INTÉGRATION ET PROBABILITÉS, DÉBUTS SUCCINCTS POUR LE M2

par

Ph Bougerol, Paris 6

Table des matières

PARTIE I
TRIBUS

1. Tribus

Un espace mesurable est un couple $(E, \mathcal{E})$ formé d'un ensemble E et d'une tribu $\mathcal{E}$ sur E où

Définition 1.1. — Une tribu $\mathcal{E}$ sur E est une famille de parties de E telle que

— $E \in \mathcal{E}$
— Si $A \in \mathcal{E}$, le complémentaire A^c est dans $\mathcal{E}$.
— Si $A_n, n \in \mathbb{N}$, sont dans $\mathcal{E}$ alors $\cup_{n\in\mathbb{N}} A_n \in \mathcal{E}$.

Par comparaison, une topologie sur E est une famille $\mathcal{O}$ de parties de E, appelés ouverts, telle que

— $E, \emptyset \in \mathcal{O}$,
— Si $A, B \in \mathcal{O}$ alors $A \cap B \in \mathcal{O}$,
— Une réunion quelconque déléments de $\mathcal{O}$ est dans $\mathcal{O}$.

Un exemple explicite de tribu sur E est l'ensemble des parties de E, surtout utilisé lorsque E est fini ou dénombrable. Dans les autres cas, cette tribu est souvent inutile car on ne peut pas y définir de mesure intéressante. La plupart du temps, il est difficile (sinon impossible) de décrire explicitement tous les éléments d'une tribu.

On utilise

Définition 1.2. — Etant donné un ensemble $\mathcal{G}$ de parties de E, on appelle tribu engendrée par $\mathcal{G}$ et on note $\sigma(\mathcal{G})$ l'intersection de toutes les tribus contenant $\mathcal{G}$.

2. Tribu borélienne

Définition 2.1. — Si E est un espace topologique muni d'une topologie $\mathcal{O}$, on appelle tribu borélienne de E, et on note $\mathcal{B}(E)$, la tribu $\sigma(\mathcal{O})$.

Prenons par exemple un espace métrique (E, d) c'est à dire un ensemble E muni d'une application

$$d : E \times E \to \mathbb{R}^+$$

appelée distance, vérifiant

$$d(x, y) = 0 \text{ ssi } x = y,$$

$$d(x, y) \leq d(x, z) + d(z, y), \text{ pour tous } x, y, z \in E.$$

La boule ouverte $B(x, \varepsilon)$ est $\{y \in E; d(x, y) < \varepsilon\}$ et par définition la topologie associée est formé des ensembles réunion quelconque de boules.

Les exemple suivants sont importants :

1. $\mathbb{R}$, $\mathbb{R}^d$ muni de la distance usuelle.

2. L'ensemble $C([0, 1], \mathbb{R})$ des fonctions continues de $[0, 1]$ dans $\mathbb{R}$ muni de la distance

$$d(f, g) = \sup_{0 \leq x \leq 1} |f(x) - g(x)|$$

Dans ces cas on ne peut pas décrire tous les éléments de la tribu borélienne explicitement. Sur un espace topologique on utilisera systématiquement la tribu borélienne sans le mentionner.

Sur $\bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$, on met la tribu engendrée par les ensembles de la forme $]a, +\infty], a \in \mathbb{R}$.

3. Application mesurable

Définition 3.1. — Etant donné deux espaces mesurables $(E_1, \mathcal{E}_1)$ et $(E_2, \mathcal{E}_2)$, une application

$$f : E_1 \to E_2$$

est dite mesurable si pour tout $A \in \mathcal{E}_2$, $\{f \in A\} = \{x \in E_1; f(x) \in A\}$ est dans $\mathcal{E}_2$.

Lemme 3.2 **(Critère de mesurabilité).** — *Si $\mathcal{E}_2 = \sigma(\mathcal{G})$, $f : E_1 \to E_2$ est mesurable dès que $\{f \in A\} \in \mathcal{E}_1$ pour tout $A \in \mathcal{G}$.*

Preuve. — Sous cette condition

$$\{A \in \mathcal{E}_2; \{f \in A\} \in \mathcal{E}_1\}$$

est une tribu contenant $\mathcal{G}$ donc contenant $\sigma(\mathcal{G})$. □

Lemme 3.3. — *Une composée d'applications mesurables est mesurable.*

Preuve. — Immédiat □

3.1. Exemples. — Les tribus boréliennes de $\mathbb{R}$ et de $\bar{\mathbb{R}}$ sont engendrées par les ensembles $]a, +\infty]$. Donc, en appliquant le lemme précédent,

Lemme 3.4. — *Une application $f : (E, \mathcal{E}) \to \mathbb{R}$ où $\bar{\mathbb{R}}$ est mesurable dès que $\{f > a\} \in \mathcal{E}$, pour tout $a \in \mathbb{R}$.*

On peut y remplacer $\{f > a\}$ par $\{f \geq a\}$. Toujours avec le lemme 3.2,

Lemme 3.5. — *Si E_1 et E_2 sont deux espaces topologiques, une application continue $f : E_1 \to E_2$ est mesurable.*

Lemme 3.6. — *Si $f_1, f_2 : E \to \mathbb{R}$ sont mesurables, $f_1 + f_2$, $f_1 f_2$, $\max(f_1, f_2)$, $\min(f_1, f_2)$, etc... sont mesurables.*

Preuve. — La fonction $f : E \to \mathbb{R}^2$ est mesurable (on applique le lemme 3.2 à $\mathcal{G} = \{] - \infty, a[\times] - \infty, b[, a, b \in \mathbb{R}\}$. L'application $\phi : \mathbb{R}^2 \to \mathbb{R}$ définie par $\phi(x, y) = x + y$ est mesurable car continue. Donc $f_1 + f_2 = \phi \circ f$ est mesurable, pareil pour les autres... □

Si (x_n) est une suite de réels on pose

$$\limsup_{n\to+\infty} x_n = \lim_{n\to+\infty} \sup_{k\geq n} x_k = \inf_{n\in\mathbb{N}} \sup_{k\geq n} x_k$$

$$\liminf_{n\to+\infty} x_n = \lim_{n\to+\infty} \inf_{k\geq n} x_k = \sup_{n\in\mathbb{N}} \inf_{k\geq n} x_k.$$

Elles existent toujours dans $\bar{\mathbb{R}}$, c'est leur intérêt. De plus elles sont égales, si et seulement si la suite x_n converge dans $\bar{\mathbb{R}}$ vers cette valeur commune.

Proposition 3.7. — *Si pour tout $n \in \mathbb{N}$, $f_n : E \to \mathbb{R}$ est mesurable alors $\sup_{n\in\mathbb{N}} f_n, \inf_{n\in\mathbb{N}} f_n, \limsup_{n\to+\infty} f_n, \liminf_{n\to+\infty} f_n$ sont mesurables.*

Preuve. — La fonction $\sup_{n\in\mathbb{N}} f_n$ définie par $(\sup_{n\in\mathbb{N}} f_n)(x) = \sup_{n\in\mathbb{N}} f_n(x)$ vérifie

$$\{\sup_{n\in\mathbb{N}} f_n > a\} = \cup_{n\in\mathbb{N}}\{f_n > a\}$$

et est donc mesurable. Pareil pour l'inf on peut par exemple utiliser que

$$\inf f_k = -\sup(-f_k).$$

Ensuite $\limsup_{n\to+\infty} f_n = \inf_{n\geq 0} \sup_{k\geq n} f_k$ est donc mesurable. □

4. Espaces produits

Donnons nous un ensemble T d'indices, en pratique $T = \{1, 2, \cdots, n\}$ ou $T = \mathbb{N}$, ou $T = \mathbb{R}^+$, et pour tout $t \in T$ un espace mesuré $(E_t, \mathcal{E}_t)$. On considère l'espace produit

$$E = \prod_{t \in T} E_t.$$

Par définition un élément $\omega \in E$ est une famille $\omega = (\omega_t, t \in T)$ où $\omega_t \in E_t$. L'application

$$\omega \mapsto \omega_t$$

est appelée la t-ième coordonnée. Si $E_t = F$ pour tout $t \in T$, on utilise la notation $E = F^T$.

Définition 4.1. — La tribu produit $\mathcal{E} = \otimes_{t \in T} \mathcal{E}_t$ est la tribu sur E engendrée par les ensembles du type

$$\{\omega \in E; \omega_t \in A_t\}$$

où $t \in T$ et $A_t \in \mathcal{E}_t$.

Il résulte du lemme 3.2 que

Proposition 4.2. — *Soit $(F, \mathcal{F})$ un espace mesuré. Une application $f : F \to \prod_{t \in T} E_t$ est mesurable si et seulement si les coordonnées f_t sont mesurables.*

5. Classe monotone

Comme on ne peut pas en général décrire explicitement les éléments d'une tribu, il nous faut un outil puissant pour utiliser les tribus engendrées. C'est le rôle du théorème de la classe monotone.

Définition 5.1. — Un sous ensemble $\mathcal{M}$ se parties de E est une classe monotone si

— $E \in \mathcal{M}$,
— Si $A, B \in \mathcal{M}$ et $A \subset B$ alors $B \setminus A \in \mathcal{M}$,
— Si, pour tout entier n, $A_n \in \mathcal{M}$ et $A_n \subset A_{n+1}$ alors $\cup_{n \geq 0} A_n \in \mathcal{M}$.

Théorème 5.2 **(Théorème de la classe monotone)**

Si $\mathcal{C}$ est un ensemble de parties de E stable par intersection finie (i.e, si $A, B \in \mathcal{C}$, alors $A \cap B \in \mathcal{C}$), une classe monotone contenant $\mathcal{C}$ contient la tribu $\sigma(\mathcal{C})$.

Preuve. — On va montrer que l'intersection $\mathcal{M}$ de toutes les classes monotones contenant $\mathcal{C}$ est une tribu donc contient $\sigma(\mathcal{C})$.

Il suffit de montrer que $\mathcal{M}$ est stable par intersection finie. On le fait en deux étapes :

Première étape : Soit $A \in \mathcal{C}$, alors

$$\mathcal{M}_A = \{B \in \mathcal{M} \text{ tel que } B \cap A \in \mathcal{M}\}$$

est une classe monotone contenant $\mathcal{C}$ donc $\mathcal{M}$. Donc si $A \in \mathcal{C}$ et $B \in \mathcal{M}$, alors $B \cap A \in \mathcal{M}$.

Deuxième étape : Soit $B \in \mathcal{M}$, alors

$$\mathcal{M}_B = \{A \in \mathcal{M} \text{ tel que } A \cap B \in \mathcal{M}\}$$

est une classe monotone contenant $\mathcal{C}$ par l'étape précédente donc $\mathcal{M}$. □

Des situations typiques de classe $\mathcal{C}$ pour appliquer le théorème :

— $E = \mathbb{R}, \mathcal{C} = \{]-\infty, x], x \in \mathbb{R}\}, \sigma(\mathcal{C}) =$ tribu borélienne de $\mathbb{R}$.

— $E = E_1 \times E_2, \mathcal{C} = \{A_1 \times A_2, A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2, \sigma(\mathcal{C}) = \mathcal{E}_1 \otimes \mathcal{E}_2\}$.

— $E = \prod_{t \in T} E_t, \sigma(\mathcal{C}) = \prod_{t \in T} \mathcal{E}_t$, où

$$\mathcal{C} = \{\omega \in \mathbb{R}; n \in \mathbb{N}, \omega_{t_i} \in T, A_i \in \mathcal{E}_{t_i}, 1 \leq i \leq n\}.$$

PARTIE II
MESURE ET INTÉGRATION

6. Mesure

Soit $(E, \mathcal{E})$ un espace muni d'une tribu.

Définition 6.1. — Une mesure sur E est une application $m : \mathcal{E} \to \mathbb{R}^+ \cup \{+\infty\}$ telle que

— $m(\emptyset) = 0$,

— Si les ensembles $A_n \in \mathcal{E}, n \in \mathbb{N}$ sont disjoints,

$$m(\cup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} m(A_n).$$

(Pour nous une mesure est toujours à valeurs positives). Le seul exemple explicite simple de mesure est essentiellement donné dans le cas où E est fini ou dénombrable, muni de la tribu des parties de E et de la mesure

$$m = \sum_{x \in E} a(x) \delta_x$$

où $a(x) \geq 0$ et δ_x, appelé la mesure de Dirac en x, est défini par

$$\delta_x(A) = \mathbf{1}_A(x)$$

où $\mathbf{1}_A$ est la fonction indicatrice de A donné par $\mathbf{1}_A(x) = 1$ si $x \in A$ et $\mathbf{1}_A(x) = 0$ sinon.

Lorsque $a(x) = 1$ pour tout $x \in E$, m s'appelle la mesure de comptage sur E. Lorsque E est fini et $a(x) = 1/\mathrm{card}(E)$, m s'appelle la mesure uniforme sur E.

Définition 6.2. — On dit que la mesure m est

— bornée, si $m(E) < +\infty$,
— une probabilité si $m(E) = 1$,
— σ-finie, si il existe $A_n \in \mathcal{E}$ tels que $m(A_n) < +\infty$ et $E = \cup_{n\in\mathbb{N}} A_n$.

7. Intégration au sens de Lebesgue

7.1. Intégration des fonctions étagées. — On travaille sur $(E, \mathcal{E}, m)$ où m est une mesure.

Définition 7.1. — Une fonction $f : E \to \mathbb{R}$ est dite étagée si elle est mesurable et ne prend qu'un nombre fini de valeurs

Soit f étagée positive, si $f(E) = \{a_1, \cdots, a_n\}$ on pose

$$\int f\,dm = \sum_{i=1}^{n} a_i m(f = a_i)$$

avec la convention $0. + \infty = 0$. Si $A \in \mathcal{E}$,

$$\int \mathbf{1}_A\,dm = m(A).$$

Lemme 7.2. — *Si f et g sont étagées positives,*

$$\int (f+g)\,dm = \int f\,dm + \int g\,dm$$

Preuve. — Posons $f(E) = \{a_i, i \le 1 \le n\} =$, $g(E) = \{b_j, 1 \le j \le m\}$, $(f+g)(E) = \{c_k, 1 \le k \le p\}$. Alors

$$\int (f+g)\,dm = \sum_{k=1}^{p} c_k\, m(f+g = c_k)$$

$$= \sum_{k=1}^{p} \sum_{\{i,j; a_i+b_j = c_k\}} (a_i + b_j)\, m(f = a_i, g = b_j)$$

$$= \sum_{k=1}^{p} \sum_{\{i,j; a_i+b_j = c_k\}} a_i\, m(f = a_i, g = b_j) + \sum_{k=1}^{p} \sum_{\{i,j; a_i+b_j = c_k\}} b_j\, m(f = a_i, g = b_j).$$

Or

$$\begin{aligned}
&\sum_{k=1}^{p} \sum_{\{i,j;a_i+b_j=c_k\}} a_i\, m(f=a_i, g=b_j) \\
&= \sum_{k=1}^{p} \sum_{i=1}^{n} \sum_{\{j;a_i+b_j=c_k\}} a_i\, m(f=a_i, g=b_j) \\
&= \sum_{i=1}^{n} a_i \sum_{k=1}^{p} \sum_{\{j;a_i+b_j=c_k\}} m(f=a_i, g=b_j) \\
&= \sum_{i=1}^{n} a_i m(f=a_i) = \int f\, dm.
\end{aligned}$$

et pareil pour l'autre bout. □

D'autre part, de façon évidente, pour $\alpha > 0$, $\int \alpha f\, dm = \alpha \int f\, dm$ donc avec le lemme précédent,

Lemme 7.3. — *Si* $0 \leq f = \sum_{i=1}^{r} \alpha_i \mathbf{1}_{A_i}$, $\int f\, dm = \sum \alpha_i m(A_i)$.

En écrivant que $f = (f-g) + g$ on obtient que

Lemme 7.4. — *Si* $0 \leq g \leq f$ *sont des fonctions étagées,*

$$0 \leq \int f\, dm \leq \int g\, dm.$$

7.2. Intégration des fonctions positives. —

Définition 7.5. — Si $f : E \to \mathbb{R}$ est une fonction mesurable positive, on pose

$$\int f\, dm = \sup_{\{g \text{ étagée}\,;\, 0 \leq g \leq f\}} \int g\, dm \leq +\infty.$$

Théorème 7.6 **(Théorème de convergence monotone)**
Si $f_n, n \in \mathbb{N}$, *est une suite croissante de fonctions mesurables positives*

$$\lim_{n\to+\infty} \int f_n\, dm = \int \lim_{n\to+\infty} f_n\, dm.$$

Preuve. — La fonction $f = \lim f_n$ étant égale à $\limsup f_n$ est mesurable. Par le lemme 7.4, si $0 \leq n \leq m$

$$0 \leq \int f_n\, dm \leq \int f_m\, dm \leq \int f\, dm$$

donc

$$\lim_{n\to+\infty} \int f_n\, dm = \int f\, dm.$$

Reste à montrer l'inégalité inverse. Prenons une fonction g étagée telle que $0 \leq g \leq f$ et un réel $0 < c < 1$. Les ensembles

$$E_n = \{f_n \geq cg\}$$

forment une suite croissante telle que $\cup_{n\geq 0} E_n = E$. On a

$$\int f_n \, dm \geq \int f_n \mathbf{1}_{E_n} \, dm \geq \int cg\mathbf{1}_{E_n} \, dm$$

Si $g = \sum_{i=1}^r \alpha_i \mathbf{1}_{A_i}$, $cg\mathbf{1}_{E_n} = \sum_{i=1}^r c\alpha_i \mathbf{1}_{A_i \cap E_n}$ donc

$$\int cg\mathbf{1}_{E_n} \, dm = \sum_{i=1}^r c\alpha_i m(A_i \cap E_n).$$

On en déduit que $\int cg\mathbf{1}_{E_n} \, dm \to c \int g \, dm$ quand $n \to +\infty$ d'où

$$\lim_{n\to+\infty} \int f_n \, dm \geq c \int g \, dm,$$

ceci étant vrai pour tout $0 < c < 1$. On fait tendre c vers 1 pour obtenir que

$$\lim_{n\to+\infty} \int f_n \, dm \geq \int g \, dm$$

ce qui entraîne que

$$\lim_{n\to+\infty} \int f_n \, dm \geq \sup_{\{g \text{ étagée}; 0 \leq g \leq f\}} \int g \, dm = \int f \, dm$$

□

Lemme 7.7. — *Toute fonction mesurable positive $f : E \to \mathbb{R}$ est la limite d'une suite croissante (f_n) de fonctions étagées positives.*

Preuve. — On peut prendre par exemple

$$f_n = \sum_{k=0}^{n(2^n-1)} \frac{k}{2^n} \mathbf{1}_{\{\frac{k}{2^n} \leq f < \frac{k+1}{2^n}\}} + n\mathbf{1}_{\{n \leq f\}}$$

□

Proposition 7.8. — *Si f et g sont deux fonctions mesurables positives et $\alpha, \beta > 0$*

$$\int (\alpha f + \beta g) \, dm = \alpha \int f \, dm + \beta \int g \, dm$$

Preuve. — On écrit $f = \lim f_n, g = \lim g_n$ où f_n et g_n sont étagées, on applique le lemme 7.2 puis le théorème de convergence monotone. □

Corollaire 7.9. — *Si (g_n) est une suite de fonctions mesurables positives*

$$\int \sum_{n=0}^{+\infty} g_n \, dm = \sum_{n=0}^{+\infty} \int g_n \, dm.$$

Preuve. — On applique le théorème de convergence monotone à $f_n = \sum_{k=0}^{n} g_k$. □

Le résultat suivant, souvent négligé, est essentiel. On l'applique souvent lorsque $\liminf f_n = \lim f_n$ est une vraie limite.

***Lemme 7.10* (Lemme de Fatou).** — *Si (f_n) est une suite de fonctions mesurables positives,*

$$\int \liminf_{n \to +\infty} f_n \, dm \leq \liminf_{n \to +\infty} \int f_n \, dm$$

Preuve. — On applique le théorème de convergence monotone à la suite croissante $g_n = \inf_{k \geq n} f_k$:

$$\int \liminf f_n \, dm = \int \lim g_n \, dm = \lim \int g_n \, dm,$$

et on remarque que, pour tout $k \geq n$

$$\int g_n \, dm \leq \int f_k \, dm$$

donc

$$\int g_n \, dm \leq \inf_{k \geq n} \int f_k \, dm.$$

□

***Proposition 7.11* (Inégalité de Markov).** — *Si f est mesurable positive, pour tout $a > 0$,*

$$m(f \geq a) \leq \frac{\int f \, dm}{a}.$$

Preuve. — On intègre des deux cotés l'inégalité $\mathbf{1}_{\{f \geq a\}} \leq f/a$. □

7.3. Fonctions intégrables. — Si $f : E \to \mathbb{R}$ ou $f : E \to \bar{\mathbb{R}}$ on pose

$$f^+ = \max(f, 0), f^- = \max(-f, 0)$$

Alors $f = f^+ - f^-$ et $|f| = f^+ + f^-$.

Définition 7.12. — On dit qu'une fonction mesurable $f : E \to \bar{\mathbb{R}}$ est intégrable si $\int |f| \, dm < +\infty$. Dans ce cas on pose

$$\int f \, dm = \int f^+ \, dm - \int f^- \, dm.$$

On emploie la notation, si $A \in \mathcal{E}$

$$\int \mathbf{1}_A f \, dm = \int_A f \, dm.$$

On a $|\int f \, dm| \leq \int |f| \, dm$. On dit qu'une propriété est vraie presque partout ou presque surement (noté p.p. ou p.s.) si elle est vraie sur le complémentaire d'un ensemble de mesure nulle (donc avec probabilité 1 lorsque m est une probabilité).

Proposition 7.13. — *Une fonction intégrable est finie p.p. Une fonction positive d'intégrale nulle est nulle p.p.*

Preuve. — Si f est intégrable on fait tendre a vers $+\infty$ pour obtenir, avec l'inégalité de Markov, que

$$m(|f| = +\infty) = \lim_{a \to +\infty} m(|f| > a) \leq \lim_{a + \infty} \frac{\int |f| \, dm}{a} = 0,$$

et si f est positive d'intégrale nulle, on fait tendre a vers 0,

$$m(|f| \neq 0) = \lim_{a \to 0} m(|f| > a) \leq \lim_{a + \infty} \frac{\int |f| \, dm}{a} = 0.$$

□

***Lemme 7.14* (de Borel Cantelli)**. — *Soit (A_n) une suite de sous ensembles mesurables de E telle que $\sum_{n=0}^{\infty} m(A_n) < +\infty$. Alors*

$$\sum_{k=0}^{+\infty} \mathbf{1}_{A_n} < +\infty \ p.p.$$

Autrement dit, presque tout $x \in E$ n'appartient qu'à un nombre fini de A_n.

Preuve. — Par convergence monotone,

$$\int \sum_{k=0}^{+\infty} \mathbf{1}_{A_n} \, dm = \sum_{k=0}^{+\infty} m(A_n) < +\infty$$

donc $\sum_{n=0}^{+\infty} \mathbf{1}_{A_n} < +\infty$ p.p. par la proposition précédente. □

***Théorème 7.15* (Théorème de convergence dominée)**

Soit $f_n, n \in \mathbb{N}$ une suite de fonctions mesurables telle que

— *$f_n \to f$ quand $n \to +\infty$, p.p.*

— *Il existe une fonction intégrable g telle que $|f_n|(x) \leq g(x)$, p.p.*

Alors

$$\lim_{n \to +\infty} \int f_n \, dm = \int f \, dm.$$

Preuve. — On a $|f_n - f| \leq 2g$ p.p. donc par le lemme de Fatou appliquéé à $2g - |f_n - f|$

$$\begin{aligned}\int 2g\, dm &\leq \int \liminf(2g - |f_n - f|)\, dm \\ &\leq \liminf \int 2g - |f_n - f|\, dm \\ &\leq \int 2g\, dm - \limsup \int |f_n - f|\, dm.\end{aligned}$$

Donc $\limsup \int |f_n - f|\, dm = 0$ et

$$|\int f_n\, dm - \int f\, dm| \leq \int |f_n - f|\, dm$$

tend vers 0. □

7.4. Dérivation sous le signe somme. — On travaille encore sur $(E, \mathcal{E})$. Etant donné un intervalle ouvert I de $\mathbb{R}$ on se donne une fonction $f = I \times \mathbb{E} \to \mathbb{R}$.

Théorème 7.16. — *On suppose que*

- *Pour tout $t \in I$, $x \mapsto f(t, x)$ est intégrable.*
- *Pour tout $x \in E$, $t \to f(t, x)$ est dérivable.*
- *Il existe une fonction intégrable $g : E \to \mathbb{R}$ telle que, pour tout $t \in I$ et $x \in E$,*

$$|\frac{d}{dt} f(t, x)| \leq g(x)$$

Alors sur I

$$\frac{d}{dt} \int f(t, x)\, dm(x) = \int \frac{d}{dt} f(t, x)\, dm(x).$$

Preuve. — Rappelons que par le théorème des accroissements finis, pour x fixé, pour tout $t, s \in I$ il existe u entre t et s tel que

$$\frac{f(t, x) - f(s, x)}{t - s} = \frac{d}{dt} f(u, x)$$

donc

$$\frac{|f(t, x) - f(s, x)|}{|t - s|} \leq g(x).$$

On a, par convergence dominée,

$$\begin{aligned}\frac{d}{dt} \int f(t, x)\, dm(x) &= \lim_{s \to t} \frac{\int f(t, x)\, dm(x) - \int f(s, x)\, dm(x)}{t - s} \\ &= \lim_{s \to t} \int \frac{f(t, x) - f(s, x)}{t - s}\, dm(x)\end{aligned}$$

$$= \int \lim_{s\to t} \frac{f(t,x)-f(s,x)}{t-s}\, dm(x) = \int \frac{d}{dt} f(t,x)\, dm(x)$$

$\square$

8. Mesure à densité

Définition 8.1. — On dit qu'une mesure μ sur $(E,\mathcal{E})$ à une densité ϕ par rapport à une mesure m et on écrit

$$d\mu = \phi\, dm$$

si pour tout $A \in \mathcal{E}$

$$\mu(A) = \int_A \phi\, dm$$

Proposition 8.2. — *Si $d\mu = \phi\, dm$ et $f : E \to \mathbb{R}^+$ est mesurable*

$$\int f d\mu = \int f\phi\, dm$$

Preuve. — Puisque f est limite d'une suite croissante de fonctions étagées on se ramène (avec le thm de convergence monotone) au cas où f est une fonction indicatrice. $\square$

Proposition 8.3. — *Une mesure m est σ-finie si et seulement si il existe une probabilité ν et une fonction $\phi : E \to \mathbb{R}^+$ mesurable telle que $dm = \phi\, d\nu$.*

Preuve. — Si $dm = \phi\, d\nu$ les ensembles $\{\phi \leq n\}, n \in \mathbb{N}$ sont de mesure finie pour m et de réunion E donc m est σ-finie. Réciproquement, si m est σ-finie, avec $m(A_n) < +\infty$ et $\cup A_n = E$, on peut supposer que $m(A_0) = 0$ et $m(A_n) > 0$ pour $n \geq 1$. Il suffit alors de poser

$$d\nu = \sum_{n=1}^{+\infty} \frac{1}{2^n m(A_n)} \mathbf{1}_{A_n}\, dm$$

$\square$

Définition 8.4. — On dit que la mesure m est absolument continue par rapport (resp. équivalente) à la mesure μ si $\mu(A) = 0$ entraîne que (resp. si et seulement si) $m(A) = 0$.

***Théorème 8.5* (Théorème de Radon Nikodym)**. — *Si m et μ sont σ-finies, μ est absolument continue par rapport à m si et seulement si μ a une densité par rapport à m.*

Preuve reportée.

On voit que μ est équivalente à m ssi $d\mu = \phi\, dm$ avec ϕ strictement positive m p.p.

9. Inégalités

Une fonction $\phi : \mathbb{R} \to \mathbb{R}$ est convexe si

$$\phi(\alpha x + (1-\alpha)y) \leq \alpha\phi(x) + (1-\alpha)\phi(y)$$

pour tous $x, y \in \mathbb{R}$ et $0 \leq \alpha \leq 1$ (on montre qu'elle est continue).

***Proposition 9.1* (Inégalité de Jensen).** — *Si $\phi : \mathbb{R} \to \mathbb{R}$ est convexe et m est une probabilité,*

$$\phi(\int f\, dm) \leq \int \phi \circ f\, dm.$$

Preuve. — L'ensemble $C = \{(x,y) \in \mathbb{R}^2; y \geq f(x)\}$ est convexe et fermé. Il résulte du théorème de projection (cf.) que si $E(\phi) = \{(a,b) \in \mathbb{R}; \phi(x) \geq ax + b, \forall x \in \mathbb{R}\}$ alors, pour tout $x \in \mathbb{R}$,

$$\phi(x) = \sup\{ax + b; (a,b) \in E(f)\}$$

Si $(a,b) \in E(f)$, $f(X) \geq aX + b$ donc $\int \phi(f(x))\, dm(x) \geq \int (af(x)+b)\, dm(x) = a\int f(x)\, dm(x) + b$ d'où

$$\int \phi(f(x))\, dm(x) \geq \sup\{a\int f(x)\, dm(x) + b\}; (a,b) \in E(\phi)\} = \phi(\int f\, dm)$$

□

***Proposition 9.2* (Inégalité de Hölder).** — *Si $0 \leq a \leq 1$ et $f, gù$ mesurables positives,*

$$\int f^a g^{1-a}\, dm \leq (\int f\, dm)^a (\int g\, dm)^{1-a}.$$

Preuve. — La fonction log étant concave, si $u, v > 0$,

$$a\log(u^{1/a}) + (1-a)\log(u^{1/1-a}) \leq a\log(au^{1/a} + (1-a)u^{1/1-a})$$

donc

$$uv \leq au^{1/a} + (1-a)v^{1/1-a}.$$

On prend $u = (\frac{f}{\int f\, dm})^a$ et $v = (\frac{g}{\int g\, dm})^{1-a}$ et on intègre. On obtient

$$\int \frac{f^a g^{1-a}}{(\int f\, dm)^a (\int g\, dm)^{1-a}}\, dm \leq a\int \frac{f}{\int f\, dm}\, dm + (1-a)\int \frac{g}{\int g\, dm}\, dm = 1$$

d'où la proposition. □

Pour $a = 1/2$ on obtient (cf. aussi une preuve directe plus bas)

***Lemme 9.3* (inégalité de Cauchy Schwarz).** — *Pour f, g mesurables*

$$(\int fg\, dm)^2 \leq \int f^2\, dm \int g^2\, dm.$$

10. Espaces $\mathcal{L}^1$ et $\mathcal{L}^2$

Une norme sur un espace vectoriel V est une application $x \mapsto \|x\|$ de V dans $\mathbb{R}$ telle que, pour tous $x, y \in V$

— $\|x+y\| \leq \|x\| + \|y\|$

— $\|x\| \geq 0$

— $\|x\| = 0$ si seulement si $x = 0$.

On lui associe une distance par la formule $d(x,y) = \|x-y\|$.

Soit $(E, \mathcal{E}, m)$ un espace muni d'une mesure m. Si $f : E \to \bar{\mathbb{R}}$ est mesurable on pose

$$\|f\|_1 = \int |f|\, dm, \ \ \|f\|_2 = (\int f^2\, dm)^{1/2}.$$

On note $\mathcal{L}^1$ ou $\mathcal{L}^1(E, \mathcal{E}, m)$,l'ensemble des fonctions f intégrables, et $\mathcal{L}^2$ l'ensemble des fonctions f de carré intégrable, i.e. telles que $\|f\|_2 < +\infty$. On a, pour $p = 1, 2$,

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

Ceci est immédiat pour $p = 1$ et résulte pour $p = 2$ de Cauchy Schwarz.

On dit qu'une suite de fonctions f_n de $\mathcal{L}^p$ converge vers f dans $\mathcal{L}^p$, si

$$\lim_{n \to +\infty} \|f_n - f\|_p = 0.$$

Théorème 10.1. — *Si une suite f_n converge vers f dans $\mathcal{L}^p$, $p = 1, 2$, il existe une sous suite d'entiers n_k telle que, pour presque tout $x \in E$*

$$f_{n_k}(x) \to f(x).$$

Preuve. — Ecrivons la preuve pour $p = 1$. Puisque $f_n \to f$ dans $\mathcal{L}^1$, pour tout $k \in \mathbb{N}$, il existe n_k tel que

$$\|f_{n_k} - f\|_1 \leq \frac{1}{2^k}.$$

Alors, par le corollaire 7.9,

$$\int \sum_{k=0}^{\infty} |f_{n_k} - f|\, dm = \sum_{k=0}^{\infty} \int |f_{n_k} - f|\, dm = \sum_{k=0}^{\infty} \|f_{n_k} - f\| \leq \sum_{k=0}^{\infty} \frac{1}{2^k} < +\infty.$$

Donc (cf. Proposition 7.13) la série $\sum_{k=0}^{\infty} |f_{n_k} - f|$ converge p.p. ce qui entraîne que son terme général $f_{n_k} - f$ tend vers 0, p.p. □

Définition 10.2. — La suite (f_n) est de Cauchy dans $\mathcal{L}^p$ si pour tout $\varepsilon > 0$ il existe $N > 0$ tel que pour $m, n \geq N$

$$\|f_n - f_m\| \leq \varepsilon.$$

***Théorème 10.3* ($\mathcal{L}^p$ est complet)**. — *Soit (f_n) une suite de Cauchy dans $\mathcal{L}^p$, $p = 1, 2$. Alors il existe $f \in \mathcal{L}^p$ tel que $f_n \to f$ dans $\mathcal{L}^p$.*

Preuve. — Ecrivons la preuve pour $p = 1$. Puisque la suite (f_n) est de Cauchy, pour tout $k \in \mathbb{N}$, il existe N_k tel que, pour $n, m \geq N_k$,

$$\|f_n - f_m\| \leq 2^{-k}.$$

Si on prend $n_k = \max(N_0, \cdots, N_k)$ on voit que

$$\|f_{n_{k+1}} - f_{n_k}\| \leq 2^{-k}.$$

Donc, un peu comme dans la preuve précédente,

$$\int \sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}| \, dm = \sum_{k=0}^{\infty} \|f_{n_{k+1}} - f_{n_k}\| \leq \sum_{k=0}^{\infty} \frac{1}{2^k} < +\infty.$$

Donc (cf. Proposition 7.13) presque partout, $\sum_{k=0}^{\infty} |f_{n_k} - f_{n_{k+1}}|$ est fini et la série $\sum_{k=0}^{\infty} (f_{n_k} - f_{n_{k+1}})$ est donc convergente (car absolument convergente). On a

$$f_{n_p} = f_{n_0} + \sum_{k=0}^{p-1} (f_{n_{k+1}} - f_{n_k})$$

donc f_{n_p} converge p.p. vers f quand $p \to +\infty$ où

$$f = f_{n_0} + \sum_{k=0}^{\infty} (f_{n_{k+1}} - f_{n_k}).$$

Pour tout $n \geq N_k, p \geq k$, $\|f_n - f_{n_p}\| \leq 2^{-k}$ donc, en appliquant le lemme de Fatou,

$$\int |f_n - f| \, dm \leq \int \liminf_{p \to +\infty} |f_n - f_{n_p}| \, dm \leq 2^{-k}$$

d'où $\|f_n - f\|$ tend vers 0. □

Le fait que $\|f\| = 0$, n'entraîne pas que f est nulle partout mais seulement presque partout (Proposition 7.13). Donc $\| \cdot \|$ n'est pas une norme, mais seulement une semi-norme, et $d(f, g) = \|f - g\|$ pas vraiment une distance. Ce n'est pas génant il suffit de remplacer l'égalité entre fonctions par l'égalité presque sure (on peut aussi plus formellement remplacer $\mathcal{L}^p$ par l'ensemble quotient L^p des classes de fonctions égales presque partout).

11. Espace de Hilbert

Un role particulier est joué par $\mathcal{L}^2$ car il a une structure géométrique plus riche que les autres. On peut y définir les angles et surtout l'orthogonalité, à partir du produit scalaire : pour $f, g \in \mathcal{L}^2$

$$\langle f, g \rangle = \int fg \, dm,$$

pour lequel $\|f\|_2 = \sqrt{\langle f, f \rangle}$.

Sur un espace vectoriel (réel) H, un produit scalaire est une application $(x, y) \mapsto \langle x, y\rangle$ de H^2 dans $\mathbb{R}$ telle que $(x, y) \in H^2 \mapsto \langle x, y\rangle$ telle que, pour tous $x, y, z \in H, \alpha, \beta \in \mathbb{R}$,

$$\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle,$$

$$\langle x, y\rangle = \langle y, x\rangle$$

$$\langle x, x\rangle \geq 0 \text{ et } \langle x, x\rangle = 0 \text{ ssi } x = 0.$$

On pose $\|x\| = \sqrt{\langle x, x\rangle}$. Il résulte du lemme suivant que c'est une norme.

***Lemme 11.1* (Cauchy Schwarz).** — *Pour tous $x, y \in H$,*

$$\langle x, y\rangle \leq \|x\|\|y\|.$$

Preuve. — Pour tout $\lambda \in \mathbb{R}$, le trinome en λ du second degré

$$\lambda^2\|x\|^2 + 2\lambda\langle x, y\rangle + \|y\|^2 = \langle \lambda x + y, \lambda x + y\rangle$$

est positif, ce qui n'est possible que si son discriminant $\langle x, y\rangle - \|x\|\|y\|$ est négatif. □

***Définition 11.2*.** — Un espace de Hilbert (réel) H est un espace vectoriel muni d'un produit scalaire, complet pour la norme associée.

La propriété fondamentale d'un espace de Hilbert est qu'on peut y définir la projection orthogonale sur un convexe fermé : rappelons d'abord qu'un convexe est un sous ensemble C tel que, si $x, y \in C$, $\lambda x + (1 - \lambda)y \in C$, pour tout λ tel que $0 \leq \lambda \leq 1$.

***Théorème 11.3*.** — *Soit C un convexe fermé d'un espace de Hilbert H. Pour tout $x \in H$, il existe un unique $v \in C$, appelé projection orthogonale de x sur C tel que*

$$\|x - v\| = \inf\{\|x - z\|, z \in C\}.$$

Lorsque C est un sous espace vectoriel fermé, il est caractérisé par les deux propriétés, (i) $v \in C$; (ii) pour tout $w \in C$, $\langle x - v, w\rangle = 0$.

Preuve. — Posons $D = \inf\{\|x - c\|, c \in C\}$. Pour tout $n \in \mathbb{N}$, il existe $c_n \in C$ tel que

$$\|x - c_n\|^2 \leq D^2 + \frac{1}{n}$$

Montrons que (c_n) est une suite de Cauchy, en utilisant l'égalité suivante immédiate à vérifier en développant :

$$\|a - b\|^2 + \|a + b\|^2 = 2(\|a\|^2 + \|b\|^2).$$

On a en prenant $a = c_n - x, b = c_m - x$

$$\|c_n - c_m\|^2 + \|c_n + c_m - 2x\|^2 = 2(\|c_n - x\|^2 + \|c_m - x\|^2)$$

que l'on écrit

$$\|c_n - c_m\|^2 = 2\|c_n - x\|^2 + 2\|c_m - x\|^2 - 4\|\frac{c_n + c_m}{2} - x\|^2$$

On a $\|\frac{c_n+c_m}{2} - x\| \geq D^2$ car $(c_n + c_m)/2 \in C$ on en déduit que, si $m \geq n$,

$$\|c_n - c_m\|^2 \leq 2(D^2 + \frac{1}{n}) + 2(D^2 + \frac{1}{m}) - 4D^2 \leq 2/n.$$

La suite c_n est donc de Cauchy, soit v sa limite. On a bien

$$\|x - v\| = D$$

Pour l'unicité, si $v' \in C$ est tel que $\|x - v'\| = D$ alors $\|x - \frac{v+v'}{2}\| < D$ ce qui est absurde. Enfin,si C est un espace vectoriel, pour tout $w \in C$ et $t \in \mathbb{R}$, $v + tw \in C$ donc

$$D^2 \leq \|x - v - tw\|^2 = \|x - v\|^2 + t^2\|w\|^2 - 2t\langle x - v, w\rangle$$

donc $2\langle x-v, w\rangle \leq t\|w\|^2$ d'où $\langle x-v, w\rangle \leq 0$. Remplacant w par $-w$ on a bien $\langle x - v, w\rangle = 0$. □

Corollaire 11.4. — *Dans un espace de Hilbert H, pour toute forme linéaire continue $f : H \to \mathbb{R}$, il existe un $z \in H$ tel que $f(x) = \langle x, z\rangle$ pour tout $x \in H$.*

Preuve. — Si f est identiquement nulle, prendre $z = 0$. Sinon, on applique le théorème au sous espace $V = f^{-1}(0)$ et à un $x \in H$ tel que $f(x) \neq 0$: il existe $z \in V$ tel que $f(y) = 0$ entraîne que $\langle x - z, y\rangle = 0$. Posons $w = x - z$. Alors, pour tout $y \in H$, $y - \frac{f(y)}{f(w)}w \in V$ donc $\langle y, w\rangle = \frac{f(y)}{f(w)}\langle w, w\rangle$ et $f(y) = \langle y, v\rangle$ pour $v = \frac{f(w)}{\langle w,w\rangle}w$. □

Proposition 11.5. — *Soit W un sous espace vectoriel d'un Hilbert H. Si le seul vecteur orthogonal à W est 0, alors W est dense dans H*

Preuve. — Projeter sur l'adhérence de W. □

Définition 11.6. — On appelle mesure de Radon sur un espace topologique E une mesure m sur les boréliens telle que $m(K) < +\infty$ pour tout compact K de E.

Corollaire 11.7. — *Si m est une mesure de Radon sur $\mathbb{R}$, l'ensemble des fonctions $\mathbf{1}_{[a,b]}, a < b$ est dense dans $\mathcal{L}^2(\mathbb{R}, m)$.*

Preuve. — Soit W le sous espace vectoriel de $\mathcal{L}^2$ engendreé par les fonctions $\mathbf{1}_{[a,b]}, a < b$. Si f est orthogonal à W,

$$0 = \langle \mathbf{1}_{[a,b]}, f\rangle = \int_a^b f\,dm$$

pour tout $a < b$. L'ensemble

$$\mathcal{M} = \{A \text{ borélien de } \mathbb{R}; \int_A f\, dm = 0\}$$

est alors une classe monotone (par le théorème de convergence dominé) contenant la classe $\mathcal{C}$ des intervalles $[a, b]$, stable par intersection finie, donc tous les boréliens par le théorème de la classe monotone. On a donc

$$\int_A f\, dm = 0, \text{ pour tout borélien } A \text{ de } \mathbb{R}$$

Prenant $A = \{f > 0\}$ puis $A = \{f < 0\}$ on en déduit que f est nul p.p. On conclut alors avec la proposition précédente. □

Pour une probabilité, par Cauchy Schwarz

$$\|f\|_1 \le \|f\|_2$$

donc $\mathcal{L}^2$ est contenu dans $\mathcal{L}^1$ et la convergence dans $\mathcal{L}^2$ entraîne celle dans $\mathcal{L}^1$

12. Mesure produit

Soit $(E_1, \mathcal{E}_1, m_1), (E_2, \mathcal{E}_2, m_2)$ deux espaces mesurés où m_1 et m_2 sont σ-finies. On note $\mathcal{E}_1 \otimes \mathcal{E}_2$ la tribu produit engendrée par les pavés

$$\mathcal{P} = \{A_1 \times A_2, A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2\}$$

qui forment une classe stable par intersection finie. Du fait qu'une mesure σ-finie a une densité par rapport à une mesure de probabilités (cf. Proposition 8.3), on pourra sans perte de généralité supposer dans les preuves qui suivent qu'on travaille avec des probabilités quand on applique le théorème de la classe monotone.

Proposition 12.1. — *Si $f : E_1 \times E_2 \to \mathbb{R}$ est une fonction $\mathcal{E}_1 \otimes \mathcal{E}_2$-mesurable positive, alors*

$$y \mapsto \int f(x, y)\, dm_1(x)$$

est une fonction sur E_2, $\mathcal{E}_2$-mesurable.

Preuve. — Puisque f est limite d'une suite croissante de fonctions étagées il suffit (avec le thm de convergence monotone) de traiter le cas où f est une fonction indicatrice. La classe

$$\mathcal{M} = \{A \in \mathcal{E}_1 \otimes \mathcal{E}_2; y \mapsto \int \mathbf{1}_A(x, y)\, dm_1(x)\}$$

est une classe monotone contenant clairement $\mathcal{P}$ donc $\sigma(\mathcal{P}) = \mathcal{E}_1 \otimes \mathcal{E}_2$, par le théorème de la classe monotone. □

Grace à ce lemme on peut définir une mesure m sur $E_1 \times E_2$ par la formule

$$m(A) = \int [\int \mathbf{1}_A(x,y)\, dm_1(x)] dm_2(y)$$

Théorème 12.2. — *Soit $f : E_1 \times E_2 \to \mathbb{R}$ une fonction mesurable, si*
-ou bien f est positive
-ou bien f est intégrable par rapport à m
alors

$$\begin{aligned}\int f\, dm &= \int \{ \int f(x,y)\, dm_1(x) \}\, dm_2(y) \\ &= \int \{ \int f(x,y)\, dm_2(y) \}\, dm_1(x)\end{aligned}$$

Preuve. — Si f est positive, comme pour la proposition au dessus on se ramène au cas où f est une fonction indicatrice et on utilise le théorème de la classe monotone. Sinon on écrit que $f = f^+ - f^-$. □

Remarquons que toujours d'après le théorème de la classe monotone, m est l'unique mesure telle que

$$m(A_1 \times A_2) = m_1(A_1) m_2(A_2)$$

pour tout $A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2$.

PARTIE III
PROBABILITÉS

13. Variables aléatoires

Dans le monde probabiliste on se donne un premier espace "abstrait" $(\Omega, \mathcal{F})$ que l'on n'aura pas besoin d'expliciter, et des espaces "concrets" $(E, \mathcal{E})$, où le plus souvent E est $\mathbb{R}$, $\mathbb{R}^d$, ou $C([0,1], \mathbb{R})$ munis de la tribu borélienne ou un espace produit $E = \prod_{t \in T} \mathbb{R}$.

Une variable aléatoire à valeurs dans $(E, \mathcal{E})$ est par définition une application mesurable de Ω dans E. Elle s'interprète comme un lien entre l'espace abstrait et l'espace concret.

La tribu $\sigma(X)$ engendrée par X est exactement la classe des ensembles $\{X \in A\}$, où A est dans $\mathcal{E}$.

Lemme 13.1. — *Considérons une application $X : \Omega \to E$. Alors une application $Z : \Omega \to \mathbb{R}$ est $\sigma(X)$-mesurable si et seulement si il existe une application mesurable $f : E \to \mathbb{R}$ telle que $Z = f(X)$.*

Preuve. — On écrit $Z = Z^+ - Z^-$ puis que $Z^+ = \lim_{n\to+\infty} Z_n$ où Z_n est étagée $\sigma(X)$-mesurable. Z_n est de la forme $Z_n = \sum_k \alpha_k^n \mathbf{1}_{A_k^n}$ où $A_k^n \in \sigma(X)$, donc s'écrivant $A_k^n = \{X \in B_k^n\}$ pour un $B_k^n \in \mathcal{E}$. Puisque $\mathbf{1}_{A_k^n} = \mathbf{1}_{B_k^n}(X)$, on a $Z_n = h_n(X)$ si

$$h_n(x) = \sum_k \alpha_k^n \mathbf{1}_{B_k^n},$$

donc $Z^+ = h_+(X)$ où $h_+ = \limsup h_n$. Pareil avec Z^-. □

On munit $(\Omega, \mathcal{F})$ d'une probabilité $\mathbb{P}$ et on note $\mathbb{E}(X) = \int X d\mathbb{P}$.

Définition 13.2. — Soit $X : \Omega \to E$ une variable aléatoire. On appelle **loi** de X la probabilité $\mu_{(X)}$ sur $(E, \mathcal{E})$ définie par $\mu_{(X)}(A) = \mathbb{P}(X \in A)$, pour tout A de $\mathcal{E}$.

La règle de calcul suivante est d'un emploi constant : il faut absolument savoir la manier. La première égalité est la définition de l'espérance.

Proposition 13.3. — *Soit X une variable aléatoire à valeurs dans E. Pour toute fonction mesurable $f : E \to \mathbb{R}$, à valeurs positives ou telle que $f(X)$ soit intégrable, on a*

$$\mathbb{E}(f(X)) = \int_\Omega f(X)\, d\mathbb{P} = \int_E f\, d\mu_{(X)}.$$

Preuve. — Si $f \geq 0$, on écrit f comme limite croissante de fonctions étagées et applique le théorème de convergence monotone. Sinon on utilise $f = f^+ - f^-$. □

14. Convergence en probabilité

Définition 14.1. — Une suite de v.a. réeeles X_n tend vers X en probabilité si pour tout $\varepsilon > 0$,

$$\mathbb{P}(|X_n - X] \geq \varepsilon) \to 0$$

quand $n \to +\infty$.

Lemme 14.2. — *$X_n \to 0$ en probabilité si et seulement si, pour un $R > 0$ $\min(|X_n|, R) \to 0$ dans L^1*

Preuve. — Si $X_n \to 0$ en probabilité, pour tout $\varepsilon > 0$,

$$\begin{aligned}\mathbb{E}(\min(|X_n|, R)) &= \mathbb{E}(\min(|X_n|, R)\mathbf{1}_{\{|X_n|\geq\varepsilon\}}) + \mathbb{E}(\min(|X_n|, R)\mathbf{1}_{\{|X_n|<\varepsilon\}}) \\ &\leq R\mathbb{P}(|X_n| \geq \varepsilon) + \varepsilon \leq 2\varepsilon\end{aligned}$$

pour n assez grand. Réciproquement, par l'inégalité de Markov, pour $0 \leq \varepsilon \leq R$

$$\mathbb{P}(|X_n| \geq \varepsilon) = \mathbb{P}(\min(R, |X_n|) \geq \varepsilon) \leq \mathbb{E}(\frac{\min(R, |X_n|)}{\varepsilon})$$

qui tend vers 0. □

On déduit donc du Théorème 10.1 que

Proposition 14.3. — *Si $X_n \to 0$ p.s. alors $X_n \to 0$ en probabilité.. Si $X_n \to X$ en probabilité il existe une sous suite (n_k) telle que $X_{n_k} \to X$ presque surement.*

Remarquons que par l'inégalité de Markov,

Proposition 14.4. — *La convergence dans $\mathcal{L}^1$ entraîne la convergence en probabilité.*

Lemme 14.5. — *SI $X_n \to X$ et $Y_n \to Y$ en probabilité, alors $X_n + Y_n \to X + Y$ en probabilité.*

Preuve. — Il suffit de remarquer que

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

□

15. Uniforme intégrabilité

On se place sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

Définition 15.1. — Une famille $\mathcal{L}$ de v.a. est uniformément intégrable si

$$\sup_{X \in \mathcal{L}} \mathbb{E}(|X| \mathbf{1}_{\{|X|>R\}}) \to 0, \text{ quand } R \to +\infty.$$

Lemme 15.2. — *Si $\{X_i, i \in I\}$ et $\{Y_i, i \in I\}$ sont uniformément intégrables, alors $\{X_i + Y_i, i \in I\}$ est uniformément intégrable.*

Preuve. — On a

$$\begin{aligned}\mathbb{E}(|X_i| \mathbf{1}_{\{|X_i+Y_i| \geq M\}}) &= \mathbb{E}(|X_i| \mathbf{1}_{\{|X_i+Y_i| \geq M\}} \mathbf{1}_{\{|X_i| \geq M/2\}}) + \mathbb{E}(|X_i| \mathbf{1}_{\{|X_i+Y_i| \geq M\}} \mathbf{1}_{\{|X_i| < M/2\}}) \\ &\leq \mathbb{E}(|X_i| \mathbf{1}_{\{|X_i| \geq M/2\}}) + \mathbb{E}(|X_i| \mathbf{1}_{\{|X_i+Y_i| \geq M\}} \mathbf{1}_{\{|X_i| < M/2\}})\end{aligned}$$

Le sup du premier terme tend vers 0 car (X_i) est uniformément intégrable. Pour le second on remarque qu'il est majoré par

$$\mathbb{E}(|Y_i| \mathbf{1}_{|Y_i| \geq M/2})$$

qui tend aussi vers 0. □

Le théorème suivant généralise le théorème de convergence dominée.

Théorème 15.3. — *Soit $\{X_n, n \in \mathbb{N}\}$ une suite uniformément intégrable de v.a. qui converge en probabilité (donc a fortiori si elle converge p.s.) vers X. Alors $X_n \to X$ dans $\mathcal{L}^1$ et en particulier, $\mathbb{E}(X_n) \to \mathbb{E}(X)$.*

Preuve. — La suite $Y_n = |X_n - X|$ est aussi uniformément intégrable par le lemme. Ensuite

$$\mathbb{E}(Y_n) = \mathbb{E}(Y_n \mathbf{1}_{\{Y_n < R\}}) + \mathbb{E}(Y_n \mathbf{1}_{\{Y_n \geq R\}})$$

Pour tout $\varepsilon > 0$, pour R assez grand, le second terme est inférieur à $\varepsilon/2$ par unif. int. Un tel R étant choisi, le premier terme est inférieur à $\varepsilon/2$ pour n assez grand, par convergence en probabilité et le lemme 14.2 □

Proposition 15.4. — *Une suite de v.a. X_n telle que $sup_n \mathbb{E}(X_n^2) < +\infty$ est uniformément intégrable.*

Preuve. — On utilise que $\mathbf{1}_{\{|X_n|>R\}} \leq \frac{|X_n|}{R}$ donc

$$\mathbb{E}[|X_n| \mathbf{1}_{\{|X_n|>R\}}] \leq \frac{\mathbb{E}[X_n^2]}{R}.$$

□

16. Processus aléatoires

Un processus aléatoire à valeurs dans E, indexé par T est une variable aléatoire à valeurs dans E^T muni de la tribu produit. Si $T = \mathbb{N}$ on parle de processus aléatoire à temps discret, si $T = \mathbb{R}^+$ de processus aléatoire à temps continu. La loi du processus est par définition la loi de cette variable aléatoire.

Proposition 16.1. — *Pour que $\{X_t, t \in T\}$ soit un processusà valeurs dans E, il suffit que chaque $X_t, t \in T$, soit une variable aléatoire.*

Preuve. — On utilise le lemme 3.2 et le fait que la tribu produit sur $\prod_{t \in T} \mathbb{R}$ est engendrée par les ensembles

$$\{\omega \in \prod_{t \in T} \mathbb{R}; \omega_t \in \mathcal{A}\}$$

où $t \in T$ et A est un borélien de $\mathbb{R}$. □

Par le théorème de la classe monotone la loi du processus est déterminée par les lois finies dimensionnelles

$$\mathbb{P}(X_{t_1} \in A_1, \cdots, X_{t_n} \in A_n), 0 \leq t_1 \leq t_2 \leq \cdots \leq t_n, n \in \mathbb{N}, A_i \in \mathcal{E},$$

En effet la classe des sous ensembles de E^T qui s'écrivent

$$\{\omega \in E^T; \omega(t_1) \in A_1, \cdots, \omega(t_n) \in A_n\}$$

est stable par intersection finie et engendre la tribu produit.

17. Filtration

En intégration le role des tribus est théorique et un peu marginal : on a besoin de la tribu borélienne de $\mathbb{R}$ car elle permet de définir la mesure de Lebesgue.

Par contre en probabilités leur role est essentiel : elles décrivent ce qu'un observateur connait. Pour faire apparaitre la notion de temps, on introduit lorsque $T = \mathbb{N}$ ou $T = \mathbb{R}^+$,

Définition 17.1. — Sur un espace $(\Omega, \mathcal{F})$ une filtration $\{\mathcal{F}_t, t \in T\}$ est une famille de sous tribus de $\mathcal{F}$ vérifiant

$$\mathcal{F}_s \subset \mathcal{F}_t, \text{ si } 0 \leq s \leq t$$

Définition 17.2. — Un processus $X_t, t \in T = \mathbb{N}$ ou $\mathbb{R}^+$, est dit

- adapté à la filtration $(\mathcal{F}_t)$ si X_t est $\mathcal{F}_t$ mesurable pour tout $t \in T$.

- progressif (ou progressivement mesurable) si $(s, \omega) \in [0, t] \times \Omega \mapsto X_s(\omega)$ est $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$ mesurable, pour tout $t \in T$.

***Exemple 17.3* (Important)**. — Etant donné un processus $X_t, t \in T$,

$$\mathcal{F}_t^{(0)} = \sigma(X_s, s \leq t)$$

est une filtration (appelée filtration naturelle du processus), et X est adapté à cette filtration.

Lemme 17.4. — *Un processus adapté continu à droite (ou gauche) est progressif.*

Preuve. — Ecrire, que pour $0 \leq s \leq t$,

$$X_s = \lim_{n \to +\infty} \sum_{k=1}^{n} X_{tk/n} \mathbf{1}_{]tk/n, tk/n]}(s)$$

et remarquer que le terme de droite est progressif. □

18. Indépendance

Des sous tribus $\mathcal{F}_1, \cdots, \mathcal{F}_n$ de $\mathcal{F}$ sont indépendantes si

$$\mathbb{P}(A_1 \cap A_2 \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2) \cdots P(A_n)$$

pour tous $A_1 \in \mathcal{F}_1, \cdots A_n \in \mathcal{F}_n$. Les tribus d'une famille $\mathcal{F}_i, i \in I$, de sous tribu de $\mathcal{F}$ sont indépendantes, si toute sous famille finie l'est. Des événements $A_i, i \in I$, sont indépendants si les tribus $\sigma(A_i) = \{\emptyset, A_i, A_i^c, \Omega\}, i \in I$, le sont. Par exemple on vérifie que A_1, A_2, A_3 sont indépendants si

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$$

pour $1 \leq i < j \leq 3$ (indépendance 2 à 2) et

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

Des v.a. $X_i, i \in I$ sont indépendantes si les tribus $\sigma(X_i), i \in I$, le sont. Par exemple, deux v.a. X, Y sont indépendantes si et seulement si

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$$

pour tout f, g mesurables positives. On a donc

Lemme 18.1. — *Deux v.a. X et Y sont indépendantes si et seulement si les lois vérifient*

$$\mu_{(X,Y)} = \mu_{(X)} \otimes \mu_{(Y)}.$$

La proposition suivante en résulte par Funini. On y voit le lien entre variables aléatoires indépendantes (X et Y) et variables indépendantes (ω et ω').

Proposition 18.2. — *Si X, Y sont deux variables aléatoires indépendantes, X à valeurs dans $(E_1, \mathcal{E}_1)$, et Y à valeurs dans $(E_2, \mathcal{E}_2)$, pour tout $F : E_1 \times E_2 \to \mathbb{R}$ mesurable positive,*

$$\mathbb{E}(F(X,Y)) = \int_\Omega \int_\Omega F(X(\omega), Y(\omega'))\, d\mathbb{P}(\omega)\, d\mathbb{P}(\omega').$$

Le résultat suivant, non trivial, est important.

***Lemme 18.3* (Lemme de regroupement)**. — *Si les tribus $\mathcal{F}_i, i \in I$, sont indépendantes et si $I_1, I_2, \cdots, I_n$ sont des parties disjointes de I, alors les tribus*

$$\mathcal{G}_k = \sigma(\mathcal{F}_i, i \in I_k), k = 1, \cdots, n,$$

sont indépendantes.

Preuve. — Il suffit de faire la preuve pour $n = 2$ (pourquoi ?). Posons

$$\mathcal{C}_1 = \{A_1 \cap A_2 \cap \cdots \cap A_r; r \in \mathbb{N}, A_i \in \mathcal{F}_i, i \in I_1\}$$

$$\mathcal{C}_2 = \{B_1 \cap B_2 \cap \cdots \cap B_r; r \in \mathbb{N}, B_i \in \mathcal{F}_i, i \in I_2\}$$

Les classes $\mathcal{C}_1$ et $\mathcal{C}_2$ sont stables par intersection finies, $\sigma(\mathcal{C}_1) = \mathcal{G}_1$ et $\sigma(\mathcal{C}_2) = \mathcal{G}_2$. Pour $C \in \mathcal{C}_1$ fixé, considérons

$$\mathcal{M}_C = \{A \in \mathcal{G}_2; \mathbb{P}(C \cap A) = \mathbb{P}(C)\mathbb{P}(A)\}$$

C'est une classe monotone contenant $\mathcal{C}_2$ donc $\mathcal{G}_2$, par le théorème de la classe monotone. On considère alors, pour $A \in \mathcal{G}_2$ fixé,

$$\mathcal{M}'_A = \{B \in \mathcal{G}_1; \mathbb{P}(B \cap A) = \mathbb{P}(B)\mathbb{P}(A)\}$$

c'est une classe monotone contenant $\sigma(\mathcal{C}_1)$ par ce qui précède, donc $\mathcal{G}_1$. □

On déduit de ce lemme que si les v.a. $X_1, \cdots, X_n$ sont indépendantes et $n_1 < n_2 < n_k$, alors les v.a.

$$Z_r = (X_{n_r}, X_{n_r+1}, \cdots, X_{n_{r+1}-1})$$

sont indépendantes. Par exemple si X_1, X_2, X_3 sont indépendantes, alors $X_1 + X_2$ est indépendant de $\cos(X_3)$.

PARTIE IV
MESURE DE LEBESGUE ET JEU DE PILE OU FACE

19. Mesure de Lebesgue sur $\mathbb{R}$

19.1. Jeu de pile ou face. — La mesure de Lebesgue sur $\mathbb{R}$ est l'unique mesure m vérifiant

$$m([a, b]) = b - a$$

L'unicité résulte du théorème de la classe monotone. L'existence par contre est difficile à montrer. Elle est équivalente à l'existence du jeu de pile ou face (que nous admettons).

Théorème 19.1. — *Soit $X_n, n \in \mathbb{N}$, un jeu de pile ou face équitable, c'est à dire une suite de v.a. indépendantes telles que*

$$\mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 1) = 1/2$$

pour tout $n \geq 0$. Alors la loi de $Z = \sum_{n=1}^{+\infty} X_n/2^n$ est la mesure de Lebesgue sur $[0, 1]$.

Preuve. — Si m est un entier, p.s.,

$$0 \leq \sum_{n=m+1}^{+\infty} X_n/2^n < 1/2^m$$

donc, pour $k_1, \cdots, k_m \in \{0, 1\}$ fixés,

$$\sum_{r=1}^{m} k_r/2^m \leq Z < (\sum_{r=1}^{m} k_r/2^m) + 1/2^m$$

si et seulement si $X_1 = k_1, \cdots, X_m = k_m$. Cet ensemble est de probabilité $1/2^m$. Donc la loi de Z donne la mesure de Lebesgue aux intervalles dyadiques. Or la mesure des intervalles dyadiques détermine la loi. $\square$

19.2. Simulation. — Partant du principe qu'un ordinateur ne peut simuler que la mesure de Lebesgue, comment simuler les autres mesures.

Par commodité d'écriture on va travailler sur $\mathbb{R}^+$ au lieu de $\mathbb{R}$. On ne perd pas en généralité car l'exponentielle établit une bijection bicontinue entre $\mathbb{R}$ et $\mathbb{R}^+\backslash\{0\}$.

Définition 19.2. — On appelle fonction de répartition d'une mesure de Radon m sur $\mathbb{R}^+$ la fonction

$$F(t) = m([0,t]), t \geq 0.$$

Une fonction de répartiton est croissante continue à droite car si $t_n \to t$ en décroissant, $\mathbf{1}_{[0,t_n]} \to \mathbf{1}_{[0,t]}$. Par contre elle peut ne pas être continue à gauche car si $s_n \to s$ en croissant $\mathbf{1}_{[0,s_n]} \to \mathbf{1}_{[0,s[}$ donc $F(s_n) \to m([0,s[)$ et

$$F(s) - F(s^-) = m(\{s\})$$

Il résulte du théorème de la classe monotone que la fonction de répartition F caractérise la mesure de Radon m.

Proposition 19.3. — *Soit $F : \mathbb{R}^+ \to \bar{\mathbb{R}}^+$ croissante et continue à droite. Alors l'application*

$$t \mapsto G(t) = \inf\{s \geq 0; F(s) > t\}$$

de $\mathbb{R}^+$ dans $\mathbb{R}^+$ est croissante, continue à droite. De plus

$$[0, G(t)[\subset \{F \leq t\} \subset [0, G(t)]$$

Preuve. — Il est clair que G est croissante. Si elle n'est pas continue à droite au point $r \geq 0$, il existe x tel que

$$G(r) < x < G(r+h)$$

pour tout $h > 0$. Puisque $F(r) < x$ il existe $s < x$ tel que $G(s) > r$ donc $G(x) > r$. Puisque $x < G(r+h)$,

$$\inf\{s \geq 0, F(s) > r+h\} > x$$

donc $G(x) \leq r+h$, pour tout $h > 0$. Une contradiction.

Pour les inclusions, si $s < G(t)$ clairement $F(s) \leq t$ et si $F(s) \leq t$, $G(t) \geq s$. □

Remarque : F est à son tour l'inverse continu à droite de G. Le théorème suivant a un intérêt théorique clair et peut servir pour simuler les v.a. réelles.

Théorème 19.4. — *Soit m un mesure de Radon sur $\mathbb{R}^+$ de fonction de répartition $F(t) = m([0,t])$ d'inverse continu à droite G. Alors m est l'image de la mesure de Lebesgue sur $\mathbb{R}^+$ par G.*

Preuve. — On a, par la proposition précédente,

$$G(t) = \text{Lebesgue}([0, G(t)]) = \text{Lebesgue}(F \leq t)$$

□

20. Extension en dimension infinie

Le théorème suivant montre l'universalité de la mesure de Lebesgue. On aurait pu utiliser aussi le jeu de pile ou face car puisque $\mathbb{N}$ est en bijection avec $\mathbb{N}^2$, on fabrique une infinité de jeu de pile ou face à partir d'un seul.

Théorème 20.1. — *Soit μ une probabilité sur l'espace produit $\mathbb{R}^{\mathbb{N}}$. Il existe une application mesurable ϕ de $[0,1]$ dans $\mathbb{R}^{\mathbb{N}}$ telle que μ soit l'image de la mesure de Lebesgue sur $[0,1]$ par ϕ.*

Preuve. — Puisque $\mathbb{R}$ est en bijection bimesurable avec $]0,1[$ (utiliser arctan par exemple) on se ramène au cas où μ est une probabilité sur l'espace produit $[0,1]^{\mathbb{N}}$. Tout réel t de $[0,1]$ s'écrit de façon unique sous forme dyadique

$$t = 0, t_1 t_2 t_3 \cdots$$

(qui signifie que $t = \sum_{k=1}^{+\infty} t_k/2^k$) où $t_k \in \{0,1\}$ et où t_k n'est pas toujours égal à 1 à partir d'un certain rang. Pour

$$x = (x^{(0)}, x^{(1)}, x^{(1)}, \cdots) \in [0,1]^{\mathbb{N}}$$

écrivons chaque $x^{(n)} = \sum_{k=1}^{+\infty} t_k^{(n)}/2^k$ comme au dessus. On considère l'application

$$\Psi : [0,1]^{\mathbb{N}} \to [0,1]$$

définie par, en écriture dyadique,

$$\Psi(x) = 0, \underbrace{t_1^{(0)}}\, \underbrace{t_2^{(0)} t_1^{(1)}}\, \underbrace{t_3^{(0)} t_2^{(1)} t_1^{(2)}}\, \underbrace{t_4^{(0)} t_3^{(1)} t_2^{(2)} t_1^{(3)}} \cdots$$

Cette application est mesurable injective et donc une bijection de $[0,1]^{\mathbb{N}}$ sur son image $\Psi([0,1]^{\mathbb{N}})$, dont on vérifie que c'est un borélien. On note Ψ^{-1} son inverse. L'image $\Psi(\mu)$ de μ par Ψ est une probabilité sur $[0,1]$ donc image de la mesure de Lebesgue $m_{[0,1]}$ sur $[0,1]$ par l'inverse G de sa fonction de répartition F.

$$\underset{m_{[0,1]}}{[0,1]} \xrightarrow{G} \underset{\Psi(\mu)}{\Psi([0,1]^{\mathbb{N}})} \subset [0,1] \xrightarrow{\Psi^{-1}} \underset{\mu}{[0,1]^{\mathbb{N}}}$$

Alors μ est l'image de $m_{[0,1]}$ par $\Psi^{-1} \circ G$. □

Conséquence : On pourrait prendre, dans toutes nos applications ultérieures, $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \text{Boréliens}, \text{Lebesgue})$.

21. Mesure de Lebesgue sur $\mathbb{R}^d$

***Théorème 21.1* (Théorème de changement de variables)**
Soient O_1 et O_2 deux ouverts de $\mathbb{R}^d$ et

$$\varphi : O_1 \to O_2$$

une application bijective de classe C^1 et d'inverse C^1. Pour toute fonction mesurable positive $f : \mathbb{R}^d \to \mathbb{R}$,

$$\int_{O_1} f(\varphi(x))\, dm(x) = \int_{O_2} f(x)|J\varphi^{-1}(x)|\, dm(x)$$

où m est la mesure de Lebesgue et $J\varphi(x) = \det \frac{\partial \varphi_i}{\partial x_j}(x)$.

Nous l'admettons. Lorsque φ est linéaire ceci se montre en utilisant que m est la seule mesure invariante par translation à une constante près et que toute matrice M s'écrit $M = K_1 D K_2$ où K_1 et K_2 sont orthogonales et D diagonale.

22. Convergence en loi

Définition 22.1. — Une suite de probabilités $\mu_n, n \in \mathbb{N}$, sur un espace topologique E converge en loi vers la probabilité μ, si pour toute fonction continue bornée $f : E \to \mathbb{R}$

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

On dit que des v.a. convergent en loi si leur loi converge (en loi).

22.1. Transformée de Fourier. —

Définition 22.2. — On appelle fonction caractéristique d'une v.a. aléatoire X à valeurs dans $\mathbb{R}^d$ ou transformée de Fourier de sa loi μ la fonction $\hat{\mu} : \mathbb{R}^d \to \mathbb{C}$ définie par

$$\hat{\mu}(\lambda) = \mathbb{E}(e^{i\langle \lambda, X\rangle}) = \int e^{i\langle \lambda, x\rangle}\, d\mu(x),\ \ \lambda \in \mathbb{R}^d,$$

où $\langle \lambda, X\rangle = \sum_{n=1}^d \lambda_n X_n$.

L'exemple essentiel est celui de la loi gaussienne $N(0,1)$. Il s'agit de la loi de densité

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

par rapport à la mesure de Lebesgue sur $\mathbb{R}$ (On vérifie que c'est une densité en faisant le changement de variables en coordonnées polaires

$$\int e^{-x^2/2} dx \int e^{-y^2/2} dy = \int_0^{+\infty} \int_0^{2\pi} e^{-r^2/2} r\, dr d\theta)$$

Théorème 22.3. — *La transformée de Fourier de la loi $\mu = N(0,1)$ est donnée par*
$$\hat{\mu}(\lambda) = \int e^{i\lambda x}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx = e^{-\lambda^2/2}, \lambda \in \mathbb{R}.$$

Preuve. — Par parité
$$\hat{\mu}(\lambda) = \frac{1}{\sqrt{2\pi}}\int e^{i\lambda x}e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}}\int \cos(\lambda x)e^{-x^2/2}\,dx.$$
Avec le théorème de dérivation sous le signe somme que l'on peut appliquer car
$$|x\sin(\lambda x)e^{-x^2/2}| \leq |x|e^{-x^2}2 \in L^1$$
on voit que
$$\hat{\mu}(\lambda)' = -\frac{1}{\sqrt{2\pi}}\int x\sin(\lambda x)e^{-x^2/2}\,dx.$$
On fait une intégration par parties $u = \sin\lambda x, v = e^{-x^2/2}$,
$$hat\mu'(\lambda) = \lambda\int \cos(\lambda x)e^{-x^2/2}\,dx$$
donc $\hat{\mu}(\lambda)' = \lambda\hat{\mu}(\lambda)$ ce qui entraîne que $\hat{\mu}(\lambda) = \hat{\mu}(0)e^{-\lambda^2/2}$. □

Remarque : Cette formule est aussi valable pour λ complexe. En particulier, pour t réel on a la transformée dite de Laplace (et c'est facile à voir directement), si $X \sim N(0,1)$,
$$\mathbb{E}(e^{tX}) = e^{t^2/2}.$$

Théorème 22.4. — *Deux v.a. ayant la même transformée de Fourier ont la même loi.*

Preuve. — Ecrivons la preuve quand $d = 1$, mais c'est la même dans le cas général. Considérons une v.a. X de loi μ et de fonction caractéristique $\hat{\mu}$ et une v.a. Z indépendante de X de loi $N(0,1)$. Pour toute fonction continue $f : \mathbb{R} \to \mathbb{R}$ à support compact et $\sigma > 0$
$$\mathbb{E}(f(X+\sigma Z)) = \frac{1}{\sqrt{2\pi}}\int f(x+\sigma y)e^{-y^2/2}\,dy\,d\mu(x)$$
$$= \frac{1}{\sqrt{2\pi}}\int f(x+\sigma y)(\int e^{iyt}\frac{e^{-t^2/2}}{\sqrt{2\pi}}\,dt)\,dy\,d\mu(x)$$
On fait le changement de variables (à x, t fixés) $u = x + \sigma y$. On obtient
$$\mathbb{E}(f(X+\sigma Z)) = \frac{1}{2\pi}\int f(u)\int e^{i(\frac{u-x}{\sigma})t}e^{-t^2/2}\frac{du}{\sigma}dtd\mu(x)$$
$$= \frac{1}{2\pi\sigma}\int \hat{\mu}(-\frac{t}{\sigma})e^{\frac{iut}{\sigma}-t^2/2}f(u)\,du\,dt$$

Ceci ne dépend que de $\hat{\mu}$. Il en est de même de

$$\mathbb{E}(f(X)) = \lim_{\sigma \to 0} \mathbb{E}(f(X + \sigma Y))$$

donc $\hat{\mu}$ détermine μ. □

Lemme 22.5. — *Une fonction continue $f : \mathbb{R} \to \mathbb{R}$ à support compact est uniformément continue, au sens où, pour tout $\varepsilon > 0$ il existe $\eta \geq 0$ tel que si $|x - y| \leq \eta$, $|f(x) - f(y)| \leq \varepsilon$.*

Preuve. — Par l'absurde. □

Théorème 22.6. — *Une suite de probabilités $\mu_n, n \in \mathbb{N}$, sur $\mathbb{R}^d$ converge en loi vers une probabilité μ si et seulement si, pour tout $\lambda \in \mathbb{R}^d$, quand $n \to +\infty$,*

$$\hat{\mu}_n(\lambda) \to \hat{\mu}(\lambda).$$

Preuve. — (écrit pour $d = 1$) Si $\mu_n \to \mu$ en loi, par définition, pour toute $f : \mathbb{R} \to \mathbb{R}$ continue bornée, $\int f\, d\mu_n \to \int f\, d\mu$. Prenant $f(t) = \sin \lambda t$ et $f(t) = \cos \lambda t$ on obtient que $\hat{\mu}_n(\lambda) \to \hat{\mu}(\lambda)$.

Réciproquement, si X_n est de loi μ_n et Z est indépendante de (X_n) de loi $N(0, 1)$ on a (cf preuve précédente), si f est continue à support compact,

$$\lim_{n \to +\infty} \mathbb{E}(f(X_n + \sigma Z)) = \lim_{n \to +\infty} \frac{1}{2\pi\sigma} \int \hat{\mu}_n(-\frac{t}{\sigma}) e^{\frac{iut}{\sigma} - t^2/2} f(u)\, du\, dt$$

$$= \frac{1}{2\pi\sigma} \int \hat{\mu}(-\frac{t}{\sigma}) e^{\frac{iut}{\sigma} - t^2/2} f(u)\, du\, dt = \mathbb{E}(f(X + \sigma Z)).$$

Appliquons le lemme à f, pour $\varepsilon > 0$ il existe $\eta > 0$ tel que, si $\|f\|_\infty = Sup_{x \in \mathbb{R}} |f(x)|$,

$$|\mathbb{E}(f(X_n + \sigma Z)) - \mathbb{E}(f(X_n))| \leq \mathbb{E}|f(X_n + \sigma Z) - f(X_n)|$$

$$\leq \mathbb{E}(\varepsilon \mathbf{1}_{\{|\sigma Z| \leq \eta\}} + 2\|f\|_\infty \mathbf{1}_{\{\sigma Z| \geq \eta\}})$$

$$\leq \varepsilon + 2\|f\|_\infty \mathbb{P}(|\sigma Z| \geq \eta) \leq 2\varepsilon$$

pour σ assez petit. La même majoration s'applique à X au lieu de X_n donc,

$$|\mathbb{E}(f(X_n) - \mathbb{E}(f(X))| \leq |\mathbb{E}(f(X_n) - \mathbb{E}(f(X_n + \sigma Z))|$$

$$+|\mathbb{E}(f(X_n + \sigma Z) - \mathbb{E}(f(X + \sigma Z))| + |\mathbb{E}(f(X + \sigma Z)) - f(X))|$$

est $< 6\varepsilon$ pour n assez grand. Donc $\mathbb{E}(f(X_n))$ tend vers $\mathbb{E}(f(X))$. Reste à passer à f continue bornée... □

PARTIE V
ESPÉRANCE CONDITIONNELLE

23. Définition comme projection généralisée

Dans un espace $(\Omega, \mathcal{F}, \mathbb{P})$ si $\mathcal{G}$ est une sous tribu de $\mathcal{F}$, l'espace $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ est un sous espace vectoriel fermé de $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ (a proprement parlé il faut travailler dans L^2, d'où les p.s. dans ce qui suit). L'espérance conditionnelle de $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ sachant $\mathcal{G}$ est par définition la projection de X sur $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$. C'est donc la seule (classe de) v.a. Z, telle que

— Z est $\mathcal{G}$ mesurable.

— $\mathbb{E}(Z\mathbf{1}_A) = \mathbb{E}(X\mathbf{1}_A)$ pour tout ensemble $A \in \mathcal{G}$.

On note $Z = \mathbb{E}(X|\mathcal{G})$. Si maintenant X est seulement une v.a. positive, $X \wedge n$ est dans $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ et la suite $Z_n = \mathbb{E}(X \wedge n|\mathcal{G})$ est croissante. On pose $\mathbb{E}(X|\mathcal{G}) = \lim_{n\to+\infty} \mathbb{E}(X \wedge n|\mathcal{G})$. On voit immédiatement qu'on a

Théorème 23.1 **(Caractérisation de l'espérance conditionnelle)**
Soit X une v.a. positive alors la (classe de) $Z = \mathbb{E}(X|\mathcal{G})$ est caractérisée par les deux propreétés suivantes

— *Z est $\mathcal{F}$ mesurable.*

— *$\mathbb{E}(Z\mathbf{1}_A) = \mathbb{E}(X\mathbf{1}_A)$ pour tout ensemble $A \in \mathcal{G}$*

On a si $X, Y \geq 0$, et Y est $\mathcal{G}$-mesurable

$$\mathbb{E}(XY) = \mathbb{E}(Y\mathbb{E}(X|\mathcal{G})).$$

Soit X une v.a. telle que $\mathbb{E}(|X||\mathcal{G}) < +\infty, p.s.$. Alors $Z = \mathbb{E}(X^+|\mathcal{G}) - \mathbb{E}(X^-|\mathcal{G})$ est notée aussi $Z = \mathbb{E}(X|\mathcal{G})$. Remarquons que l'espérance conditionnelle n'est définie que p.s. Toute égalité où elle intervient n'a de sens que p.s., de ce fait on oubliera souvent le terme p.s. Les inégalités suivantes montrent que l'application $T : \mathcal{L}^p \to \mathcal{L}^p$ donnée par $T(X) = \mathbb{E}(X|\mathcal{F})$ est continue pour $p = 1, 2$.

Proposition 23.2 **(Continuité de l'espérance conditionelle)**

$$\|\mathbb{E}(X|\mathcal{F})\|_{L^1} \leq \|X\|_{L^1}; \quad \|\mathbb{E}(X|\mathcal{F})\|_{L^2} \leq \|X\|_{L^2}.$$

Les propriétés suivantes sont faciles à démontrer et essentielles :

Proposition 23.3. — *Si les expressions ont un sens, par exemple pour les v.a. positives, p.s. :*

1. $\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X)$;
2. $|\mathbb{E}(X|\mathcal{G})| \leq \mathbb{E}(|X||\mathcal{G})$;
3. *Si Y est $\mathcal{G}$-mesurable,* $\mathbb{E}(XY|\mathcal{G}) = Y\mathbb{E}(X|\mathcal{G})$.

4. *Si* $\mathcal{G}'$ *est une sous-tribu de* $\mathcal{G}$, $\mathbb{E}(X|\mathcal{G}') = \mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{G}')$;
5. $\mathbb{E}(1|\mathcal{G}) = 1$;
6. $\mathbb{E}(aX + bY|\mathcal{G}) = a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G})$;
7. *Si* $X \leq Y$, $\mathbb{E}(X|\mathcal{G}) \leq \mathbb{E}(Y|\mathcal{G})$;
8. $|\mathbb{E}(XY|\mathcal{G})|^2 \leq \mathbb{E}(X^2|\mathcal{G})\mathbb{E}(Y^2|\mathcal{G})$.

On a l'analogue des théorèmes de convergence de Lebesgue et de Fatou,

Proposition 23.4. — *Soit* (X_n) *une suite de v.a.*

(Thm de convergence monotone conditionnel) Si $0 \leq X_n \uparrow X$, $\mathbb{E}(X_n|\mathcal{G}) \to \mathbb{E}(X|\mathcal{G})$, *p.s.*

(Lemme de Fatou conditionnel) Si $0 \leq X_n$, $\mathbb{E}(\liminf X_n|\mathcal{G}) \leq \liminf \mathbb{E}(X_n|\mathcal{G})$,

(Thm de convergence dominée conditionnel) Si $|X_n| \leq V$ *où* $V \in L^1$ *et si* $X_n \to X$, $\mathbb{E}(X_n|\mathcal{G}) \to \mathbb{E}(X|\mathcal{G})$, *p.s.*

Preuve. — On commence par montrer le théorème de convergence monotone : si $0 \leq X_n \uparrow X$ la suite $\mathbb{E}(X_n|\mathcal{G})$ est croissante donc a une limite, que nous notons Z. Comme limite de fonctions $\mathcal{G}$-mesurables, Z est $\mathcal{G}$-mesurable. Et pour toute v.a. U positive et $\mathcal{G}$-mesurable, par le théorème de convergence monotone usuel appliqué deux fois,

$$\mathbb{E}(UZ) = \lim_{n\to+\infty} \mathbb{E}(U\mathbb{E}(X_n|\mathcal{G})) = \lim_{n\to+\infty} \mathbb{E}(UX_n) = \mathbb{E}(UX)$$

Donc $Z = \mathbb{E}(X|\mathcal{G})$ par la propriété caractéristique de l'espérance conditionnelle (Thm 23.1). Les autres énoncés s'en déduisent comme dans le cas non conditionnel. □

Très souvent les calculs explicites d'espérance conditionnelle utilisent le théorème suivant :

***Proposition 23.5* (Essentiel)**. — *Si* $X : \Omega \to (E_1, \mathcal{E}_1)$ *est* $\mathcal{G}$*-mesurable et* $Y : \Omega \to (E_2, \mathcal{E}_2)$ *indépendant de* $\mathcal{G}$ *et si* $\phi : E_1 \times E_2 \to \mathbb{R}^+$ *est mesurable, alors, p.s.,*

$$\mathbb{E}(\phi(X,Y)|\mathcal{G})(\omega) = \int \phi(X(\omega), Y(\omega'))\, d\mathbb{P}(\omega')$$

ce que l'on peut écrire aussi $\mathbb{E}(\phi(x,Y))$ *pour* $x = X(\omega)$.

Preuve. — Par Fubini, on sait que l'application $x \mapsto \mathbb{E}(\phi(x,Y))$ est mesurable. Donc $Z(\omega) = \mathbb{E}(\phi(X(\omega),Y))$ est $\mathcal{G}$-mesurable, puisque c'est une fonction mesurable de X, lui même $\mathcal{G}$-mesurable. Si $A \in \mathcal{G}$, les deux v.a. $(\mathbf{1}_A, X)$ et Y sont indépendantes. Donc, en appliquant la proposition 18.2,

$$\mathbb{E}(\phi(X,Y)\mathbf{1}_A)) = \int_\Omega \int_\Omega \phi(X(\omega), Y(\omega')\mathbf{1}_A(\omega)\, d\mathbb{P}(\omega)\, d\mathbb{P}(\omega')$$

$$= \int_\Omega [\int_\Omega \phi(X(\omega), Y(\omega')\mathbf{1}_A(\omega)\, d\mathbb{P}(\omega')]\, d\mathbb{P}(\omega) = \mathbb{E}(Z\mathbf{1}_A)$$

par Fubini. On a donc bien que $Z = \mathbb{E}(\phi(X,Y)|\mathcal{G})$ par la propriété caractéristique rappelée au dessus. □

24. Un théorème de Bayes

On se donne un espace $(\Omega, \mathcal{F})$ muni de deux probabilités $\mathbb{P}$ et $\mathbb{Q}$ et $\mathcal{G}$ une sous tribu de $\mathcal{F}$.

Proposition 24.1. — *On suppose que $d\mathbb{Q} = Z\, d\mathbb{P}$. Alors, en notant $\mathbb{E}$ l'espérance par rapport à $\mathbb{P}$ et $\mathbb{E}_\mathbb{Q}$ l'espérance par rapport à $\mathbb{Q}$. Pour toute v.a. $Y \geq 0$,*

$$\mathbb{E}_\mathbb{Q}(Y|\mathcal{G}) = \frac{\mathbb{E}(YZ|\mathcal{G})}{\mathbb{E}(Z|\mathcal{G})}.$$

Preuve. — La fonction $X = \frac{\mathbb{E}(YZ|\mathcal{G})}{\mathbb{E}(Z|\mathcal{G})}$ est $\mathcal{G}$-mesurable et, pour toute v.a. $\mathcal{G}$-mesurable bornée U

$$\mathbb{E}_\mathbb{Q}(XU) = \mathbb{E}(XUZ) = \mathbb{E}(XU\mathbb{E}(Z|\mathcal{G}))$$
$$= \mathbb{E}(\mathbb{E}(YZ|\mathcal{G})U) = \mathbb{E}(YZU) = \mathbb{E}_\mathbb{Q}(YU)$$

□

Remarque : $\mathbb{E}(Z|\mathcal{G}) > 0, \mathbb{Q}$-p.s. car si $A = \{\mathbb{E}(Z|\mathcal{G}) = 0\}$,

$$\mathbb{Q}(A) = \mathbb{E}(\mathbf{1}_A Z) = \mathbb{E}(\mathbf{1}_A \mathbb{E}(Z|\mathcal{G})) = 0.$$

PARTIE VI
COMPLÉMENTS "CULTURELS"

Cette partie ne fait plus partie des "choses à savoir". On y établit d'abord l'existence de lois conditionnelles. Puis on démontre les deux théorèmes limites fondamentaux des probabilités : la loi des grands nombres et le théorème central limite, mais dans un cadre plus général qu'habituellement : le théorème ergodique de Birkhoff (avec une démonstration due à Neveu) et le TCL pour martingales avec une preuve due à McLeish adaptée par Sethuraman).

25. Version régulière de l'espérance conditionnelle

Définition 25.1. — Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (E, \mathcal{E})$ une v.a. et $\mathcal{G}$ une sous tribu de $\mathcal{F}$. On appelle loi conditionnelle de X sachant $\mathcal{G}$ une famille $\{P(\omega, \cdot), \omega \in \Omega\}$ de probabilités sur $(E, \mathcal{E})$ telle que, pour tout $A \in \mathcal{E}$, p.s.,

$$\mathbb{E}(\mathbf{1}_A(X)|\mathcal{G})(\omega) = P(\omega, A).$$

Théorème 25.2. — *Si $E = \mathbb{R}^{\mathbb{N}}$ et $\mathcal{E}$ est la tribu produit, toute v.a. X à valeurs dans E possède une loi conditionnelle.*

Preuve. — On traite d'abord le cas où $E = \mathbb{R}$. Soit $\Omega_0 \in \mathcal{F}$ tel que $\mathbb{P}(\Omega_0) = 1$ et tel que pour tout $\omega \in \Omega_0$, les conditions suivantes (qui sont en nombre dénombrable) soient réalisées, pour tous t, s rationnels tels que $t \leq s$,
- — $F_t(\omega) = \mathbb{E}(\mathbf{1}_{\{X \leq t\}}|\mathcal{G})(\omega)$ est bien défini ;
- — $F_t(\omega) \leq F_s(\omega)$;
- — $F_t(\omega) = \lim_{n \to +\infty} F_{t+1/n}(\omega)$.

On pose, pour tout $t \in \mathbb{R}$ $F_t(\omega) = \inf_{1 \geq s \geq t, s \in \mathbb{Q}} F_s(\omega)$. Cette fonction est croissante continue à droite avec $F_{-\infty} = 0, F_{+\infty} = 1$ donc il lui correspond une probabilité que l'on note $P(\omega, \cdot)$, pour laquelle, pour tout $t \in \mathbb{R}$

$$\mathbb{E}(\mathbf{1}_{\{X \leq t\}}|\mathcal{G})(\omega) = P(\omega,]-\infty, t]).$$

Vérifions que $P(\omega, A) = \mathbb{E}(\mathbf{1}_A(X)|\mathcal{G})(\omega)$. Si U est une v.a. $\mathcal{G}$-mesurable positive, posons

$$\mathcal{M} = \{A \in \mathcal{B}(\mathbb{R}); \omega \mapsto P(\omega, A) \text{ est } \mathcal{G} \text{ mesurable et } \mathbb{E}(1_A U) = \mathbb{E}(P(., A)U).\}$$

$\mathcal{M}$ est une classe monotone contenant les ensembles $]-\infty, t], t \in \mathbb{Q}$, classe stable par intersection finie engendrant la tribu. Donc par le théorème de la classe monotone, $\mathcal{M}$ est la tribu borélienne.

Considérons maintenant le cas où $E = \mathbb{R}^{\mathbb{N}}$. On a vu qu'il existe une bijection bimesurable ϕ de $\mathbb{R}^{\mathbb{N}}$ sur un borélien B de $[0, 1]$. On écrit que

$$\mathbf{1}_A(X) = \mathbf{1}_{\phi(A)}(\phi(X))$$

et on est ramené au cas précédent. □

Applications : Hölder, théorèmes limites, probabilité de transition pour les processus de Markov,etc...

26. Théorème ergodique et loi des grands nombres

26.1. Théorème ergodique. — On suppose que sur $(\Omega, \mathcal{F}, \mathbb{P})$ il existe une application mesurable $T : \Omega \to \Omega$ qui préserve $\mathbb{P}$, c'est à dire que

$$\mathbb{P}(T \in A) = \mathbb{P}(A), \text{ pour tout } A \in \mathcal{F}.$$

Définition 26.1. — La tribu des invariants est

$$\mathcal{I} = \{A \in \Omega; T^{-1}(A) = A, p.s.\}.$$

Une v.a. X est $\mathcal{I}$-mesurable si et seulement si $X = X \circ T$.

Lemme 26.2. — . *Si Z est intégrable et $\mathbb{E}(Z|\mathcal{I}) < 0$ p.s. alors $Z^* = \sup(0, Z, Z + Z \circ T, \cdots, \sum_{k=0}^{n-1} Z \circ T^k, \cdots)$ est fini p.s.*

Preuve. — $Z_n^* = \sup(0, Z, Z + Z \circ T, \cdots, \sum_{k=0}^{n-1} Z \circ T^k)$ est croissante, tend vers Z^* et vérifie

$$Z_{n+1}^* = \sup(0, Z + Z_n^* \circ T).$$

Pour toute v.a. $X \geq 0$, $\mathbb{E}(X \circ T|\mathcal{I}) = \mathbb{E}(X|\mathcal{I})$ car pour toute v.a. U, $\mathcal{I}$-mesurable positive,

$$\mathbb{E}(\mathbb{E}(X|\mathcal{I}))U) = \mathbb{E}(XU) = \mathbb{E}((X \circ T)U).$$

On a donc

$$0 \leq \mathbb{E}(Z_{n+1}^*|\mathcal{I}) - \mathbb{E}(Z_n^*|\mathcal{I}) \leq \mathbb{E}((Z_{n+1}^* - Z_n^* \circ T)|\mathcal{I}) \leq \mathbb{E}((\sup(-Z_n^* \circ T, Z)|\mathcal{I})).$$

Remarquons que puisque $Z^* \geq 0$,

$$|\sup(-Z^* \circ T, Z)| \leq |Z|.$$

On peut donc appliquer le théorème de convergence dominée pour obtenir que

$$0 \leq \mathbb{E}(\sup(-Z^* \circ T, Z)|\mathcal{I})).$$

L'ensemble $A = \{Z^* = +\infty\}$ est invariant, puisque $Z^* = \sup(0, Z + Z^* \circ T)$ et

$$0 \leq \mathbf{1}_A \mathbb{E}(\sup(-Z^* \circ T, Z)|\mathcal{I}) = \mathbf{1}_A \mathbb{E}(Z|\mathcal{I}),$$

donc $\mathbb{P}(A) = 0$ car on a supposé que $\mathbb{E}(Z|\mathcal{I}) < 0$. □

Théorème 26.3 **(Théorème ergodique de Birkhoff)**
Pour toute v.a. Z intégrable, p.s.,

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{k=0}^{n-1} Z \circ T^k = \mathbb{E}(Z|\mathcal{I}).$$

Preuve. — On applique le lemme à $Z - \mathbb{E}^{\mathcal{I}}(Z) - \varepsilon$, où $\varepsilon > 0$. On en déduit que $\sup_n \sum_{k=0}^{n-1}[Z \circ T^k - \mathbb{E}^{\mathcal{I}}(Z) - \varepsilon] < +\infty$ donc, p.s.

$$\limsup_{n \to +\infty} \frac{1}{n} \sum_{k=0}^{n-1} Z \circ T^k \leq \mathbb{E}(Z|\mathcal{I}) + \varepsilon.$$

On en déduit l'énoncé en faisant tendre ε vers 0 puis en appliquant ceci aussi à $-Z$. □

26.2. Loi des grands nombres. — Soit μ une probabilité sur $\mathbb{R}$. On choisit $\Omega = \mathbb{R}^{\mathbb{N}}$, muni de la tribu produit et de la probabilité produit $\mathbb{P} = \mu^{\otimes\mathbb{N}}$. Les variables aléatoires $Z_n(\omega) = \omega_n$ sont indépendantes et de loi μ. La transformation de décalage ("shift" en anglais) est définie par, si $\omega = (\omega_n, n \geq 0)$,

$$T(\omega)_n = \omega_{n+1}.$$

Elle conserve la probabilité $\mathbb{P}$. On a

$$Z_n = Z_0 \circ T^n$$

Il résulte donc du théorème de Birkhoff que, si Z_0 est intégrable, p.s.,

$$\frac{Z_0 + \cdots + Z_{n-1}}{n} \to \mathbb{E}(Z_0|\mathcal{I}).$$

Pour terminer remarquons que, pour tout $k \geq 0$

$$\limsup_{n\to+\infty} \frac{Z_0 + \cdots + Z_{n-1}}{n} = \limsup_{n\to+\infty} \frac{Z_{k+1} + Z_{k+2} + \cdots + Z_{n-1}}{n}$$

est indépendant de $\sigma(Z_p, p \leq k)$ pour tout $k \geq 0$ donc de $\sigma(Z_p, p \geq 0)$ (par classe monotone). Or elle est visiblement mesurable par rapport à cette tribu. Elle est donc indépendante d'elle même donc constante p.s. Finalement $\mathbb{E}(Z_0|\mathcal{I})$ est constante, donc égale à son espérance $\mathbb{E}(Z_0)$. On a donc montré

***Théorème 26.4* (Loi des grands nombres).** — *Pour toute suite Z_n de v.a. intégrables, indépendantes et de même loi Z, p.s.,*

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{k=0}^{n-1} Z_k = \mathbb{E}(Z_0).$$

27. Un théorème central limite pour martingales

Lemme 27.1. — *Pour x réel, au voisinage de 0,*

$$\exp(ix) = (1 + ix)exp(-\frac{x^2}{2} + r(x))$$

où $r(x) = O(x^3)$.

Preuve. — On a, en développant, pour x assez petit,

$$\frac{exp(ix + \frac{x^2}{2})}{1 + ix} = (1 + ix + \frac{x^2}{2} + \frac{1}{2}(ix + \frac{x^2}{2})^2 + O(x^3))(1 - ix - \frac{x^2}{2} + O(x^3))$$
$$= 1 + O(x^3) = \exp(\log(1 + h(x)))$$

où $h(x) = O(x^3)$. Alors $r(x) = \log(1 + h(x)) = O(x^3)$. □

Lemme 27.2. — *Soit $S_n = X_{n1} + \cdots + X_{nn}$, où $\sum_{j=1}^n X_{nj}^2 \to 1$ et $\max_{1\le j\le n} |X_{nj}| \to 0$ en probabilité. On suppose que $T_n = \prod_{j=1}^n (1 + itX_{nj})$ est uniformément intégrable et que $\mathbb{E}(T_n) \to 1$. Alors S_n converge en loi vers $N(0,1)$.*

Preuve. — Posons

$$U_n = \exp(-\frac{t^2}{2}\sum_{j=1}^n X_{nj}^2 + \sum_{j=1}^n r(tX_{nj})).$$

Alors si $|r(x)| \le Cx^3$ pour x petit,

$$|\sum_{j=1}^n r(tX_{nj}))| \le Ct^3 \max_j |X_{nj}| \sum_{j=1}^n X_{nj}^2$$

tend vers 0 quand $n \to +\infty$ donc $U_n \to e^{-t^2/2}$, en probabilité. On écrit que, avec le lemme,

$$exp(itS_n) = T_n U_n = T_n(U_n - e^{-t^2/2}) + T_n e^{-t^2/2}.$$

Alors

$$|T_n(U_n - e^{-t^2/2})| \le 1 + e^{-t^2/2}|T_n|$$

est uniformément intégrable, et tend vers 0 en probabilité donc

$$\lim_n \mathbb{E}(exp(itS_n)) = e^{-t^2/2} + \lim_n \mathbb{E}(T_n(U_n - e^{-t^2/2})) = e^{-t^2/2}$$

ce qui montre le résultat par Fourier. □

Théorème 27.3. — *Soit $\{M_k^{(n)}, k \ge 0\}$ des v.a. intégrables telles que, si*

$$Y_{nj} = M_j^{(n)} - M_{j-1}^{(n)},$$

— $M_0^{(n)} = 0$
— *Pour tout n, $M_k^{(n)}, k \ge 0$, est une martingale, i.e. pour tout $k \ge 1$,*

$$\mathbb{E}(M_k^{(n)} | \sigma(M_j^{(n)}, j \le k-1) = M_{k-1}^{(n)}.$$

— $\mathbb{E}(\max_{j\le n} |Y_{nj}|) \to 0$,
— $\sum_{j=1}^n Y_{nj}^2 \to 1, p.s$

alors $M_n^{(n)} \to N(0,1)$ en loi.

Preuve. — Première étape : troncature. On pose $X_{n1} = Y_{n1}$, pour $j \ge 2$,

$$X_{nj} = Y_{nj} \mathbf{1}_{\{\sum_{r\le j-1} Y_{nr}^2 \le 2\}}$$

et $S_k^{(n)} = \sum_{j=1}^k X_{nj}$ Alors

$$\mathbb{P}(M_n^{(n)} \neq S_n^{(n)}) \leq \mathbb{P}(\sum_{r=1}^{n-1} Y_{nr}^2 > 2) \to 0.$$

Donc $M_n^{(n)}$ et $S_n^{(n)}$ ont la même limite en loi.

Montrons que $S_n^{(n)}$ satisfait aux hypothèses du lemme. Puisque $M^{(n)}$ est une martingale, pour $\mathcal{F}_{j-1} = \sigma(M_k^{(n)}, k \leq j-1)$,

$$\begin{aligned}\mathbb{E}(X_{nj}|\mathcal{F}_{j-1}) &= \mathbb{E}(Y_{nj}\mathbf{1}_{\{\sum_{r\leq j-1} Y_{nr}^2 \leq 2\}}|\mathcal{F}_{j-1}) \\ &= \mathbf{1}_{\{\sum_{r\leq j-1} Y_{nr}^2 \leq 2\}}\mathbb{E}(Y_{nj}|\mathcal{F}_{j-1}) = 0.\end{aligned}$$

Donc

$$\mathbb{E}(T_n) = \mathbb{E}(\prod_{j=1}^n (1 + itX_{nj})) = \mathbb{E}(\mathbb{E}(\prod_{j=1}^n (1 + itX_{nj})|\mathcal{F}_{n-1})) = \mathbb{E}(\prod_{j=1}^{n-1} (1 + itX_{nj}))$$

et, en continuant, par récurrence, $\mathbb{E}(T_n) = 1$.

Montrons que T_n est uniformément intégrable. On a, pour t réel,

$$|1 + itx|^2 = 1 + t^2x^2 \leq e^{t^2x^2}$$

donc si $J = \inf\{j; \sum_{r=1}^j Y_{nr}^2 > 2\}$,

$$\begin{aligned}|T_n| &= \prod_{j=1}^J |1 + itX_{nj}| = (\prod_{j=1}^{J-1} |1 + itX_{nj}|)|1 + itX_{nJ}| \\ &\leq \exp(t^2 \sum_{j\leq J-1} X_{nj}^2/2)(1 + |t \max_p X_{np}|) \leq e^{t^2}(1 + |t \max_{p\leq n} X_{np}|)\end{aligned}$$

Puisque $\mathbb{E}(\max_{p\leq n} |X_{np}|) \to 0$, $\max_{p\leq n} |X_{np}|$ est uniformément intégrable, donc T_n l'est aussi. On déduit du lemme que $S_n^{(n)}$ et donc $M_n^{(n)}$ convergent en loi vers $N(0, 1)$. □

27.1. Cas des v.a. indépendantes. — Soit $X_k, k \geq 0$ une suite de v.a. indépendantes et de même loi, centrées, telles que $\mathbb{E}(X_k^2) = 1$. Les hypothèses du théorème sont vérifiées pour les suites

$$X_{nk} = \frac{X_k}{\sqrt{n}}$$

par la loi des grands nombres d'une part et par le fait suivant appliqué à $Z_k = X_k^2$,

Lemme 27.4. — *Si (Z_k) sont des v.a. positives intégrables de même loi,*

$$\lim_{n\to+\infty} \frac{1}{n}\mathbb{E}(\max_{1\leq k\leq n} Z_k) = 0.$$

Preuve. — Pour $M > 0$ assez grand,

$$\frac{1}{n}\mathbb{E}(\max_{k\leq n} Z_k \mathbf{1}_{\{Z_k>M\}}) \leq \frac{1}{n}\mathbb{E}(\sum_{k\leq n} Z_k \mathbf{1}_{\{Z_k>M\}}) \leq \mathbb{E}(Z_1 \mathbf{1}_{\{Z_1>M\}}) < \varepsilon/2$$

Et ce M étant choisi, pour n assez grand

$$\frac{1}{n}\mathbb{E}(\max_{1\leq k\leq n} Z_k \mathbf{1}_{\{Z_k\leq M\}}) \leq \frac{M}{n} < \varepsilon/2.$$

□

Il est remarquable que ceci s'applique aussi aux accroissements de martingale stationnaires ergodiques.

Ph Bougerol, Paris 6