □

Exemple 33: Filtration dyadique

Soit $\Omega =]0, 1]$, $\mathcal{A} = \mathcal{B}(]0, 1])$, $\mathbb{P} = \lambda$ (mesure de Lebesgue). Soit $n \in \mathbb{N}^*$. On définit la tribu engendrée par les intervalles dyadiques :

$$\mathcal{B}_n = \sigma\left(\left]\frac{i-1}{2^n}, \frac{i}{2^n}\right], i = 1, \ldots, 2^n\right)$$

Si $f \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, alors l'espérance conditionnelle est constante sur chaque intervalle dyadique et vaut la moyenne de la fonction sur cet intervalle :

$$\mathbb{E}[f \mid \mathcal{B}_n] = \sum_{k=1}^{2^n} 2^n \left(\int_{]\frac{k-1}{2^n}, \frac{k}{2^n}]} f \, d\lambda\right) \mathbb{1}_{]\frac{k-1}{2^n}, \frac{k}{2^n}]}$$

Cas d'une fonction positive non intégrable : Considérons $g : \Omega \to \mathbb{R}^+, \omega \mapsto \frac{1}{\omega}$. g est une v.a. positive, mais $\int_\Omega g \, d\lambda = [\ln(\omega)]_0^1 = +\infty$, donc $g \notin L^1$. Calculons $\mathbb{E}[g \mid \mathcal{B}_n]$ via la limite des troncatures $g \wedge m$ ($m \to \infty$).

Sur le premier intervalle $]\frac{0}{2^n}, \frac{1}{2^n}]$, l'intégrale de g diverge.

$$\mathbb{E}[g \wedge m \mid \mathcal{B}_n] = \sum_{k=1}^{2^n} 2^n \left(\int_{]\frac{k-1}{2^n}, \frac{k}{2^n}]} (g \wedge m)\, d\lambda \right) \mathbb{1}_{]\frac{k-1}{2^n}, \frac{k}{2^n}]}$$

En passant à la limite $m \to \infty$:

$$\mathbb{E}[g \mid \mathcal{B}_n] = \lim_{m \to \infty} \mathbb{E}[g \wedge m \mid \mathcal{B}_n] = \infty \cdot \mathbb{1}_{]0, \frac{1}{2^n}]} + \sum_{k=2}^{2^n} \left(2^n \int_{\frac{k-1}{2^n}}^{\frac{k}{2^n}} \frac{1}{\omega} d\omega \right) \mathbb{1}_{]\frac{k-1}{2^n}, \frac{k}{2^n}]}$$

Pour $k \geq 2$, l'intégrale vaut $\ln(\frac{k}{2^n}) - \ln(\frac{k-1}{2^n}) = \ln(\frac{k}{k-1})$. Donc :

$$\mathbb{E}[g \mid \mathcal{B}_n] = +\infty \cdot \mathbb{1}_{]0, \frac{1}{2^n}]} + \sum_{k=2}^{2^n} 2^n \ln\left(\frac{k}{k-1}\right) \mathbb{1}_{]\frac{k-1}{2^n}, \frac{k}{2^n}]}$$