Révisions de M2 : intégration et probabilités

Nicolas Fournier, d'après Thomas Duquesne

2016-2017

Table des matières

Chapitre I

Théorie de la mesure et de l'intégration

I.1 Tribus, classes monotones

Définition 1 *Soit E un ensemble non-vide. La classe de tous les sous-ensembles de E est notée $\mathscr{P}(E)$, c'est-à-dire que $\mathscr{P}(E) = \{A : A \subset E\}$.*

(i) $\mathscr{E} \subset \mathscr{P}(E)$ est une tribu (ou sigma-algèbre) si $E \in \mathscr{E}$ et si $\mathscr{E}$ est stable par passage au complémentaire (i.e. $A \in \mathscr{E}$ implique $A^c \in E$) et par union dénombrable (i.e. $A_n \in \mathscr{E}$, $n \geq 1$ implique $\cup_{n\geq 1} A_n \in \mathscr{E}$).

(ii) $\mathscr{P} \subset \mathscr{P}(E)$ est un pi-système si $E \in \mathscr{P}$ et si $\mathscr{P}$ est stable par intersection finie (i.e. $A, B \in \mathscr{P}$ implique $A \cap B \in \mathscr{P}$).

(iii) $\mathscr{L} \subset \mathscr{P}(E)$ est une classe monotone (ou lambda-système) si $E \in \mathscr{L}$, si $\mathscr{L}$ est stable par différence propre (i.e. pour tous $A, B \in \mathscr{L}$ tels que $A \subset B$, on a $B \backslash A \in \mathscr{L}$) et par union dénombrable croissante (i.e. si $A_n \in \mathscr{L}$, $n \in \mathbb{N}$ vérifient $A_n \subset A_{n+1}$ alors $\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{L}$).

Si $\mathscr{E}$ est une tribu sur E, on dit que $(E, \mathscr{E})$ est un espace mesurable.

Remarque 2 *(i) Une tribu est stable par union finie et par intersection finie ou dénombrable.*
(ii) Une tribu est une classe monotone.
(iii) Une classe monotone stable par intersection finie est une tribu (montrer d'abord que $A, B \in \mathscr{L}$ implique $A \cup B \in \mathscr{L}$ puis écrire une union dénombrable quelconque comme une union croissante).

Lemme 3 *Soit $(\mathscr{E}_i, i \in I)$, une famille de tribus (resp. de classes monotones) sur E. Alors,*

$$\bigcap_{i\in I} \mathscr{E}_i = \{A \subset E \; : \; \forall i \in I \, , \; A \in \mathscr{E}_i \}$$

est une tribu (resp. une classe monotone).

La preuve est immédiate. Ce lemme autorise la définition suivante.

Définition 4 *Soit $\mathscr{R} \subset \mathscr{P}(E)$. On définit la tribu (resp. la classe monotone) engendrée par $\mathscr{R}$ comme la plus petite tribu (resp. la plus petite classe monotone) contenant $\mathscr{R}$ et on la note $\sigma(\mathscr{R})$ (resp. $\lambda(\mathscr{R})$). Plus formellement, on a*

$$\sigma(\mathscr{R}) = \bigcap_{\substack{\mathscr{E} \text{ tribu telle que} \\ \mathscr{R}\subset\mathscr{E}}} \mathscr{E} \qquad \text{et} \qquad \lambda(\mathscr{R}) = \bigcap_{\substack{\mathscr{L} \text{ classe monotone telle que} \\ \mathscr{R}\subset\mathscr{L}}} \mathscr{L}$$

Théorème 5 **(Théorème de la classe monotone)** *Soient E un ensemble non-vide, $\mathscr{L}$ une classe monotone sur E et $\mathscr{P}$ un pi-système sur E. Si $\mathscr{P} \subset \mathscr{L}$, alors $\sigma(\mathscr{P}) \subset \mathscr{L}$.*

Pour montrer qu'une propriété est satisfaite pour tout élément d'une tribu $\mathscr{F}$ engendrée par une classe $\mathscr{C}$ (i.e. $\mathscr{F} = \sigma(\mathscr{C})$), on introduit $\mathscr{G} = \{A \in \mathscr{F} \ : \ A \text{ vérifie la propriété}\}$, puis on montre que $\mathscr{G}$ contient $\mathscr{C}$. Pour conclure, il y a deux méthodes :

- soit on parvient à montrer directement que $\mathscr{G}$ est une tribu ;
- soit on montrer que $\mathscr{G}$ est une classe monotone (c'est plus facile) ET que $\mathscr{C}$ est un pi-système.

Preuve : on pose $\mathscr{E} = \lambda(\mathscr{P})$. Bien sûr, $\mathscr{E} \subset \mathscr{L}$. On va montrer que $\mathscr{E}$ est une tribu, ce qui impliquera le résultat (car alors $\sigma(\mathscr{P}) \subset \mathscr{E}$). Comme $\mathscr{E}$ est une classe monotone, il suffit de montrer que $\mathscr{E}$ est stable par intersection (finie).

Pour $A \in \mathscr{E}$, on pose $\mathscr{L}_A = \{B \subset E : \ A \cap B \in \mathscr{E}\}$. Montrons que $\mathscr{L}_A$ est une classe monotone. Tout d'abord, il est clair que $E \in \mathscr{L}_A$. Montrons la stabilité de $\mathscr{L}_A$ par différence propre : pour cela on se donne $B, C \in \mathscr{L}_A$ tels que $B \subset C$, on observe que $A \cap (C \backslash B) = (A \cap C) \backslash (A \cap B)$ qui est donc dans $\mathscr{E}$ puisque $A \cap C$ et $A \cap B$ sont dans $\mathscr{E}$ (car $B, C \in \mathscr{L}_A$), puisque $A \cap B \subset A \cap C$, et puisque $\mathscr{E}$ est stable par différence propre (en tant que classe monotone). Soit ensuite $B_n \in \mathscr{L}_A$, $n \in \mathbb{N}$ t.q. $B_n \subset B_{n+1}$. Puisque $B_n \cap A \in \mathscr{E}$ pour tout n, puisque la suite $B_n \cap A$ est croissante et puisque $\mathscr{E}$ est stable par union dénombrable croissante (en tant que classe monotone), on a bien $A \cap (\cup_{n \in \mathbb{N}} B_n) = \cup_{n \in \mathbb{N}} (A \cap B_n) \in \mathscr{E}$. Donc $\cup B_n \in \mathscr{L}_A$.

Si maintenant $A \in \mathscr{P}$, il est clair que $\mathscr{P} \subset \mathscr{L}_A$, donc $\mathscr{L}_A$ est une classe monotone qui contient $\mathscr{P}$ et donc $\mathscr{E} \subset \mathscr{L}_A$.

Ceci montre que pour tout $A \in \mathscr{P}$, tout $B \in \mathscr{E}$, $A \cap B \in \mathscr{E}$.

Du coup, pour tout $B \in \mathscr{E}$, on a $\mathscr{P} \subset \mathscr{L}_B$. Donc $\mathscr{L}_B$ est une classe monotone qui contient $\mathscr{P}$ et donc $\mathscr{E} \subset \mathscr{L}_B$. Autrement dit, pour tout $A, B \in \mathscr{E}$, $A \cap B \in \mathscr{E}$. ■

I.2 Tribus boréliennes

Une topologie $\mathcal{O}$ sur un ensemble E est une classe de parties de E stable par union quelconque, intersection finie et telle que $E, \emptyset \in \mathcal{O}$. Les éléments de $\mathcal{O}$ sont appelés *ouverts*.

Définition 6 *Soit E un ensemble muni d'une topologie $\mathcal{O}$. On appelle tribu borélienne $\mathscr{B}(E)$ la tribu $\sigma(\mathcal{O})$ (lorsqu'il n'y a pas d'ambiguité sur la topologie).*

Remarque 7 *(i) Si E est un espace vectoriel normé, on a une topologie $\mathcal{O}$ découlant la la norme ($O \in \mathcal{O}$ ssi pour tout $x \in O$, il existe $\epsilon > 0$ tel que pour tout $y \in E$ avec $||x - y|| < \epsilon$, on ait $y \in O$) et donc une tribu borélienne naturelle.*

(ii) Si E est un espace vectoriel de dimension finie, alors toutes les normes induisent la même topologie et donc la même tribu borélienne.

Remarque 8 *Soit $\mathcal{O}$ une topologie sur E et $A \subset E$ quelconque. Alors $\mathcal{O}_A = \{A \cap O \ : \ O \in \mathcal{O}\}$ est une topologie sur A, on pose donc $\mathscr{B}(A) = \sigma(\mathcal{O}_A)$. On a $\mathscr{B}(A) = \{A \cap B \ : \ B \in \mathscr{B}(E)\}$.*

Définition 9 *On note $\mathscr{B}(\mathbb{R}^d)$ la tribu borélienne sur $\mathbb{R}^d$ (muni de sa topologie naturelle, associée à n'importe quelle norme).*

Lemme 10 *On note $\mathscr{R}$ l'ensembles des pavés ouverts à extrémités rationnelles de $\mathbb{R}^d$, i.e.*

$$\mathscr{R} = \big\{]a_1, b_1[\times \ldots \times]a_d, b_d[: \ a_i < b_i \ , \ a_i, b_i \in \mathbb{Q} \ , \ 1 \leq i \leq d \ \big\}.$$

Alors (i) $\mathscr{R}$ est dénombrable, (ii) tout ouvert de $\mathbb{R}^d$ est union dénombrable d'éléments de $\mathscr{R}$ et (iii) $\mathscr{B}(\mathbb{R}^d) = \sigma(\mathscr{R})$.

Preuve : Le point (i) est évident. Pour (ii), il suffit d'écrire un ouvert O comme l'union de tous les éléments de $\mathscr{R}$ inclus dans O (utiliser que O est ouvert pour montrer l'égalité). Enfin, (iii) découle de (ii), puisqu'alors $\mathcal{O} \subset \sigma(\mathscr{R})$ et donc $\mathscr{B}(\mathbb{R}^d) = \sigma(\mathcal{O}) \subset \sigma(\mathscr{R})$, l'autre inclusion étant triviale puisque $\mathscr{R} \subset \mathcal{O}$. ■

Remarque 11 *On a $\mathscr{B}(\mathbb{R}) = \sigma(\{]-\infty, x] \ : \ x \in \mathbb{R}\}) = \sigma(\{[a,b] \ : \ a,b \in \mathbb{R}, a \leq b\})$, on a aussi $\mathscr{B}(\mathbb{R}) = \sigma(\{]a,b[: \ a,b \in \mathbb{Q}, a < b\})$, etc.*

I.3 Mesures (positives)

On dit qu'une famille $(A_n)_{n\geq 1}$ de parties de E est 2.2.d. (pour deux à deux disjoints) si $A_n \cap A_m = \emptyset$ dès que $n \neq m$.

Définition 12 *Soit $(E, \mathscr{E})$ un espace mesurable. Une application $\mu : \mathscr{E} \to [0, \infty]$ est une mesure (positive) sur $(E, \mathscr{E})$ ssi*
(i) $\mu(\emptyset) = 0$,
(ii) pour tout $(A_n)_{n\geq 1} \subset \mathscr{E}$ 2.2.d., on a $\mu\Big(\bigcup_{n\geq 1} A_n\Big) = \sum_{n\geq 1} \mu(A_n)$.

Cette seconde propriété est appelée *propriété de sigma-additivité*. Si μ est une mesure sur $(E, \mathscr{E})$, on dit que $(E, \mathscr{E}, \mu)$ est un espace mesuré. Une mesure de masse $\mu(E) = 1$ est appelée *probabilité* (ou *mesure de probabilité*).

Remarque 13 *On montre alors facilement que si $A, B \in \mathscr{E}$ sont disjoints, alors $\mu(A \cup B) = \mu(A) + \mu(B)$ et que si $A, B \in \mathscr{E}$ avec $A \cup B$, alors $\mu(B) = \mu(A) + \mu(B \setminus A)$.*

Proposition 14 *Soit $(E, \mathscr{E})$ un espace mesurable et μ, une mesure sur $\mathscr{E}$. Soit $A_n \in \mathscr{E}$, $n \in \mathbb{N}$, une suite d'ensembles mesurables.*

(i) Si $A_n \subset A_{n+1}$, $n \in \mathbb{N}$, alors $\lim_n \uparrow \mu(A_n) = \mu(A)$, où $\bigcup A_n = A$.

(ii) Si $A_{n+1} \subset A_n$, $n \in \mathbb{N}$, et si $\mu(A_0) < \infty$, alors $\lim_n \downarrow \mu(A_n) = \mu(B)$, où $\bigcap A_n = B$.

(iii) La mesure μ est sigma-sous-additive : $\mu\left(\bigcup A_n\right) \leq \sum \mu(A_n)$.

Preuve : on suppose tout d'abord que la suite A_n , $n \in \mathbb{N}$, est croissante pour l'inclusion et on prouve (i). On pose $B_0 = A_0$ et $B_n = A_n \backslash A_{n-1}$ pour tout $n \geq 1$. Il est clair que les B_n sont deux-à-deux disjoints. De plus, $\bigcup_{0\leq k\leq n} B_k = A_n$ et $\bigcup_{n\in\mathbb{N}} B_k = A$. La sigma-additivité de μ entraîne alors

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

ce qui prouve (i). Le point (ii) se déduit de (i) par passage au complémentaire : on pose $A'_n = A_0 \backslash A_n$ pour tout $n \geq 1$. Comme la suite A_n , $n \in \mathbb{N}$, est décroissante, la suite des A'_n est croissante pour l'inclusion. Par ailleurs, puisque $\mu(A_0) < \infty$, il en est de même pour les quantités $\mu(A_n)$ et on a : $\mu(A'_n) = \mu(A_0) - \mu(A_n)$. Le point (i) donne alors

$$\lim_{n\to\infty} \mu(A_n) = \mu(A'_0) - \lim_{n\to\infty} \mu(A'_n) = \mu(A_0) - \mu(A') ,$$

où $\bigcup A'_n = A'$. Il suffit ensuite de remarquer que $A' = A_0 \backslash A$ et donc que $\mu(A) = \mu(A_0) - \mu(A')$.

Il reste a démontrer (iii). Pour cela on remarque tout d'abord que pour tous $B, C \in \mathscr{E}$, on a $B \cup C = B \cup (C \cap B^c)$ et que B et $C \cap B^c$ sont disjoints. Donc

$$\mu(B \cup C) = \mu(B) + \mu(C \cap B^c) \leq \mu(B) + \mu(C).$$

En appliquant de façon répétée cette inégalité on trouve $\mu(\cup_{0\leq k\leq n} A_k) \leq \sum_{0\leq k\leq n} \mu(A_k)$. Or la suite $\cup_{0\leq k\leq n} A_k$ est croissante donc (i) implique que

$$\mu\Big(\cup_{n\in\mathbb{N}} A_n \Big) = \lim_{n\to\infty} \mu\Big(\cup_{0\leq k\leq n} A_k \Big) \leq \lim_{n\to\infty} \sum_{0\leq k\leq n} \mu(A_k) ,$$

qui entraîne bien le résultat désiré. ∎

Définition 15 *Soit $(E, \mathscr{E}, \mu)$ un espace mesuré. La mesure μ est dite finie si $\mu(E) < \infty$. Elle est dite sigma-finie s'il existe une suite d'ensembles mesurables $E_n \in \mathscr{E}$, $n \in \mathbb{N}$, t.q. $\bigcup_{n\in\mathbb{N}} E_n = E$ et $\mu(E_n) < \infty$, pour tout $n \in \mathbb{N}$.*

Lemme 16 *Soit $(E, \mathscr{E}, \mu)$ un espace mesuré. Il y a équivalence entre les assertions suivantes.*

(i) La mesure μ est sigma-finie.

(ii) Il existe $E_n \in \mathscr{E}$, $n \in \mathbb{N}$, tels que $E_n \subset E_{n+1}$, $\mu(E_n) < \infty$ et $\bigcup_{n\in\mathbb{N}} E_n = E$.

(iii) Il existe $E_n \in \mathscr{E}$, $n \in \mathbb{N}$, deux-à-deux disjoints tels que $\mu(E_n) < \infty$ et $\bigcup_{n\in\mathbb{N}} E_n = E$.

Preuve : on suppose que μ est sigma-finie. Il existe donc $E_n \in \mathscr{E}$, $n \in \mathbb{N}$, tels que $\mu(E_n) < \infty$ et $\bigcup_{n\in\mathbb{N}} E_n = E$. On pose $E'_n = \bigcup_{0\leq k\leq n} E_k$. On a bien $E'_n \in \mathscr{E}$,

$$\mu(E'_n) = \mu\Big(\bigcup_{0\leq k\leq n} E_k \Big) \leq \sum_{0\leq k\leq n} \mu(E_k) < \infty$$

et $\bigcup_{n\in\mathbb{N}} E'_n = E$. Cela montre que (i) implique (ii). Avec les notations précédédentes, on pose $E''_0 = E_0$ et pour tout $n \geq 1$, on pose également $E''_n = E'_n \backslash E'_{n-1}$. On voit que $E''_n \in \mathscr{E}$, que les E''_n sont deux-à-deux disjoints, que $\mu(E''_n) \leq \mu(E'_n) < \infty$ et que $\bigcup_{n\in\mathbb{N}} E''_n = E$. Cela montre que (ii) implique (iii). Clairement (iii) implique (i), ce qui termine la preuve du lemme. ∎

Proposition 17 *Soit $(E, \mathscr{E})$ un ensemble mesurable.*

(i) Soit $\mu_1, \mu_2 : \mathscr{E} \to [0, \infty]$, deux mesures, et soit $a \in]0, \infty[$. On définit $\mu : \mathscr{E} \to [0, \infty]$, par $\mu(A) = \mu_1(A) + a\mu_2(A)$, pour tout $A \in \mathscr{E}$. Alors μ est une mesurenotée $\mu = \mu_1 + a\mu_2$.

(ii) Plus généralement, soit $\mu_n : \mathscr{E} \to [0, \infty]$, $n \in \mathbb{N}$, une suite de mesures. Soit $a_n \in]0, \infty[$, $n \in \mathbb{N}$. Pour tout $A \in \mathscr{E}$, on pose $\mu(A) = \sum_{n\in\mathbb{N}} a_n\mu_n(A)$. Alors $\mu : \mathscr{E} \to [0, \infty]$ est une mesure notée $\mu = \sum a_n\mu_n$.

(iii) Soit $\mu : \mathscr{E} \to [0, \infty]$ une mesureet soit $B \in \mathscr{E}$. Pour tout $A \in \mathscr{E}$, on pose $\nu(A) = \mu(B \cap A)$. Alors ν est une mesure appellée mesure-trace sur B de μ.

Le preuve est assez facile.

Théorème 18 *(i) Il existe une unique mesure* $\ell : \mathscr{B}(\mathbb{R}) \to [0, \infty]$*, appelée mesure de Lebesgue sur* $\mathbb{R}$*, telle que pour tout* $a < b$*, on ait* $\ell(]a, b]) = b - a$*.*

(ii) Il existe une unique mesure $\ell_d : \mathscr{B}(\mathbb{R}^d) \to [0, \infty]$*, appelée mesure de Lebesgue sur* $\mathbb{R}^d$*, telle que pour tout pavé* $R =]a_1, b_1] \times \ldots \times]a_d, b_d]$*, on ait* $\ell_d(R) = (b_1 - a_1) \times \ldots \times (b_d - a_d)$*.*

L'existence est difficile et sera prouvée à la fin du chapitre. L'unicité découle du théorème suivant avec $(E, \mathscr{E}) = (\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ et $\mathscr{P} = \{]a_1, b_1] \times \ldots \times]a_d, b_d] \ : -\infty < a_i < b_i < \infty\} \cup \{\mathbb{R}^d\}$ (exercice).

Théorème 19 (Unicité de prolongement des mesures) *Soit* $(E, \mathscr{E})$ *un espace mesurable. Soit* $\mathscr{P}$*, un pi-système tel que* $\sigma(\mathscr{P}) = \mathscr{E}$*. Soient* μ_1 *et* μ_2 *deux mesures sigma-finies. On fait les deux hypothèses suivantes (si* $\mu_1(E) = \mu_2(E) < \infty$*, la seconde est inutile)*

(i) Pour tout $A \in \mathscr{P}$*,* $\mu_1(A) = \mu_2(A)$*.*

(ii) Il existe $E_n \in \mathscr{P}$*,* $n \in \mathbb{N}$*, tels que*

$$E = \cup_{n \geq 0} E_n \ , \quad E_n \subset E_{n+1} \quad et \quad \mu_1(E_n) = \mu_2(E_n) < \infty \ .$$

Alors $\mu_1 = \mu_2$*, i.e. pour tout* $B \in \mathscr{E}$*, on a* $\mu_1(B) = \mu_2(B)$*.*

Preuve : on suppose tout d'abord que $\mu_1(E) = \mu_2(E) < \infty$. On pose alors

$$\mathscr{L} = \{B \in \mathscr{E} \ : \ \mu_1(B) = \mu_2(B) \ \} \ .$$

Par hypothèse, on a $\mathscr{P} \subset \mathscr{L}$. Montrons ensuite que $\mathscr{L}$ est une classe monotone : comme on a supposé que μ_1 et μ_2 ont même masse, on a $E \in \mathscr{L}$. Soient $B, C \in \mathscr{L}$ tels que $B \subset C$. Comme μ_1 et μ_2 sont de *masse finie*, on peut écrire :

$$\mu_1(C \backslash B) = \mu_1(C) - \mu_1(B) = \mu_2(C) - \mu_2(B) = \mu_2(C \backslash B),$$

qui entraîne bien que $C \backslash B \in \mathscr{L}$. Il reste à prouver la stabilité par union dénombrable croissante, ce qui est assez facile.

Ainsi, $\mathscr{L}$ est une classe monotone qui contient $\mathscr{P}$, donc par le théorème 5, on a $\mathscr{E} = \sigma(\mathscr{P}) \subset \mathscr{L} \subset \mathscr{E}$. Donc $\mathscr{L} = \mathscr{E}$, i.e. $\mu_1(B) = \mu_2(B)$ pour tout $B \in \mathscr{E}$.

On déduit le cas général facilement : pour $n \geq 1$, on introduit les mesures $\mu_{1,n}(A) = \mu_1(A \cap E_n)$ et $\mu_{2,n}(A) = \mu_2(A \cap E_n)$ qui sont finies et vérifient (i). Elles sont donc égales. On conclut aisément, puisque pour tout $A \in \mathscr{E}$, $\mu_1(A) = \lim_n \mu_{1,n}(A)$ et $\mu_2(A) = \lim_n \mu_{2,n}(A)$. ■

I.4 Applications mesurables

Soit $f : E \to F$ une application de E dans F. Pour $B \subset F$, on note

$$f^{-1}(B) = \{x \in E \ : \ f(x) \in B\} \ \subset \ E.$$

Soit $B_j \subset F$, $j \in J$, une famille (quelconque) de sous-ensembles de F. Alors, on a :

$$f^{-1}\Big(\bigcup_{j \in J} B_j\Big) = \bigcup_{j \in J} f^{-1}(B_j) \quad \text{et} \quad f^{-1}\Big(\bigcap_{j \in J} B_j\Big) = \bigcap_{j \in J} f^{-1}(B_j) \ .$$

On a aussi $f^{-1}(F) = E$ et $f^{-1}(B^c) = (f^{-1}(B))^c$ pour tout $B \subset F$.

Définition 20 *Soient* $(E, \mathscr{E})$ *et* $(F, \mathscr{F})$*, deux espaces mesurables. Une application* $f : E \to F$ *est dite* $(\mathscr{E}, \mathscr{F})$*-mesurable ssi*

$$\forall B \in \mathscr{F} \ , \quad f^{-1}(B) \in \ \mathscr{E} \ . \tag{I.1}$$

Convention. *Lorsque* f *est à valeurs dans* $\mathbb{R}$*, on dit simplement que* $f \in \mathscr{L}(E, \mathscr{E})$ *ou que* f *est* $\mathscr{E}$*-mesurable en sous-entendant qu'elle est* $(\mathscr{E}, \mathscr{B}(\mathbb{R}))$*-mesurable.*

Proposition 21 *Soient* $(E, \mathscr{E})$ *et* $(F, \mathscr{F})$*, deux espaces mesurables. Soit* $\mathscr{C} \subset \mathscr{F}$ *une classe d'ensembles t.q.* $\sigma(\mathscr{C}) = \mathscr{F}$*. Soit* $f : E \to F$*. Alors* f *est* $(\mathscr{E}, \mathscr{F})$*-mesurable ssi*

$$\forall C \in \mathscr{C} \ , \quad f^{-1}(C) \ \in \ \mathscr{E} \ .$$

Preuve : on pose $\mathscr{G} = \{B \in \mathscr{F} : f^{-1}(B) \in \mathscr{E}\}$ et on veut montrer que $\mathscr{F} \subset \mathscr{G}$. On a $\mathscr{C} \subset \mathscr{G}$ par hypothèse, il suffit donc de montrer que $\mathscr{G}$ est une tribu (car alors $\mathscr{F} = \sigma(\mathscr{C}) \subset \sigma(\mathscr{G}) = \mathscr{G}$). C'est facile. ■

Donnons plusieurs applications de ce résultat.

Remarque 22 *(i) Pour montrer que* $f : E \mapsto \mathbb{R}$ *est* $\mathscr{E}$*-mesurable, il suffit de voir que pour tout* $a \in \mathbb{R}$*,* $f^{-1}(]-\infty, a]) \in \mathscr{E}$ *(car* $\sigma\big(\{]-\infty, a]; a \in \mathbb{R}\}\big) = \mathscr{B}(\mathbb{R})$*).*

(ii) Si E *et* F *sont munis de topologies* $\mathcal{O}_E$ *et* $\mathcal{O}_F$*, alors toute fonction continue de* E *dans* F *est* $(\mathscr{B}(E), \mathscr{B}(F))$*-mesurable.*

Pour le point (ii), on rappelle que f est dite continue si pour tout $O \in \mathcal{O}_F$, on a $f^{-1}(O) \in \mathcal{O}_E$, que $\mathscr{B}(E) = \sigma(\mathcal{O}_E)$ et que $\mathscr{B}(F) = \sigma(\mathcal{O}_F)$.

Proposition 23 *Soient* $(E, \mathscr{E})$*,* $(F, \mathscr{F})$ *et* $(G, \mathscr{G})$*, trois espaces mesurables. Soit* $f : E \to F$ *et* $g : F \to G$ *qui sont supposées respectivement* $(\mathscr{E}, \mathscr{F})$*-mesurable et* $(\mathscr{F}, \mathscr{G})$*-mesurable. Alors* $g \circ f$ *est* $(\mathscr{E}, \mathscr{G})$*-mesurable.*

La preuve est immédiate.

Proposition 24 *Soit* $(E, \mathscr{E})$*, un espace mesurable.*

(i) Soit $f, g \in \mathscr{L}(E, \mathscr{E})$ *et* $a \in \mathbb{R}$*. Alors* $f + ag$ *et* fg *sont également dans* $\mathscr{L}(E, \mathscr{E})$*.*

(ii) Soit $f_n \in \mathscr{L}(E, \mathscr{E})$*,* $n \in \mathbb{N}$*. Alors* $\inf f_n$*,* $\sup f_n$*,* $\liminf f_n$ *et* $\limsup f_n$ *sont dans* $\mathscr{L}(E, \mathscr{E})$*.*

(iii) Soit $f_n \in \mathscr{L}(E, \mathscr{E})$*,* $n \in \mathbb{N}$*. On suppose que pour tout* $x \in E$*,* $\lim_{n\to\infty} f_n(x) = f(x)$ *existe. Alors* f *est dans* $\mathscr{L}(E, \mathscr{E})$*.*

Preuve : pour (i), on prouve d'abord que $h := (f, g) : E \to \mathbb{R}^2$ est $(\mathscr{E}, \mathscr{B}(\mathbb{R}^2))$-mesurable. Il suffit de montrer que pour tout ouvert U de $\mathbb{R}^2$, $h^{-1}(U) \in \mathscr{E}$. Mais on sait (voir Lemme 10) que U est une union dénombrable de rectangles ouverts, i.e. $U = \cup_{n\geq 1}(]a_n, b_n[\times]c_n, d_n[)$. Par conséquent,

$$h^{-1}(U) = \cup_{n\geq 1} h^{-1}(]a_n, b_n[\times]c_n, d_n[) = \cup_{n\geq 1}(f^{-1}(]a_n, b_n[) \cap g^{-1}(]c_n, d_n[)) \ \in \ \mathscr{E} \ .$$

On introduit ensuite $\psi_1(x, y) = x + ay$, $\psi_2(x, y) = xy$, qui sont continues et donc $(\mathscr{B}(\mathbb{R}^2), \mathscr{B}(\mathbb{R}))$-mesurables. On conclut qu'en effet, $f + cg = \psi_1 \circ h$ et $fg = \psi_2 \circ h$ sont $(\mathscr{E}, \mathscr{B}(\mathbb{R}))$-mesurables.

Pour (ii), on remarque qu'il suffit de traiter $\sup f_n$ (car $\inf f_n = -\sup(-f_n)$, car $\limsup f_n = \inf_n \sup_{p\geq n} f_p$ et $\liminf f_n = \sup_n \inf_{p\geq n} f_p$). On pose donc $g = \sup_n f_n$ et on remarque que pour tout $a \in \mathbb{R}$, on a

$$g^{-1}(]-\infty, a]) = \cap_{n\in\mathbb{N}} f_n^{-1}(]-\infty, a]) \in \mathscr{E}.$$

Enfin, (iii) découle de (ii) car alors $f = \limsup f_n$. ■

Exemple 25 *Soit $A \subset E$ et $\mathbf{1}_A : E \to \{0,1\}$ sa fonction indicatrice. Alors $\mathbf{1}_A$ est $\mathscr{E}$-mesurable ssi $A \in E$.*

Définition 26 *Soit $(E, \mathscr{E})$, un espace mesurable. Une fonction $s : E \to \mathbb{R}_+$ est dite $\mathscr{E}$-mesurable étagée (positive) s'il existe $A_1, \ldots, A_n \in \mathscr{E}$, deux-à-deux disjoints et $a_1, \ldots, a_n \in \mathbb{R}_+$ tels que $s = \sum_{k=1}^n a_k \mathbf{1}_{A_k}$.*

Une fonction $\mathscr{E}$-mesurable $s : E \mapsto \mathbb{R}_+$ est étagée ssi elle prend un nombre fini de valeurs (exercice).

Lemme 27 *Soit $(E, \mathscr{E})$ un espace mesurable et $f : E \mapsto \mathbb{R}_+$ une fonction $\mathscr{E}$-mesurable.*

(i) Il existe une famille $(s_n)_{n\geq 1}$ de fonction $\mathscr{E}$-mesurables étagées telle que pour tout $x \in E$, $s_n(x)$ est croissante et tend vers $f(x)$ (quand $n \to \infty$).

(ii) Il existe $a_n \geq 0$ et $B_n \in \mathscr{E}$, $n \geq 1$, tels que $f = \sum_{n\geq 1} a_n \mathbf{1}_{B_n}$.

Preuve : Pour (i), on définit $\psi_n(a) = \min\{n, \lfloor 10^n a \rfloor / 10^n\}$ pour tout $a \in \mathbb{R}_+$, qui est bien sûr mesurable (de $\mathbb{R}_+$ dans $\mathbb{R}_+$). Comme $\psi_n(a)$ est croissante (en n) et tend vers a (quand $n \to \infty$) pour tout $a \geq 0$ et comme ψ_n prend un nombre fini de valeurs, on conclut aisément que $s_n = \psi_n \circ f$ convient.

Pour (ii), on écrit $f = \lim_n s_n = s_1 + \sum_{k\geq 1}(s_{k+1} - s_k)$. Comme s_1, $s_2 - s_1$, $s_3 - s_2$, etc. sont mesurables, positives et prennent un nombre fini de valeurs, on peut les écrire sous la forme $s_1 = \sum_{i=1}^{n_1} a_{1,i} \mathbf{1}_{B_{1,i}}$, $s_2 - s_1 = \sum_{i=1}^{n_2} a_{2,i} \mathbf{1}_{B_{2,i}}$, etc. Ainsi, $f = \sum_{k\geq 1} \sum_{i=1}^{n_k} a_{k,i} \mathbf{1}_{B_{k,i}}$, qu'on peut bien sûr réécrire avec une seule somme. ■

I.5 Construction de l'intégrale

Dans toute cette section, on considère un espace mesuré $(E, \mathscr{E}, \mu)$.

Intégration des fonctions étagées. On note $\mathbf{S}_+$ l'ensemble des fonctions *$\mathscr{E}$-mesurables* étagées (positives).

Définition 28 *Soit $s \in \mathbf{S}_+$, on a donc $\#(s(E)) < \infty$ et on peut écrire $s = \sum_{a\in s(E)} \mathbf{1}_{s^{-1}(\{a\})}$. On définit, pour tout $A \in \mathscr{E}$,*

$$\int_A s d\mu = \sum_{a\in s(E)} a\mu(A \cap s^{-1}(\{a\})).$$

Proposition 29 *(i) Si $s \in \mathbf{S}_+$ et $\alpha \geq 0$, alors $\alpha s \in \mathbf{S}_+$ et $\int_A (\alpha s) d\mu = \alpha \int_A s d\mu$, $\forall\ A \in \mathscr{E}$.*

(ii) Si $s_1, s_2 \in \mathbf{S}_+$, alors $s_1 + s_2 \in \mathbf{S}_+$ et $\int_A (s_1 + s_2) d\mu = \int_A s_1 d\mu + \int_A s_2 d\mu$, $\forall\ A \in \mathscr{E}$.

(iii) Si $s_1, s_2 \in \mathbf{S}_+$ et si $s_1 \leq s_2$, alors $\int_A s_1 d\mu \leq \int_A s_2 d\mu$, $\forall\ A \in \mathscr{E}$.

(iv) Si $s \in \mathbf{S}_+$, alors ν définie par $\nu(A) = \int_A s d\mu$, $\forall\ A \in \mathscr{E}$, est une mesure sur $(E, \mathscr{E})$.

Preuve : (i) est évident, (iii) découle de (ii) (écrire $s_2 = s_1 + t$ avec $t = s_2 - s_1 \geq 0$) et (iv) n'est pas difficile. Concentrons-nous sur (ii), qui est relativement délicat. On pose $I_1 = s_1(E)$, $I_2 = s_2(E)$ et $I = s(E)$, où $s = s_1 + s_2$. Pour tout $a \in I$, on peut écrire l'union disjointe

$$s^{-1}(\{a\}) = \bigcup_{a_1 \in I_1, a_2 \in I_2 \text{ tels que } a_1 + a_2 = a} (s_1^{-1}(\{a_1\}) \cap s_2^{-1}(\{a_2\})).$$

De plus, $\{s_1^{-1}(\{a_1\}), a_1 \in I_1\}$ est une partition de E, ainsi que $\{s_2^{-1}(\{a_2\}), a_2 \in I_2\}$. Donc par définition,

$$\int_A s d\mu = \sum_{a \in I} a\mu(A \cap s^{-1}(\{a\})) = \sum_{a \in I} a \sum_{a_1 \in I_1, a_2 \in I_2 \text{ tels que } a_1 + a_2 = a} \mu(A \cap s_1^{-1}(\{a_1\}) \cap s_2^{-1}(\{a_2\})),$$

puis, en rentrant a dans la seconde somme et en le remplaçant par $a_1 + a_2$, on trouve

$$\begin{aligned} \int_A s d\mu = & \sum_{a \in I} \sum_{a_1 \in I_1, a_2 \in I_2 \text{ tels que } a_1 + a_2 = a} a_1 \mu(A \cap s_1^{-1}(\{a_1\}) \cap s_2^{-1}(\{a_2\})) \\ & + \sum_{a \in I} \sum_{a_1 \in I_1, a_2 \in I_2 \text{ tels que } a_1 + a_2 = a} a_2 \mu(A \cap s_1^{-1}(\{a_1\}) \cap s_2^{-1}(\{a_2\})). \end{aligned}$$

Le premier terme peut se réécrire simplement

$$\sum_{a_1 \in I_1, a_2 \in I_2} \mu(A \cap s_1^{-1}(\{a_1\}) \cap s_2^{-1}(\{a_2\})) = \sum_{a_1 \in I_1} a_1 \mu(A \cap s_1^{-1}(\{a_1\}))) = \int_A s_1 d\mu.$$

La seconde égalité utilise que $\{s_2^{-1}(\{a_2\}), a_2 \in I_2\}$ est une partition de E. De même, le second terme vaut $\int_A s_2 d\mu$. ■

Remarque 30 *On déduit aisément de la proposition précédente que si $s = \sum_1^n a_i \mathbf{1}_{B_i}$, avec $a_i \geq 0$ et $B_i \in \mathscr{E}$ (non forcément 2.2.d.), alors pour tout $A \in E$, on a*

$$\int_A s d\mu = \sum_1^n a_i \mu(A \cap B_i).$$

En effet, on a clairement $\int_A \mathbf{1}_{B_i} d\mu = \mu(A \cap B_i)$ par définition, et il suffit ensuite d'utiliser le point (ii) de la proposition.

On aurait pu prendre cette formule comme définition : ç'aurait été une définition plus simple, mais il aurait fallu montrer que le résultat obtenu ne dépendait pas de la décomposition choisie.

Intégration des fonctions positives.

Définition 31 *Pour toute application $f : E \to [0, \infty]$, $\mathscr{E}$-mesurable et tout $A \in \mathscr{E}$ on pose*

$$\int_A f\, d\mu = \sup \left\{ \int_A s\, d\mu \ : \ s \in \mathbf{S}_+ \ t.q.\ 0 \leq s(x) \leq f(x) \ \forall\, x \in A \right\}.$$

En utilisant le point (iii) de la proposition précédente, on voit que si $s \in \mathbf{S}_+$, cette nouvelle définition donne bien le même résultat, i.e. que cette définition est bien une extension de la définition 28.

Proposition 32 *Soient $f, g : E \to [0, \infty]$, deux fonctions $\mathscr{E}$-mesurables.*

(i) Si $f \leq g$ alors $\int_A f\, d\mu \leq \int_A g\, d\mu$, pour tout $A \in \mathscr{E}$.

(ii) Soit $A \in \mathscr{E}$, Si $f(x) = 0$ pour tout $x \in A$, alors $\int_A f\, d\mu = 0$.

(iii) Soit $A \in \mathscr{E}$ tel que $\mu(A) = 0$. Alors $\int_A f\, d\mu = 0$.

(iv) Soit $A \in \mathscr{E}$. Alors $\int_A f\, d\mu = \int_E \mathbf{1}_A f\, d\mu$.

Preuve : on remarque que $\{s \in \mathbf{S}_+ :\ s(x) \leq f(x), x \in A\} \subset \{s \in \mathbf{S}_+ :\ s(x) \leq g(x), x \in A\}$ si $f \leq g$, ce qui entraîne (i). Les autres points sont faciles (immédiat si f est dans $\mathbf{S}_+$, puis extension facile). ■

Théorème 33 **(Convergence monotone)** *Soit $f_n : E \to [0, \infty]$, $n \in \mathbb{N}$, une suite croissante d'applications $\mathscr{E}$-mesurables, i.e. $f_n(x) \leq f_{n+1}(x)$ $\forall\ x \in E$, $\forall\ n \in \mathbb{N}$. On pose $f = \sup_{n\in\mathbb{N}} f_n$, qui est bien définie à valeurs dans $[0, \infty]$ et $\mathscr{E}$-mesurable. On a alors*

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

Preuve : par le point (i) de la proposition 32, la suite $n \mapsto \int_E f_n\, d\mu$ est croissante. De plus $f_n \leq f$ et donc $\int_E f_n\, d\mu \leq \int_E f\, d\mu$, puis

$$\lim_{n\to\infty} \uparrow \int_E f_n\, d\mu = \sup_{n\in\mathbb{N}} \int_E f_n\, d\mu\ \leq \int_E f\, d\mu\ .$$

Pour l'inégalité contraire, on fixe $a \in]0, 1[$ et $s \in \mathbf{S}_+$ telle que $s \leq f$. Pour tout entier $n \geq 1$, on pose

$$E_n = \{x \in E\ :\ f_n(x) \geq as(x)\}\ .$$

Comme f_n et s sont $\mathscr{E}$-mesurables, on a $E_n \in \mathscr{E}$. On voit que $E_n \subset E_{n+1}$ et, comme f_n tend simplement vers $f \geq s > as$, on a $E = \cup_{n\geq 1} E_n$.

On remarque ensuite que pour tout $x \in E$, on a $f_n(x) \geq \mathbf{1}_{E_n}(x) f_n(x) \geq \mathbf{1}_{E_n}(x) as(x)$, d'où l'inégalité $\int_E f_n d\mu \geq a \int_{E_n} s d\mu$. En utilisant le fait que $A \in \mathscr{E} \mapsto \int_A s\, d\mu$ est une mesure positive (voir Proposition 29 (iv)), on conclut que $\lim_{n\to\infty} \uparrow \int_{E_n} s\, d\mu = \int_E s\, d\mu$ et donc que $\lim_n \int_E f_n d\mu \geq a \int_E s\, d\mu$. Comme a peut être choisi arbitrairement proche de 1, on conclut que $\lim_n \int_E f_n d\mu \geq \int_E s\, d\mu$. En prenant finalement le supremum sur toutes les fonctions étagées positives majorées par f, on obtient $\lim_{n\to\infty} \int_E f_n\, d\mu \geq \int_E f\, d\mu$. ■

Lemme 34 *Soient $f, g : E \to [0, \infty]$, deux applications $\mathscr{E}$-mesurables. Soit $a \in [0, \infty[$. On a alors*

$$\int_E (f + ag)\, d\mu = \int_E f\, d\mu + a \int_E g\, d\mu\ .$$

Preuve : on considère deux suites croissantes $s_n, s'_n \in S_+$ telles que $\lim_n s_n = f$ et $\lim_n s'_n = g$ (voir le lemme 27). On sait déjà que $\int_E (s_n + as'_n)\, d\mu = \int_E s_n\, d\mu + a \int_E s'_n\, d\mu$. Le théorème de convergence monotone permet de passer à limite dans cette expression.

Proposition 35 **(Interversion positive)** *Soit $f_n : E \to [0, \infty]$, $n \in \mathbb{N}$, des applications $\mathscr{E}$-mesurables. On a*

$$\int_E \Big(\sum_{n\in\mathbb{N}} f_n \Big)\, d\mu = \sum_{n\in\mathbb{N}} \int_E f_n\, d\mu\ .$$

Preuve : découle immédiatement du théorème de convergence monotone appliqué à $g_n = \sum_{k=0}^{n} f_k$. ■

Théorème 36 **(Lemme de Fatou)** *Soit $f_n : E \to [0, \infty]$, $n \in \mathbb{N}$, une suite d'applications $\mathscr{E}$-mesurables. On a*

$$\int_E (\liminf_{n\to\infty} f_n)\, d\mu \;\le\; \liminf_{n\to\infty} \int_E f_n\, d\mu \;.$$

Preuve : on pose $g_n = \inf_{p\ge n} f_p$, $n \in \mathbb{N}$. On remarque que $0 \le g_n \le g_{n+1}$ et que $g_n \le f_n$. Par ailleurs, on a $\lim_n g_n = \liminf_n f_n$. Le théorème de convergence monotone implique alors que

$$\int_E (\liminf_{n\to\infty} f_n)\, d\mu = \lim_{n\to\infty} \int_E g_n\, d\mu \;.$$

Or on a

$$\lim_{n\to\infty} \int_E g_n\, d\mu = \liminf_{n\to\infty} \int_E g_n\, d\mu \le \liminf_{n\to\infty} \int_E f_n\, d\mu \;,$$

ce qui montre bien le résultat voulu. ■

Intégration des fonctions quelconques.

Définition 37 *Soit $f : E \to \mathbb{R}$, une application $\mathscr{E}$-mesurable. On dit que f est μ-intégrable ou que $f \in \mathscr{L}^1(E, \mathscr{E}, \mu)$ ssi $\int_E |f|\, d\mu < \infty$. On définit alors, pour tout $A \in \mathscr{E}$, $\int_A f\, d\mu = \int_A f_+\, d\mu - \int_A f_-\, d\mu$, où les fonctions $\mathscr{E}$-mesurables positives f_+ et f_- sont définies par $f_+(x) = \max\{f(x), 0\}$ et $f_-(x) = \max\{-f(x), 0\}$.*

On observe que $f = f_+ - f_-$, ce qui justifie cette définition. On observe aussi que $|f| = f_+ + f_-$. Donc $\int_E |f|\, d\mu < \infty$ implique que $\int_A f_+\, d\mu$ et $\int_A f_-\, d\mu$ sont deux nombres réels positifs finis (pour $A \in \mathscr{E}$ fixé), il n'y a donc pas de problème dans la définition.

Proposition 38 *Soient $f, g \in \mathscr{L}^1(E, \mathscr{E}, \mu)$.*

(i) $\int_A f\, d\mu = \int_E \mathbf{1}_A f\, d\mu$, pour tout $A \in \mathscr{E}$.

(ii) $|\int_E f\, d\mu| \le \int_E |f|\, d\mu$.

(iii) $\mathscr{L}^1(E, \mathscr{E}, \mu)$ est un $\mathbb{R}$-espace vectoriel et pour tout $a \in \mathbb{R}$, $\int_E (f + ag)\, d\mu = \int_E f\, d\mu + a \int_E g\, d\mu$.

(iv) Si $f \le g$ alors $\int_E f\, d\mu \le \int_E g\, d\mu$.

(v) Si $A \in \mathscr{E}$ est tel que $\mu(A) = 0$, alors $\int_A f\, d\mu = 0$.

Preuve : (i) et (v) se déduisent du cas positif. Pour (ii), on écrit $\left|\int_E f_+\, d\mu - \int_E f_-\, d\mu\right| \le \int_E f_+\, d\mu + \int_E f_-\, d\mu = \int_E |f|\, d\mu$. (iv) découle de (iii). Par contre, (iii) est un peu subtile, puisqu'il est faux par exemple que $(f+g)_+ = f_+ + g_+$. Supposons par exemple $a \ge 0$ (sinon, on remplace g par $-g$). Comme $|f + ag| \le |f| + a|g|$, il est évident (voir proposition 32-(i) et lemme 34) que $f + ag \in \mathscr{L}^1(E, \mathscr{E}, \mu)$. On pose $h = f + ag$, ce qui se réécrit $h_+ - h_- = f_+ - f_- + ag_+ - ag_-$. Donc on a $h_+ + f_- + ag_- = h_- + f_+ + ag_+$ et, par le lemme 34 :

$$\int_E h_+\, d\mu + \int_E f_-\, d\mu + a \int_E g_-\, d\mu = \int_E h_-\, d\mu + \int_E f_+\, d\mu + a \int_E g_+\, d\mu \;.$$

On conclut facilement. ■

Théorème 39 **(Convergence dominée, V1)** *Soient $f_n : E \to \mathbb{R}$, $n \in \mathbb{N}$, une suite de fonctions $\mathscr{E}$-mesurables telle que*

(a) Il existe $f : E \to \mathbb{R}$ telle que $\lim_n f_n(x) = f(x)$, pour tout $x \in E$.

(b) Il existe $g \in \mathscr{L}^1(E, \mathscr{E}, \mu)$ telle que $|f_n(x)| \le g(x)$, pour tout $n \in \mathbb{N}$ et tout $x \in E$.

Alors (i) $f_n, f \in \mathscr{L}^1(E, \mathscr{E}, \mu)$, (ii) $\lim_{n\to\infty} \int_E |f - f_n| \, d\mu = 0$ et (iii) $\lim_{n\to\infty} \int_E f_n \, d\mu = \int_E f \, d\mu$.

Preuve : en passant à la limite en n dans les inégalités de l'hypothèse (b), on obtient $|f| \le g$ et on a donc $\int_E |f| \, d\mu < \infty$, par le point (iv) de la proposition 38, ce qui prouve (i). On remarque ensuite que $|f_n - f| \le |f| + |f_n| \le 2g$. Si on pose $h_n = 2g - |f - f_n|$, on observe que les h_n sont des fonctions positives $\mathscr{E}$-mesurables et μ-intégrables. De plus $\liminf_n h_n = \lim_n h_n = 2g$, d'après l'hypothèse (a). En appliquant le lemme de Fatou à la suite h_n, on obtient

$$\int_E 2g \, d\mu = \int_E (\liminf_n h_n) \, d\mu \le \liminf_{n\to\infty} \int_E (2g - |f - f_n|) \, d\mu = \int_E 2g \, d\mu - \limsup_{n\to\infty} \int_E |f - f_n| \, d\mu \, ,$$

ce qui implique que $\limsup_{n\to\infty} \int_E |f - f_n| \, d\mu \le 0$. Cela prouve (ii), qui implique bien sûr (iii). ■

I.6 Mesure image

Proposition 40 **(Changement de variable/Transfert)** *Soit $(E, \mathscr{E}, \mu)$ un espace mesuré et $(F, \mathscr{F})$ un espace mesurable. Soit $f : E \to F$ une application $(\mathscr{E}, \mathscr{F})$-mesurable. Pour tout $B \in \mathscr{F}$, on pose $\nu(B) = \mu(f^{-1}(B))$.*

(i) ν est une mesure sur $(F, \mathscr{F})$ (appelée mesure image de μ par f).

(ii) Pour tout $h : F \mapsto \mathbb{R}_+$, $\mathscr{F}$-mesurable, on a $\int_F h d\nu = \int_E (h \circ f) d\mu$.

(ii) Pour tout $h : F \mapsto \mathbb{R}$, $\mathscr{F}$-mesurable, $h \in \mathscr{L}^1(F, \mathscr{F}, \mu)$ ssi $h \circ f \in \mathscr{L}^1(E, \mathscr{E}, \mu)$ et dans ce cas, $\int_F h d\nu = \int_E (h \circ f) d\mu$.

Preuve : Le (i) est très facile. Pour (ii), on écrit $h = \sum_{n\ge1} a_n \mathbf{1}_{C_n}$, avec $a_n \ge 0$ et $C_n \in \mathscr{F}$ (voir Lemme 27-(ii)), puis

$$\int_F h d\nu = \sum_{n\ge1} a_n \nu(C_n) = \sum_{n\ge1} a_n \mu(f^{-1}(C_n)) = \sum_{n\ge1} a_n \int_E \mathbf{1}_{f^{-1}(C_n)} d\mu.$$

Mais $\mathbf{1}_{f^{-1}(C_n)} = \mathbf{1}_{C_n} \circ f$. Du coup,

$$\int_F h d\nu = \sum_{n\ge1} a_n \int_E (\mathbf{1}_{C_n} \circ f) d\mu = \int_E (h \circ f) d\mu.$$

On a utilisé un peu partout la proposition 35 pour échanger les sommes et les intégrales. Le point (iii) se déduit classiquement du point (ii), à faire en exercice. ■

I.7 Ensembles négligeables

Définition 41 *Soit $(E, \mathscr{E}, \mu)$, un espace mesuré. Un sous-ensemble N de E est dit μ-négligeable ssi il existe $B \in \mathscr{E}$ tel que $N \subset B$ et $\mu(B) = 0$.*

On note $\mathscr{N}_\mu$ l'ensemble de tous les μ-négligeables de E. On dit que l'espace mesuré $(E, \mathscr{E}, \mu)$ est complet ssi $\mathscr{E}$ contient tous les μ-négligeables, c'est-à-dire $\mathscr{N}_\mu \subset \mathscr{E}$.

Remarque 42 *Soient* $(E, \mathscr{E}, \mu)$*, un espace mesuré et* $N_n \in \mathscr{N}_\mu$*,* $n \in \mathbb{N}$*. Alors* $\bigcup_{n\in\mathbb{N}} N_n \in \mathscr{N}_\mu$*.*

Cela découle de la σ-additivité de la mesure.

Définition 43 *(Propriété vraie presque partout) Soit* $(E, \mathscr{E}, \mu)$*, un espace mesuré. Une propriété est dite vraie* μ*-p.p. si elle est vraie en dehors d'un ensemble* μ*-négligeable.*

Par exemple, pour deux fonctions réelles $f, g : E \mapsto \mathbb{R}$, on dit que $f = g$ μ-p.p. si l'ensemble $\{x \in E \ : \ f(x) \neq g(x)\}$ est μ-négligeable.

Proposition 44 *Soit* $(E, \mathscr{E}, \mu)$ *un espace mesuré. Soit* $f, g : E \to \overline{\mathbb{R}}$ *deux applications* $\mathscr{E}$*-mesurables.*

(i) Si $\int_E |f|\, d\mu = 0$*, alors* $f = 0$ μ*-p.p.*

(ii) Si $\int_E |f|\, d\mu < \infty$*, alors* $|f| < \infty$ μ*-p.p.*

(iii) Si $f = g$ μ*-p.p., alors* $\int_E f d\mu = \int_E g d\mu$*.*

Preuve : (i) Pour tout $n \geq 1$, on pose $E_n = \{x \in E \ : \ |f(x)| \geq 1/n\} = |f|^{-1}([1/n, \infty]) \in \mathscr{E}$. On a donc $f^{-1}(\overline{\mathbb{R}}\backslash\{0\}) = \cup_{n\in\mathbb{N}} E_n$. D'autre part, $|f| \geq \mathbf{1}_{E_n}|f| \geq \frac{1}{n}\mathbf{1}_{E_n}$. Par conséquent, $0 = \int_E |f|\, d\mu \geq \mu(E_n)/n$. Donc, $\mu(E_n) = 0$. On en déduit que $f^{-1}(\overline{\mathbb{R}}\backslash\{0\}) = \cup_n E_n$ est μ-négligeable comme désiré.

(ii) Soit $E_n = \{|f| \geq n|\} = |f|^{-1}([n, \infty[\} \in \mathscr{E}$. Alors $\mathbf{1}_{E_n}$ décroit vers $\mathbf{1}_{\{|f|=\infty\}}$, donc par Fatou,

$$\mu(\{|f| = \infty\}) = \int_E \mathbf{1}_{\{|f|=\infty\}} d\mu \leq \liminf_n \int_E \mathbf{1}_{E_n} d\mu \leq \liminf_n n^{-1} \int_E |f| d\mu = 0.$$

(iii) Soit $A = \{f \neq g\} \in \mathscr{E}$. On a $\mu(A) = 0$, et donc $\int_E |f - g| d\mu = \int_A |f - g| d\mu = 0$. On conclut que $|\int_E f d\mu - \int_E g d\mu| \leq \int_E |f - g| d\mu = 0$. ∎

Théorème 45 **(Théorèmes limites, versions définitives)** *Soit* $(E, \mathscr{E}, \mu)$ *un espace mesuré. Soit* f_n*,* $n \in \mathbb{N}$*, une suite d'applications* $\mathscr{E}$*-mesurables de* E *dans* $\overline{\mathbb{R}}$*.*

(Convergence monotone) *Si pour tout* $n \in \mathbb{N}$*,* $0 \leq f_n \leq f_{n+1}$ μ*-p.p., alors* $\sup_n f_n = f$ *est bien définie, positive* μ*-p.p., et on a* $\lim_{n\to\infty}\uparrow \int_E f_n\, d\mu = \int_E f\, d\mu$*.*

(Lemme de Fatou) *Si pour tout* $n \in \mathbb{N}$*, on a* $f_n \geq 0$ μ*-p.p., alors* $\liminf_n f_n = f$ *est positive* μ*-p.p. et on a* $\int_E f\, d\mu \leq \liminf_{n\to\infty} \int_E f_n\, d\mu$*.*

(Convergence dominée) *S'il existe une application* f*,* $\mathscr{E}$*-mesurable, telle que* $\lim_n f_n = f$*,* μ*-p.p. et s'il existe une application* $g \in \mathscr{L}^1(E, \mathscr{E}, \mu)$ *telle que* $|f_n| \leq g$ μ*-p.p. pour tout* $n \in \mathbb{N}$*. Alors,* f *est* μ*-intégrable, on a* $\lim_{n\to\infty} \int_E |f - f_n|\, d\mu = 0$ *et* $\lim_{n\to\infty} \int_E f_n\, d\mu = \int_E f\, d\mu$*.*

(Interversion) *Si* $\sum_n \int_E |f_n| d\mu < \infty$*, alors la série* $\sum_n |f_n|$ *converge* μ*-p.p. Il existe une fonction* $h : E \mapsto \mathbb{R}$*,* $\mathscr{E}$*-mesurable, telle que* $\sum_{n\geq 1} f_n = h$ μ*-p.p. De plus, on a* $\int_E h d\mu = \sum_{n\geq 1} \int_E f_n d\mu$*.*

Preuve : Les trois premiers points se déduisent des versions précédentes (sans les presque partout) de la même manière, étudions par exemple la convergence dominée. On trouve facilement un ensemble A, dont le complémentaire est μ-négligeable, sur lequel on a $|f_n| \leq g$ pour tout n et $\lim_n f_n = f$. On introduit alors $f'_n = f_n \mathbf{1}_A$, $f' = f\mathbf{1}_A$ et $g' = g\mathbf{1}_A$, qui satisfont les hypothèses du théorème de convergence dominée. On en déduit que $f' \in \mathscr{L}^1(E, \mathscr{E}, \mu)$ et que $\lim_n \int_E |f'_n - f'| d\mu = 0$. Mais bien sûr, $f = f'$ μ-p.p., donc $f \in \mathscr{L}^1(E, \mathscr{E}, \mu)$. De plus, $f'_n - f' = f_n - f$ μ-p.p. (pour tout n), donc $\int_E |f_n - f| d\mu = \int_E |f'_n - f'| d\mu$. Ainsi, $\lim_n \int_E |f - f_n|\, d\mu = 0$, d'où on déduit que $\lim_n \int_E f_n\, d\mu = \int_E f\, d\mu$.

Pour l'interversion, on utilise que $\int_E(\sum_n |f_n|)d\mu = \sum_n \int_E |f_n|d\mu < \infty$, ce qui implique que $\sum_n |f_n| < \infty$ μ-p.p. Soit $A = \{\sum_n |f_n| < \infty\} \in \mathscr{E}$, on sait que A^c est μ-négligeable. Sur A, $\sum_n f_n$ converge et on pose $h = \sum_n f_n$. Enfin, on pose $h = 0$ sur A^c. On a donc bien $\sum_n f_n = h$ μ-p.p. Enfin, on peut utiliser la convergence dominée (avec $g = \sum_n |f_n|$, $h_n = \sum_1^n f_k$ et $h = \sum_{k\geq 1} f_k$) pour conclure que $\lim_n \int_E h_n d\mu = \int_E h d\mu$. Comme $\int_E h_n d\mu = \sum_1^n \int_E f_k d\mu$, on en déduit qu'en effet, $\int_E h d\mu = \sum_{n\geq 1} \int_E f_n d\mu$. ∎

I.8 Continuité et dérivabilité

Voici un théorème sur la continuité des intégrales à paramètres.

Théorème 46 *Soit (Y, d) un espace métrique et $(E, \mathscr{E}, \mu)$ un espace mesuré. Soit $f : E \times Y \to \overline{\mathbb{R}}$. On fait les hypothèses suivantes.*

(a) Pour tout $y \in Y$, $f(\cdot, y)$ est $\mathscr{E}$-mesurable.

(b) Pour μ-presque tout $x \in E$, $y \mapsto f(x, y)$ est continue sur Y.

(c) Il existe une fonction $g : E \to [0, \infty]$ qui est $\mathscr{E}$-mesurable et μ-intégrable telle que $|f(x, y)| \leq g(x)$, pour μ-presque tout $x \in E$, ceci pour tout $y \in Y$.

Alors la fonction $F : Y \to \mathbb{R}$ donnée par $F(y) = \int_E f(x, y)\mu(dx)$ est bien définie et continue.

Preuve : observons tout d'abord que $F(y)$ est bien définie pour tout $y \in Y$ par (a) et (c). Soit $y_n \in Y$, $n \in \mathbb{N}$, une suite de points de Y qui converge vers $y \in Y$ fixé (i.e. $\lim_n d(y_n, y) = 0$). Pour tout $n \in \mathbb{N}$, on pose $h_n(x) = f(x, y_n)$, $x \in E$. L'hypothèse (a) entraîne que les h_n sont des applications $\mathscr{E}$-mesurables. De plus, l'hypothèse (b) montre que pour μ-presque tout x, on a $\lim_n h_n(x) = h(x)$, où $h(x) = f(x, y)$, $x \in E$. Enfin, on a $|h_n| \leq g$ μ-p.p. pour tout n d'après l'hypothèse (c). Le théorème de convergence dominée implique alors que $\lim_n F(y_n) = \lim_n \int_E h_n \, d\mu = \int_E h \, d\mu = F(y)$. ∎

Le théorème suivant concerne la dérivabilité des intégrales à paramètres.

Théorème 47 *Soit I un intervalle ouvert de $\mathbb{R}$, soit $(E, \mathscr{E}, \mu)$ un espace mesuré, soit $f : E \times I \to \mathbb{R}$. On fait les hypothèses suivantes.*

(a) Pour tout $y \in I$, $f(\cdot, y)$ est $\mathscr{E}$-mesurable et intégrable.

(b) pour μ-p.t. $x \in \mathscr{E}$, $y \mapsto f(x, y)$ est dérivable sur I.

(c) Il existe $g : E \to [0, \infty]$, $\mathscr{E}$-mesurable et μ-intégrable, telle que pour μ-p.t. $x \in \mathscr{E}$, pour tout $y \in I$, on ait $\left|\frac{\partial f}{\partial y}(x, y)\right| \leq g(x)$.

Alors $F(y) = \int_E f(x, y) \, d\mu(x)$ est bien définie, continue et dérivable sur I. De plus, on a $F'(y) = \int_E \frac{\partial f}{\partial y}(x, y)\mu(dx)$ pour tout $y \in I$.

Preuve : on fixe $y_n \in I$, $n \in \mathbb{N}$, une suite de points distincts de y qui converge vers $y \in I$ fixé. Pour tout n, on pose $h_n(x) = \frac{f(x,y_n)-f(x,y)}{y_n-y}$. Par (a), les h_n sont intégrables. Par ailleurs, l'hypothèse (b) implique que pour μ-p.t. $x \in \mathscr{E}$, $\lim_n h_n(x) = \frac{\partial f}{\partial y}(x, y)$. Enfin, l'inégalité des accroissements finis combinée avec l'hypothèse (c) entraîne que pour μ-p.t. $x \in \mathscr{E}$, $|h_n(x)| \leq \sup_{y\in I} \left|\frac{\partial f}{\partial y}(x, y)\right| \leq g(x)$. Le théorème de convergence dominée s'applique donc à la suite des h_n et montre que

$$\lim_{n\to\infty} \frac{F(y_n)-F(y)}{y_n-y} = \lim_{n\to\infty} \int_E h_n \, d\mu = \int_E \frac{\partial f}{\partial y}(x,y) \, d\mu(x) \ ,$$

ce qui implique le résultat désiré. ■

I.9 Espaces $\mathscr{L}^p$ et L^p

Dans toute la section, l'espace mesuré $(E, \mathscr{E}, \mu)$ est fixé. Soit $f : E \to \mathbb{R}$, une application $\mathscr{E}$-mesurable. Pour $p \in [1, \infty[$, on pose

$$\|f\|_p := \Big(\int_E |f|^p \, d\mu \Big)^{\frac{1}{p}} \ .$$

On introduit aussi le *supremum essentiel* de f

$$\|f\|_\infty = \inf \big\{ a \in [0,\infty] \ : \ \mu(\{x \in E : |f(x)| > a\}) = 0 \ \big\}.$$

Il est facile de voir que $|f| \leq \|f\|_\infty$ μ-p.p. (exercice).

Définition 48 *Pour $p \in [1,\infty]$, on note $\mathscr{L}^p(E, \mathscr{E}, \mu)$ l'ensemble des applications $f : E \mapsto \overline{\mathbb{R}}$, $\mathscr{E}$-mesurables, telles que $||f||_p < \infty$.*

On commence par montrer les inégalités de Hölder et de Minkowski. On rappelle que $p, q \in [1, \infty]$ sont conjugués ssi $\frac{1}{p} + \frac{1}{q} = 1$ (avec $1/0 = \infty$ et $1/\infty = 0$).

Proposition 49 (Hölder) *Soient p, q, conjugués et $f, g : E \to [0, \infty]$ des applications $\mathscr{E}$-mesurables. On a alors : $\int_E fg \, d\mu \leq ||f||_p ||g||_q$ avec la convention $0 \times \infty = 0$.*

Preuve : Le cas $p = 1, q = \infty$ (ou le contraire) est facile, on suppose donc que $p, q \in]1, \infty[$. On a $s^{1/p} t^{1/q} \leq s/p + t/q$ pour tout $s, t \geq 0$ (utiliser la concavité du log ou, plus simplement, fixer s et étudier la fonction en t). On pose $c = ||f||_p$ et $d = ||g||_q$. On applique l'inégalité précédente à $s = (f(x)/c)^p$, $t = (g(x)/d)^q$ et on obtient $(cd)^{-1} f(x) g(x) \leq c^{-p} f(x)^p/p + d^{-q} g(x)^q/q$. En intégrant cette inégalité, on obtient alors $(cd)^{-1} \int_E fg \, d\mu \leq 1/p + 1/q = 1$, ce qui entraîne le résultat désiré. ■

Proposition 50 (Minkowski) *Soit $p \in [1, \infty]$ et soient $f, g : E \to [0, \infty]$, deux applications $\mathscr{E}$-mesurables. On a alors : $||f + g||_p \leq ||f||_p + ||g||_p$.*

Preuve : c'est trivial si $p = 1$ et facile si $p = \infty$, on suppose donc que $p \in]1, \infty[$ et on note q son conjugué. Par Hölder, on a

$$\int_E f(f+g)^{p-1} \, d\mu \leq ||f||_p \Big(\int_E (f+g)^{q(p-1)} d\mu \Big)^{\frac{1}{q}} \quad \text{et} \quad \int_E g(f+g)^{p-1} \, d\mu \leq ||g||_p \Big(\int_E (f+g)^{q(p-1)} d\mu \Big)^{\frac{1}{q}}.$$

En additionnant ces deux inégalités et en remarquent que $p = q(p-1)$, on obtient

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

soit encore $||f+g||_p^p \leq (||f||_p + ||g||_p) ||f+g||_p^{p/q}$. Comme $p/q = p - 1$, on conclut aisément. ■

Remarque 51 *Fixons* $p \in [1,\infty]$. *Alors* $\mathscr{L}^p(E,\mathscr{E},\mu)$ *est un* $\mathbb{R}$*-espace vectoriel (bien sûr,* $f,g \in \mathscr{L}^p(E,\mathscr{E},\mu)$ *et* $a \in \mathbb{R}$ *implique que* $f+ag \in \mathscr{L}^p(E,\mathscr{E},\mu)$*). De plus,* $||\cdot||_p$ *vérifie* $||f+g||_p \leq ||f||_p+||g||_p$*,* $||af||_p = |a|||f||_p$*, et* $||0||_p = 0$*. Par contre,* $||f||_p = 0$ *implique que* $f = 0$ μ*-p.p., mais pas que* $f = 0$*. Ce n'est donc pas une "vraie" norme.*

Définition 52 *On introduit la relation* $f \sim g$ *si* $f = g$ μ*-p.p. C'est une relation d'équivalence. On introduit ensuite, pour* $p \in [1,\infty]$*,* $\mathrm{L}^p(E,\mathscr{E},\mu)$ *l'espace quotient* $\mathscr{L}^p(E,\mathscr{E},\mu)/\sim$*. C'est un* $\mathbb{R}$*-e.v. Enfin, pour* $u \in \mathrm{L}^p(E,\mathscr{E},\mu)$*, on définit* $||u||_p = ||f||_p$ *pour n'importe quel* f *représentant du* u *(cette définition ne dépend pas du choix de* f *puisque si* f,g *sont deux représentants de* u*, alors* $f = g$ μ*-p.p. et donc* $||f||_p = ||g||_p$*).*

Remarque 53 *Soit* $p \in [1,\infty]$*. Alors* $||\cdot||_p$ *est une (vraie) norme sur* $\mathrm{L}^p(E,\mathscr{E},\mu)$*.*

Les espaces L^p sont complets. On montre d'abord le lemme suivant.

Lemme 54 *Soit* $p \in [1,\infty]$ *et soit* $f_n \in \mathscr{L}^p(E,\mathscr{E},\mu)$*,* $n \in \mathbb{N}$*, une suite de Cauchy (i.e. pour tout* $\epsilon > 0$*, il existe* $n(\epsilon) \geq 1$ *tel que pour tout* $m,n \geq n(\epsilon)$*,* $||f_n - f_m||_p \leq \epsilon$*). Alors il existe* $f \in \mathscr{L}^p(E,\mathscr{E},\mu)$ *et une suite strictement croissante d'indices* $(n_k, k \in \mathbb{N})$ *tels que* $\lim_k f_{n_k} = f$ μ*-p.p. et* $\lim_k ||f_{n_k} - f||_p = 0$*.*

Preuve : Le cas $p = \infty$ est plus facile, on traite donc le cas où $p \in [1,\infty[$. Pour tout $k \in \mathbb{N}$, on pose $n_k = n(2^{-0}) + \cdots + n(2^{-k})$ qui est strictement croissante, et on a

$$\|f_{n_{k+1}} - f_{n_k}\|_p \leq 2^{-k} , \quad k \in \mathbb{N} .$$

Pour tout $k \in \mathbb{N}$, on pose $g_k = \sum_{0\leq i\leq k} |f_{n_{i+1}} - f_{n_i}|$ et $g = \sum_{i\geq 0} |f_{n_{i+1}} - f_{n_i}|$. Par Minkowski, on montre aisément que $||g||_p \leq \sum_{i\geq 0} ||f_{n_{i+1}} - f_{n_i}||_p \leq 2$ et donc $g < \infty$ μ-p.p., ce qui implique que la série $f_{n_0}(x) + \sum_{i\geq 0}(f_{n_{i+1}}(x) - f_{n_i}(x))$ est absolument convergente, et donc convergente, μ-p.p., on appelle f sa somme (en posant $f = 0$ là où elle n'existe pas). De plus $f_{n_k} = f_{n_0} + \sum_{i=0}^{k}(f_{n_{i+1}} - f_{n_i})$, d'où $\lim_k f_{n_k} = f$ μ-p.p. Enfin, $||f_{n_k} - f||_p = ||\sum_{i\geq k}(f_{n_{i+1}} - f_{n_i})||_p \leq \sum_{i\geq k} ||f_{n_{i+1}} - f_{n_i}||_p \to 0$. ■

Théorème 55 *Pour tout* $p \in [1,\infty]$*,* $\mathrm{L}^p(E,\mathscr{E},\mu)$ *muni de la norme* $\|\cdot\|_p$ *est complet.*

La preuve découle immédiatement du lemme précédent et du suivant.

Lemme 56 *Soit* (Y,d) *un espace métrique. Si* $(x_n)_{n\geq 1}$ *est une suite de Cauchy dans* (Y,d) *et qu'elle admet une sous-suite convergente, alors elle converge.*

Preuve : Soit $(x_{n_k})_{k\geq 1}$ la sous-suite qui converge et soit $x = \lim_k x_{n_k}$. Fixons $\epsilon > 0$. Il existe n_ϵ t.q. pour tout $n,m \geq n_\epsilon$, $d(x_n,x_m) < \epsilon$. En particulier, si $n \geq n_\epsilon$, on a $\limsup_k d(x_n,x_{n_k}) < \epsilon$. Or pour tout $n \geq 1$, on a $d(x_n,x) = \lim_k d(x_n,x_{n_k})$. Donc pour tout $n \geq n_\epsilon$, on a $d(x_n,x) < \epsilon$. ■

I.10 Théorème de Fubini

Dans cette section on fixe deux espaces mesurables $(E_1,\mathscr{E}_1,\mu_1)$ et $(E_2,\mathscr{E}_2,\mu_2)$. On cherche à définir une mesure sur l'espace produit $E_1 \times E_2$ telle que la mesure d'un rectangle $A_1 \times A_2$ (où $A_1 \in \mathscr{E}_1$ et $A_2 \in \mathscr{E}_2$) soit $\mu_1(A_1)\mu_2(A_2)$.

Définition 57 *(Tribu produit) Soient* $(E_1,\mathscr{E}_1)$ *et* $(E_2,\mathscr{E}_2)$*, deux espaces mesurables. On note* $\mathscr{E}_1 \otimes \mathscr{E}_2 = \sigma(\mathscr{P})$ *la tribu sur* $E_1 \times E_2$ *engendrée par* $\mathscr{P} = \{A_1 \times A_2 \ : \ A_1 \in \mathscr{E}_1, A_2 \in \mathscr{E}_2\}$*.*

Lemme 58 *On note π_1 et π_2 les projections canoniques de $E_1 \times E_2$:*

$$\forall (x_1, x_2) \in E_1 \times E_2, \quad \pi_1(x_1, x_2) = x_1 \quad \text{et} \quad \pi_2(x_1, x_2) = x_2 .$$

La tribu produit $\mathscr{E}_1 \otimes \mathscr{E}_2$ est la plus petite tribu $\mathscr{G}$ sur $E_1 \times E_2$ qui rende $(\mathscr{G}, \mathscr{E}_1)$-mesurable la projection π_1 et qui rende $(\mathscr{G}, \mathscr{E}_2)$-mesurable la projection π_2.

Preuve : On note $\mathscr{G}$ la plus petite tribu $E_1 \times E_2$ qui rende π_1 et π_2 mesurables.

Pour tout $A_1 \in \mathscr{E}_1$, on a $\pi_1^{-1}(A_1) = A_1 \times E_2 \in \mathscr{P} \subset \mathscr{E}_1 \otimes \mathscr{E}_2$, donc π_1 est $(\mathscr{E}_1 \otimes \mathscr{E}_2, \mathscr{E}_1)$-mesurable. De même π_2 est $(\mathscr{E}_1 \otimes \mathscr{E}_2, \mathscr{E}_2)$-mesurable. Donc $\mathscr{G} \subset \mathscr{E}_1 \otimes \mathscr{E}_2$.

Pour tout $A_1 \in \mathscr{E}_1$ et tout $A_2 \in \mathscr{E}_2$, on a $A_1 \times A_2 = \pi_1^{-1}(A_1) \cap \pi_2^{-1}(A_2) \in \mathscr{G}$. Donc $\mathscr{P} \subset \mathscr{G}$, et donc $\mathscr{E}_1 \otimes \mathscr{E}_2 = \sigma(\mathscr{P}) \subset \mathscr{G}$. ■

Proposition 59 *(i) Soit $A \in \mathscr{E}_1 \otimes \mathscr{E}_2$. Pour tout $x_1 \in E_1$, $A_{x_1} = \{x_2 \in E_2 \ : (x_1, x_2) \in A\}$ appartient à $\mathscr{E}_2$. Pour tout $x_2 \in E_2$, $A_{x_2} = \{x_1 \in E_1 \ : (x_1, x_2) \in A\}$ appartient à $\mathscr{E}_1$.*

(ii) Si $f : E_1 \times E_2 \mapsto \mathbb{R}$ est $\mathscr{E}_1 \otimes \mathscr{E}_2$-mesurable, alors $x_2 \mapsto f(x_1, x_2)$ est $\mathscr{E}_2$-mesurable pour tout $x_1 \in E_1$ et $x_1 \mapsto f(x_1, x_2)$ est $\mathscr{E}_1$-mesurable pour tout $x_2 \in E_1$.

Preuve : (i) On fixe $x_1 \in E_1$ et on considère $\mathscr{C} = \{A \in \mathscr{E}_1 \otimes \mathscr{E}_2 \ : \ A_{x_1} \in \mathscr{E}_2\}$. On veut montrer que $\mathscr{E}_1 \otimes \mathscr{E}_2 \subset \mathscr{C}$. On montre facilement que $\mathscr{C}$ est une tribu (car $(E \times F)_{x_1} = F$, car $(A^c)_{x_1} = (A_{x_1})^c$, etc.) De plus, pour $A = A_1 \times A_2 \in \mathscr{P}$, on a $A_{x_1} = A_2$ si $x_1 \in A_1$ et $A_{x_1} = \emptyset$ sinon, donc dans tous les cas, $A_{x_1} \in \mathscr{E}_2$. Ainsi, $\mathscr{P} \subset \mathscr{C}$ et donc $\mathscr{E}_1 \otimes \mathscr{E}_2 = \sigma(\mathscr{P}) \subset \mathscr{C}$.

(ii) On fixe $x_1 \in E_1$ et on pose $g(x_2) = f(x_1, x_2)$. Pour tout $B \in \mathscr{B}(\mathbb{R})$, on a $g^{-1}(B) = \{x_2 \in E_2 \ : (x_1, x_2) \in f^{-1}(B)\} = (f^{-1}(B))_{x_1}$. Mais $f^{-1}(B) \in \mathscr{E}_1 \otimes \mathscr{E}_2$ et donc, par (i), $g^{-1}(B) \in \mathscr{E}_2$. ■ .

Théorème 60 **(Théorème de Fubini)** *Soient $(E_1, \mathscr{E}_1, \mu_1)$ et $(E_2, \mathscr{E}_2, \mu_2)$, deux espaces mesurés avec μ_1 et μ_2 σ-finies.*

(i) Il existe une unique mesure μ sur $(E_1 \times E_2, \mathscr{E}_1 \otimes \mathscr{E}_2)$ telle que $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ pour tout $A_1 \in \mathscr{E}_1$ et tout $A_2 \in \mathscr{E}_2$. Cette mesure se note $\mu_1 \otimes \mu_2$ et se nomme mesure produit.

(ii) Soit $f : E_1 \times E_2 \mapsto \overline{\mathbb{R}}_+$, $\mathscr{E}_1 \otimes \mathscr{E}_2$-mesurable. Alors

$$(*) \quad \int_{E_1 \times E_2} f d(\mu_1 \otimes \mu_2) = \int_{E_2} \Big[\int_{E_1} f(x_1, x_2) \mu_1(dx_1) \Big] \mu_2(dx_2) = \int_{E_1} \Big[\int_{E_2} f(x_1, x_2) \mu_2(dx_2) \Big] \mu_1(dx_1).$$

(iii) Soit $f : E_1 \times E_2 \mapsto \overline{\mathbb{R}}$, $\mathscr{E}_1 \otimes \mathscr{E}_2$-mesurable. Alors (a), (b) et (c) sont équivalents et impliquent qu'on a encore $(*)$:
(a) $f \in \mathscr{L}^1(E_1 \times E_2, \mathscr{E}_1 \otimes \mathscr{E}_2, \mu_1 \otimes \mu_2)$,
(b) $x_1 \mapsto \int_{E_2} |f(x_1, x_2)| \mu_2(dx_2) \in \mathscr{L}^1(E_1, \mathscr{E}_1, \mu_1)$,
(c) $x_2 \mapsto \int_{E_1} |f(x_1, x_2)| \mu_1(dx_1) \in \mathscr{L}^1(E_2, \mathscr{E}_2, \mu_2)$.

Preuve : Le point (iii) se déduit immédiatement du point (ii) (appliqué à $|f|$, f_+ et f_-). On suppose d'abord que μ_1 et μ_2 sont finies et on procède en 5 étapes. On étudiera le cas général à l'étape 6.

Etape 1 : On montre que pour tout $A \in \mathscr{E}_1 \otimes \mathscr{E}_2$, $x_1 \mapsto \mu_2(A_{x_1})$ est $\mathscr{E}_1$-mesurable. Pour cela, on introduit $\mathscr{L} = \{A \in \mathscr{E}_1 \otimes \mathscr{E}_2 \text{ t.q. } x_1 \mapsto \mu_2(A_{x_1}) \text{ est } \mathscr{E}_1\text{-mesurable}\}$. On montre facilement que $\mathscr{L}$ est une classe monotone. De plus, elle contient $\mathscr{P}$ (car si $A = A_1 \times A_2 \in \mathscr{P}$, alors $\mu_2(A_{x_1}) = \mathbf{1}_{A_1}(x_1)\mu_2(A_2)$ qui est bien $\mathscr{E}_1$-mesurable). Enfin, $\mathscr{P}$ est stable par intersection (finie). Le théorème de la classe monotone nous dit que $\sigma(\mathscr{P}) \subset \mathscr{L}$, ce qui termine l'étape puisque $\mathscr{E}_1 \otimes \mathscr{E}_2 = \sigma(\mathscr{P})$.

Etape 2 : On introduit alors $\mu : \mathscr{E}_1 \otimes \mathscr{E}_2 \mapsto [0, \infty]$ définie par

$$\mu(A) = \int_{E_1} \mu_2(A_{x_1})\mu_1(dx_1) = \int_{E_1} \Big[\int_{E_2} \mathbf{1}_A(x_1, x_2)\mu_2(dx_2)\Big]\mu_1(dx_1).$$

C'est bien défini grâce à l'étape 1. On vérifie facilement que pour tout $A = A_1 \times A_2 \in \mathscr{P}$, on a $\mu(A) = \mu_1(A_1)\mu_2(A_2)$. De plus μ est une mesure sur $(E_1 \times E_2, \mathscr{E}_1 \otimes \mathscr{E}_2)$, car $\mu(\emptyset) = 0$ et car si on a une famille 2.2.d. $A_n \in \mathscr{E}_1 \otimes \mathscr{E}_2$, alors

$$\mu(\cup_n A_n) = \int_{E_1} \Big[\int_{E_2} \sum_n \mathbf{1}_{A_n}(x_1, x_2)\mu_2(dx_2)\Big]\mu_1(dx_1) = \sum_n \int_{E_1} \Big[\int_{E_2} \mathbf{1}_{A_n}(x_1, x_2)\mu_2(dx_2)\Big]\mu_1(dx_1),$$

qui n'est autre que $\sum_n \mu(A_n)$. On a utilisé deux fois l'interversion somme/intégrale (tout est positif).

Etape 3 : Pour $f : E_1 \times E_2 \mapsto \overline{\mathbb{R}}_+$, $\mathscr{E}_1 \otimes \mathscr{E}_2$-mesurable, on a

$$(**) \qquad \int_{E_1 \times E_2} f d\mu = \int_{E_1} \Big[\int_{E_2} f(x_1, x_2)\mu_2(dx_2)\Big]\mu_1(dx_1).$$

En effet, il suffit d'écrire $f = \sum_n a_n \mathbf{1}_{A_n}$ avec $a_n \geq 0$ et $B_n \in \mathscr{E}_1 \otimes \mathscr{E}_2$ (voir Lemme 27) et d'utiliser l'interversion somme/intégrale pour écrire

$$\int_{E_1 \times E_2} f d\mu = \sum_n a_n \mu(A_n) = \sum_n a_n \int_{E_1} \Big[\int_{E_2} \mathbf{1}_{A_n}(x_1, x_2)\mu_2(dx_2)\Big]\mu_1(dx_1)$$

par définition de μ. On réutilise alors l'interversion somme/intégrale pour conclure.

Etape 4 : On montre ici qu'il existe au plus une mesure μ sur $(E_1 \times E_2, \mathscr{E}_1 \otimes \mathscr{E}_2)$ telle que $\mu(A) = \mu_1(A_1)\mu_2(A_2)$ pour tout $A = A_1 \times A_2 \in \mathscr{P}$: cela découle directement du théorème d'unicité de prolongement des mesures, car si on a deux telles mesures μ et ν, on a $\mu(E_1 \times E_2) = \nu(E_1 \times E_2) < \infty$, on a $\mu(A) = \nu(A)$ pour tout $A \in \mathscr{P}$, et car $\mathscr{P}$ est un pi-système qui engendre $\mathscr{E}_1 \otimes \mathscr{E}_2$.

Etape 5 : Nous avons donc montré (i) et la moitié des égalités dans le point (ii). L'autre moitié se démontre de la même façon, en reprenant la preuve depuis le début (étapes 1 à 4) en échangeant les rôles de E_1 et E_2 et en utilisant l'unicité prouvée à l'étape 5 (on obtient la même mesure μ).

Etape 6 : On ne suppose plus que μ_1 et μ_2 sont finies. Comme elles sont σ-finies, on écrit $E_1 = \cup_p E_1^p$ et $E_2 = \cup_q E_2^q$, avec les E_1^p 2.2.d., éléments de $\mathscr{E}_1$ et pour lesquels $\mu_1(E_1^p) < \infty$, et avec les E_2^q 2.2.d., éléments de $\mathscr{E}_2$ et pour lesquels $\mu_2(E_2^q) < \infty$. On introduit les mesures $\mu_1^p(A_1) = \mu_1(A_1 \cap E_1^p)$ sur $(E_1, \mathscr{E}_1)$ et $\mu_2^q(A_2) = \mu_2(A_2 \cap E_2^q)$ sur $(E_2, \mathscr{E}_2)$.

On montre alors facilement que $\mu = \sum_{p,q}(\mu_{1,p} \otimes \mu_{2,q})$ vérifie toutes les conclusions du théorème.

Pour l'unicité, considérons deux mesures μ, ν sur $(E_1 \times E_2, \mathscr{E}_1 \otimes \mathscr{E}_2)$ telle que $\mu(A) = \mu_1(A_1)\mu_2(A_2) = \nu(A)$ pour tout $A = A_1 \times A_2 \in \mathscr{P}$. Posons $\mu_{p,q}(A) = \mu(A \cap (E_1^p \times E_2^q))$ et définissons $\nu_{p,q}$ de la même façon. En appliquant le théorème déjà montré dans le cas fini, on voit que $\mu_{p,q} = \nu_{p,q}$ pour tout p, q. On conclut facilement que $\mu = \sum_{p,q} \mu_{p,q} = \sum_{p,q} \nu_{p,q} = \nu$. ■

Remarque 61 *Si on a trois (ou plus) espaces mesurés $(E_1, \mathscr{E}_1, \mu_1)$, $(E_2, \mathscr{E}_2, \mu_2)$ et $(E_3, \mathscr{E}_3, \mu_3)$ avec μ_1, μ_2, μ_3 σ-finis, alors on identifie $(E_1 \times E_2) \times E_3$ et $E_1 \times (E_2 \times E_3)$ et on montre facilement que (à* *identification près)* $(\mathscr{E}_1 \otimes \mathscr{E}_2) \otimes \mathscr{E}_3 = \mathscr{E}_1 \otimes (\mathscr{E}_2 \otimes \mathscr{E}_3)$ *et* $(\mu_1 \otimes \mu_2) \otimes \mu_3 = \mu_1 \otimes (\mu_2 \otimes \mu_3)$*, ce qui permet de justifier la notation suivante :*

$$(E_1 \times E_2 \times E_3\,,\, \mathscr{E}_1 \otimes \mathscr{E}_2 \otimes \mathscr{E}_3\,,\, \mu_1 \otimes \mu_2 \otimes \mu_3).$$

Remarque 62 *(i) On peut vérifier que* $\mathscr{B}(\mathbb{R}^d) = \mathscr{B}(\mathbb{R})^{\otimes d}$.

(ii) Si la mesure de Lebesgue ℓ *sur* $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ *existe, la mesure de Lebesgue* ℓ_d *sur* $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}))$ *existe aussi.*

Le point (i) est à vérifier soigneusement. Pour (ii), il suffit de poser $\ell_d = \ell \otimes \cdots \otimes \ell$, qui est bien une mesure sur $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ vérifiant $\ell_d(A_1 \times \cdots \times A_d) = \ell(A_1) \times \cdots \times \ell(A_d)$ pour tout $A_1, \dots, A_d \in \mathscr{B}(\mathbb{R})$. Donc en particulier, pour tout pavé $R =]a_1, b_1] \times \ldots \times]a_d, b_d]$, on a bien $\ell_d(R) = (b_1 - a_1) \times \ldots \times (b_d - a_d)$ (puisque $\ell(]a, b]) = b - a$).

I.11 Espaces de Hilbert

Nous allons avoir besoin des résultats classiques suivants, appliqués à l'espace $\mathrm{L}^2(E, \mathscr{E}, \mu)$, qui est un espace de Hilbert : nous avons vu que muni de la norme $||\cdot||_2$, c'est un espace vectoriel normé complet. On vérifie facilement que $\langle f, g\rangle = \int_E fgd\mu$ (bien défini pour $f, g \in \mathrm{L}^2(E, \mathscr{E}, \mu)$ par Cauchy-Schwarz) est un produit scalaire, qui de plus vérifie $\langle f, f\rangle = ||f||_2^2$. Le théorème de projection suivant est ultra classique.

Théorème 63 *Soit* H *un espace de Hilbert et* $C \subset H$ *un s.e.v. fermé. Pour tout* $x \in H$*, il existe un unique* $v = \pi_C(x)$*, élément de* C*, tel que* $||x - v|| = \min\{||x - w|| \,:\, w \in C\}$*. De plus* $v = \pi_C(x)$ *est caractérisé par (a)* $v \in C$ *et (b) pour tout* $w \in C$*,* $\langle x - v, w\rangle = 0$.

Preuve : On fixe $x \in H$ et on pose $\alpha = \inf\{||x - w|| \,:\, w \in C\}$.

(i) Existence de $\pi_C(x)$. Pour tout $n \geq 1$, on peut trouver $v_n \in C$ t.q. $||x - v_n||^2 \leq \alpha^2 + 1/n$. On vérifie par le calcul (développer tous les termes)

$$||v_n - v_m||^2 = 2\Big(||x - v_n||^2 + ||x - v_m||^2 - 2||x - (v_n + v_m)/2||^2\Big).$$

Mais comme $(v_n + v_m)/2 \in C$, on a $||x - (v_n + v_m)/2||^2 \geq \alpha^2$. Ainsi, $||v_n - v_m||^2 \leq 2/n + 2/m$, et la suite v_n est de Cauchy. Elle converge donc vers un certain $v \in H$, qui de plus appartient à C puisque C est fermé. Et bien sûr, on a $||x - v|| = \lim_n ||x - v_n|| \leq \alpha$ et donc (par définition de α) $||x - v|| = \alpha$.

(ii) Unicité. Si $||x - v|| = ||x - v'|| = \alpha$ avec $v, v' \in C$, on écrit

$$||v - v'||^2 = 2\Big(||x - v||^2 + ||x - v'||^2 - 2||x - (v + v')/2||^2\Big).$$

Comme de plus $||x - (v + v')/2|| \geq \alpha$, on conclut que $||v - v'|| \leq 0$, donc $v = v'$.

(iii) Caractérisation, sens 1. Soit $v = \pi_C(x)$. Alors $v \in C$ bien sûr. De plus, pour $w \in C$ et $t \in \mathbb{R}$, on a $v - tw \in C$ et donc $\alpha^2 \leq ||x - (v - tw)||^2 = ||x - v||^2 + t^2||w||^2 + 2t\langle x - v, w\rangle$. Comme $||x - v|| = \alpha$, on conclut que $2t\langle x - v, w\rangle + t^2||w||^2 \geq 0$ pour tout t. Donc pour tout $t > 0$, $2\langle x - v, w\rangle + t||w||^2 \geq 0$, et on conclut que $\langle x - v, w\rangle \geq 0$ (en faisant $t \to 0$). Pour tout $t < 0$, $2\langle x - v, w\rangle + t||w||^2 \leq 0$, et on conclut que $\langle x - v, w\rangle \leq 0$ (en faisant $t \to 0$).

(iv) Caractérisation sens 2. Soit $v \in C$ t.q. $\langle x-v, w\rangle = 0$ pour tout $w \in C$. Alors pour tout $w \in C$, $||w-x||^2 = ||w-v+v-x||^2 = ||x-v||^2 + ||v-w||^2$ car $\langle w-v, v-x\rangle = 0$ (puisque $w-v \in C$). Donc pour tout $w \in C$, $||w-x||^2 \geq ||x-v||^2$. ■

Théorème 64 *Soit H un espace de Hilbert et $F : H \mapsto \mathbb{R}$ une forme linéaire continue. Alors il existe $z_0 \in H$ t.q. pour tout $x \in H$, on ait $F(x) = \langle x, z_0\rangle$.*

Preuve : Si Ker $F = H$, on peut choisir $z_0 = 0$. On suppose donc que $C =$Ker $F \neq H$. On considère $x_0 \in H \setminus C$, on pose $v_0 = \pi_C(x_0)$ et $w_0 = x_0 - v_0$ (C est un s.e.v. fermé car F est continue). On vérifie que $F(w_0) = F(x_0) - F(v_0) = F(x_0) \neq 0$ et donc aussi $w_0 \neq 0$. Pour tout $x \in H$, on a $x - [F(x)/F(w_0)]w_0 \in C$ (car F est linéaire). Donc (comme $w_0 = x_0 - \pi_C(x_0)$ est orthogonal à C),

$$\left\langle w_0, x - \frac{F(x)}{F(w_0)} w_0 \right\rangle = 0, \quad \text{i.e.} \quad \frac{||w_0||^2}{F(w_0)} F(x) = \langle w_0, x\rangle.$$

On conclut avec $z_0 = [F(w_0)/||w_0||^2]w_0$. ■

I.12 Théorème de Radon-Nikodym

Définition 65 *Soit μ et ν deux mesures sur un espace mesurable $(E, \mathscr{E})$. On dit que*

(a) μ est absolument continue par rapport à ν (ou $\mu << \nu$) si pour tout $A \in \mathscr{E}$, $\nu(A) = 0$ implique $\mu(A) = 0$;

(b) μ admet une densité f par rapport à ν (ou $d\mu = f d\nu$) si $f : E \mapsto \mathbb{R}_+$ est $\mathscr{E}$-mesurable et si pour tout $A \in \mathscr{E}$, $\mu(A) = \int_A f d\nu$.

Proposition 66 *Si $d\mu = f d\nu$, alors pour tout $g : E \mapsto \mathbb{R}_+$, $\mathscr{E}$-mesurable, on a $\int_E g d\mu = \int_E g f d\nu$. Si maintenant $g : E \mapsto \mathbb{R}$ est $\mathscr{E}$-mesurable, on a $g \in \mathscr{L}^1(E, \mathscr{E}, \mu)$ ssi $fg \in \mathscr{L}^1(E, \mathscr{E}, \nu)$ et dans ce cas, on a aussi $\int_E g d\mu = \int_E g f d\nu$.*

Preuve : Comme d'habitude, il suffit de traiter le cas positif. Soit donc $g : E \mapsto \mathbb{R}_+$, $\mathscr{E}$-mesurable, qu'on écrit $g = \sum_n a_n \mathbf{1}_{A_n}$ avec $a_n \geq 0$ et $A_n \in \mathscr{E}$, voir le lemme 27. On a alors

$$\int_E g d\mu = \sum_n a_n \mu(A_n) = \sum_n a_n \int_E \mathbf{1}_{A_n} f d\nu = \int_E g f d\nu,$$

On a utilisé plusieurs fois l'interversion somme/intégrale (tout est positif). ■

Lemme 67 *Soit $(E, \mathscr{E}, \mu)$ un espace mesuré. Alors (a) μ est σ-finie si et seulement si (b) il existe une probabilité ν sur $(E, \mathscr{E})$ par rapport à laquelle μ admet une densité f.*

Preuve : Pour (b) implique (a), on introduit $E_n = \{f \leq n\}$. Bien sûr, $\cup_n E_n = E$ et $\mu(E_n) = \int_{E_n} f d\nu \leq n\nu(E_n) = n$.

Si μ est σ-finie, on considère une famille $A_n \in \mathscr{E}$, 2.2.d., telle que $\mu(A_n) < \infty$ et $\cup_{n\geq 1} A_n = E$. On introduit la mesure $\nu(B) = \sum_{n\geq 1} [2^n \mu(A_n)]^{-1} \mu(A_n \cap B)$ et la fonction $f = \sum_{n\geq 1} 2^n \mu(A_n) \mathbf{1}_{A_n}$. On vérifie que ν est une probabilité (c'est une mesure par la proposition 17-(ii)-(iii) et on a $\nu(E) = \sum_{n\geq 1} 2^{-n} = 1$). Et on a bien $d\mu = f d\nu$, car pour $B \in \mathscr{E}$, comme les A_n forment une partition de E,

$$\int_B f d\nu = \sum_{n\geq 1} \int_{B\cap A_n} f d\nu = \sum_{n\geq 1} \int_{B\cap A_n} 2^n \mu(A_n) d\nu = \sum_{n\geq 1} 2^n \mu(A_n) [2^n \mu(A_n)]^{-1} \mu(A_n \cap B) = \sum_{n\geq 1} \mu(A_n \cap B),$$

qui vaut bien $\mu(B)$. ■

Le résultat suivant est important.

Théorème 68 **(Radon-Nikodym)** *Soit $(E, \mathscr{E})$ un espace mesurable. Soient μ et ν, deux mesures σ-finies sur $\mathscr{E}$. Alors $\mu << \nu$ si et seulement si μ a une densité f par rapport à ν.*

Preuve : Si μ a une densité f par rapport à ν, alors pour tout $A \in \mathscr{E}$, si $\nu(A) = 0$, alors $\mu(A) = \int_A f d\nu = 0$, voir proposition 32-(iii). L'autre sens est plus délicat. En utilisant le lemme précédent, on peut se ramener au cas où μ et ν sont des probabilités (exercice).

On pose $\lambda = \mu + \nu$ et, pour $\varphi \in \mathscr{L}^2(E, \mathscr{E}, \lambda)$, $F(\varphi) = \int_E \varphi d\mu$. Ce F est une forme linéaire sur $\mathscr{L}^2(E, \mathscr{E}, \lambda)$. Elle est de plus continue, car pour $\varphi, \varphi' \in \mathscr{L}^2(E, \mathscr{E}, \lambda)$, on a $|F(\varphi) - F(\varphi')| \leq \int_E |\varphi - \varphi'| d\mu \leq ||\varphi - \varphi'||_{\mathscr{L}^2(E,\mathscr{E},\mu)} \leq ||\varphi - \varphi'||_{\mathscr{L}^2(E,\mathscr{E},\lambda)}$. Par le théorème 64, il existe donc $h \in \mathscr{L}^2(E, \mathscr{E}, \lambda)$ tel que pour tout $\varphi \in \mathscr{L}^2(E, \mathscr{E}, \lambda)$, $F(\varphi) = \int_E h\varphi d\lambda$, ce qui se réécrit $\int_E \varphi d\mu = \int_E \varphi h d\lambda$, soit encore

$$(*) \qquad \int_E \varphi(1-h) d\mu = \int_E \varphi h d\nu.$$

(Pour les puristes, le paragraphe ci-dessus doit être traité dans $\mathrm{L}^2(E, \mathscr{E}, \lambda)$).

Montrons maintenant que $h \in [0, 1[$ λ-p.p. Comme $\lambda = \mu + \nu$ et comme $\mu << \nu$, il suffit de voir que $h \in [0, 1[$ ν-p.p. En appliquant $(*)$ avec $\varphi = \mathbf{1}_{\{h \geq 1\}}$, on obtient

$$\nu(\{h \geq 1\}) \leq \int_E h \mathbf{1}_{\{h\geq 1\}} d\nu = \int_E (1-h)\mathbf{1}_{\{h \geq 1\}} d\mu \leq 0,$$

donc $h < 1$ ν-p.p. En appliquant $(*)$ avec $\varphi = \mathbf{1}_{\{h<0\}}$, on obtient

$$0 \leq \int_E \mathbf{1}_{\{h<0\}}(1-h) d\mu = \int_E \mathbf{1}_{\{h<0\}} h d\nu \leq 0,$$

donc $\int_E \mathbf{1}_{\{h<0\}} h d\nu = 0$, ce qui n'est possible que si $h \geq 0$ ν-p.p.

On pose maintenant $f = h/(1-h)$, qui est bien $\mathscr{E}$-mesurable et à valeurs dans $\mathbb{R}_+$ (quitte à la modifier sur un ν-négligeable). Reste à verifier que $d\mu = f d\nu$. Soit donc $A \in \mathscr{E}$. En appliquant $(*)$ avec $\varphi = \mathbf{1}_A/[1 - h + 1/n]$ (qui est bornée et donc dans $\mathscr{L}^2(E, \mathscr{E}, \lambda)$), on trouve

$$\int_A \frac{1-h}{1-h+1/n} d\mu = \int_A \frac{h}{1-h+1/n} d\nu.$$

En faisant tendre n vers l'infini (par convergence monotone), on conclut que $\mu(A) = \int_A f d\nu$. ■

I.13 Construction de la mesure de Lebesgue

On veut construire la mesure de Lebesgue ℓ sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. C'est difficile et assez peu intuitif. Nous allons montrer péniblement les trois propositions suivantes.

Proposition 69 *Soit E un ensemble et $\ell : \mathscr{P}(E) \mapsto [0, \infty]$ vérifiant*

(a) $\ell(\emptyset) = 0$;
(b) Si $A, B \in \mathscr{P}(E)$ et si $A \subset B$, alors $\ell(A) \leq \ell(B)$;

(c) $\ell(\cup_n A_n) \leq \sum_n \ell(A_n)$ *pour toute famille dénombrable de parties de* E.

Alors $\mathscr{L} = \{B \in \mathscr{P}(E) \ : \ \ell(A) = \ell(A \cap B^c) + \ell(A \cap B) \ \forall\, A \in \mathscr{P}(E)\}$ *est une tribu et* ℓ*, restreinte à* $\mathscr{L}$*, est une mesure sur* $(E, \mathscr{L})$.

Une application vérifiant (a), (b), (c) s'appelle une mesure extérieure.

Proposition 70 *On définit* $\ell : \mathscr{P}(\mathbb{R}) \mapsto [0, \infty]$ *par*

$$\ell(A) = \inf \Big\{ \sum_{i \geq 1} (b_i - a_i) \ : \ a_i \leq b_i \ \forall\, i \geq 1 \text{ et } A \subset \cup_{i \geq 1}]a_i, b_i[\Big\},$$

avec la convention que $]a, a[= \emptyset$*. Alors* ℓ *vérifie les trois points (a),(b) et (c) de la proposition 69 et la tribu* $\mathscr{L}$ *en résultant contient* $\mathscr{B}(\mathbb{R})$.

Proposition 71 *Le* ℓ *de la proposition 70 vérifie* $\ell([a, b]) = b - a$ *pour tout* $a < b$.

Avec ces trois propositions, on a construit une mesure ℓ sur $(\mathbb{R}, \mathscr{L})$, avec $\mathscr{B}(\mathbb{R}) \subset \mathscr{L}$ et $\ell(]a, b]) = b - a$: on a construit la mesure de Lebesgue sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

Preuve de la proposition 69 : (i) On montre facilement que pour tout $A, B \in \mathscr{P}(E)$, on a $\ell(A) \leq \ell(A \cap B^c) + \ell(A \cap B)$: il suffit d'utiliser (c) avec $A_1 = A \cap B^c$, $A_2 = A \cap B$, $A_k = \emptyset$ pour $k \geq 3$, et (a).

(ii) Il est évident que $B \in \mathscr{L}$ implique que $B^c \in \mathscr{L}$, et que $E, \emptyset \in \mathscr{L}$.

(iii) On montre ici que $B_1, B_2 \in \mathscr{L}$ implique que $B_1 \cup B_2 \in \mathscr{L}$. Soit donc $A \in \mathscr{P}(E)$. Comme $B_1 \in \mathscr{L}$ on a $\ell(A \cap (B_1 \cup B_2)) = \ell([A \cap (B_1 \cup B_2)] \cap B_1^c) + \ell([A \cap (B_1 \cup B_2)] \cap B_1)$, ce qui se réécrit

$$(*1) \qquad \ell(A \cap (B_1 \cup B_2)) = \ell(A \cap B_2 \cap B_1^c) + \ell(A \cap B_1).$$

Comme $B_2 \in \mathscr{L}$, on a $\ell(A \cap B_1^c) = \ell([A \cap B_1^c] \cap B_2) + \ell([A \cap B_1^c] \cap B_2^c)$, soit encore

$$(*2) \qquad \ell(A \cap B_1^c) = \ell(A \cap B_1^c \cap B_2) + \ell(A \cap (B_1 \cup B_2)^c).$$

Enfin, comme $B_1 \in \mathscr{L}$, on a

$$(*3) \qquad \ell(A \cap B_1^c) + \ell(A \cap B_1) = \ell(A).$$

En combinant $(*1)$, $(*2)$ et $(*3)$, on arrive à montrer que $\ell(A) = \ell(A \cap (B_1 \cup B_2)^c) + \ell(A \cap (B_1 \cup B_2))$, ce qui achève l'étape.

(iv) On montre ici que si $B_1, \ldots, B_n \in \mathscr{L}$ sont 2.2.d, alors $\cup_1^n B_i \in \mathscr{L}$ (on le sait déjà par (iii)) et pour tout $A \in \mathscr{P}(E)$,

$$\ell(A) = \sum_1^n \ell(A \cap B_i) + \ell(A \cap (\cap_1^n B_i^c)).$$

Comme $\cup_1^n B_i \in \mathscr{L}$, on a $\ell(A) = \ell(A \cap (\cup_1^n B_i)) + \ell(A \cap (\cap_1^n B_i^c))$, et il suffit de montrer que $\ell(A \cap (\cup_1^n B_i)) = \sum_1^n \ell(A \cap B_i)$. Comme $B_1 \in \mathscr{L}$, on a $\ell(A \cap (\cup_1^n B_i)) = \ell(A \cap (\cup_1^n B_i) \cap B_1^c) + \ell(A \cap (\cup_1^n B_i) \cap B_1)$, ce qui se réécrit $\ell(A \cap (\cup_1^n B_i)) = \ell(A \cap (\cup_2^n B_i)) + \ell(A \cap B_1)$. On conclut en itérant ce procédé.

(v) On montre ici que si $B_1, B_2, \cdots \in \mathscr{L}$ sont 2.2.d, alors $\cup_1^\infty B_i \in \mathscr{L}$ et $\ell(\cup_1^\infty B_i) = \sum_1^\infty \ell(B_i)$. Pour tout $A \in \mathscr{P}(E)$ et tout n, par (iv) (et (b)), on a

$$\ell(A) = \sum_1^n \ell(A \cap B_i) + \ell(A \cap (\cap_1^n B_i^c)) \geq \sum_1^n \ell(A \cap B_i) + \ell(A \cap (\cap_1^\infty B_i^c)).$$

En faisant tendre n vers l'infini, on conclut que (en utilisant aussi (c), car $A \cap (\cup_1^\infty B_i) = \cup_1^\infty (A \cap B_i)$)

$$\ell(A) \geq \sum_1^\infty \ell(A \cap B_i) + \ell(A \cap (\cap_1^\infty B_i^c)) \geq \ell(A \cap (\cup_1^\infty B_i)) + \ell(A \cap (\cup_1^\infty B_i)^c).$$

Mais en se rappelant le point (i), on conclut que

$$\ell(A) = \sum_1^\infty \ell(A \cap B_i) + \ell(A \cap (\cap_1^\infty B_i^c)) = \ell(A \cap (\cup_1^\infty B_i)) + \ell(A \cap (\cup_1^\infty B_i)^c).$$

Donc $\cup_1^\infty B_i \in \mathscr{L}$. Et la première égalité avec $A = \cup_1^\infty B_i$ donne $\ell(\cup_1^\infty B_i) = \sum_1^\infty \ell(B_i)$.

(vi) Il ne reste qu'à montrer que $\mathscr{L}$ est stable par union dénombrable (non nécessairement 2.2.d.) Cela découle facilement de la stabilité par union finie (iii) et par union dénombrable disjointe (v). ■

Preuve de la proposition 69 : (a) est évident, puisque $\emptyset \subset \cup_i]a_i, b_i[$ avec par exemple $a_i = b_i = 0$, ce qui donne $\sum_i (b_i - a_i) = 0$.

(b) est évident : si $A \subset B$, $\ell(B)$ est défini par un infimum d'un ensemble plus grand que celui qui défini A, donc $\ell(B) \leq \ell(A)$.

(c) Soit donc $A_1, A_2, \dots$ des parties de $\mathbb{R}$. On veut montrer que $\ell(\cup_n A_n) \leq \sum_n \ell(A_n)$. On suppose tous les $\ell(A_n) < \infty$, sinon c'est évident. Fixons $\epsilon > 0$. Pour tout n, on peut trouver des $a_i^n \leq b_i^n$ tels que $A_n \subset \cup_i]a_i^n, b_i^n[$ et tels que $\sum_i (b_i^n - a_i^n) \leq \ell(A_n) + \epsilon 2^{-n}$. Du coup, $\cup_n A_n$ est inclus dans $\cup_{i,n}]a_i^n, b_i^n[$, et donc $\ell(\cup_n A_n) \leq \sum_{i,n} (b_i^n - a_i^n) \leq \sum_n (\ell(A_n) + \epsilon 2^{-n}) = \sum_n \ell(A_n) + \epsilon$.

Reste à voir que $\mathscr{L}$ contient $\mathscr{B}(\mathbb{R})$. Comme $\mathscr{L}$ est une tribu, il suffit de montrer que $]-\infty, x] \in \mathscr{L}$ pour tout $x \in \mathbb{R}$. On veut donc montrer (on se rappelle le (i) de la preuve de la proposition 69) que pour tout $A \subset \mathbb{R}$, $\ell(A) \geq \ell(A \cap]-\infty, x]) + \ell(A \cap]x, +\infty[)$. Par définition de ℓ, il suffit de montrer que si $A \subset \cup_i]a_i, b_i[$, alors $\sum_i (b_i - a_i) \geq \ell(A \cap]-\infty, x]) + \ell(A \cap]x, +\infty[)$.

• Comme $A \cap]x, +\infty[\subset (\cup_i]a_i, b_i[) \cap]x, \infty[= \cup_i]a_i \vee x, b_i \vee x[$, on a $\ell(A \cap]x, +\infty[) \leq \sum_i (b_i \vee x - a_i \vee x)$.

• Pour $\epsilon > 0$, comme $A \cap]-\infty, x] \subset]x - \epsilon, x + \epsilon[\cup [(\cup_i]a_i, b_i[) \cap]-\infty, x[]$, ce qui se réécrit $A \subset]x - \epsilon, x + \epsilon[\cup (\cup_i]a_i \wedge x, b_i \wedge x[)$, on a $\ell(A \cap]-\infty, x]) \leq 2\epsilon + \sum_i (b_i \wedge x - a_i \wedge x)$.

Ainsi, $\ell(A \cap]-\infty, x]) + \ell(A \cap]x, +\infty[) \leq 2\epsilon + \sum_i (b_i \wedge x - a_i \wedge x + b_i \vee x - a_i \vee x) = 2\epsilon + \sum_i (b_i - a_i)$. Comme $\epsilon > 0$ est arbitraire, on conclut qu'en effet, $\ell(A \cap]-\infty, x]) + \ell(A \cap]x, +\infty[) \leq \sum_i (b_i - a_i)$. ■

Preuve de la proposition 71. Déjà, $\ell([a, b]) \leq b - a$: en effet, $[a, b] \subset \cup_i]a_i, b_i[$ avec $a_1 = a - \epsilon$, $b_1 = b + \epsilon$ et $a_i = b_i = 0$ pour tout $i \geq 2$, donc $\ell(]a, b]) \leq \sum_i (b_i - a_i) = b - a + 2\epsilon$.

Reste à voir que $\ell([a, b]) \geq (b - a)$. Il faut donc montrer que si $[a, b] \subset \cup_i]a_i, b_i[$, alors $\sum_i (b_i - a_i) \geq b - a$. Comme $[a, b]$ est compact et que $\cup_i]a_i, b_i[$ en est un recouvrement par des ouverts, on peut en extraire un recouvrement fini : on a donc $[a, b] \subset \cup_1^N]a_i, b_i[$, et on doit démontrer que $\sum_1^N (b_i - a_i) \geq b - a$. Ceci peut se faire par récurrence sur N (c'est évident si $N = 1$, très facile si $N = 2$), c'est un peu fastidieux, sans plus. ■

I.14 Une partie réelle non borélienne

Proposition 72 *(i) Il existe une partie $A \subset \mathbb{R}$ qui n'appartient pas à $\mathscr{B}(\mathbb{R})$.*

(ii) Il est impossible de construire une mesure μ sur $(\mathbb{R}, \mathscr{P}(\mathbb{R}))$, invariante par translation, telle que $\mu([0,2])$ soit non trivial (i.e. ni nul, ni infini).

Par *invariante par translation*, on veut dire que si $A \subset \mathscr{P}(\mathbb{R})$ et si $a \in \mathbb{R}$, alors $\mu(A_a) = \mu(A)$, où $A_a = \{x + a \ : \ x \in A\}$. Comme on souhaite que la mesure de Lebesgue soit invariante par translation, le point (ii) dit en gros qu'on ne peut pas la construire sur toutes les parties de $\mathbb{R}$: on est obligé de considérer la tribu borélienne. Enfin, notons que la mesure nulle peut bien sûr se construire sur $(\mathbb{R}, \mathscr{P}(\mathbb{R}))$ et est invariante par translation (mais pas très intéressante), ainsi que la mesure de comptage ($\mu(A)$ = cardinal de A), qui n'est pas très intéressante non plus puisque $\mu([0,1]) = \infty$.

Preuve : Soit la relation d'équivalence sur $\mathbb{R}$ définie par $x \sim y$ ssi $x - y \in \mathbb{Q}$. Pour chaque classe d'équivalence, on considère un représentant (axiôme du choix non dénombrable), qu'on peut choisir dans $[0,1[$ (car $x \sim x - \lfloor x \rfloor$). On note A l'ensemble de ces représentants.

On a alors $\mathbb{R} = \cup_{q\in\mathbb{Q}} A_q$ et cette union est disjointe (c'est précisément la partition de $\mathbb{R}$ en classes d'équivalence). On a aussi bien sûr $\cup_{q\in\mathbb{Q}\cap[0,1]} A_q \subset [0,2]$.

Soit μ une mesure sur $(\mathbb{R}, \mathscr{P}(\mathbb{R}))$ invariante par translation. Alors, par σ-additivité, $\mu(\mathbb{R}) = \sum_{q\in\mathbb{Q}} \mu(A_q) = \sum_{q\in\mathbb{Q}} \mu(A)$. Donc si $\mu(A) = 0$, alors $\mu(\mathbb{R}) = 0$ et donc $\mu([0,2]) = 0$. D'autre part, $\mu([0,2]) \geq \sum_{q\in\mathbb{Q}\in[0,1]} \mu(A_q) = \sum_{q\in\mathbb{Q}\in[0,1]} \mu(A)$. Donc si $\mu(A) > 0$, alors $\mu([0,2]) = \infty$. Ceci montre le point (ii).

Enfin, A ne peut être borélien, sinon, sa mesure de Lebesgue existait, et on aurait $\ell(A_q) = \ell(A)$ pour tout $q \in \mathbb{Q}$, et on pourrait conclure comme ci-dessus que $\ell([0,2])$ est soit nul, soit infini, ce qui est bien sûr faux. ■

I.15 Complétion d'espaces mesurés

Il est agréable que tous les négligeables soient mesurables. Si ce n'est pas le cas, voici une manière d'y remédier.

Théorème 73 *Soit $(E, \mathscr{E}, \mu)$, un espace mesuré et $\mathscr{N}_\mu$ l'ensemble des μ-négligeables. On définit $\mathscr{E}^\mu = \sigma(\mathscr{E}, \mathscr{N}_\mu)$. Il existe une unique mesure $\bar{\mu}$ sur $(E, \mathscr{E}^\mu)$ telle que $\bar{\mu}(A) = \mu(A)$ pour tout $A \in \mathscr{E}$. De plus,*

(i) on a $\mathscr{N}_{\bar{\mu}} = \mathscr{N}_\mu \subset \mathscr{E}^\mu$,

(ii) pour $B \in \mathscr{E}^\mu$, il existe $A_1, A_2 \in \mathscr{E}$ t.q. $A_1 \subset B \subset A_2$ et $\mu(A_1) = \bar{\mu}(B) = \mu(A_2)$,

(iii) pour $f : E \mapsto \mathbb{R}_+$, $\mathscr{E}^\mu$-mesurable, il existe $h_1, h_2 : E \mapsto \mathbb{R}_+$, $\mathscr{E}$-mesurables, t.q. $h_1 \leq f \leq h_2$, t.q. $h_1 = h_2$ μ-p.p. et t.q. $\int_E h_1 d\mu = \int_E f d\bar{\mu} = \int_E h_2 d\mu$.

Preuve : (a) On montre que pour tout $B \in \mathscr{E}^\mu$, il existe $A \in \mathscr{E}$ t.q. $A \subset B$ et $B \setminus A \in \mathscr{N}_\mu$. Pour cela, on pose $\mathscr{F} = \{B \subset E \ : \ \exists A \in \mathscr{E}$ tel que $A \subset B$ et $B \setminus A \in \mathscr{N}_\mu\}$. On a bien sûr $\mathscr{E} \subset \mathscr{F}$ et $\mathscr{N}_\mu \subset \mathscr{F}$. Il suffit donc de montrer que $\mathscr{F}$ est une tribu. La seule difficulté est de montrer que si $B \in \mathscr{F}$, alors $B^c \in \mathscr{F}$. On écrit $B = A \cup N$, avec $A \in \mathscr{E}$ et $N = B \setminus A \in \mathscr{N}_\mu$. Il existe $N_0 \in \mathscr{N}_\mu \cap \mathscr{E}$ tel que $N \subset N_0$. On écrit alors que $B^c = A^c \cap N^c = (A^c \cap N^c \cap N_0^c) \cup (A^c \cap N^c \cap N_0)$, ce qui montre que $B^c \in \mathscr{F}$, car l'union est disjointe, car $A^c \cap N^c \cap N_0^c = A^c \cap N_0^c \in \mathscr{N}$, et car bien sûr $A^c \cap N^c \cap N_0 \in \mathscr{N}_\mu$ (puisqu'inclus dans N_0).

(b) Pour $B \in \mathscr{E}^\mu$, on écrit $B = A \cup N$ avec $A \in \mathscr{E}$ et $N \in \mathscr{N}_\mu$ (voir (a)) et on pose $\bar{\mu}(B) = \mu(A)$. Cette définition ne dépend pas de la décomposition, car si $A \cup N = A' \cup N'$ avec $A, A' \in \mathscr{E}$ et

$N, N' \in \mathscr{N}_\mu$, on a bien sûr $\mu(A) = \mu(A')$ (il suffit par symétrie de montrer que $\mu(A) = \mu(A \cap A')$; pour cela on écrit $A = (A \cap A') \cup (A \cap (A')^c)$ et on observe que $A \cap (A')^c \in \mathscr{E} \cap \mathscr{N}_\mu$ car $A \cap (A')^c \subset (A \cup N) \cap (A')^c = (A' \cup N') \cap (A')^c \subset N'$; on a donc $\mu(A) \leq \mu(A \cap A') + \mu(A \cap (A')^c) = \mu(A \cap A')$).

(c) On vérifie (exercice facile) que $\bar{\mu}$ est une mesure sur $(E, \mathscr{E}^\mu)$: $\bar{\mu}(\emptyset) = 0$ et la σ-additivité découle essentiellement de celle de μ et du fait qu'une union dénombrable de négligeables l'est encore. De plus, on a bien $\bar{\mu}(A) = \mu(A)$ pour tout $A \in \mathscr{E}$.

(d) Pour l'unicité, supposons que ν soit une mesure positive sur l'espace mesurable $(E, \mathscr{E}^\mu)$ telle que $\nu(A) = \mu(A)$, pour tout $A \in \mathscr{E}$. Soit $B \in \mathscr{E}^\mu$. Par (a), on peut écrire $B = A \cup N$ avec $A \in \mathscr{E}$ et $N \in \mathscr{N}_\mu$. Il existe donc $N_0 \in \mathscr{E}$ tel que $N \subset N_0$ et $\mu(N_0) = 0$. On remarque alors que $A \subset B \subset A \cup N_0$. Donc forcément, $\mu(A) = \nu(A) \leq \nu(B) \leq \nu(A \cup N_0) = \mu(A \cup N_0) = \mu(A)$, ce qui détermine $\nu(B)$.

Montrons maintenant le point (i) : soit donc $N \in \mathscr{N}_{\bar{\mu}}$. Il existe $N_0 \in \mathscr{E}^\mu$ tel que $\bar{\mu}(N_0) = 0$ et $N \subset N_0$. Par (a), il existe $N_1 \in \mathscr{E}$ t.q. $N_1 \subset N_0$ et $N_0 \setminus N_1 \in \mathscr{N}_\mu$. Mais $N_1 \in \mathscr{N}_\mu$, car $\mu(N_1) = \bar{\mu}(N_1) \leq \bar{\mu}(N_0) = 0$. Donc $N_0 = N_1 \cup (N_0 \setminus N_1) \in \mathscr{N}_\mu$ et, comme $N \subset N_0$, $N \in \mathscr{N}_\mu$.

Pour (ii), on considère $B \in \mathscr{E}^\mu$. Par (a), il existe $A \in \mathscr{E}$ et $N \in \mathscr{N}_\mu$ t.q. $B = A \cup N$. Soit $N_0 \in \mathscr{E}$ t.q. $\mu(N_0) = 0$ et $N \subset N_0$. On conclut facilement avec $A_1 = A$ et $A_2 = A \cup N_0$: on a $A_1, A_2 \in \mathscr{E}$, on a $A_1 \subset B \subset A_2$, et on a $\mu(A_1) = \bar{\mu}(A_1) \leq \bar{\mu}(B) \leq \bar{\mu}(A_2) = \mu(A_2) = \mu(A_1)$.

Soit enfin $f : E \mapsto \mathbb{R}_+$, $\mathscr{E}^\mu$-mesurable. Par le lemme 27, on peut écrire $f = \sum_n a_n \mathbf{1}_{B_n}$, avec $a_n \geq 0$ et $B_n \in \mathscr{E}^\mu$. On applique ensuite (ii) pour construire, pour chaque n, $A_{n,1}, A_{n,2} \in \mathscr{E}$ t.q. $A_{n,1} \subset B_n \subset A_{n,2}$ et $\mu(A_{n,1}) = \bar{\mu}(B_n) = \mu(A_{n,2})$. On conclut facilement que les fonctions $h_1 = \sum_n a_n \mathbf{1}_{A_{n,1}}$ et $h_2 = \sum_n a_n \mathbf{1}_{A_{n,2}}$ vérifient les conclusions du (iii). ■

Chapitre II

Probabilités

II.1 Vocabulaire et notations

Espace de probabilités. Un espace de probabilité est espace mesuré $(\Omega, \mathscr{F}, \mathbf{P})$ où la mesure $\mathbf{P}$ est une probabilité : $\mathbf{P}(\Omega) = 1$.

Ω est un ensemble qui modélise l'aléa. Les éléments de $\mathscr{F}$ sont des *évenements*. La probabilité de $A \in \mathscr{F}$ est $\mathbf{P}(A)$. Pour $A, B \in \mathscr{F}$ on note $\mathbf{P}(A \text{ ou } B) = \mathbf{P}(A \cup B)$ et $\mathbf{P}(A \text{ et} B) = \mathbf{P}(A, B) = \mathbf{P}(A \cap B)$.

Variables aléatoires. Soit $(E, \mathscr{E})$, un espace mesurable. Une application $X : \Omega \mapsto E$ qui est $(\mathscr{F}, \mathscr{E})$-mesurable est appelée une *variable aléatoire*. C'est donc une fonction de l'aléa.

Soit $X : \Omega \to E$, une variable aléatoire (v.a.). Pour $C \in \mathscr{E}$, on note

$$\{X \in C\} = X^{-1}(C) = \{\omega \in \Omega : X(\omega) \in C\} .$$

On omet le plus souvent les accolades : $\mathbf{P}(X \in C) = \mathbf{P}(\{X \in C\}) = \mathbf{P}(X^{-1}(C))$. Si Y est une autre variable aléatoire, on note $\mathbf{P}(X \in C\,,\, Y \in D) = \mathbf{P}(X^{-1}(C) \cap Y^{-1}(D))$.

Espérance. Soit $X : \Omega \to \mathbb{R}$ une v.a. réelle (v.a.r.).

- Si $X \geq 0$, on définit son espérance par $\mathbf{E}[X] = \int_\Omega X(\omega)\mathbf{P}(d\omega)$.
- Sinon, on dit que X est intégrable si $\mathbf{E}[|X|] < \infty$ et on pose alors aussi $\mathbf{E}[X] = \int_\Omega X(\omega)\mathbf{P}(d\omega)$.

Loi. La loi d'une v.a. X à valeurs dans E (muni d'une tribu $\mathscr{E}$) est la mesure image μ_X de $\mathbf{P}$ par X, voir proposition 40. Autrement dit, pour tout $A \in \mathscr{E}$, on a $\mu_X(A) = \mathbf{P}(X \in A)$. On a alors, pour toute fonction $\varphi : E \mapsto \mathbb{R}$, $\mathscr{E}$-mesurable, $\mathbf{E}[\varphi(X)] = \int_E \varphi d\mu_X$, dès que φ est positive ou $\mathbf{E}[|\varphi(X)|] < \infty$. On remarque que $\varphi(X)$ est bien une v.a.r. Bien sûr, μ_X est toujours une probabilité.

Espace canonique. Si on a une mesure de probabilité μ sur un espace mesurable $(E, \mathscr{E})$, on peut toujours construire une v.a. X de loi μ. Il suffit de prendre $(\Omega, \mathscr{F}, \mathbf{P}) = (E, \mathscr{E}, \mu)$ et $X : \Omega \mapsto E$ définie par $X(\omega) = \omega$. Pour $A \in \mathscr{F} = \mathscr{E}$, on a bien $\mathbf{P}(X \in A) = \mu(\{\omega \ : \ \omega \in A\}) = \mu(A)$.

Mais dans toute la suite, sauf mention contraire, $(\Omega, \mathscr{F}, \mathbf{P})$ est un espace de probabilité abstrait sur lequel sont définies toutes les v.a.

Variance. Si X est une v.a.r. telle que $\mathbf{E}[X^2] < \infty$, on pose

$$\mathrm{Var}\, X = \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2] - \mathbf{E}[X]^2.$$

II.2 Premières propriétés

Inégalités. On rappelle les inégalités du premier chapitre.

• Hölder. Si $p, q \in]1, \infty[$ avec $1/p + 1/q = 1$ et X, Y deux v.a.r., $\mathbf{E}[|XY|] \leq \mathbf{E}[|X|^p]^{1/p}\mathbf{E}[|Y|^q]^{1/q}$. Avec $Y = 1$, comme $\mathbf{E}[1] = 1$, ça donne $\mathbf{E}[|X|] \leq \mathbf{E}[|X|^p]^{1/p}$.

• Cauchy-Schwarz. Si X, Y sont deux v.a.r., alors $\mathbf{E}[|XY|] \leq \mathbf{E}[|X|^2]^{1/2}\mathbf{E}[|Y|^2]^{1/2}$.

• Minkowski. Si $p \in]1, \infty[$ et X, Y deux v.a.r., alors $\mathbf{E}[|X + Y|^p]^{1/p} \leq \mathbf{E}[|X|^p]^{1/p} + \mathbf{E}[|Y|^p]^{1/p}$.

Lemme 74 **(Inégalité de Jensen)** *Soit $\varphi : \mathbb{R} \to \mathbb{R}$, une fonction convexe et X une v.a.r. telle que X et $\varphi(X)$ soient intégrables. Alors $\varphi\big(\mathbf{E}[X]\big) \leq \mathbf{E}\big[\varphi(X)\big]$.*

Preuve : Comme φ est convexe, pour tout $x_0 \in \mathbb{R}$, il existe $a \in \mathbb{R}$ tel que $\varphi(x) \geq \varphi(x_0) + a(x - x_0)$ pour tout $x \in \mathbb{R}$. En choisissant $x_0 = \mathbf{E}[X]$, on trouve a tel que $\varphi(x) \geq \varphi(\mathbf{E}[X]) + a(x - \mathbf{E}[X])$ pour tout $x \in \mathbb{R}$. Donc $\varphi(X) \geq \varphi(\mathbf{E}[X]) + a(X - \mathbf{E}[X])$. On conclut en prenant l'espérance. ■

Lemme 75 **(Inégalité de Markov)** *Soit X une v.a.r. positive. Pour tout $a > 0$, on a $\mathbf{P}(X \geq a) \leq \mathbf{E}[X]/a$.*

Preuve : on remarque que $\mathbf{1}_{\{X \geq a\}} \leq X/a$ et on prend l'espérance. ■

Remarque 76 *On peut décliner cette inégalité. Par exemple, si X v.a.r. (pas forcément positive), et si $a \in \mathbb{R}$, alors pour tout $\theta > 0$, on peut écrire $\mathbf{P}(X \geq a) = \mathbf{P}(e^{\theta X} \geq e^{\theta a}) \leq e^{-\theta a}\mathbf{E}[e^{\theta X}]$, et $\mathbf{P}(X \leq a) = \mathbf{P}(e^{-\theta X} \leq e^{-\theta a}) \leq e^{\theta a}\mathbf{E}[e^{-\theta X}]$.*

Lemme 77 **(Inégalité de Tchebychev)** *Soit X une v.a.r. telle que $\mathbf{E}[X^2] < \infty$. Alors pour tout $a > 0$, $\mathbf{P}(|X - \mathbf{E}[X]| \geq a) \leq (\mathrm{Var}\, X)/a^2$.*

Preuve : on applique Markov à $Y = |X - \mathbf{E}[X]|^2$. ■

Fonction de répartition. Soit X, une variable réelle $\mathscr{F}$-mesurable. Sa *fonction de répartition* est la fonction $F_X : \mathbb{R} :\to [0, 1]$ donnée par $F_X(a) = \mathbf{P}(X \leq a) = \mu_X(] - \infty, a])$.

Lemme 78 *Soit X une v.a.r. La fonction F_X est croissante, continue à droite, $\lim_{a \to -\infty} F_X(a) = 0$ et $\lim_{a \to \infty} F_X(a) = 1$. De plus, pour tout $a \in \mathbb{R}$, on a $\mathbf{P}(X = a) = F_X(a) - F_X(a-)$.*

Preuve : F_X est bien sûr croissante. La continuité à droite découle de la proposition 14-(ii) : si a_n décroit vers a, on a $\cap] - \infty, a_n] =] - \infty, a]$, ce qui implique que $\lim_n \mu_X(] - \infty, a_n]) = \mu_X(] - \infty, a])$. On montre de même que $\lim_{a \to -\infty} F_X(a) = 0$ et $\lim_{a \to \infty} F_X(a) = 1$. Enfin, il est facile de voir (par la proposition 14-(i)) que $F_X(a-) = \mu_X(] - \infty, a[)$, d'où $F_X(a) - F_X(a-) = \mu_X(\{a\})$. ■

Proposition 79 *Soient X et Y, deux v.a.r. Alors, $F_X = F_Y$ ssi $\mu_X = \mu_Y$. Autrement dit, la loi d'une variable est caractérisée par sa fonction de répartition.*

Preuve : Si $\mu_X = \mu_Y$, alors bien sûr $F_X = F_Y$. Supposons que $F_X = F_Y$. On pose $\mathscr{P} = \{\mathbb{R}\} \cup \{\,]-\infty, a]; a \in \mathbb{R}\}$. C'est un pi-système (car il contient $\mathbb{R}$ et car si $A, B \in \mathscr{P}$, on a $A \cap B \in \mathscr{P}$) qui de plus engendre $\mathscr{B}(\mathbb{R})$. Comme on a $\mu_X(A) = \mu_Y(A)$ pour tout $A \in \mathscr{P}$, le théorème d'unicité de prolongement des mesures garantit donc que $\mu_X = \mu_Y$. ■

Densité. Soit X une v.a.r. Si pour tout $A \in \mathscr{B}(\mathbb{R})$ de mesure de Lebesgue nulle, on a $\mu_X(A) = \mathbf{P}(X \in A) = 0$, alors le théorème de Radon-Nikodym garantit l'existence d'une densité $f_X : \mathbb{R} \mapsto \mathbb{R}_+$, $\mathscr{B}(\mathbb{R})$-mesurable, telle que pour tout $A \in \mathscr{B}(\mathbb{R})$,

$$\mathbf{P}(X \in A) = \mu_X(A) = \int_A f(x)dx.$$

Cette dernière intégrale est une notation pour $\int_A f_X d\ell$, où ℓ est la mesure de Lebesgue sur $\mathbb{R}$. Par la Proposition 66, on a alors aussi, pour toute fonction $\varphi : \mathbb{R} \mapsto \mathbb{R}$, $\mathscr{B}(\mathbb{R})$-mesurable, positive ou telle que $\mathbf{E}[|\varphi(X)|] < \infty$,

$$\mathbf{E}[\varphi(X)] = \int_{\mathbb{R}} \varphi(x) f_X(x)dx.$$

Lois usuelles sur $\mathbb{R}$. Voici une liste de lois usuelles, dont on pourra retrouver facilement les définitions et propriétés élémentaires dans n'importe quel livre (ou ailleurs).

• Lois discrètes : loi de Bernoulli $\mathrm{Ber}(p)$, loi binômiale $\mathrm{B}(n,p)$, loi de Poisson $\mathscr{P}(\lambda)$, loi géométrique $\mathrm{Geo}(p)$.

• Lois à densité : loi exponentielle $\mathscr{E}(\lambda)$, loi gamma $\Gamma(a,p)$, loi normale $\mathscr{N}(m,\sigma^2)$, loi uniforme $\mathscr{U}([a,b])$. On dit parfois que $X \sim \mathscr{N}(m,0)$ si $\mathbf{P}(X = m) = 1$.

Lois ni discrètes ni à densité. Il est très facile de trouver des variables aléatoires dont la loi n'est ni discrète, ni à densité. Si par exemple on a $U \sim \mathscr{N}(0,1)$ et $V \sim \mathrm{Ber}(1/2)$ indépendantes, on vérifie (exercice) la loi de $X = VU$ est donnée par $\mu_X(dx) = \frac{1}{2}\delta_0(dx) + \frac{1}{2\sqrt{2\pi}} e^{-x^2/2}dx$, qui a donc une partie à densité et une partie discrète.

Plus intéressant, il existe des variables aléatoire *continues*, i.e. dont la fonction de répartition est continue, i.e. telles que $\mathbf{P}(X = x) = 0$ pour tout $x \in \mathbb{R}$, mais qui n'admettent pas de densité. Voici un exemple qui peut s'avérer utile. Nous ne donnons que les étapes principales.

On montre facilement que si U suit une loi uniforme sur $[0,1]$, alors les chiffres X_i de U sont i.i.d. de loi uniforme sur $\{0,1,...,9\}$.

Soit $u \in [0,1]$ et $u = \sum_{n \geq 1} x_n 10^{-n}$ son écriture en base 10. On dit que u est *1-normal* si pour tout $i \in \{0,...,9\}$, $\lim_n n^{-1} N_n(u,i) = 1/10$, où $N_n(u,i) = \#\{k \in \{1,...,n\} \; : x_k = i\}$. On note $\mathscr{N}$ l'ensemble des $u \in [0,1]$ qui sont *1-normaux*.

On déduit des deux paragraphes précédent que si $U \sim \mathscr{U}([0,1])$, alors $\mathbf{P}(U \in \mathscr{N}) = 1$ (car par la LGN, $\lim_n n^{-1} N_n(U,i) = \lim_n n^{-1} \sum_1^N \mathbf{1}_{X_k = i} = 1/10$ p.s.). Autrement dit, la mesure de Lebesgue sur $[0,1]$ ne charge que les nombres 1-normaux, i.e. $\ell(\mathscr{N} \cap [0,1]) = 1$. *Pourtant, on ne connaît quasiment aucun nombre* 1*-normal, e.g. on ne sait pas si* $\sqrt{2} - 1$, $e - 2$, $\pi - 3$, *etc. le sont.*

Considérons maintenant $Y_1, Y_2, ...$ i.i.d. de loi uniforme sur $\{0,1,...,8\}$ et soit $V = \sum_{n \geq 1} Y_n 10^{-n}$. Les Y_i sont donc les chiffres (après la virgule) de V en base 10. On voit que $\lim_n n^{-1} N_n(V,i) =$

$\lim_n n^{-1} \sum_1^N \mathbf{1}_{Y_k=1} = 1/9$ p.s., donc V est p.s. non 1-normal. La loi de V n'admet donc pas de densité par rapport à la mesure de Lebesgue (puisqu'on a $\ell(\mathscr{N}^c) = 0$ et pourtant $\mu_V(\mathscr{N}^c) = 1$, et par le théorème de Radon-Nikodym). Pourtant, il est clair V est une v.a. continue : pour $v \in [0,1]$, on écrit $v = \sum_{n\geq 1} y_n 10^{-n}$, et on a $\mathbf{P}(V = v) = \prod_1^\infty \mathbf{P}(V_n = y_n) = 0$, car c'est un produit infini de nombres qui valent soit 0 (si $y_n = 9$) soit 1/8 (sinon).

Vecteurs aléatoires. Une v.a. $\mathbf{X} = (X_1, \ldots, X_d)$ à valeurs dans $\mathbb{R}^d$ est appelée vecteur aléatoire (tous les vecteurs sont des vecteurs colonne). Pour $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$, on pose

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbf{P}\big(X_1 \leq x_1, \ldots, X_d \leq x_d\big) = \mu_{\mathbf{X}}\big(\,]-\infty, x_1] \times \ldots \times]-\infty, x_d]\big).$$

On appelle parfois $F_{\mathbf{X}}$ la *fonction de répartition de* $\mathbf{X}$. En raisonnant comme pour les v.a.r., on voit que deux vecteurs aléatoires ont même loi ssi ils ont même fonction de répartition.

On note $\|\cdot\|$ la norme Euclidienne de $\mathbb{R}^d$ et $\langle \cdot, \cdot \rangle$ le produit scalaire Euclidien. Soit $\mathbf{X} = (X_1, \ldots, X_d)$ un vecteur aléatoire tel que $\mathbf{E}\big[\|\mathbf{X}\|^2\big] = \mathbf{E}[X_1^2] + \ldots + \mathbf{E}[X_d^2] < \infty$. On note

$$\mathbf{E}[\mathbf{X}] = \big(\mathbf{E}[X_1], \ldots, \mathbf{E}[X_1]\big) \in \mathbb{R}^d$$

son *vecteur moyenne*. On note aussi

$$\Gamma_{\mathbf{X}} = \mathbf{E}[(\mathbf{X} - \mathbf{E}[\mathbf{X}])(\mathbf{X} - \mathbf{E}[\mathbf{X}])^*] = \Big(\mathbf{E}\big[(X_k - \mathbf{E}[X_k])(X_\ell - \mathbf{E}[X_\ell])\big]\Big)_{1\leq k,\ell\leq d}$$

sa *matrice de covariance de* $\mathbf{X}$, qui est une matrice réelle $d \times d$ symétrique (et positive). On montre aisément que pour tout $\mathbf{u} = (u_1, \ldots, u_d) \in \mathbb{R}^d$, la v.a.r. $\langle \mathbf{u}, \mathbf{X} \rangle$ vérifie

$$\mathbf{E}[\langle \mathbf{u}, \mathbf{X} \rangle] = \langle \mathbf{u}, \mathbf{E}[\mathbf{X}] \rangle \quad \text{et} \quad \mathrm{Var}\,(\langle \mathbf{u}, \mathbf{X} \rangle) = \mathbf{u}^* \Gamma_{\mathbf{X}} \mathbf{u}.$$

Si M est une matrice $d \times d$, le vecteur aléatoire $M\mathbf{X}$ vérifie enfin

$$\mathbf{E}[M\mathbf{X}] = M\mathbf{E}[\mathbf{X}] \quad \text{et} \quad \Gamma_{M\mathbf{X}} = M \Gamma_{\mathbf{X}} M^*.$$

Presque sûr. On dit qu'une propriété est vraie p.s. si elle est vraie en dehors d'un ensemble $\mathbf{P}$-négligeable. Par exemple, deux v.a. X, Y sont égales p.s. si $\mathbf{P}(X \neq Y) = 0$, i.e. si $\mathbf{P}(X = Y) = 1$.

II.3 Convergence de variables aléatoires.

Dans cette section, $(\Omega, \mathscr{F}, \mathbf{P})$ désigne l'espace de probabilité sur lequel toutes les variables aléatoires mentionnées sont définies. On note $\mathscr{L}(\Omega, \mathscr{F})$, l'ensemble des v.a.r., i.e. des application de Ω dans $\mathbb{R}$ qui sont $(\mathscr{F}, \mathscr{B}(\mathbb{R}))$-mesurables. Par la proposition 24, si $X, Y \in \mathscr{L}(\Omega, \mathscr{F})$, alors $X + aY, XY \in \mathscr{L}(\Omega, \mathscr{F})$ (pour $a \in \mathbb{R}$).

On introduit la relation d'équivalence $\sim$ sur $\mathscr{L}(\Omega, \mathscr{F})$ en posant $X \sim Y$ ssi $X - Y = 0$ p.s., et on considère le quotient

$$\mathrm{L}(\Omega, \mathscr{F}, \mathbf{P}) = \mathscr{L}(\Omega, \mathscr{F}) / \sim \ .$$

Dans la suite, on travaille sur $\mathrm{L}(\Omega, \mathscr{F}, \mathbf{P})$ plutôt que sur $\mathscr{L}(\Omega, \mathscr{F})$ mais par commodité, on confond dans les notations un élément de $\mathrm{L}(\Omega, \mathscr{F}, \mathbf{P})$, qui est une classe d'équivalence de variables, avec n'importe lequel de ses représentants.

Rappelons quelques résultats sur les espaces L^p qui sont prouvés dans un cadre général à la section I.9 du chapitre I. Pour X une v.a.r., on pose

$$\forall p \in [1, \infty[\,,\ \|X\|_p = \big(\mathbf{E}[\,|X|^p\,]\big)^{\frac{1}{p}} \quad \text{et} \quad \|X\|_\infty = \inf\big\{a \in [0,\infty] \ :\ \mathbf{P}(\,|X| > a) = 0\big\}.$$

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

$$\mathrm{L}^p(\Omega, \mathscr{F}, \mathbf{P}) = \big\{X \in \mathrm{L}(\Omega, \mathscr{F}, \mathbf{P}) : \|X\|_p < \infty\big\}\ .$$

qui, muni de la norme $\|\cdot\|_p$, est un *espace vectoriel normé complet.*

Définition 80 *Soient $X_n \in \mathrm{L}(\Omega, \mathscr{F}, \mathbf{P})$, $n \in \mathbb{N} \cup \{\infty\}$.*

(i) La suite $(X_n)_{n\geq 0}$ converge en probabilité vers X_∞ ssi $\forall \varepsilon > 0$, $\lim_{n\to\infty} \mathbf{P}(|X_n - X_\infty| > \varepsilon) = 0$.

(ii) La suite $(X_n)_{n\geq 0}$ converge p.s. vers X_∞ ssi $\mathbf{P}(\lim_{n\to\infty} X_n = X_\infty) = 1$.

(iii) Soit $p \in [1, \infty]$. On suppose que $X_n \in \mathrm{L}^p$, $n \in \mathbb{N} \cup \{\infty\}$. On dit que la suite $(X_n)_{n\geq 0}$ converge vers X_∞ dans L^p ssi $\lim_{n\to\infty} \|X_n - X_\infty\|_p = 0$.

On rappelle les théorèmes de passage à la limite sous l'espérance qui sont la traduction, avec les notations probabilistes, des énoncés des théorèmes d'intégration.

Théorème 81 *Soit $(X_n)_{n\geq 0}$ une suite de v.a.r. positives p.s.*

(Convergence monotone) *On suppose que $X_n \leq X_{n+1}$ p.s., pour tout $n \geq 0$. On pose $X_\infty = \sup_n X_n$. Alors $\lim_n \mathbf{E}[X_n] = \mathbf{E}[X_\infty]$.*

(Interversion positive) $\mathbf{E}\Big[\sum_{n\geq 0} X_n\Big] = \sum_{n\geq 0} \mathbf{E}[X_n]$.

(Fatou) $\mathbf{E}[\liminf_n X_n] \leq \liminf_n \mathbf{E}[X_n]$.

Théorème 82 **(Convergence dominée)** *Soit $(X_n)_{n\geq 0}$ une suite de v.a.r. Si*

- *Il existe une v.a. réelle X_∞ telle que $\lim_n X_n = X_\infty$ p.s.,*
- *Il existe une v.a.r. Z telle que $\mathbf{E}[Z] < \infty$ et $|X_n| \leq Z$ p.s., pour tout $n \geq 0$.*

alors, X_∞ est intégrable, $\lim_n \mathbf{E}\big[\,|X_n - X_\infty|\,\big] = 0$, et $\lim_n \mathbf{E}[X_n] = \mathbf{E}[X_\infty]$.

Corollaire 83 **(Interversion L^1)** *Soit $(X_n)_{n\geq 0}$, une suite de v.a.r. telle que $\sum_{n\geq 0} \mathbf{E}\big[|X_n|\big] < \infty$. Alors $\sum_{n\geq 0} X_n$ est une v.a.r. intégrable et $\mathbf{E}[\sum_{n\geq 0} X_n] = \sum_{n\geq 0} \mathbf{E}[X_n]$.*

Borel-Cantelli. Soient $A_n \in \mathscr{F}$, $n \in \mathbb{N}$. On pose

$$\begin{aligned}
\limsup_n A_n = \cap_{n\geq 0} \cup_{k\geq n} A_k =& \{\omega \in \Omega \ :\ \omega \in A_n \text{ pour une infinité de } A_n\} \\
=& \{\text{une infinité de } A_n \text{ sont réalisés}\}, \\
\liminf_n A_n = \cup_{n\geq 0} \cap_{k\geq n} A_k =& \{\omega \in \Omega \ :\ \omega \in A_n \text{ à partir d'un certain rang}\} \\
=& \{\text{tous les } A_n \text{ sont réalisés à partir d'un certain rang}\}.
\end{aligned}$$

On remarque que $(\limsup_n A_n)^c = \liminf_n A_n^c$.

Lemme 84 (Borel-Cantelli) *Soit $A_n \in \mathscr{F}$, $n \in \mathbb{N}$. On suppose que $\sum_{n\geq 0} \mathbf{P}(A_n) < \infty$. On a alors $\mathbf{P}(\limsup_{n\to\infty} A_n) = 0$. Par passage au complémentaire, on voit que*

$$\mathbf{P}\big(\textit{à partir d'un certain, aucun } A_n \textit{ n'est réalisé}\big) = 1.$$

Preuve : on pose $Z = \sum_{n\geq 0} \mathbf{1}_{A_n}$. On a $\mathbf{E}[Z] = \sum_{n\geq 0} \mathbf{P}(A_n) < \infty$. Donc $\mathbf{P}(Z = \infty) = 0$. Or $\{Z = \infty\} = \{\text{une infinité de } A_n \text{ sont réalisés}\} = \limsup_n A_n$. ■

Une application du lemme de Borel-Cantelli est le critère de convergence p.s. suivant.

Proposition 85 *Soit $\varepsilon_n \in]0, \infty[$, $n \in \mathbb{N}$, telle que $\sum_{n\geq 0} \varepsilon_n < \infty$. Soit $(X_n)_{n\geq 0}$, une suite de v.a.r. telles que $\sum_{n\geq 0} \mathbf{P}\big(|X_{n+1} - X_n| > \varepsilon_n\big) < \infty$. Il existe une v.a.r. X_∞ telle que $\lim_{n\to\infty} X_n = X_\infty$ p.s.*

Preuve : on pose $A_n = \{|X_{n+1} - X_n| > \varepsilon_n\}$ et $B_n = A_n^c = \{|X_{n+1} - X_n| \leq \varepsilon_n\}$. Le lemme de Borel-Cantelli implique que $\mathbf{P}(\limsup_n A_n) = 0$, c'est-à-dire que $\mathbf{P}(\liminf_n B_n) = 1$. On pose $B_* = \liminf_n B_n$: par définition, pour tout $\omega \in B_*$, il existe $N(\omega) \in \mathbb{N}$, tel que pour tout $n \geq N(\omega)$, $|X_{n+1}(\omega) - X_n(\omega)| \leq \varepsilon_n$. Comme $\sum_n \epsilon_n < \infty$, on conclut que pour tout $\omega \in B_*$, la série $\sum_{n\geq 0} |X_{n+1}(\omega) - X_n(\omega)|$ converge, et donc la série de terme général $X_{n+1}(\omega) - X_n(\omega)$ est convergente. Par téléscopage, cela implique que $\lim_n X_n(\omega)$ existe et on note $X_\infty(\omega)$ cette limite. Si $\omega \notin B_*$, on pose $X_\infty(\omega) = 0$ (par exemple), et on conclut. ■

Proposition 86 *Soit X_n, $n \in \mathbb{N} \cup \{\infty\}$, des v.a.r.*

(a) Si $X_n \to X_\infty$ dans L^p, alors $X_n \to X_\infty$ dans L^q pour tout $1 \leq q \leq p \leq \infty$.

(b) Si $X_n \to X_\infty$ dans L^1, alors $X_n \to X$ en probabilité.

(c) Si $X_n \to X_\infty$ p.s., alors $X_n \to X$ en probabilité.

Preuve : (a) est évident puisque par Hölder, $||X_n - X_\infty||_q \leq ||X_n - X_\infty||_p$. Pour (b), on utilise Markov : pour tout $\epsilon > 0$, $\mathbf{P}(|X_n - X_\infty| \geq \epsilon) \leq \epsilon^{-1}\mathbf{E}[|X_n - X_\infty|] \to 0$. Pour (c), on fixe ϵ et on écrit $\mathbf{P}(|X_n - X_\infty| \geq \epsilon) = \mathbf{E}[\mathbf{1}_{\{|X_n - X_\infty| \geq \epsilon\}}]$, qui tend vers 0 par convergence dominée. ■ .

Proposition 87 *Soit X_n, $n \in \mathbb{N} \cup \{\infty\}$, des v.a.r. On suppose que $X_n \to X_\infty$ en probabilité. Alors, il existe une suite strictement croissante d'indices $(n_k)_{k\geq 0}$ telle que $\lim_k X_{n_k} = X_\infty$ p.s.*

Preuve : par définition, pour tout $k \in \mathbb{N}$, il existe $p_k \in \mathbb{N}^*$, tel que pour tout $n \geq p_k$, on ait $\mathbf{P}(|X_n - X_\infty| > 2^{-k}) \leq 2^{-k}$. On pose $n_k = p_1 + \ldots + p_k$ et on a bien $\mathbf{P}(|X_{n_k} - X_\infty| > 2^{-k}) \leq 2^{-k}$. On pose $B_k = \{|X_{n_k} - X_\infty| \leq 2^{-k}\}$ et $B_* = \liminf_k B_k$. On a $\mathbf{P}(B_*) = 1$, par Borel-Cantelli. Pour tout $\omega \in B_*$, il existe un rang $k(\omega)$, tel que pour tout $k \geq k(\omega)$, on ait $\omega \in B_k$, c'est-à-dire $|X_{n_k}(\omega) - X_\infty(\omega)| \leq 2^{-k}$ et donc $\lim_k X_{n_k}(\omega) = X_\infty(\omega)$, ce qui montre la proposition. ■

Remarque 88 *La convergence p.s. n'est pas une vraie convergence : elle n'est associée à aucune topologie sur $\mathrm{L}(\Omega, \mathscr{F}, \mathbf{P})$ (donc* a fortiori *à aucune norme).*

Preuve : Soit $(H, \mathscr{T})$ un espace topologique, soit (x_n) une suite d'éléments de H et $x \in H$. Alors $\lim_n x_n = x$ ssi de toute sous-suite de (x_n) on peut extraire une sous-sous-suite qui converge vers x. Le premier sens est évident (si $\lim_n x_n = x$ alors de toute sous-suite de (x_n) on peut extraire une sous-sous-suite qui converge vers x : n'importe quelle sous-sous-suite convient). Pour montrer l'autre sens, il suffit de montrer que si x_n ne converge pas vers x, alors il existe une sous-suite de x_n dont on ne peut pas extraire une sous-sous-suite qui converge vers x. Mais comme x_n ne converge pas vers x, on peut trouver $O \in \mathscr{T}$ avec $x \in O$ et une sous-suite (x_{n_k}) à valeurs dans O^c. De cette sous-suite, on ne peut clairement pas extraire une sous-sous-suite convergeant vers x.

Si $X_n \to X$ en probabilité, alors de toute sous-suite de X_n, on peut extraire une sous-sous-suite convergeant p.s. vers X. Or il est possible que $X_n \to X$ en probabilité sans que $X_n \to X$ p.s. ■

Théorème 89 *Pour tout $X, Y \in \mathrm{L}(\Omega, \mathscr{F}, \mathbf{P})$, on pose $d(X,Y) = \mathbf{E}[1 \wedge |X - Y|]$.*

(i) d est une distance sur $\mathrm{L}(\Omega, \mathscr{F}, \mathbf{P})$.

(ii) $d(X_n, X_\infty) \to 0$ ssi $X_n \to X_\infty$ en probabilité.

(iii) $\mathrm{L}(\Omega, \mathscr{F}, \mathbf{P})$, muni de d, est complet.

Preuve : (i) On a $d(X,Y) \leq d(X,Z) + d(Y,Z)$ car $(a+b) \wedge 1 \leq a \wedge 1 + b \wedge 1$ pour tout $a, b \geq 0$. On a $d(X,Y) = d(Y,X)$ et $d(X,X) = 0$. Enfin, $d(X,Y) = 0$ implique que $|X - Y| \wedge 1 = 0$ p.s. et donc $X = Y$ p.s., soit encore $X = Y$ au sens de $\mathrm{L}(\Omega, \mathscr{F}, \mathbf{P})$.

(ii) Supposons que $d(X_n, X_\infty) \to 0$ et fixons $\epsilon \in]0,1[$. Alors $\mathbf{P}(|X_n - X_\infty| \geq \epsilon) = \mathbf{P}(|X_n - X_\infty| \wedge 1 \geq \epsilon) \leq \epsilon^{-1} d(X_n, X_\infty)$ par Markov, et $X_n \to X_\infty$ en probabilité. Si maintenant $X_n \to X_\infty$ en probabilité, alors pour $\epsilon > 0$ arbitrairement petit, on écrit

$$
\begin{aligned}
d(X_n, X_\infty) =& \mathbf{E}\big[\mathbf{1}_{\{|X_n - X_\infty| \leq \varepsilon\}}(1 \wedge |X_n - X_\infty|)\big] + \mathbf{E}\big[\mathbf{1}_{\{|X_n - X_\infty| > \varepsilon\}}(1 \wedge |X_n - X_\infty|)\big] \\
\leq& \varepsilon + \mathbf{P}\big(|X_n - X_\infty| > \varepsilon\big),
\end{aligned}
$$

d'où $\limsup_n d(X_n, X_\infty) \leq \epsilon$, puis $\lim_{n \to \infty} d(X_n, X_\infty) = 0$.

(iii) Soit $X_n \in \mathrm{L}(\Omega, \mathscr{F}, \mathbf{P})$ une suite de Cauchy pour d. Pour montrer qu'elle converge, il suffit de montrer qu'elle admet une sous-suite d-convergente (voir le lemme 56). Pour tout $\epsilon \in]0,1]$, il existe N_ϵ, pour tout $n, m \geq N_\epsilon$, $d(X_n, X_m) \leq \epsilon^2$ et donc

$$
\mathbf{P}(|X_n - X_m| \geq \epsilon) = \mathbf{P}(|X_n - X_m| \wedge 1 \geq \epsilon) \leq \epsilon^{-1} d(X_n, X_m) \leq \epsilon.
$$

Pour tout $k \in \mathbb{N}$, on pose $n_k = N(2^{-0}) + \ldots + N(2^{-k})$. On a donc

$$
\forall k \in \mathbb{N}, \quad \mathbf{P}\big(|X_{n_{k+1}} - X_{n_k}| > 2^{-k}\big) \leq 2^{-k} .
$$

Par la proposition 85, la suite X_{n_k} converge donc p.s., et donc en probabilité, et donc pour d. ■

Lemme 90 *Si on a deux suites de v.a.r. telles que $X_n \to X_\infty$ et $Y_n \to Y_\infty$ en probabilité, alors $f(X_n, Y_n) \to f(X_\infty, Y_\infty)$ en probabilité pour toute fonction $f : \mathbb{R}^2 \mapsto \mathbb{R}$ continue.*

Preuve : Soit $Z_n = f(X_n, Y_n)$ et $Z_\infty = f(X_\infty, Y_\infty)$. On fixe $\epsilon > 0$. Soit $A > 0$ (grand). f est uniformément continue sur le compact $[-A-1, A+1]^2$. Donc il existe $\eta > 0$ tel que si x, x', y, y' sont tous dans $[-A-1, A+1]$ avec de plus $|x - x'| < \eta$ et $|y - y'| < \eta$, alors $|f(x,y) - f(x',y')| < \epsilon$. Donc

$$
\begin{aligned}
\mathbf{P}(|Z_n - Z_\infty| > \epsilon) \leq& \mathbf{P}(|X_\infty| > A \text{ ou } |Y_\infty| > A \text{ ou } |X_n| > A+1 \text{ ou } |Y_n| > A+1) \\
&+ \mathbf{P}(X_n, X_\infty, Y_n, Y_\infty \in [-A-1, A+1] \text{ et } (|X_n - X_\infty| \geq \eta \text{ ou } |Y_n - Y_\infty| \geq \eta)) \\
\leq& \mathbf{P}(|X_\infty| > A) + \mathbf{P}(|Y_\infty| > A) + \mathbf{P}(|X_n - X_\infty| > 1) + \mathbf{P}(|Y_n - Y_\infty| > 1) \\
&+ \mathbf{P}(|X_n - X_\infty| \geq \eta) + \mathbf{P}(|Y_n - Y_\infty| \geq \eta).
\end{aligned}
$$

On conclut donc que $\limsup_n \mathbf{P}(|Z_n - Z_\infty| > \epsilon) \leq \mathbf{P}(|X_\infty| > A) + \mathbf{P}(|Y_\infty| > A)$. Il ne reste plus qu'à faire tendre A vers l'infini. ■

II.4 Uniforme intégrabilité

Définition 91 *Une famille* $(X_i)_{i\in I}$ *de v.a.r. intégrables est dite uniformément intégrable (ou U.I.) si* $\lim_{a\to\infty}\sup_{i\in I}\mathbf{E}[|X_i|\mathbf{1}_{\{|X_i|\geq a\}}]=0$.

Remarque 92 *Une famille bornée dans* L^1 *n'est pas nécessairement U.I. : si* $\mathbf{P}(X_n=n)=1/n$ *et* $\mathbf{P}(X_n=0)=1-1/n$, *alors* $\mathbf{E}[|X_n|]=1$, *mais pour tout* $a>0$, $\sup_{n\geq 1}\mathbf{E}[|X_n|\mathbf{1}_{|X_n|\geq a}]=1$.

Lemme 93 *Soit* $(X_i)_{i\in I}$ *une famille de v.a.r. intégrables.*

(i) Si I *est fini, alors* $(X_i)_{i\in I}$ *est U.I.*

(ii) S'il existe une v.a.r. Z *intégrable telle que pour tout* $i\in I$, $|X_i|\leq Z$ *p.s., alors* $(X_i)_{i\in I}$ *est U.I.*

(iii) Si $\sup_{i\in I}\mathbf{E}[|X_i|^p]<\infty$ *pour un* $p>1$, *alors* $(X_i)_{i\in I}$ *est U.I.*

Preuve : Pour (ii), on écrit $\sup_{i\in I}\mathbf{E}[|X_i|\mathbf{1}_{\{|X_i|\geq a\}}]\leq\mathbf{E}[Z\mathbf{1}_{Z>a}]$ qui tend vers 0 quand $a\to\infty$ par convergence dominée. Le (i) découle du (ii) avec $Z=\sum_{i\in I}|X_i|$. Le (iii) découle du théorème suivant avec $\varphi(x)=x^p$. ■

Théorème 94 (de de la Vallée Poussin) *Soit* $(X_i)_{i\in I}$, *une famille de v.a.r. On a l'équivalence*

(i) $(X_i)_{i\in I}$ *est U.I.*

(ii) Il existe $\varphi:\mathbb{R}_+\mapsto\mathbb{R}_+$ *(qu'on peut choisir croissante et convexe) telle que* $\lim_\infty\frac{\varphi(x)}{x}=\infty$ *et telle que* $\sup_{i\in I}\mathbf{E}[\varphi(|X_i|)]<\infty$.

Preuve : (ii) implique (i) est facile : on pose $w(a)=\sup_{x>a}[x/(1+\varphi(x))]$, qui est décroissante et tend vers 0 quand $a\to\infty$. On pose $K=\sup_{i\in I}\mathbf{E}\big[1+\varphi(|X_i|)\big]$. Pour tout $a>0$ et tout $i\in I$, on a

$$\mathbf{E}[|X_i|\mathbf{1}_{\{|X_i|>a\}}]=\mathbf{E}\Big[(1+\varphi(|X_i|))\frac{|X_i|}{1+\varphi(|X_i|)}\mathbf{1}_{\{|X_i|>a\}}\Big]\leq w(a)\mathbf{E}\big[1+\varphi(|X_i|)\big]\leq Kw(a),$$

ce qui implique bien que $\lim_{a\to\infty}\sup_{i\in I}\mathbf{E}[|X_i|\mathbf{1}_{\{|X_i|>a\}}]=0$.

Supposons maintenant (i). Soit $(a_k)_{k\geq 0}$ une suite de réels positifs t.q. $\sup_{i\in I}\mathbf{E}[|X_i|\mathbf{1}_{\{|X_i|>a_k\}}]\leq 2^{-k}$. On peut supposer que $a_k\to\infty$ (en remplaçant a_k par a_k+k si nécessaire). Soit ensuite $\varphi(x)=\sum_{k\geq 0}(x-a_k)_+$. Cette fonction est croissante, convexe, et on a $\lim_\infty(\varphi(x)/x)=\infty$ (exercice). De plus,

$$\forall\, i\in I,\qquad \mathbf{E}[\varphi(|X_i|)]=\sum_{k\geq 0}\mathbf{E}[(|X_i|-a_k)_+]\leq\sum_{k\geq 0}\mathbf{E}[|X_i|\mathbf{1}_{\{|X_i|\geq a_k\}}]\leq\sum_{k\geq 0}2^{-k}=2.$$

La preuve est complète. ■

Théorème 95 (Convergence dominée, version optimale) *Soit* X_n *une suite de v.a.r. et* X_∞ *une v.a.r., toutes intégrables. On a l'équivalence*

(i) $X_n\to X_\infty$ *dans* L^1.

(ii) $X_n\to X_\infty$ *en probabilité et la famille* $(X_n)_{n\geq 1}$ *est U.I.*

Preuve : Comme $X_\infty \in \mathrm{L}^1$, on montre facilement que $(X_n)_{n\geq 1}$ est U.I. ssi $(X_n - X_\infty)_{n\geq 1}$ est U.I. (exercice). De plus, $X_n \to X_\infty$ dans L^1 (ou en probabilité) ssi $X_n - X_\infty$ dans L^1 (ou en probabilité). On peut donc se ramener au cas où $X_\infty = 0$.

Pour montrer que (ii) implique (i). Fixons $\epsilon > 0$ petit et $a > 0$ grand. On écrit

$$\mathbf{E}[|X_n|] \leq \mathbf{E}[\epsilon \mathbf{1}_{\{|X_n|\leq \epsilon\}}] + \mathbf{E}[|X_n|\mathbf{1}_{\{|X_n|>\epsilon\}}] \leq \epsilon + a\mathbf{P}(|X_n| > \epsilon) + \mathbf{E}[|X_n|\mathbf{1}_{\{|X_n|\geq a\}}].$$

En faisant tendre n vers l'infini, on trouve donc

$$\limsup_n \mathbf{E}[|X_n|] \leq \epsilon + \sup_{n\geq 0} \mathbf{E}[|X_n|\mathbf{1}_{\{|X_n|\geq a\}}]$$

Comme $(X_n)_{n\geq 0}$ est U.I., on trouve $\limsup_n \mathbf{E}[|X_n|] \leq \epsilon$ en faisant tendre $a \to \infty$, et il n'y a plus qu'à faire tendre $\epsilon \to 0$.

Supposons maintenant (i). Bien sûr, $X_n \to 0$ en probabilité, et il suffit de montrer que la famille $(X_n)_{n\geq 1}$ est U.I. On fixe donc $\epsilon > 0$, et on cherche $A_\epsilon > 0$ tel que $\sup_{n\geq 0} \mathbf{E}[|X_n|\mathbf{1}_{\{|X_n|\geq A_\epsilon\}}] \leq \epsilon$. Il suffit de considérer N_ϵ tel que $\sup_{n\geq \epsilon} \mathbf{E}[|X_n|] \leq \epsilon$ puis $A_\epsilon > 0$ tel que $\sup_{n\in\{0,\dots,N_\epsilon\}} |X_n|\mathbf{1}_{\{|X_n|\geq A_\epsilon\}}] \leq \epsilon$ (ce A_ϵ existe car la famille $X_0, \dots, X_{N_\epsilon}$ est finie et donc U.I.). ■

II.5 Indépendance

Définition 96 *Soit I un ensemble quelconque et, pour tout $i \in I$, une v.a. X_i à valeurs dans E_i (muni d'une tribu $\mathscr{E}_i$). On note $\sigma(X_i, i \in I)$ la plus petite tribu $\mathscr{G}$ sur Ω t.q. pour tout $i \in I$, X_i soit $(\mathscr{G}, \mathscr{E}_i)$-mesurable.*

Il est facile de vérifier que $\sigma(X_i, i \in I) = \sigma(\{\{X_i \in C_i\} : C_i \in \mathscr{E}_i, i \in I\})$. La situation est plus simple lorsqu'on ne considère qu'une seule variable, puisqu'alors $\sigma(X) = \{\{X \in C\} : C \in \mathscr{E}\}$ (vérifier que le membre de droite est une tribu sur Ω).

Le résultat suivant est crucial.

Théorème 97 *(i) Soit X une v.a. à valeurs dans E (muni d'une tribu $\mathscr{E}$) et Y une v.a.r. Si Y est $\sigma(X)$-mesurable (i.e. $(\sigma(X), \mathscr{B}(\mathbb{R}))$-mesurable), alors il existe $\varphi : E \mapsto \mathbb{R}$, $(\mathscr{E}, \mathscr{B}(\mathbb{R}))$-mesurable, telle que $Y = \varphi(X)$.*

(ii) Soient $X_1, \dots, X_n$ des v.a. à valeurs dans $E_1, \dots, E_n$, munis de tribus $\mathscr{E}_1, \dots, \mathscr{E}_n$. Si Y est une v.a.r. $\sigma(X_1, \dots, X_n)$-mesurable, alors il existe $\varphi : E_1 \times \cdots \times E_N \mapsto \mathbb{R}$, $(\mathscr{E}_1 \otimes \cdots \otimes \mathscr{E}_n, \mathscr{B}(\mathbb{R}))$-mesurable, telle que $Y = \varphi(X_1, \dots, X_n)$.

Preuve : Pour (i), on suppose d'abord $Y \geq 0$, et on utilise le Lemme 27 pour écrire $Y = \sum_{n\geq 0} a_n \mathbf{1}_{B_n}$, avec $a_n \geq 0$ et $B_n \in \sigma(X)$. Mais $\sigma(X) = \{\{X \in C\} : C \in \mathscr{E}\}$, donc chaque B_n est de la forme $B_n = \{X \in C_n\}$, pour un $C_n \in \mathscr{E}$. Du coup, $Y = \sum_{n\geq 0} a_n \mathbf{1}_{\{X\in C_n\}} = \varphi(X)$, avec $\varphi = \sum_{n\geq 0} a_n \mathbf{1}_{C_n}$, qui est bien sûr $(\mathscr{E}, \mathscr{B}(\mathbb{R}))$-mesurable. Si maintenant Y est quelconque, on écrit $Y_+ = \varphi_1(X)$, $Y_- = \varphi_2(X)$, et on conclut avec $\varphi = \varphi_1 - \varphi_2$.

Le point (ii) n'est que l'application du (i) à la variable $X = (X_1, \dots, X_n)$, qui est une v.a. à valeurs dans $E_1 \times \cdots \times E_N$ (muni de la tribu $\mathscr{E}_1 \otimes \cdots \otimes \mathscr{E}_n$). ■

Définition 98 *(a) Soit $A_i \in \mathscr{F}$, $i \in I$, une famille d'événements. On dit qu'ils sont indépendants si pour tout sous-ensemble* **fini** *d'indices $J \subset I$, on a $\mathbf{P}(\cap_{j\in J} A_j) = \prod_{j\in J} \mathbf{P}(A_j)$.*

(b) Soit $\mathscr{R}_i$, $i \in I$, une famille de classes de sous-ensembles telles que $\mathscr{R}_i \subset \mathscr{F}$, pour tout $i \in I$ (on ne suppose rien d'autre sur les $\mathscr{R}_i$). On dit que les classes $\mathscr{R}_i$, $i \in I$, sont indépendantes si toute famille d'événements $A_i \in \mathscr{R}_i$, $i \in I$, est indépendante sous $\mathbf{P}$, c'est-à-dire que $\forall J \subset I$ fini, $\forall A_j \in \mathscr{R}_j, j \in J$, $\mathbf{P}(\cap_{j\in J} A_j) = \prod_{j\in J} \mathbf{P}(A_j)$.

(c) Soit $(E_i, \mathscr{E}_i)$, $i \in I$, une famille d'espaces mesurables. Pour tout $i \in I$, soit X_i une v.a. à valeurs dans E_i. On dit que les v.a. X_i, $i \in I$ sont indépendantes si les tribus $\sigma(X_i)$, $i \in I$ sont indépendantes, i.e. si pour tout sous-ensemble fini d'indices $J \subset I$, pour tous $B_j \in \mathscr{E}_j$, $j \in J$, on a $\mathbf{P}(\cap_{j\in J}\{X_j \in B_j\}) = \prod_{j\in J} \mathbf{P}(X_j \in B_j)$.

Remarque 99 *Il est possible que 3 événements A_1, A_2, A_3 soient indépendants 2 à 2 sans être indépendants. Par exemple, on lance 2 pièces, $A_1 = \{$la première tombe sur pile$\}$, $A_2 = \{$la deuxième tombe sur face$\}$, $A_3 = \{$les 2 pièces tombent sur le même motif$\}$. On vérifie facilement que A_1, A_2, A_3 sont indépendants 2 à 2. Pourtant, $\mathbf{P}(A_1 \cap A_2 \cap A_3) = 0 \neq \mathbf{P}(A_1)\mathbf{P}(A_2)\mathbf{P}(A_3)$.*

Le résultat suivant peut se voir comme une réciproque du lemme de Borel-Cantelli.

Lemme 100 **(Lemme de Borel)** *Soit $A_n \in \mathscr{F}$, $n \in \mathbb{N}$, une suite d'événements indépendants. Si on a $\sum_{n\geq 0} \mathbf{P}(A_n) = \infty$, alors $\mathbf{P}(\limsup_{n\to\infty} A_n) = 1$.*

Preuve : Comme $\limsup_{n\to\infty} A_n = \cap_n B_n$, où $B_n = \cup_{k\geq n} A_k$ (qui est décroissante), il suffit (voir proposition 14-(ii)) de montrer que $\mathbf{P}(B_n) = 1$ pour tout n, i.e. que $\mathbf{P}(B_n^c) = 0$ pour tout n. Mais $B_n^c = \cap_{k\geq n} A_k^c$, et on vérifie facilement que les A_k^c sont indépendants. Du coup, $\mathbf{P}(B_n^c) = \lim_{N\to\infty} \mathbf{P}(\cap_{k=n}^N A_k^c) = \lim_{N\to\infty} \prod_{k=n}^N \mathbf{P}(A_k^c) = \prod_{k=n}^\infty \mathbf{P}(A_k^c)$. On écrit ensuite que $\prod_{k=n}^\infty \mathbf{P}(A_k^c) = \exp(\sum_{k=n}^\infty \log(1-\mathbf{P}(A_k)))$. En utilisant que $\sum_{k=n}^\infty \mathbf{P}(A_k) = \infty$, on conclut que $\sum_{k=n}^\infty \log(1-\mathbf{P}(A_k)) = -\infty$, et donc $\mathbf{P}(B_n^c) = 0$. ∎

Le théorème suivant est d'un usage très fréquent.

Théorème 101 *Soit $\mathscr{P}_i \subset \mathscr{F}$, $i \in I$, une famille de pi-systèmes. On suppose que les pi-systèmes $\mathscr{P}_i$, $i \in I$, sont indépendants. Alors les tribus $\sigma(\mathscr{P}_i)$, $i \in I$, le sont également.*

Preuve : par définition de l'indépendance, on peut supposer I fini, donc $I = \{1, \ldots, n\}$.

Montrons déjà que $\{\sigma(\mathscr{P}_1), \mathscr{P}_2, \ldots, \mathscr{P}_n\}$ sont indépendants. Pour cela, fixons $A_2 \in \mathscr{P}_2, \ldots, A_n \in \mathscr{P}_n$ et considérons $\mathscr{L} = \{B \in \sigma(\mathscr{P}_1) \ : \ \mathbf{P}(B \cap A_2 \cap \cdots \cap A_n) = \mathbf{P}(B)\mathbf{P}(A_2)\ldots\mathbf{P}(A_n)\}$. Il suffit de montrer que $\mathscr{L} = \sigma(\mathscr{P}_1)$. On montre facilement que $\mathscr{L}$ est une classe monotone. D'autre part, $\mathscr{L}$ contient $\mathscr{P}_1$ (par hypothèse). Par le théorème de la classe monotone, on a donc $\sigma(\mathscr{P}_1) \subset \mathscr{L}$, ce qu'on désirait.

En appliquant l'étape précédente à $\mathscr{P}_1^* = \mathscr{P}_2$, $\mathscr{P}_2^* = \sigma(\mathscr{P}_1)$, $\mathscr{P}_k^* = \mathscr{P}_k$ pour $k = 3, \ldots, n$, on voit que $\{\sigma(\mathscr{P}_1^*), \mathscr{P}_2^*, \ldots, \mathscr{P}_n^*\}$ sont indépendants, ce qui se réécrit $\{\sigma(\mathscr{P}_1), \sigma(\mathscr{P}_2), \mathscr{P}_3, \ldots, \mathscr{P}_n\}$ sont indépendants, etc. ∎.

Corollaire 102 **(Indépendance par paquets)** *Soit I, un ensemble d'indices et $I_j \subset I$, $j \in J$, une partition de I. Soit $\mathscr{F}_i \subset \mathscr{F}$, $i \in I$, une famille de tribus indépendantes. Les tribus $\mathscr{G}_j = \sigma(\mathscr{F}_i, i \in I_j)$, $j \in J$ sont alors indépendantes.*

Preuve : pour tout $j \in J$, on note

$$\mathscr{P}_j = \{A_1 \cap \ldots \cap A_n\,;\, n \in \mathbb{N},\, A_1, \ldots, A_n \in \cup_{i \in I_j} \mathscr{F}_i\}.$$

Il est facile de voir que $\mathscr{P}_j$ est un pi-système tel que $\sigma(\mathscr{P}_j) = \mathscr{G}_j$. L'indépendance des tribus $\mathscr{F}_i$, $i \in I$, entraîne facilement celle des $\mathscr{P}_j$, $j \in J$. Le théorème 101 permet de conclure. ■

Proposition 103 *Soit I un ensemble d'indices et, pour tout $i \in I$, un espace mesurable $(E_i, \mathscr{E}_i)$, un pi-système $\mathscr{C}_i \subset \mathscr{E}_i$ tel que $\sigma(\mathscr{C}_i) = \mathscr{E}_i$ et une v.a. X_i qui est $(\mathscr{F}, \mathscr{E}_i)$-mesurable. Les assertions suivantes sont équivalentes.*

(i) Les variables X_i, $i \in I$, sont indépendantes.

(ii) Pour tout $J \subset I$ fini, pour tous $B_j \in \mathscr{E}_j$, $j \in J$, $\mathbf{P}(\cap_{j \in J}\{X_j \in B_j\}) = \prod_{j \in J} \mathbf{P}(X_j \in B_j)$.

(iii) Pour tout $J \subset I$ fini, pour tous $B_j \in \mathscr{C}_j$, $j \in J$, $\mathbf{P}(\cap_{j \in J}\{X_j \in B_j\}) = \prod_{j \in J} \mathbf{P}(X_j \in B_j)$.

(iv) Pour tout $J = \{j_1, \ldots, j_n\} \subset I$, où les j_k sont distincts, $\mu_{(X_{j_1}, \ldots, X_{j_n})} = \mu_{X_{j_1}} \otimes \ldots \otimes \mu_{X_{j_n}}$.

(v) Pour tout $J \subset I$ fini, toutes $h_j : E_j \to \mathbb{R}$, $\mathscr{E}_j$-mesurables bornées (ou positives), $j \in J$,

$$\mathbf{E}\Big[\prod_{j \in J} h_j(X_j)\Big] = \prod_{j \in J} \mathbf{E}[h_j(X_j)].$$

Preuve : (i) est équivalent à (ii) par définition, (ii) implique (iii) est trivial, ainsi que (v) implique (ii) (prendre $h_j = \mathbf{1}_{B_j}$).

Montrons (iii) implique (i). Soit $\mathscr{P}_i = \{\{X_i \in B_i\} \;:\; B_i \in \mathscr{C}_i\}$. C'est un pi-système (car $\mathscr{C}_i$ en est un). (iii) dit précisément que les $\mathscr{P}_i, i \in I$ sont indépendants. Donc par la proposition 101, les $\sigma(\mathscr{P}_i), i \in I$ sont indépendants. Il nous faut donc montrer que $\sigma(\mathscr{P}_i) = \sigma(X_i)$. Mais $\{B \in \mathscr{E}_i \;:\; \{X_i \in B\} \in \sigma(\mathscr{P}_i)\}$ est une tribu (exercice) qui contient $\mathscr{C}_i$, donc $\mathscr{E}_i$. Ainsi, pour tout $B \in \mathscr{E}_i$, $\{X_i \in B\} \in \sigma(\mathscr{P}_i)$. Comme $\sigma(X_i) = \{\{X_i \in B\} \;:\; B \in \mathscr{E}_i\}$, on conclut que $\sigma(\mathscr{P}_i) = \sigma(X_i)$.

Montrons que (ii) implique (iv). $\mu_{X_{j_1}} \otimes \ldots \otimes \mu_{X_{j_n}}$ est l'unique mesure μ sur $E_{j_1} \times \cdots \times E_{j_n}$ t.q. pour tout $B_1 \in \mathscr{E}_{j_1}, \ldots, B_n \in \mathscr{E}_{j_n}$, on ait $\mu(B_1 \times \cdots \times B_n) = \mu_{X_{j_1}}(B_1) \ldots \mu_{X_{j_n}}(B_n) = \prod_1^n \mathbf{P}(X_{j_k} \in B_k)$. Or $\mu_{(X_{j_1}, \ldots, X_{j_n})}(B_1 \times \cdots \times B_n) = \mathbf{P}(X_{j_1} \in B_1, \ldots, X_{j_n} \in B_n) = \prod_1^n \mathbf{P}(X_{j_k} \in B_k)$ par (ii). Donc en effet, $\mu_{(X_{j_1}, \ldots, X_{j_n})} = \mu_{X_{j_1}} \otimes \ldots \otimes \mu_{X_{j_n}}$.

Enfin, montrons que (iv) implique (v) : on suppose $J = \{j_1, \ldots, j_n\}$ et on écrit

$$\mathbf{E}\Big[\prod_1^n h_{j_k}(X_{j_k})\Big] = \int_{E_{j_1} \times \cdots \times E_{j_n}} (\prod_1^n h_{j_k}) d\mu_{(X_{j_1}, \ldots, X_{j_n})} = \int_{E_{j_1} \times \cdots \times E_{j_n}} (\prod_1^n h_{j_k}) d(\mu_{X_{j_1}} \otimes \ldots \otimes \mu_{X_{j_n}}).$$

Donc, par Fubini,

$$\mathbf{E}\Big[\prod_1^n h_{j_k}(X_{j_k})\Big] = \prod_1^n \int_{E_{j_k}} h_{j_k} d\mu_{X_{j_k}} = \prod_1^n \mathbf{E}[h_{j_k}(X_{j_k})].$$

Cela termine la peuve. ■

Lemme 104 *Soient $X_1, \ldots, X_n$ des v.a.r. indépendantes.*

(a) Si les X_i sont positives, alors $\mathbf{E}[X_1 \times \ldots \times X_n] = \mathbf{E}[X_1] \times \ldots \times \mathbf{E}[X_n]$.

(b) Si les X_i sont intégrables, alors on a aussi $\mathbf{E}[X_1 \times \ldots \times X_n] = \mathbf{E}[X_1] \times \ldots \times \mathbf{E}[X_n]$.

Preuve : découle facilement du point (v) de la proposition précédente. ■

II.6 Loi du $0-1$ et loi des grands nombres.

Définition 105 *Soit X_n, $n \in \mathbb{N}$ une suite de v.a. La* tribu asymptotique *de cette suite est $\mathscr{T}_\infty = \cap_{n\geq 0}\sigma(X_k, k \geq n)$.*

Exemple 106 *Soient X_n, $n \in \mathbb{N}$, une suite de v.a.r. Alors les v.a. $Y = \limsup_n X_n$ et $Z = \liminf_n X_n$ sont $\mathscr{T}_\infty$-mesurables. Pour montrer que Y est $\mathscr{T}_\infty$-mesurable, il faut montrer que pour tout $B \in \mathscr{B}(\mathbb{R})$, $\{Y \in B\} \in \mathscr{T}_\infty$, i.e. que pour tout $n_0 \geq 0$, $\{Y \in B\} \in \sigma(X_k, k \geq n_0)$. C'est évident, puisque $\limsup_n X_n = \limsup_n X_{n_0+n}$.*

Définition 107 *Soit $\mathscr{T}$ une tribu sur Ω telle que $\mathscr{T} \subset \mathscr{F}$. On dit que $\mathscr{T}$ est $\mathbf{P}$-triviale si pour tout $A \in \mathscr{T}$, $\mathbf{P}(A) = 0$ ou $\mathbf{P}(A) = 1$.*

Lemme 108 *Soit $\mathscr{T}$, une sous-tribu de $\mathscr{F}$. Soit X une v.a.r. $\mathscr{T}$-mesurable. On suppose que $\mathscr{T}$ est $\mathbf{P}$-triviale. Alors, il existe $x_0 \in \mathbb{R}$ tel que $X = x_0$ p.s.*

Preuve : pour tout $x \in \mathbb{R}$, $F_X(x) = \mathbf{P}(X \leq x) = 0$ ou 1. Comme de plus $F_X(-\infty) = 0$, $F_X(+\infty) = 1$ et comme F_X est croissante, on en déduit que F_X saute de 0 à 1 en un certain x_0. Ainsi, comme F_X est continue à droite, $\mathbf{P}(X = x_0) = F_X(x_0) - F_X(x_0-) = 1$. ■

Théorème 109 **(Kolmogorov : loi du $0-1$)** *Soient X_n, $n \in \mathbb{N}$, une suite de v.a. indépendantes. Alors leur tribu asymptotique $\mathscr{T}_\infty$ est $\mathbf{P}$-triviale.*

Preuve : on pose $\mathscr{F}_n = \sigma(X_0, \ldots, X_n)$, $\mathscr{T}_n = \sigma(X_p, p > n)$. On a donc $\mathscr{T}_\infty = \bigcap_{n\geq 0} \mathscr{T}_n$. On voit que $\mathscr{F}_n$ est indépendante de $\mathscr{T}_n$ puis, comme $\mathscr{T}_\infty \subset \mathscr{T}_n$, que $\mathscr{F}_n$ est indépendante de $\mathscr{T}_\infty$. Donc $\mathscr{T}_\infty$ est indépendante de $\cup_{n\geq 0}\mathscr{F}_n$ et donc de $\sigma(\cup_{n\geq 0}\mathscr{F}_n)$ (car $\cup_{n\geq 0}\mathscr{F}_n$ est un pi-système). Mais bien sûr, $\sigma(\cup_{n\geq 0}\mathscr{F}_n) = \sigma(X_0, X_1, \ldots)$ et donc $\mathscr{T}_\infty \subset \sigma(\cup_{n\geq 0}\mathscr{F}_n)$. Donc $\mathscr{T}_\infty$ est indépendante de $\mathscr{T}_\infty$. En particulier, pour tout $A \in \mathscr{T}_\infty$, A est indépendant de A, i.e. $\mathbf{P}(A) = \mathbf{P}(A)^2$, puis $\mathbf{P}(A) \in \{0, 1\}$.

Théorème 110 **(Loi des grands nombres)** *Soit X_n une suite de v.a.r. i.i.d. et intégrables. Alors $\bar{X}_n = n^{-1}(X_1 + \cdots + X_n) \to \mathbf{E}[X_1]$ p.s.*

Preuve : Il suffit de montrer que pour tout $a > \mathbf{E}[X_1]$, p.s., $\limsup_n \bar{X}_n \leq a$. En effet, on en déduira que p.s., pour tout $k \geq 1$, $\limsup_n \bar{X}_n \leq \mathbf{E}[X_1] + 1/k$, puis p.s. $\limsup_n \bar{X}_n \leq \mathbf{E}[X_1]$, puis, en appliquant le résultat à la suite $-X_n$, que $\limsup_n(-\bar{X}_n) \leq -\mathbf{E}[X_1]$ p.s., i.e. $\liminf_n \bar{X}_n \geq \mathbf{E}[X_1]$ p.s. Ainsi, $\lim_n \bar{X}_n = \mathbf{E}[X_1]$ p.s.

On fixe donc $a > \mathbf{E}[X_1]$. On pose $S_0 = S_0^* = 0$ et, pour $n \geq 1$, $S_n = X_1 + \ldots + X_n$ et $S_n^* = X_2 + \ldots + X_{n+1}$, $M_n = \max_{0\leq k\leq n}(S_k - ak)$ et $M_n^* = \max_{0\leq k\leq n}(S_k^* - ak)$, puis

$$M_\infty = \lim_{n\to\infty} M_n = \sup_{n\geq 0} M_n \quad \text{et} \quad M_\infty^* = \lim_{n\to\infty} M_n^* = \sup_{n\geq 0} M_n^* .$$

On remarque ensuite que

$$\begin{aligned}
M_{n+1} &= \max\left(0, X_1 - a, X_1 - a + S_1^* - a, X_1 - a + S_2^* - 2a, \ldots, X_1 - a + S_n^* - na\right) \\
&= \max(0, X_1 - a + M_n^*) \\
&= M_n^* + X_1 + \max(-X_1 - M_n^*, -a) \\
&= M_n^* + X_1 - \min(X_1 + M_n^*, a).
\end{aligned}$$

Clairement M_n et M_n^* ont même loi. Donc

$$\mathbf{E}[X_1] = \mathbf{E}[\min(X_1 + M_n^*, a)] + \mathbf{E}[M_{n+1}] - \mathbf{E}[M_n] \geq \mathbf{E}[\min(X_1 + M_n^*, a)].$$

On remarque que $|\min(X_1 + M_n^*, a)| \leq |X_1| + a$ (car $M_n^* \geq 0$, séparer les cas (i) $a \leq X_1 + M_n^*$, (ii) $0 \leq X_1 + M_n^* \leq a$ et (iii) $X_1 + M_n^* \leq 0$ et $X_1 + M_n^* \leq a$). Donc par convergence dominée,

$$\mathbf{E}[X_1] \geq \mathbf{E}[\min(X_1 + M_\infty^*, a)]. \qquad (*)$$

Mais $\{M_\infty^* = \infty\}$ est dans la tribu asymptotique des X_n (exercice). Comme elle est $\mathbf{P}$-triviale par la loi du $0-1$, on a donc $\mathbf{P}(M_\infty^* = \infty) \in \{0, 1\}$.

Si on avait $M_\infty^* = \infty$ p.s., on aurait, par (*), $\mathbf{E}[X_1] \geq a$. Donc $M_\infty^* < \infty$ p.s. Du coup, $M_\infty < \infty$ p.s. (car de même loi que M_∞^*), ce qui signifie que pour tout $n \geq 0$, $S_n - an \leq M_\infty < \infty$ et donc $\limsup_n (S_n/n) \leq a$. ∎

II.7 Convergence en loi

Dans toute la section, on considère un ensemble E muni d'une distance d (et donc d'une tribu borélienne associée $\mathscr{B}(E)$). On note $\mathscr{M}_1(E)$ l'ensemble des probabilités sur $(E, \mathscr{B}(E))$. Soit $C_b(E)$ l'ensemble des fonctions continues et bornées de E dans $\mathbb{R}$. et soit $Lip_b(E)$ l'ensemble des fonctions Lipschitz et bornées de E dans $\mathbb{R}$. On rappelle que $f : E \mapsto \mathbb{R}$ est Lipschitz s'il existe une constante C telle que pour tout $x, y \in E$, on ait $|f(x) - f(y)| \leq Cd(x, y)$.

Définition 111 *(a) Soient $\mu_n \in \mathscr{M}_1(E)$, $n \in \mathbb{N} \cup \{\infty\}$. On dit que la suite $(\mu_n)_{n\in\mathbb{N}}$ converge étroitement vers μ_∞, et on note $\mu_n \Longrightarrow \mu_\infty$, si*

$$\forall\, \varphi \in C_b(E), \quad \lim_n \int_E \varphi\, d\mu_n = \int_E \varphi\, d\mu_\infty.$$

(b) Soient X_n, $n \in \mathbb{N} \cup \{\infty\}$, des v.a. à valeurs dans E. On dit que la suite de variables $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X_∞ si $\mu_{X_n} \Longrightarrow \mu_{X_\infty}$, i.e. si

$$\forall\, \varphi \in C_b(E), \quad \lim_n \mathbf{E}[\varphi(X_n)] = \mathbf{E}[\varphi(X)].$$

L'unicité de la limite étroite se déduit du lemme suivant (car $C_b(E) \subset Lip_b(E)$).

Lemme 112 *Soient $\mu, \nu \in \mathscr{M}_1(E)$. Si $\int_E \varphi d\mu = \int_E \varphi d\nu$ pour tout $\varphi \in Lip_b(E)$, alors $\mu = \nu$.*

Preuve : L'ensemble $\mathscr{P}$ des fermés de E est un pi-système qui engendre $\mathscr{B}(E)$. Par le théorème d'unicité du prolongement des mesures, il suffit de montrer que $\mu(F) = \nu(F)$ pour tout $F \in \mathscr{P}$. Pour $F \in \mathscr{P}$ et $k \geq 1$, on considère $\varphi_{k,F}(x) = (1 - kd(x, F))_+$, qui est Lipschitz bornée et qui décroit vers $\mathbf{1}_F(x)$ (exercice, utiliser que F est fermé). On a donc par hypothèse $\int_E \varphi_{k,F} d\mu = \int_E \varphi_{k,F} d\nu$, puis, par convergence dominée, $\mu(F) = \nu(F)$. ∎

Remarque 113 *Par contre, si X_n converge en loi vers X et vers Y, on n'a pas forcément $X = Y$ (du tout). On a seulement $\mu_X = \mu_Y$.*

Théorème 114 **(Théorème du porte-manteau incomplet)** *Soient X_n, $n \in \mathbb{N} \cup \{\infty\}$, des v.a. à valeurs dans E. Les assertions suivantes sont équivalentes.*

(i) X_n converge en loi vers X_∞.

(ii) $\forall \varphi \in Lip_b(E)$, $\lim_n \mathbf{E}[\varphi(X_n)] = \mathbf{E}[\varphi(X_\infty)]$.

(iii) $\forall F$ fermé de E, $\limsup_n \mathbf{P}(X_n \in F) \leq \mathbf{P}(X_\infty \in F)$.

(iv) $\forall O$ ouvert de E, $\liminf_n \mathbf{P}(X_n \in O) \geq \mathbf{P}(X_\infty \in O)$.

Preuve : (i) entraîne (ii) est évident, (iii) et (iv) sont équivalents (par passage au complémentaire).

Montrons que (ii) implique (iii). Pour F fermé et $k \geq 1$, on considère $\varphi_{k,F}(x) = (1 - kd(x, F))_+$, qui est Lipschitz bornée et décroît vers $\mathbf{1}_F(x)$. On peut donc écrire, pour chaque k,

$$\limsup_n \mathbf{P}(X_n \in F) \leq \limsup_n \mathbf{E}[\varphi_{k,F}(X_n)] = \mathbf{E}[\varphi_{k,F}(X_\infty)].$$

Mais, par convergence dominée, $\lim_k \mathbf{E}[\varphi_{k,F}(X_\infty)] = \mathbf{P}(X_\infty \in F)$.

Montrons que (iii) implique (i). Soit $\varphi : E \mapsto \mathbb{R}$ continue bornée. On veut $\lim_n \mathbf{E}[\varphi(X_n)] = \mathbf{E}[\varphi(X_\infty)]$. Il suffit de montrer que $\limsup_n \mathbf{E}[\varphi(X_n)] \leq \mathbf{E}[\varphi(X_\infty)]$ (on conclura en utilisant le même résultat pour $-\varphi$). Supposons que $a \leq \varphi \leq b$. On écrit $\mathbf{E}[\varphi(X_n)] = a + \int_a^b \mathbf{P}(\varphi(X_n) \geq r)dr$. Mais pour tout r, $\mathbf{P}(\varphi(X_n) \geq r) = \mathbf{P}(X_n \in \varphi^{-1}([r, \infty[))$. Comme $\varphi^{-1}([r, \infty[)$ est fermé (car φ est continue), on a par hypothèse $\limsup_n \mathbf{P}(X_n \in \varphi^{-1}([r, \infty[)) \leq \mathbf{P}(X_\infty \in \varphi^{-1}([r, \infty[)) = \mathbf{P}(\varphi(X_\infty) \geq r)$. On conclut facilement que $\limsup_n \int_a^b \mathbf{P}(\varphi(X_n) \geq r)dr \leq \int_a^b \mathbf{P}(\varphi(X_\infty) \geq r)dr$. Ainsi, $\limsup_n \mathbf{E}[\varphi(X_n)] \leq a + \int_a^b \mathbf{P}(\varphi(X_\infty) \geq r)dr = \mathbf{E}[\varphi(X_\infty)]$. ■

Proposition 115 *Soient (E, d) et (E', d'), deux espaces métriques. Soit $\Phi : E \to E'$, une fonction continue et X_n, $n \in \mathbb{N} \cup \{\infty\}$, des v.a. à valeurs dans E. Si X_n converge en loi vers X_∞, alors $\Phi(X_n)$ converge en loi vers $\Phi(X_\infty)$.*

Preuve : On veut montrer que pour tout $\varphi : E' \mapsto \mathbb{R}$ continue bornée, on a $\mathbf{E}[\varphi(\Phi(X_n))] \to \mathbf{E}[\varphi(\Phi(X_\infty))]$. Cela découle de la convergence en loi de X_n vers X_∞, puisque $\varphi \circ \Phi : E \mapsto \mathbb{R}$ est continue bornée. ■

Proposition 116 *Soient X_n, $n \in \mathbb{N} \cup \{\infty\}$, des v.a. à valeurs dans E. Si $X_n \to X_\infty$ en probabilité, alors $X_n \to X_\infty$ en loi.*

Preuve : il suffit de montrer que pour tout $\varphi \in Lip_b(E)$, $\lim_n \mathbf{E}[\varphi(X_n)] = \mathbf{E}[\varphi(X_\infty)]$. Soit donc $\varphi \in Lip_b(E)$ et $C > 0$ tel que $|\varphi(x) - \varphi(y)| \leq Cd(x, y)$. Pour $\epsilon > 0$ fixé, on écrit

$$\big|\mathbf{E}[\varphi(X_n)] - \mathbf{E}[\varphi(X_\infty)]\big| \leq \mathbf{E}[|\varphi(X_n) - \varphi(X_\infty)|] \leq C\epsilon + 2||\varphi||_\infty \mathbf{P}(d(X_n, X_\infty) > \epsilon).$$

Donc $\limsup_n |\mathbf{E}[\varphi(X_n)] - \mathbf{E}[\varphi(X_\infty)]| \leq C\epsilon$ pour tout $\epsilon > 0$, puis $\lim_n |\mathbf{E}[\varphi(X_n)] - \mathbf{E}[\varphi(X_\infty)]| = 0$. ■

Proposition 117 *Soient X_n, $n \in \mathbb{N}$, des v.a. à valeurs dans E et $a \in E$ (déterministe). Si $X_n \to a$ en loi, alors $X_n \to a$ en probabilités.*

Preuve : Pour $\epsilon > 0$, on introduit la fonction continue bornée $\varphi_\epsilon(x) = \min\{1, d(x, a)/\epsilon\}$, on sait que $\mathbf{E}[\varphi_\epsilon(X_n)] \to \mathbf{E}[\varphi_\epsilon(a)] = 0$. Mais $\mathbf{1}_{\{d(x,a)>\epsilon\}} \leq \varphi_\epsilon(x)$, d'où $P(d(X_n, a) > \epsilon) \leq \mathbf{E}[\varphi_\epsilon(X_n)] \to 0$. ■

On peut aussi montrer le résultat suivant, dont la preuve est assez technique.

Remarque 118 *Soient X_n, $n \in \mathbb{N} \cup \{\infty\}$, des v.a. à valeurs dans E. Soit $\varphi : E \mapsto \mathbb{R}$ bornée et $(\mathscr{E}, \mathscr{B}(\mathbb{R}))$-mesurable. Soit $C_\varphi \subset E$ l'ensemble des points de continuité de φ. Si X_n converge en loi vers X_∞ et si $X_\infty \in C_\varphi$ p.s., alors $\lim_n \mathbf{E}[\varphi(X_n)] = \mathbf{E}[\varphi(X_\infty)]$.*

C'est en fait très utile. Si par exemple, si $E = \mathbb{R}$, si $X_n \to X_\infty$, si la loi de X_∞ a une densité, alors $\lim_n \mathbf{E}[\varphi(X_n)] = \mathbf{E}[\varphi(X_\infty)]$ dès que $\varphi : \mathbb{R} \mapsto \mathbb{R}$ est bornée et continue ℓ-p.p.

II.8 Convergence en loi dans $\mathbb{R}^d$.

On note $C_c(\mathbb{R}^d)$ l'ensemble des fonctions continues à support compact $\mathbb{R}^d$ dans $\mathbb{R}$.

Lemme 119 *(a) Soient $\mu, \nu \in \mathscr{M}_1(\mathbb{R}^d)$. Si $\int_E \psi d\mu = \int_E \psi d\nu$ pour tout $\psi \in C_c(E)$, alors $\mu = \nu$.*

(b) La suite $\mu_n \in \mathscr{M}_1(\mathbb{R}^d)$, $n \in \mathbb{N}$ converge étroitement $\mu_\infty \in \mathscr{M}_1(\mathbb{R}^d)$ dès que

$$\forall \psi \in C_c(\mathbb{R}^d)\,, \quad \lim_{n\to\infty} \int_{\mathbb{R}^d} \psi \, d\mu_n = \int_{\mathbb{R}^d} \psi \, d\mu_\infty.$$

Preuve : Le point (b) (et l'unicité de la limite étroite) implique le point (a) (choisir $\mu_n = \mu$ et $\mu_\infty = \nu$). Montrons (b). Pour $R > 0$, soit $\phi_R(x) = (1 - d(x, B(0,R)))_+$. On a alors $\mathbf{1}_{B(0,R)} \le \phi_R \le \mathbf{1}_{B(0,R+1)}$ et $\phi_R \in C_c(\mathbb{R}^d)$. Pour $\varphi \in C_b(\mathbb{R}^d)$ et $n \in \mathbb{N} \cup \{\infty\}$,

$$\Big| \int_{\mathbb{R}^d} \varphi \, d\mu_n - \int_{\mathbb{R}^d} \varphi\phi_R \, d\mu_n \Big| \le \|\varphi\|_\infty \int_{\mathbb{R}^d} (1 - \phi_R) d\mu_n = \|\varphi\|_\infty \Big(1 - \int_{\mathbb{R}^d} \phi_R \, d\mu_n\Big),$$

puis

$$\Big| \int_{\mathbb{R}^d} \varphi \, d\mu_n - \int_{\mathbb{R}^d} \varphi \, d\mu_\infty \Big| \le \Big| \int_{\mathbb{R}^d} \varphi\phi_R \, d\mu_n - \int_{\mathbb{R}^d} \varphi\phi_R \, d\mu_\infty \Big| + \|\varphi\|_\infty \Big(2 - \int_{\mathbb{R}^d} \phi_R \, d\mu_n - \int_{\mathbb{R}^d} \phi_R \, d\mu_\infty\Big).$$

Comme $\varphi\phi_R$ et ϕ_R sont dans $C_c(\mathbb{R}^d)$, on en déduit que

$$\limsup_n \Big| \int_{\mathbb{R}^d} \varphi \, d\mu_n - \int_{\mathbb{R}^d} \varphi \, d\mu_\infty \Big| \le 2\|\varphi\|_\infty \Big(1 - \int_{\mathbb{R}^d} \phi_R \, d\mu_\infty\Big) \le 2\|\varphi\|_\infty \mu_\infty((B(0,R))^c).$$

Comme $\lim_{R\to\infty} \mu_\infty(B(0,R)) = 1$, on conclut facilement. ■

On note $\langle \cdot, \cdot \rangle$ le produit scalaire eucilidien sur $\mathbb{R}^d$.

Définition 120 *(a) On appelle transformée de Fourier de $\mu \in \mathscr{M}_1(\mathbb{R}^d)$ la fonction de $\hat{\mu} : \mathbb{R}^d \mapsto \mathbb{C}$ définie par*

$$\widehat{\mu}(\xi) = \int_{\mathbb{R}^d} e^{i\langle \xi, x\rangle} \mu(dx) = \int_{\mathbb{R}^d} \cos(\langle \xi, x\rangle) \mu(dx) + i \int_{\mathbb{R}^d} \sin(\langle \xi, x\rangle) \mu(dx).$$

(b) Pour X une v.a. à valeurs dans $\mathbb{R}^d$, on appelle fonction caractéristique de X la transformée de Fourier de sa loi $\widehat{\mu}_X(\xi) = \mathbf{E}[e^{i\langle \xi, X\rangle}]$.

Remarque 121 *Pour toute v.a. X à valeurs dans $\mathbb{R}^d$, $\widehat{\mu}_X$ est continue de $\mathbb{R}^d$ dans $\mathbb{C}$, bornée (en module) par 1 et $\widehat{\mu}_X(0) = 1$.*

Remarque 122 *Si $X \sim \mathscr{N}(0, \sigma^2 I_d)$, alors $\widehat{\mu}_X(\xi) = e^{-\sigma^2 ||\xi||^2/2}$.*

Preuve : comme $X = (X_1, \dots, X_d)$ avec les X_k i.i.d. $\mathscr{N}(0, \sigma^2)$, on a clairement, pour $\xi = (\xi_1, \dots, \xi_d)$, $\widehat{\mu}_X(\xi) = \widehat{\mu}_{X_1}(\xi_1) \times \cdots \times \widehat{\mu}_{X_d}(\xi_d)$. Il suffit donc de traiter le cas $d = 1$. On a par définition, puis en dérivant en ξ,

$$\widehat{\mu}_X(\xi) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\xi x} e^{-x^2/(2\sigma^2)} dx \quad \text{puis} \quad \widehat{\mu}'_X(\xi) = \frac{i}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\xi x} x e^{-x^2/(2\sigma^2)} dx.$$

En intégrant par parties, on trouve $\widehat{\mu}'_X(\xi) = -\xi\sigma^2 \widehat{\mu}_X(\xi)$, dont la solution est $\widehat{\mu}_X(\xi) = \widehat{\mu}_X(0) e^{-\sigma^2\xi^2/2}$. Comme $\widehat{\mu}_X(0) = 1$, ceci termine la preuve. ■

Voici les principaux résultats de cette section.

Théorème 123 *(i)* **(Injectivité)** *Si $\mu, \nu \in \mathscr{M}_1(\mathbb{R}^d)$ et si $\widehat{\mu}(\xi) = \widehat{\nu}(\xi)$ pour tout $\xi \in \mathbb{R}^d$, alors $\mu = \nu$.*

(ii) **(Théorème de Lévy)** *Si $\mu_n \in \mathscr{M}_1(\mathbb{R}^d)$, $n \in \mathbb{N} \cup \{\infty\}$ et si $\lim_n \widehat{\mu_n}(\xi) = \widehat{\mu_\infty}(\xi)$ pour tout $\xi \in \mathbb{R}^d$, alors $\mu_n \Longrightarrow \mu_\infty$.*

Preuve : Bien sûr (ii) implique (i). On divise la preuve (ii) en plusieurs étapes.

Etape 1. On note $g_\sigma(x) = (\sigma\sqrt{2\pi})^{-d} e^{-||x||^2/(2\sigma^2)}$ la densité de la loi $\mathscr{N}(0, \sigma^2 I_d)$. Pour $\mu \in \mathscr{M}_1(\mathbb{R}^d)$ et $x \in \mathbb{R}^d$, on pose $(g_\sigma \star \mu)(x) = \int_{\mathbb{R}^d} g_\sigma(x - y)\mu(dy)$. On vérifie ici que

$$(g_\sigma \star \mu)(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-\sigma^2||\xi||^2/2} e^{i\langle \xi, x\rangle} \widehat{\mu}(-\xi) d\xi.$$

Comme $\widehat{g_{1/\sigma}}(\xi) = (\sigma\sqrt{2\pi})^d g_\sigma(\xi)$ (voir la remarque 122), on a

$$(g_\sigma \star \mu)(x) = \frac{1}{(\sigma\sqrt{2\pi})^d} \int_{\mathbb{R}^d} \widehat{g_{1/\sigma}}(x - y)\mu(dy) = \frac{1}{(\sigma\sqrt{2\pi})^d} \int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} \frac{\sigma^d}{(\sqrt{2\pi})^d} e^{i\langle x-y, \xi\rangle} e^{-\sigma^2||\xi||^2/2} d\xi \Big) \mu(dy),$$

puis, par Fubini,

$$(g_\sigma \star \mu)(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} e^{i\langle x-y, \xi\rangle} \mu(dy) \Big) e^{-\sigma^2||\xi||^2/2} d\xi = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \Big(e^{i\langle x, \xi\rangle} \widehat{\mu}(-\xi) \Big) e^{-\sigma^2||\xi||^2/2} d\xi.$$

Etape 2. Soit $\psi \in C_c(\mathbb{R}^d)$. Elle est uniformément continue, et on définit son *module d'uniforme continuité* : pour $\delta > 0$, $\omega(\psi; \delta) = \sup\big\{|\psi(x) - \psi(y)| \ : \ \|x - y\| \le \delta\big\}$. On montre ici que pour tout $\psi \in C_c(\mathbb{R}^d)$, tout $\mu \in \mathscr{M}_1(\mathbb{R})$ et tout $\sigma > 0$,

$$\Delta(\psi, \mu, \sigma) = \Big| \int_{\mathbb{R}^d} \psi d\mu - \int_{\mathbb{R}^d} \psi(x)(g_\sigma \star \mu)(x) dx \Big| \le \int_{\mathbb{R}^d} g_1(z)\, \omega(\psi; \sigma\|z\|) dz.$$

Soit $X \sim \mu$ et $G \sim \mathscr{N}(0, I_d)$ indépendantes. Comme $\sigma G \sim \mathscr{N}(0, \sigma^2 I_d)$, on conclut que $X + \sigma G$ suit la loi de densité $g_\sigma \star \mu$. Du coup, on peut réécrire

$$\Delta(\psi, \mu, \sigma) = \Big| \mathbf{E}[\psi(X) - \psi(X + \sigma G)] \Big| \le \mathbf{E}[\omega(\psi; \sigma||G||)].$$

Comme enfin $G \sim g_1$, le membre de droite se réécrit bien $\int_{\mathbb{R}^d} \omega(\psi; \sigma||z||) g_1(z) dz$.

Etape 3. On conclut. On suppose donc que $\mu_n \in \mathscr{M}_1(\mathbb{R}^d)$ pour $n \in \mathbb{N} \cup \{\infty\}$ et que $\widehat{\mu_n}(\xi) \to \widehat{\mu_\infty}(\xi)$ pour tout $\xi \in \mathbb{R}^d$. On considère $\psi \in C_c(\mathbb{R}^d)$, et on doit montrer que $I_n = |\int_{\mathbb{R}^d} \psi d\mu_n - \int_{\mathbb{R}^d} \psi d\mu_\infty|$ tend vers 0. Pour $\sigma > 0$, on écrit $I_n \leq I_n^{1,\sigma} + I^{2,\sigma} + I_n^{3,\sigma}$, où

$$\begin{aligned}
I_n^{1,\sigma} &= \Big| \int_{\mathbb{R}^d} \psi d\mu_n - \int_{\mathbb{R}^d} \psi(x)(g_\sigma \star \mu_n)(x)dx \Big|, \\
I^{2,\sigma} &= \Big| \int_{\mathbb{R}^d} \psi d\mu_\infty - \int_{\mathbb{R}^d} \psi(x)(g_\sigma \star \mu_\infty)(x)dx \Big|, \\
I_n^{3,\sigma} &= \Big| \int_{\mathbb{R}^d} \psi(x) \Big[(g_\sigma \star \mu_n)(x) - (g_\sigma \star \mu_\infty)(x) \Big] dx \Big|.
\end{aligned}$$

En utilisant l'étape 2, on voit que pour tout n, $I_n^{1,\sigma} + I^{2,\sigma} \leq \epsilon(\sigma)$, où $\epsilon(\sigma) = 2 \int_{\mathbb{R}^d} g_1(z)\, \omega(\psi; \sigma \|z\|)(dz)$. Comme $\omega(\psi;\delta)$ est borné (par $2||\psi||_\infty$) et tend vers 0 quand $\delta \to 0$, on conclut, par convergence dominée, que $\epsilon(\sigma) \to 0$ quand $\sigma \to 0$.

Par l'étape 1, on a

$$I_n^{3,\sigma} = \frac{1}{(2\pi)^d} \Big| \int_{\mathbb{R}^d} \psi(x) \int_{\mathbb{R}^d} e^{-\sigma^2 ||\xi||^2/2} e^{i\langle \xi, x\rangle} \Big[\widehat{\mu_n}(-\xi) - \widehat{\mu_\infty}(-\xi) \Big] d\xi dx \Big|.$$

Donc à $\sigma > 0$ fixé, $\lim_n I_n^{3,\sigma} = 0$ par convergence dominée, car

- $\lim_n [\widehat{\mu_n}(-\xi) - \widehat{\mu_\infty}(-\xi)] = 0$ pour tout $\xi \in \mathbb{R}^d$ par hypothèse,
- $|\psi(x) e^{-\sigma^2||\xi||^2/2} e^{i\langle\xi,x\rangle} [\widehat{\mu_n}(-\xi) - \widehat{\mu_\infty}(-\xi)]| \leq 2|\psi(x)| e^{-\sigma^2||\xi||^2/2} \in \mathscr{L}^1(\mathbb{R}^{2d}, \mathscr{B}(\mathbb{R}^{2d}), d\xi dx)$ car ψ est à support compact.

Nous avons donc montré que pour tout $\sigma > 0$, on a $I_n \leq \epsilon(\sigma) + I_n^{3,\sigma}$, avec $\lim_n I_n^{3,\sigma} = 0$ (pour tout $\sigma > 0$) et $\lim_{\sigma \to 0} \epsilon(\sigma) = 0$. On conclut que $\lim_n I_n = 0$. ■

II.9 Vecteurs gaussiens

On rappelle que si $Z \sim \mathscr{N}(0,1)$, alors $m + \sigma Z \sim \mathscr{N}(m, \sigma)$, et donc $\widehat{\mu_Z}(\xi) = e^{i\xi m - \sigma^2 \xi^2/2}$. On rappelle aussi que $X \sim \mathscr{N}(m, 0)$ si $X = m$ p.s.

Définition 124 *Soit $\mathbf{X} = (X_1, \ldots, X_d)$ une v.a. à valeurs dans $\mathbb{R}^d$. C'est un vecteur gaussien si pour tout $\mathbf{u} = (u_1, \ldots, u_d) \in \mathbb{R}^d$, $\langle u, \mathbf{X} \rangle = u_1 X_1 + \ldots + u_d X_d$ est une v.a. gaussienne.*

Remarque 125 *Si $\mathbf{X}$ est un vecteur gaussien de dimension d et si M est une matrice réelle $d \times d$, alors $M\mathbf{X}$ est un vecteur gaussien.*

Exemple 126 *Si $X_1, \ldots, X_d$ sont i.i.d. $\mathscr{N}(0,1)$, alors $\mathbf{X} = (X_1, \ldots, X_n)$ est un vecteur gaussien. En effet, pour $\mathbf{u} = (u_1, \ldots, u_d) \in \mathbb{R}^d$ et $\xi \in \mathbb{R}$, par indépendance,*

$$\widehat{\mu}_{\langle \mathbf{u}, \mathbf{X} \rangle}(\xi) = \mathbf{E}\Big[\prod_{1 \leq k \leq d} e^{i\xi u_k X_k} \Big] = \prod_{1 \leq k \leq d} \mathbf{E}[e^{i\xi u_k X_k}] = e^{-\xi^2(u_1^2 + \ldots + u_d^2)/2} = e^{-\xi^2 \|\mathbf{u}\|^2/2}.$$

Donc $\langle \mathbf{u}, \mathbf{X} \rangle \sim \mathscr{N}(0, \|\mathbf{u}\|^2)$ par injectivité de la transformée de Fourier.

Rappelons-nous les définitions du vecteur moyenne et de la matrice de covariance.

Lemme 127 *Soit* $\mathbf{X}$*, un vecteur gaussien. Pour tout* $u \in \mathbb{R}^d$*,* $\langle u, \mathbf{X}\rangle \sim \mathscr{N}(\langle u, \mathbf{E}[\mathbf{X}]\rangle, \langle u, \Gamma_{\mathbf{X}} u\rangle)$*. Et pour tout* $\xi \in \mathbb{R}^d$*,* $\widehat{\mu_{\mathbf{X}}}(\xi) = \exp(i\langle \xi, \mathbf{E}[\mathbf{X}]\rangle - \langle \xi, \Gamma_{\mathbf{X}}\xi\rangle/2)$*.*

Preuve : On sait que $\langle u, \mathbf{X}\rangle$ est une v.a. gaussienne, que $\mathbf{E}[\langle u, \mathbf{X}\rangle] = \langle u, \mathbf{E}[\mathbf{X}]\rangle$ et Var $(\langle u, \mathbf{X}\rangle) = \langle u, \Gamma_{\mathbf{X}} u\rangle$. Donc forcément, $\langle u, \mathbf{X}\rangle \sim \mathscr{N}(\langle u, \mathbf{E}[\mathbf{X}]\rangle, \langle u, \Gamma_{\mathbf{X}} u\rangle)$. Ensuite, on voit que $\widehat{\mu}_{\mathbf{X}}(\xi) = \widehat{\mu}_{\langle \xi, \mathbf{X}\rangle}(1)$, et on conclut aisément en utilisant que $\langle \xi, \mathbf{X}\rangle \sim \mathscr{N}(\langle \xi, \mathbf{E}[\mathbf{X}]\rangle, \langle \xi, \Gamma_{\mathbf{X}}\xi\rangle)$. ■

Définition 128 *Le lemme précédent montre que la loi d'un vecteur Gaussien* $\mathbf{X}$ *est entièrement caractérisée par son vecteur moyenne* $\mathbf{v}$ *et sa matrice de covariance* Γ*. On note* $\mathscr{N}\big(\mathbf{v}, \Gamma\big)$ *cette loi.*

Proposition 129 *Soient* $\mathbf{v} \in \mathbb{R}^d$ *et* Γ*, une matrice réelle de taille* $d \times d$ *symétrique positive. Il existe un vecteur Gaussien* $\mathbf{X} \sim \mathscr{N}(\mathbf{v}, \Gamma)$*.*

Preuve : Γ est symétrique positive, donc admet une racine carrée S (i.e. une matrice symétrique positive $d \times d$ telle que $S^2 = \Gamma$). On considère $Y_1, \ldots, Y_d$ i.i.d. de loi $\mathscr{N}(0, 1)$, on pose $\mathbf{Y} = (Y_1, \ldots, Y_d)$ puis $\mathbf{X} = \mathbf{v} + S\mathbf{Y}$. On vérifie ensuite que $\mathbf{X}$ est un vecteur gaussien (comme dans l'exemple), que sa moyenne est $\mathbf{v}$, et que sa matrice de covariance est $S^2 = \Gamma$. ■

II.10 Théorème de la limite centrale

Théorème 130 *Soient* $\mathbf{X}_n$*,* $n \in \mathbb{N}$*, une suite de v.a. i.i.d. à valeurs dans* $\mathbb{R}^d$*,* $n \in \mathbb{N}$*. On suppose que* $\mathbf{E}[||\mathbf{X}_1||^2] < \infty$ *et on note* $\mathbf{v} = \mathbf{E}[\mathbf{X}_1]$ *et* $\Gamma = \Gamma_{X_1}$*. Alors, quand* $n \to \infty$*,*

$$\sqrt{n}\Big(\frac{1}{n}(\mathbf{X}_1 + \cdots + \mathbf{X}_n) - \mathbf{v}\Big) \quad \textit{converge en loi vers} \quad \mathscr{N}\big(0, \Gamma\big).$$

Lemme 131 *Soit* X *est une v.a.r. t.q.* $\mathbf{E}[|X|^2] < \infty$*. Il existe* $\epsilon : \mathbb{R} \mapsto \mathbb{C}$ *telle que* $\lim_{\xi \to 0} \epsilon(\xi) = 0$ *et*

$$\forall\, \xi \in \mathbb{R}^d, \quad \widehat{\mu}_X(\xi) = 1 + i\xi\mathbf{E}[X] - \frac{\xi^2}{2}\mathbf{E}[X^2] + |\xi|^2\epsilon(\xi).$$

Preuve : On vérifie (exercice) que pour $y \in \mathbb{R}$

$$\Big| e^{iy} - 1 - iy + \frac{y^2}{2} \Big| \leq y^2 \min\{|y|, 1\}.$$

Ainsi,

$$\Big|\widehat{\mu}_X(\xi) - 1 - i\xi\mathbf{E}[X] + \frac{\xi^2}{2}\mathbf{E}[X^2]\Big| = \Big|\mathbf{E}\Big[e^{i\langle \xi, X\rangle} - 1 - i\xi X + \frac{\xi^2}{2}X^2\Big]\Big| \leq |\xi|^2\mathbf{E}\big[X^2 \min\{1, |\xi X|\}\big].$$

Par convergence dominée, $\lim_{\xi \to 0} \mathbf{E}[X^2 \min\{2, |\xi X|\}] = 0$, ce qui achève la preuve. ■

Preuve du théorème : On suppose sans perte de généralité que $\mathbf{v} = 0$ on on procède en deux étapes.

On vérifie d'abord le théorème quand $d = 1$. On doit montrer que $Z_n = n^{-1/2}(X_1 + \cdots + X_n)$ converge en loi vers $\mathscr{N}(0, \mathbf{E}[X_1^2])$. Par le théorème de Lévy, il suffit de voir que pour tout $\xi \in \mathbb{R}$, quand $n \to \infty$, on a $\widehat{\mu}_{Z_n}(\xi) \to \exp(-\mathbf{E}[X_1^2]\xi^2/2)$. Mais par indépendance et identité des lois,

$$\widehat{\mu}_{Z_n}(\xi) = \mathbf{E}\Big[e^{i\xi(X_1 + \cdots + X_n)/\sqrt{n}}\Big] = \mathbf{E}\Big[e^{i\xi X_1/\sqrt{n}}\Big]^n = [\widehat{\mu}_{X_1}(\xi/\sqrt{n})]^n.$$

Par le lemme, on a de plus l'existence d'une fonction $\epsilon : \mathbb{R} \mapsto \mathbb{C}$ telle que $\lim_{\xi\to 0} \epsilon(\xi) = 0$ et

$$\forall\, \xi \in \mathbb{R}, \quad \widehat{\mu}_{X_1}(\xi) = 1 + i\xi\mathbf{E}[X_1] - \frac{\xi^2}{2}\mathbf{E}[X_1^2] + \xi^2\epsilon(\xi) = 1 - \frac{\xi^2}{2}\mathbf{E}[X_1^2] + \xi^2\epsilon(\xi).$$

En combinant ces deux résultats, on conclut qu'en effet, quand $n \to \infty$,

$$\widehat{\mu}_{Z_n}(\xi) = \Big(1 - \frac{\xi^2}{2n}\mathbf{E}[X_1^2] + \frac{\xi^2}{n}\epsilon(\xi)\Big)^2 \to \exp(-\mathbf{E}[X_1^2]\xi^2/2).$$

On suppose maintenant $d \geq 2$. On doit montrer que $\mathbf{Z}_n = n^{-1/2}(\mathbf{X}_1 + \cdots + \mathbf{X}_n)$ converge en loi vers $\mathscr{N}(0, \Gamma)$. Par le théorème de Lévy, il suffit de voir que pour tout $\xi \in \mathbb{R}^d$, quand $n \to \infty$, on a $\widehat{\mu_{Z_n}}(\xi) \to \exp(-\langle \xi, \Gamma\xi\rangle/2)$. Fixons donc $\xi \in \mathbb{R}^d$. Les v.a. $Y_k^\xi = \langle \xi, \mathbf{X}_k\rangle$ sont i.i.d., de dimension 1, et on a $\mathbf{E}[Y_1^\xi] = 0$ et Var $(Y_1^\xi) = \langle \xi, \Gamma\xi\rangle$. On sait donc par l'étape précédente que $Z_n^\xi = n^{-1/2}(Y_1^\xi + \ldots Y_n^\xi)$ converge en loi vers $\mathscr{N}(0, \langle \xi, \Gamma\xi\rangle)$. Donc en particulier, $\widehat{\mu}_{Z_n^\xi}(1)$ tend vers $\exp(-\langle \xi, \Gamma\xi\rangle/2)$. Comme finalement $\widehat{\mu}_{Z_n}(\xi) = \widehat{\mu}_{Z_n^\xi}(1)$, la preuve est complète. ■

Chapitre III

Espérance conditionnelle

Dans tout ce chapitre $(\Omega, \mathscr{F}, \mathbf{P})$ désigne un espace de probabilité. La tribu $\mathscr{F}$ est la *tribu de référence*. On dit que $\mathscr{G}$ est une sous-tribu de $\mathscr{F}$ si $\mathscr{G}$ est une tribu sur Ω et si $\mathscr{G} \subset \mathscr{F}$. Dans ce qui suit, il sera commode d'interpréter une sous-tribu $\mathscr{G}$ de $\mathscr{F}$ comme *une famille d'événements dont on sait déjà s'ils sont réalisés ou non.* La notion d'espérance conditionnelle est une façon de formaliser le problème suivant : *comment calculer la moyenne d'une variable aléatoire sachant que les évenements de $\mathscr{G}$ sont réalisés ou non ?*

Définition 132 *Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$. Soit X une v.a.r. positive (resp. intégrable). Une v.a.r. Y positive (resp. intégrable) est une version de l'espérance conditionnelle de X sachant $\mathscr{G}$ si*

$$\text{(a) } Y \text{ est } \mathscr{G}\text{-mesurable et (b) pour tout } B \in \mathscr{G}, \text{ on a } \mathbf{E}[Y\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_B].$$

Rappelons que $H = \mathrm{L}^2(\Omega, \mathscr{F}, \mathbf{P})$, muni du produit scalaire $\langle X_1, X_2 \rangle = \mathbf{E}[X_1 X_2]$, est un espace de Hilbert. Pour $\mathscr{G}$ une sous-tribu de $\mathscr{F}$, $F = \mathrm{L}^2(\Omega, \mathscr{G}, \mathbf{P})$ est un s.e.v. fermé de H. Si $X \in \mathrm{L}^2(\Omega, \mathscr{F}, \mathbf{P})$, on peut voir $Y = \mathbf{E}[X|\mathscr{G}]$ comme la projection orthogonale de X sur F. Voir la section I.11. Nous utiliserons ici Radon-Nikodym, ce qui n'est pas très différent, puisque la preuve du théorème de Radon-Nikodym utilise la projection orthogonale dans l'espace de Hilbert $\mathrm{L}^2(\Omega, \mathscr{F}, \mathbf{P})$.

On montre d'abord l'unicité p.s. de l'espérance conditionnelle.

Proposition 133 *Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$. Soient Y et Y' deux v.a.r. $\mathscr{G}$-mesurables positives (ou intégrables). Si*

$$\forall B \in \mathscr{G}, \quad \mathbf{E}[Y\mathbf{1}_B] \leq \mathbf{E}[Y'\mathbf{1}_B],$$

alors $Y \leq Y'$ p.s.

Preuve : pour $a < b$, on pose $C_{a,b} = \{Y' \leq a < b \leq Y\} \in \mathscr{G}$. On a $Y'\mathbf{1}_{C_{a,b}} \leq a\mathbf{1}_{C_{a,b}} \leq b\mathbf{1}_{C_{a,b}} \leq Y\mathbf{1}_{C_{a,b}}$. En prenant l'espérance, on trouve $\mathbf{E}[Y'\mathbf{1}_{C_{a,b}}] \leq a\mathbf{P}(C_{a,b}) \leq b\mathbf{P}(C_{a,b}) \leq \mathbf{E}[Y\mathbf{1}_{C_{a,b}}]$. Mais par hypothèse, $\mathbf{E}[Y\mathbf{1}_{C_{a,b}}] \leq \mathbf{E}[Y'\mathbf{1}_{C_{a,b}}]$. Donc $a\mathbf{P}(C_{a,b}) = b\mathbf{P}(C_{a,b})$. Comme $a < b$, on conclut que $\mathbf{P}(C_{a,b}) = 0$. Comme $\{Y' < Y\} = \bigcup_{a,b\in\mathbb{Q}, a<b} C_{a,b}$, on voit que $\mathbf{P}(Y' < Y) = 0$. ■

Corollaire 134 *Si Y et Y' sont deux versions de l'espérance conditionnelle de X sachant $\mathscr{G}$, alors $Y = Y'$ presque sûrement. On emploie donc l'abus suivant : on parle de l' espérance conditionnelle de X sachant $\mathscr{G}$, au lieu de "version de l'espérance conditionnelle", et on note toutes les versions de l'espérance conditionnelle de X sachant $\mathscr{G}$ de la même manière, $\mathbf{E}[X|\mathscr{G}]$.*

Théorème 135 **(Existence de l'espérance conditionnelle)** *Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$ et X une v.a.r. positive (ou intégrable). Alors $\mathbf{E}[X|\mathscr{G}]$ existe.*

Preuve : Soit d'abord X une v.a.r. positive. Pour $B \in \mathscr{G}$, on pose $\nu(B) = \mathbf{E}[X\mathbf{1}_B] = \int_B X\, d\mathbf{P}$. Alors ν est une mesure sur $\mathscr{G}$, et elle est σ-finie (voir le lemme 67). Donc le théorème de Radon-Nikodym implique l'existence de $Y : \Omega \mapsto [0, \infty[$, $\mathscr{G}$-mesurable, telle que pour tout $B \in \mathscr{F}$, $\nu(B) = \int_B Y d\mathbf{P} = \mathbf{E}[Y\mathbf{1}_B]$. On conclut donc en posant $\mathbf{E}[X|\mathscr{G}] = Y$: c'est une v.a.r. positive, $\mathscr{G}$-mesurable, et on a $\mathbf{E}[Y\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_B]$ pour tout $B \in \mathscr{G}$. Remarquons d'ores et déjà que $\mathbf{E}[\mathbf{E}[X|\mathscr{G}]] = \mathbf{E}[Y] = \mathbf{E}[X]$: il suffit de choisir $B = \Omega$.

Si maintenant X est une v.a.r. intégrable, on introduit X_+ et X_-, et on montre aisément que $Y = \mathbf{E}[X_+|\mathscr{G}] - \mathbf{E}[X_-|\mathscr{G}]$ convient. En particulier, Y est intégrable car $\mathbf{E}[Y] \leq \mathbf{E}[\mathbf{E}[X_+|\mathscr{G}]] + \mathbf{E}[\mathbf{E}[X_-|\mathscr{G}]] = \mathbf{E}[X_+] + \mathbf{E}[X_-] = \mathbf{E}[|X|] < \infty$. ■

Proposition 136 *Soit $\mathscr{G}$, une sous-tribu de $\mathscr{F}$ et X, Y deux v.a.r. positives (resp. intégrables).*

(i) Si $\mathscr{G} = \{\emptyset, \Omega\}$ (tribu grossière), alors $\mathbf{E}[X|\mathscr{G}] = \mathbf{E}[X]$ p.s.

(ii) Si X est $\mathscr{G}$-mesurable, $\mathbf{E}[X|\mathscr{G}] = X$ p.s.

(iii) Pour tout $a \in \mathbb{R}_+$ (resp. tout $a \in \mathbb{R}$), $\mathbf{E}[X + aY|\mathscr{G}] = \mathbf{E}[X|\mathscr{G}] + a\mathbf{E}[Y|\mathscr{G}]$ p.s.

(iv) Si $X \leq Y$ p.s., alors $\mathbf{E}[X|\mathscr{G}] \leq \mathbf{E}[Y|\mathscr{G}]$ p.s.

(v) $|\mathbf{E}[X\,|\,\mathscr{G}]| \leq \mathbf{E}[|X|\,\big|\,\mathscr{G}]$ p.s.

(vi) $\mathbf{E}\big[\mathbf{E}[X|\mathscr{G}]\big] = \mathbf{E}[X]$.

Preuve : (i) il suffit de vérifier que $\mathbf{E}[X]$ est $\mathscr{G}$-mesurable (évident) et que pour $B \in \{\emptyset, \Omega\}$, on a bien $\mathbf{E}[X\mathbf{1}_B] = \mathbf{E}[\mathbf{E}[X]\mathbf{1}_B]$ (évident, traiter les cas $B = \emptyset$ puis $B = \Omega$).

(ii) il suffit de vérifier que X est $\mathscr{G}$-mesurable (hypothèse) et que pour $B \in \mathscr{G}$, on a bien $\mathbf{E}[X\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_B]$ (trivial).

(iii) On pose $X' = \mathbf{E}[X|\mathscr{G}]$ et $Y' = \mathbf{E}[X|\mathscr{G}]$, puis $Z' = X' + aY'$. C'est une v.a. positive (resp. intégrable), $\mathscr{G}$-mesurable, et pour $B \in \mathscr{G}$, $\mathbf{E}[Z'\mathbf{1}_B] = \mathbf{E}[\mathbf{E}[X|\mathscr{G}]\mathbf{1}_B] + a\mathbf{E}[\mathbf{E}[Y|\mathscr{G}]\mathbf{1}_B]$, qui n'est autre que $\mathbf{E}[X\mathbf{1}_B] + a\mathbf{E}[Y\mathbf{1}_B]$.

(iv) Posons $X' = \mathbf{E}[X|\mathscr{G}]$ et $Y' = \mathbf{E}[X|\mathscr{G}]$. Pour tout $B \in \mathscr{G}$, on a $\mathbf{E}[X'\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_B] \leq \mathbf{E}[Y\mathbf{1}_B] = \mathbf{E}[Y'\mathbf{1}_B]$. Comme de plus Y, Y' sont $\mathscr{G}$-mesurables, la proposition 133 implique que $Y \leq Y'$.

Les points (v) et (vi) ont plus ou moins déjà été vus dans la preuve du théorème précédent. ■

Lemme 137 *Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$, soit X une v.a.r. positive (resp. intégrable). Pour toute v.a.r. Z $\mathscr{G}$-mesurable et positive (resp. bornée), on a $\mathbf{E}[Z\mathbf{E}[X|\mathscr{G}]] = \mathbf{E}[ZX]$.*

Preuve : traitons e.g. le cas $X, Z \geq 0$. Par le lemme 27, on écrit $Z = \sum_n a_n \mathbf{1}_{B_n}$, avec $a_n \geq 0$ et $B_n \in \mathscr{G}$. Alors, en notant $Y = \mathbf{E}[X|\mathscr{G}]$,

$$\mathbf{E}[ZY] = \mathbf{E}\Big[\sum_{n\geq 0} a_n \mathbf{1}_{B_n} Y\Big] = \sum_{n\geq 0} a_n \mathbf{E}[\mathbf{1}_{B_n} Y] = \sum_{n\geq 0} a_n \mathbf{E}[\mathbf{1}_{B_n} X] = \mathbf{E}\Big[\sum_{n\geq 0} a_n \mathbf{1}_{B_n} X\Big] = \mathbf{E}[ZX].$$

On a utilisé l'interversion série/espérance positive et la définition de l'espérance conditionnelle. ■

Proposition 138 *Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$ et X une v.a.r. positive (resp. intégrable). Alors pour toute v.a.r. Z $\mathscr{G}$-mesurable positive (resp. bornée),* $\mathbf{E}[ZX|\mathscr{G}] = Z\mathbf{E}[X|\mathscr{G}]$ *p.s.*

Preuve : on suppose e.g. $X, Z \geq 0$. Il faut vérifier que $Z\mathbf{E}[X|\mathscr{G}]$ est $\mathscr{G}$-mesurable (évident) et que pour $B \in \mathscr{G}$, $\mathbf{E}[Z\mathbf{E}[X|\mathscr{G}]\mathbf{1}_B] = \mathbf{E}[XZ\mathbf{1}_B]$. Cela découle du lemme précédent appliqué à X et à $Z = Z\mathbf{1}_B$ (abus de notation). ■

Lemme 139 *Soit $\mathscr{G}$ une sous-tribu de $\mathscr{F}$ et X une v.a.r. positive (resp. intégrable) indépendante de $\mathscr{G}$. Alors* $\mathbf{E}[X|\mathscr{G}] = \mathbf{E}[X]$ *p.s.*

Preuve : on suppose e.g. $X \geq 0$. Il faut vérifier que $\mathbf{E}[X]$ est $\mathscr{G}$-mesurable (évident) et que pour $B \in \mathscr{G}$, $\mathbf{E}[X\mathbf{1}_B] = \mathbf{E}[\mathbf{E}[X]\mathbf{1}_B]$ (évident). ■

Proposition 140 *Soient $\mathscr{G}_1 \subset \mathscr{G}_2$ deux sous-tribus de $\mathscr{F}$ et X une v.a.r. positive (resp. intégrable). Alors* $\mathbf{E}[\mathbf{E}[X|\mathscr{G}_2]\,|\mathscr{G}_1] = \mathbf{E}[X|\mathscr{G}_1]$ *p.s.*

Preuve : On suppose e.g. $X \geq 0$. On doit montrer que $Y = \mathbf{E}[\mathbf{E}[X|\mathscr{G}_2]\,|\mathscr{G}_1]$ est $\mathscr{G}_1$-mesurable (évident) et que pour $B \in \mathscr{G}_1$, on a $\mathbf{E}[Y\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_B]$. Mais

$$\mathbf{E}[Y\mathbf{1}_B] = \mathbf{E}[\mathbf{E}[\mathbf{E}[X|\mathscr{G}_2]\,|\mathscr{G}_1]\mathbf{1}_B] = \mathbf{E}[\mathbf{E}[\mathbf{E}[X|\mathscr{G}_2]\mathbf{1}_B|\mathscr{G}_1]] = \mathbf{E}[\mathbf{E}[\mathbf{E}[X\mathbf{1}_B|\mathscr{G}_2]|\mathscr{G}_1]].$$

On a utilisé que $\mathbf{1}_B$ est $\mathscr{G}_1$-mesurable (on peut donc le rentrer dans l'espérance conditionnelle sachant $\mathscr{G}_1$ par la proposition 138), puis que $\mathbf{1}_B$ est $\mathscr{G}_2$-mesurable (on peut donc le rentrer dans l'espérance conditionnelle sachant $\mathscr{G}_2$ par la proposition 138). En utilisant deux fois la proposition 136-(vi) (l'espérance de l'espérance conditionnelle est l'espérance), on conclut que $\mathbf{E}[Y\mathbf{1}_B] = \mathbf{E}[X\mathbf{1}_B]$. ■

Proposition 141 *Soit $\mathscr{G}_i$, $i \in I$, une famille de sous-tribus de $\mathscr{F}$ et $\mathscr{G} = \sigma(\mathscr{G}_i, i \in I)$. Soit X une v.a.r. positive (ou intégrable) et Y une v.a.r. positive (resp. intégrable) et $\mathscr{G}$-mesurable. Alors $Y = \mathbf{E}[X|\mathscr{G}]$ p.s. dès que pour tout sous-ensemble* **fini** *d'indices $J \subset I$, pour tous $B_j \in \mathscr{G}_j$, $j \in J$, on a* $\mathbf{E}\big[X\mathbf{1}_{\bigcap_{j\in J} B_j}\big] = \mathbf{E}\big[Y\mathbf{1}_{\bigcap_{j\in J} B_j}\big]$.

Preuve : Supposons X positive, considérons Y comme dans l'énoncé et posons $\mathscr{L} = \{B \in \mathscr{G} : \mathbf{E}[X\mathbf{1}_B] = \mathbf{E}[Y\mathbf{1}_B]\}$. On veut montrer que $\mathscr{L} = \mathscr{G}$. Par hypothèse, on sait que $\mathscr{P} \subset \mathscr{L}$, où $\mathscr{P} = \{\bigcap_{j\in J} B_j; B_j \in \mathscr{G}_j\,,\ J \subset I \text{ fini}\}$. Mais $\mathscr{P}$ est un pi-système (stable par intersection finie et contient Ω). De plus, $\mathscr{L}$ est une classe monotone, car

• $\Omega \in \mathscr{P} \subset \mathscr{L}$;

• si $A, B \in \mathscr{L}$ et si $A \subset B$, alors $\mathbf{E}[X\mathbf{1}_{B\setminus A}] = \mathbf{E}[X\mathbf{1}_B] - \mathbf{E}[X\mathbf{1}_A] = \mathbf{E}[Y\mathbf{1}_B] - \mathbf{E}[Y\mathbf{1}_A] = \mathbf{E}[Y\mathbf{1}_{B\setminus A}]$, donc $B \setminus A \in \mathscr{L}$;

• si $B_n \in \mathscr{L}$ et $B_n \subset B_{n+1}$, alors en passant à la limite, par convergence monotone, dans $\mathbf{E}[X\mathbf{1}_{B_n}] = \mathbf{E}[Y\mathbf{1}_{B_n}]$, on trouve que $\mathbf{E}[X\mathbf{1}_{\cup_n B_n}] = \mathbf{E}[Y\mathbf{1}_{\cup B_n}]$. Autrement dit, $\cup_n B_n \in \mathscr{L}$.

Le théorème de la classe monotone nous dit que $\sigma(\mathscr{P}) \subset \mathscr{L}$, i.e. que $\mathscr{G} \subset \mathscr{L}$. ■

Proposition 142 *Soient $\mathscr{G}_1$ et $\mathscr{G}_1$, deux sous-tribus de $\mathscr{F}$. Soit X, une v.a.r. positive (ou intégrable). On suppose que $\mathscr{G}_2$ est indépendante de $\sigma(\sigma(X), \sigma(\mathscr{G}_1))$. Alors* $\mathbf{E}[X|\sigma(\mathscr{G}_1, \mathscr{G}_2)] = \mathbf{E}[X|\mathscr{G}_1]$ *p.s.*

Preuve : par la proposition précédente, il suffit de vérifier que $\mathbf{E}[X|\mathscr{G}_1]$ est $\sigma(\mathscr{G}_1, \mathscr{G}_2)$-mesurable (évident) et que pour tout $B_1 \in \mathscr{G}_1$ et $B_2 \in \mathscr{G}_2$, on a $\mathbf{E}[\mathbf{E}[X|\mathscr{G}_1]\mathbf{1}_{B_1}\mathbf{1}_{B_2}] = \mathbf{E}[X\mathbf{1}_{B_1}\mathbf{1}_{B_2}]$. Mais en utilisant que $B_1 \in \mathscr{G}_1$ puis que B_2 est indépendant de $\mathbf{E}[X|\mathscr{G}_1]$ et de $X\mathbf{1}_{B_1}$, on peut écrire

$$\mathbf{E}[\mathbf{E}[X|\mathscr{G}_1]\mathbf{1}_{B_1}\mathbf{1}_{B_2}] = \mathbf{E}[\mathbf{E}[X\mathbf{1}_{B_1}|\mathscr{G}_1]\mathbf{1}_{B_2}] = \mathbf{E}[\mathbf{E}[X\mathbf{1}_{B_1}|\mathscr{G}_1]]\mathbf{E}[\mathbf{1}_{B_2}] = \mathbf{E}[X\mathbf{1}_{B_1}]\mathbf{E}[\mathbf{1}_{B_2}] = \mathbf{E}[X\mathbf{1}_{B_1}\mathbf{1}_{B_2}].$$

C'est ce qu'on désirait. ■

Proposition 143 *Soit $\mathscr{G}$, une sous-tribu de $\mathscr{F}$ et $(X_n)_{n\geq 0}$, une suite de v.a.r.*

(Convergence monotone) *Si $X_{n+1} \geq X_n \geq 0$ p.s. $\forall\, n \geq 0$, alors $\mathbf{E}[\sup_n X_n|\mathscr{G}] = \lim_n \mathbf{E}[X_n|\mathscr{G}]$ p.s.*

(Interversion positive) *Si $X_n \geq 0$ p.s. $\forall\, n \geq 0$, alors $\mathbf{E}[\sum_n X_n|\mathscr{G}] = \sum_n \mathbf{E}[X_n|\mathscr{G}]$ p.s.*

(Fatou) *Si $X_n \geq 0$ p.s. $\forall\, n \geq 0$, alors $\mathbf{E}[\liminf_n X_n\,|\mathscr{G}] \leq \liminf_n \mathbf{E}[X_n|\mathscr{G}]$ p.s.*

(Interversion L^1) *Si $\sum_n \mathbf{E}[|X_n|] < \infty$, alors $\mathbf{E}[\sum_n X_n|\mathscr{G}] = \sum_n \mathbf{E}[X_n|\mathscr{G}]$ p.s.*

(Convergence dominée) *Si $\lim_n X_n$ existe p.s. et s'il existe Z intégrable telle que $|X_n| \leq Z$ p.s. $\forall\, n \geq 0$, alors $\mathbf{E}[\lim_n X_n|\mathscr{G}] = \lim_n \mathbf{E}[X_n|\mathscr{G}]$ p.s.*

Preuve : Tout se déduit de la convergence monotone exactement comme dans le cas non conditionnel. Montrons donc juste le théorème de convergence monotone. On voit que $X_n \leq X_{n+1}$ implique que $\mathbf{E}[X_n|\mathscr{G}] \leq \mathbf{E}[X_{n+1}|\mathscr{G}]$ (par la proposition 136 (*iv*)). On peut donc poser $Y = \sup_n \mathbf{E}[X_n|\mathscr{G}]$, qui est une v.a.r. positive. On doit montrer que Y est $\mathscr{G}$-mesurable (évident) et que pour tout $B \in \mathscr{G}$, on a $\mathbf{E}[Y\mathbf{1}_B] = \mathbf{E}[(\sup_n X_n)\mathbf{1}_B]$. En appliquant (deux fois) le théorème de convergence monotone usuel,

$$\mathbf{E}[Y\mathbf{1}_B] = \lim_n \mathbf{E}[\mathbf{E}[X_n|\mathscr{G}]\mathbf{1}_B] = \lim_n \mathbf{E}[X_n\mathbf{1}_B] = \mathbf{E}[(\sup_n X_n)\mathbf{1}_B].$$

L'égalité du milieu utilise la définition de l'espérance conditionnelle. ■

Les preuves usuelles des inégalités s'étendent. Par exemple,

(Jensen conditionnelle) Soit X, une v.a.r. et $\varphi : \mathbb{R} \to \mathbb{R}$ convexe. Si X et $\varphi(X)$ sont intégrables, alors pour toute sous-tribu $\mathscr{G}$ de $\mathscr{F}$, on a $\varphi(\mathbf{E}[X|\mathscr{G}]) \leq \mathbf{E}[\varphi(X)|\mathscr{G}]$ p.s.

(Hölder conditionnelle) Soit $1 < p, q < \infty$ tels que $1/p + 1/q = 1$. Pour toutes v.a.r. X, Y et toute sous-tribu $\mathscr{G}$ de $\mathscr{F}$, on a $\mathbf{E}[|XY||\mathscr{G}] \leq (\mathbf{E}[|X|^p|\mathscr{G}])^{1/p}(\mathbf{E}[|Y|^q|\mathscr{G}])^{1/q}$ p.s.

Lemme 144 *Soit X une v.a.r. intégrable et $\mathscr{G}_i$, $i \in I$ une famille de sous-tribus de $\mathscr{F}$. Alors la famille $\{\mathbf{E}[X|\mathscr{G}_i], i \in I\}$ est U.I.*

Preuve : $X \in L^1$, donc par le théorème 94 de de la Vallée Poussin, il existe $\varphi : \mathbb{R}_+ \mapsto \mathbb{R}_+$ convexe avec $\lim_\infty \frac{\varphi(x)}{x} = \infty$ telle que $\mathbf{E}[\varphi(|X|)] < \infty$. Par Jensen, on a $\mathbf{E}[\varphi(|\mathbf{E}[X|\mathscr{G}_i]|)] \leq \mathbf{E}[\mathbf{E}[\varphi(|X|)|\mathscr{G}_i]] = \mathbf{E}[\varphi(|X|)]$. Donc $\sup_{i\in I} \mathbf{E}[\varphi(|\mathbf{E}[X|\mathscr{G}_i]|)] < \infty$ et, on conclut avec de la Vallée Poussin. ■

Proposition 145 *Soit $(E, \mathscr{E})$ un espace mesurable, soit V une v.a. à valeurs dans E et soit X une v.a.r. positive (resp. intégrable). On note alors $\mathbf{E}[X|V] = \mathbf{E}[X|\sigma(V)]$.*

(a) Il existe $h : E \mapsto \mathbb{R}$, $\mathscr{E}$-mesurable, telle que $\mathbf{E}[X|V] = h(V)$.

(b) Si $h : E \to \mathbb{R}$ est $\mathscr{E}$-mesurable, alors, $\mathbf{E}[X|V] = h(V)$ p.s. ssi pour toute fonction $\varphi : E \to \mathbb{R}$, $\mathscr{E}$-mesurable positive (resp. bornée), on a $\mathbf{E}[X\varphi(V)] = \mathbf{E}[h(V)\varphi(V)]$.

Preuve : le point (a) découle du théorème 97. Le second point découle du lemme 137 et du fait que les v.a. Z qui sont $\sigma(V)$-mesurables positives (resp. bornées) sont exactement les v.a. $\varphi(V)$, avec $\varphi : E \to \mathbb{R}$, $\mathscr{E}$-mesurables et positive (resp. bornée). ■

Lemme 146 *Soient X_1, X_2 deux v.a. indépendantes à valeurs dans E_1 et E_2, munis de tribus $\mathscr{E}_1$ et $\mathscr{E}_2$. Alors pour toute fonction $\Phi : E_1 \times E_2 \mapsto \mathbb{R}$, $\mathscr{E}_1 \otimes \mathscr{E}_2$-mesurable et positive (ou bornée), on a*

$$\mathbf{E}[\Phi(X_1, X_2)|X_1] = \Psi(X_1), \quad \textit{où } \Psi : E_1 \mapsto \mathbb{R} \textit{ est définie par } \quad \Psi(x_1) = \mathbf{E}[\Phi(x_1, X_2)] \quad \forall\, x_1 \in E_1.$$

Preuve : Il faut vérifier que pour tout $\varphi : E_1 \mapsto \mathbb{R}_+$, on a $\mathbf{E}[\Phi(X_1, X_2)\varphi(X_1)] = \mathbf{E}[\Psi(X_1)\varphi(X_1)]$. Mais on sait (voir proposition 103-(iv)) que $\mu_{(X_1,X_2)} = \mu_{X_1} \otimes \mu_{X_2}$. On a donc, par Fubini positif,

$$\mathbf{E}[\Phi(X_1, X_2)\varphi(X_1)] = \int_{E_1} \Big(\int_{E_2} \Phi(x_1, x_2)\varphi(x_1)\mu_{X_2}(dx_2) \Big) \mu_{X_1}(dx_1) = \int_{E_1} \Psi(x_1)\varphi(x_1)\mu_{X_1}(dx_1),$$

puisque $\Psi(x_1) = \int_{E_2} \Phi(x_1, x_2)\mu_{X_2}(dx_2)$ par définition. Pour conclure, il n'y a plus qu'à remarquer que $\int_{E_1} \Psi(x_1)\varphi(x_1)\mu_{X_1}(dx_1) = \mathbf{E}[\Psi(X_1)\varphi(X_1)]$. ■

Quelques exemples. On pose $V = \max(U_1, U_2)$ avec U_1, U_2 indépendantes et de loi $\mathscr{U}([0,1])$.

A. Calculons $\mathbf{E}[U_1|V]$. On cherche donc $h : [0,1] \mapsto \mathbb{R}_+$ telle que pour toute $\varphi : [0,1] \mapsto \mathbb{R}_+$ mesurable, on ait $\mathbf{E}[h(V)\varphi(V)] = \mathbf{E}[U_1\varphi(V)]$. On aura alors $\mathbf{E}[U_1|V] = h(V)$ p.s.

Déjà, $\mathbf{E}[h(V)\varphi(V)] = \mathbf{E}[h(U_1)\varphi(U_1)\mathbf{1}_{\{U_1 \geq U_2\}}] + \mathbf{E}[h(U_2)\varphi(U_2)\mathbf{1}_{\{U_1 < U_2\}}] = \int_0^1 2h(u)u\varphi(u)du$.

D'autre part, $\mathbf{E}[U_1\varphi(V)] = \mathbf{E}[U_1\varphi(U_1)\mathbf{1}_{\{U_1 \geq U_2\}}] + \mathbf{E}[U_1\varphi(U_2)\mathbf{1}_{\{U_2 \geq U_1\}}]$, ce qui donne $\mathbf{E}[U_1\varphi(V)] = \int_0^1 u^2\varphi(u)du + \int_0^1 \frac{u^2}{2}\varphi(u)du = \int_0^1 \frac{3}{2}u^2\varphi(u)du$.

Il faut donc $2h(u)u = \frac{3}{2}u^2$, i.e. $h(u) = \frac{3}{4}u$. Ainsi, $\mathbf{E}[U_1|V] = \frac{3}{4}V$ p.s.

B. On fixe maintenant $f : [0,1] \mapsto \mathbb{R}_+$ mesurable et souhaite calculer $\mathbf{E}[f(U_1)|V]$. On cherche donc $h : \mathbb{R}_+ \mapsto \mathbb{R}_+$ telle que pour toute $\varphi : [0,1] \mapsto \mathbb{R}_+$ mesurable, on ait $\mathbf{E}[h(V)\varphi(V)] = \mathbf{E}[f(U_1)\varphi(V)]$. On aura alors $\mathbf{E}[f(U_1)|V] = h(V)$ p.s.

Comme précédemment (A), on a $\mathbf{E}[h(V)\varphi(V)] = \int_0^1 2h(u)u\varphi(u)du$.

D'autre part, $\mathbf{E}[f(U_1)\varphi(V)] = \mathbf{E}[f(U_1)\varphi(U_1)\mathbf{1}_{\{U_1 \geq U_2\}}] + \mathbf{E}[f(U_1)\varphi(U_2)\mathbf{1}_{\{U_2 \geq U_1\}}]$, ce qui donne $\mathbf{E}[f(U_1)\varphi(V)] = \int_0^1 uf(u)\varphi(u)du + \int_0^1 (\int_0^u f(x)dx)\varphi(u)du$.

Il faut donc que $2h(u)u = uf(u) + \int_0^u f(x)dx$, i.e. $h(u) = \frac{1}{2}f(u) + \frac{1}{2u}\int_0^u f(x)dx$.

Ainsi, $\mathbf{E}[f(U_1)|V] = \frac{1}{2}f(V) + \frac{1}{2V}\int_0^V f(x)dx$.

Ceci étant vrai pour toute fonction f mesurable positive, on conclut que la loi conditionnelle de U_1 sachant V est $\frac{1}{2}\delta_V(dx) + \frac{1}{2V}\mathbf{1}_{[0,V]}(x)dx$, ce qui peut se comprendre ainsi, avec un abus de langage : sachant V, avec probabilité $\frac{1}{2}$, on a $U_1 = V$, et avec probabilité $\frac{1}{2}$, U_1 suit la loi uniforme sur $[0, V]$.

C. On fixe encore $f : [0,1] \mapsto \mathbb{R}_+$ mesurable et souhaite calculer $\mathbf{E}[f(V)|U_1]$. On cherche donc $h : \mathbb{R}_+ \mapsto \mathbb{R}_+$ telle que pour toute $\varphi : [0,1] \mapsto \mathbb{R}_+$ mesurable, on ait $\mathbf{E}[h(U_1)\varphi(U_1)] = \mathbf{E}[f(V)\varphi(U_1)]$. On aura alors $\mathbf{E}[f(V)|U_1] = h(U_1)$ p.s.

Déjà, $\mathbf{E}[h(U_1)\varphi(U_1)] = \int_0^1 h(u)\varphi(u)du$.

D'autre part, $\mathbf{E}[f(V)\varphi(U_1)] = \mathbf{E}[f(U_1)\varphi(U_1)\mathbf{1}_{\{U_1 \geq U_2\}}] + \mathbf{E}[f(U_2)\varphi(U_1)\mathbf{1}_{\{U_2 \geq U_1\}}]$, ce qui donne $\mathbf{E}[f(V)\varphi(U_1)] = \int_0^1 uf(u)\varphi(u)du + \int_0^1 (\int_u^1 f(x)dx)\varphi(u)du$.

Il faut donc que $h(u) = uf(u) + \int_u^1 f(x)dx$.

Ainsi, $\mathbf{E}[f(V)|U_1] = U_1 f(U_1) + \int_{U_1}^1 f(x)dx$.

Ceci étant vrai pour toute fonction f mesurable positive, on conclut que la loi conditionnelle de V sachant U_1 est $U_1\delta_{U_1}(dx) + \mathbf{1}_{\{[U_1,1]\}}(x)dx$, ce qui peut se comprendre ainsi : sachant U_1, avec probabilité U_1, on a $V = U_1$, et avec probabilité $(1 - U_1)$, V suit la loi uniforme sur $[U_1, 1]$.

Chapitre IV

Martingales

IV.1 Définitions et propriétés

Définition 147 *Soit $\mathscr{F}_n$, $n \geq 0$, une suite de sous-tribus de $\mathscr{F}$. On dit que $(\mathscr{F}_n)_{n\geq 0}$ est une filtration si $\mathscr{F}_n \subset \mathscr{F}_{n+1}$ pour tout $n \geq 0$. On utilise la notation $\mathscr{F}_\infty := \sigma(\mathscr{F}_n, n \geq 0)$.*

Définition 148 *Soit $(\mathscr{F}_n)_{n\geq 0}$ une filtration. Une suite de v.a. $(X_n)_{n\geq 0}$ est dite*
- *$(\mathscr{F}_n)_{n\geq 0}$-adaptée si pour tout $n \geq 0$, X_n est $\mathscr{F}_n$-mesurable.*
- *$(\mathscr{F}_n)_{n\geq 0}$-prévisible si pour tout $n \geq 1$, X_n est $\mathscr{F}_{n-1}$-mesurable et si X_0 est $\mathscr{F}_0$-mesurable.*

Définition 149 *Soit $(\mathscr{F}_n)_{n\geq 0}$ une filtration et $(X_n)_{n\geq 0}$ une suite $(\mathscr{F}_n)_{n\geq 0}$-adaptée de v.a.r. intégrables. La suite $(X_n)_{n\geq 0}$ est une*
- *$(\mathscr{F}_n)_{n\geq 0}$-martingale si $\forall\, n \geq 0$, $\mathbf{E}[X_{n+1}|\mathscr{F}_n] = X_n$;*
- *$(\mathscr{F}_n)_{n\geq 0}$-sous-martingale si $\forall\, n \geq 0$, $\mathbf{E}[X_{n+1}|\mathscr{F}_n] \geq X_n$;*
- *$(\mathscr{F}_n)_{n\geq 0}$-sur-martingale si $\forall\, n \geq 0$, $\mathbf{E}[X_{n+1}|\mathscr{F}_n] \leq X_n$.*

Remarque 150 *Soit $(X_n)_{n\geq 0}$, une $(\mathscr{F}_n)_{n\geq 0}$-martingale (resp. sur/sous-martingale). Alors pour tout $m, n \in \mathbb{N}$, $\mathbf{E}[X_{n+m}|\mathscr{F}_n] = X_n$, (resp. $\leq X_n$, $\geq X_n$). De plus la suite $n \mapsto \mathbf{E}[X_n]$ est constante (resp. décroissante, resp. croissante).*

En utilisant Jensen conditionnel, on voit facilement que, sous réserve d'intégrabilité, si $(X_n)_{n\geq 0}$ est une martingale et si $\varphi : \mathbb{R} \mapsto \mathbb{R}$ est convexe, alors $(\varphi(X_n))_{n\geq 0}$ est une sous-martingale. Si $\varphi : \mathbb{R} \mapsto \mathbb{R}$ est concave, $(\varphi(X_n))_{n\geq 0}$ est une sur-martingale. Si maintenant $(X_n)_{n\geq 0}$ est une sous-martingale et si $\varphi : \mathbb{R} \mapsto \mathbb{R}$ est convexe et croissante, alors $(\varphi(X_n))_{n\geq 0}$ est une sous-martingale, etc.

Proposition 151 (Décomposition de Doob) *Soit $(X_n)_{n\geq 0}$ une suite de v.a.r. $(\mathscr{F}_n)_{n\geq 0}$-adaptées intégrables. Il existe une unique paire de suites $(M_n)_{n\geq 0}$, $(V_n)_{n\geq 0}$, telle que*

(i) $X_n = X_0 + M_n + V_n$, $n \geq 0$.

(ii) $(M_n)_{n\geq 0}$ est une $(\mathscr{F}_n)_{n\geq 0}$-martingale.

(iii) $(V_n)_{n\geq 0}$ est $(\mathscr{F}_n)_{n\geq 0}$-prévisible et $V_0 = 0$.

Preuve : Unicité. Considérons une paire $(M_n)_{n\geq 0}$, $(V_n)_{n\geq 0}$ satisfaisant les trois points. Par (i), on a $X_{n+1} - X_n = M_{n+1} - M_n + V_{n+1} - V_n$. En prenant l'espérance conditionnelle sachant $\mathscr{F}_n$, on trouve

$\mathbf{E}[X_{n+1}-X_n|\mathscr{F}_n] = V_{n+1}-V_n$. Comme $V_0 = 0$, on conclut que forcément, $V_n = \sum_{k=0}^{n-1} \mathbf{E}[X_{k+1}-X_k|\mathscr{F}_k]$. Puis forcément, $M_n = X_n - X_0 - V_n$.

Existence. On pose $V_0 = 0$, $V_n = \sum_{k=0}^{n-1} \mathbf{E}[X_{k+1} - X_k|\mathscr{F}_k]$ (pour $n \geq 1$), puis $M_n = X_n - X_0 - V_n$ (pour $n \geq 0$). Les points (i) et (iii) sont évidemment vérifiés. Pour (ii), on remarque que $M_{n+1} - M_n = (X_{n+1} - X_n) - (V_{n+1} - V_n) = (X_{n+1} - X_n) - \mathbf{E}[X_{n+1} - X_n|\mathscr{F}_n]$, d'où $\mathbf{E}[M_{n+1} - M_n|\mathscr{F}_n] = 0$. ■

IV.2 Convergence des martingales

Les martingales (resp. sur-martingales, sous-martingales) sont en quelque sorte des versions aléatoires des suites constantes (resp. décroissantes, croissantes). C'est pourquoi elles jouissent de bonnes propriétés de convergence quand $n \to \infty$ qui leur confèrent un rôle central en théorie des processus.

Inégalités de Doob. Commençons par montrer deux inégalités de Doob. La seconde est fort utile.

Lemme 152 **(Inégalité maximale)** *Soit $(X_n)_{n\geq 0}$ une $(\mathscr{F}_n)_{n\geq 0}$-sous-martingale constituée de variables positives et L^1. Alors pour tout $a > 0$, tout $n \in \mathbb{N}$,*

$$a\mathbf{P}\big(\sup_{0\leq m\leq n} X_m > a \big) \leq \mathbf{E}\big[X_n \mathbf{1}_{\{\sup_{0\leq m\leq n} X_m > a\}}\big].$$

(Inégalité de Doob) *Soit $(X_n)_{n\geq 0}$ une $(\mathscr{F}_n)_{n\geq 0}$-martingale et $p \in]1, \infty[$.*

$$\textit{Si} \quad \sup_{n\in\mathbb{N}} \mathbf{E}\big[\, |X_n|^p \,\big] < \infty, \quad \textit{alors} \quad \mathbf{E}\big[\sup_{n\in\mathbb{N}} |X_n|^p\big] \leq \big(\tfrac{p}{p-1}\big)^p \sup_{n\in\mathbb{N}} \mathbf{E}\big[\, |X_n|^p \,\big].$$

Preuve : on fixe $n \in \mathbb{N}$ et $a > 0$, et on introduit $T_a = \inf\{n \geq 0 : X_n > a\}$ (avec $\inf \emptyset = \infty$). Pour $m \leq n$, on a (comme $\{T_a = m\} = \{X_0 \leq a, \dots, X_{m-1} \leq a, X_m > a\} \in \mathscr{F}_m$)

$$a\mathbf{1}_{\{T_a=m\}} \leq X_m \mathbf{1}_{\{T_a=m\}} \leq \mathbf{E}[X_n|\mathscr{F}_m]\mathbf{1}_{\{T_a=m\}} = \mathbf{E}[X_n \mathbf{1}_{\{T_a=m\}}|\mathscr{F}_m].$$

Donc $a\mathbf{P}(T_a = m) \leq \mathbf{E}[X_n \mathbf{1}_{\{T_a=m\}}]$ puis, en sommant sur $m \leq n$, $a\mathbf{P}(T_a \leq n) \leq \mathbf{E}[X_n \mathbf{1}_{\{T_a\leq n\}}]$. L'inégalité maximale est prouvée, puisque $\{\sup_{0\leq m\leq n} X_m > a\} = \{T_a \leq n\}$.

Comme $x \mapsto |x|$ est convexe, $(|X_n|)_{n\geq 0}$ est une sous-martingale, et on peut appliquer l'inégalité maximale : en posant $S_n = \sup_{0\leq m\leq n} |X_m|$, on trouve $a\mathbf{P}(S_n > a) \leq \mathbf{E}[|X_n|\mathbf{1}_{\{S_n>a\}}]$, puis

$$\int_0^\infty a\mathbf{P}(S_n > a)pa^{p-2}da \leq \int_0^\infty \mathbf{E}[|X_n|\mathbf{1}_{\{S_n>a\}}]pa^{p-2}da.$$

Mais par Fubini positif,

- $\int_0^\infty a\mathbf{P}(S_n > a)pa^{p-2}da = \mathbf{E}[\int_0^{S_n} pa^{p-1}da] = \mathbf{E}[S_n^p]$,
- $\int_0^\infty \mathbf{E}[|X_n|\mathbf{1}_{\{S_n>a\}}]pa^{p-2}da = \mathbf{E}[|X_n| \int_0^{S_n} pa^{p-2}da] = \frac{p}{p-1}\mathbf{E}[|X_n|S_n^{p-1}]$.

Donc

$$\mathbf{E}[S_n^p] \leq \frac{p}{p-1}\mathbf{E}[|X_n|S_n^{p-1}] \leq \frac{p}{p-1}\mathbf{E}[|X_n|^p]^{1/p}\mathbf{E}[|S_n|^p]^{1-1/p}$$

par Hölder, d'où

$$\mathbf{E}[S_n^p] \leq \Big(\frac{p}{p-1}\Big)^p \mathbf{E}[|X_n|^p] \leq \Big(\frac{p}{p-1}\Big)^p \sup_{k\geq 0} \mathbf{E}[|X_k|^p].$$

Il n'y a plus qu'à faire tendre $n \to \infty$ à l'aide de la convergence monotone. ■

Convergence p.s. Les deux résultats principaux de convergence sont les suivants.

Théorème 153 *(i) Toute sur-martingale positive $(X_n)_{n\geq 0}$ converge p.s.*

(ii) Toute sous-martingale $(X_n)_{n\geq 0}$ bornée dans L^1 *(i.e.* $\sup_n \mathbf{E}[|X_n|] < \infty$*) converge p.s. vers une v.a.r. intégrable X_∞ quand $n \to \infty$.*

Remarque 154 *(a) Si $(X_n)_{n\geq 0}$ est une suite $(\mathscr{F}_n)_{n\geq 0}$-adaptée de v.a.r. positives telle que $\forall\ n \geq 0$, $\mathbf{E}[X_{n+1}|\mathscr{F}_n] \leq X_n$, alors X_n converge p.s. (En effet, la preuve du point (i) n'utilise pas l'intégrabilité des X_n).*

(b) Dans (ii), on peut remplacer sous-martingale par sur-martingale (appliquer le résultat à $-X$).

En fait, tous les résultats de convergence de martingales ne sont que des raffinements du lemme fondammental suivant.

Lemme 155 *Soit $(M_n)_{n\geq 0}$ une $(\mathscr{F}_n)_{n\geq 0}$-martingale bornée dans* L^2 *(i.e.* $\sup_n \mathbf{E}[M_n^2] < \infty$*). Alors M_n converge p.s. et dans L^2.*

Preuve : On définit $\Delta M_k = M_k - M_{k-1}$ pour $k \geq 1$. On remarque que pour $n > k \geq 1$,

$$\mathbf{E}[\Delta M_k \Delta M_n] = \mathbf{E}[\mathbf{E}[\Delta M_k \Delta M_n|\mathscr{F}_{n-1}]] = \mathbf{E}[\Delta M_k \mathbf{E}[\Delta M_n|\mathscr{F}_{n-1}]] = 0,$$

car ΔM_k est $\mathscr{F}_k \subset \mathscr{F}_{n-1}$-mesurable et car $\mathbf{E}[\Delta M_n|\mathscr{F}_{n-1}] = 0$. 0n en déduit que

$$\forall\ n+m > n \geq 0, \quad \mathbf{E}[(M_{n+m} - M_n)^2] = \mathbf{E}\Big[\Big(\sum_{n+1}^{n+m} \Delta_k\Big)^2\Big] = \sum_{n+1}^{n+m} \mathbf{E}[(\Delta M_k)^2]. \qquad (*)$$

En particulier $\sup_n \mathbf{E}[M_n^2] < \infty$ implique $\sup_n \mathbf{E}[(M_n - M_0)^2] < \infty$ et donc $\sum_{k\geq 1} \mathbf{E}[(\Delta M_k)^2] < \infty$.

On a $\sup_{i,j\geq n} \mathbf{E}[|M_i - M_j|^2] = \sup_{i,j\geq n} \sum_{i+1}^{j} \mathbf{E}[(\Delta M_k)^2] \leq \sum_{n+1}^{\infty} \mathbf{E}[(\Delta M_k)^2] \to 0$ quand $n \to \infty$: la suite $(M_n)_{n\geq 0}$ est de Cauchy dans L^2 et donc converge dans L^2.

On introduit maintenant $A_n = \sup_{i,j\geq n} |M_i - M_j|$. On va montrer que $\lim_n A_n = 0$ p.s., ce qui impliquera que p.s., la suite $(M_n)_{n\geq 0}$ est une suite de Cauchy (et donc convergente) dans $\mathbb{R}$, ce qui achèvera la preuve. Comme A_n est p.s. décroissante, sa limite existe p.s., et il suffit de montrer que $\lim_n \mathbf{E}[A_n^2] = 0$. On introduit $B_n = \sup_{i\geq n} |M_i - M_n|$ et on remarque que $A_n \leq 2B_n$ (car $|M_i - M_j| \leq |M_i - M_n| + |M_j - M_n|$). Il suffit donc de montrer que $\lim_n \mathbf{E}[B_n^2] = 0$.

En utilisant l'inégalité de Doob avec $p = 2$ (pour la martingale $(M_{n+i} - M_n)_{i\geq 0}$), on voit que

$$\mathbf{E}[B_n^2] \leq 4 \sup_{i\geq n} \mathbf{E}[(M_i - M_n)^2] = 4 \sum_{k\geq n+1} \mathbf{E}[(\Delta M_k)^2].$$

C'est le reste d'une série convergente. Ainsi, on a bien $\lim_n \mathbf{E}[B_n^2] = 0$. ■

Preuve du point (i) du théorème 153 : Soit donc X_n une sur-martingale positive. On pose $Y_n = e^{-X_n}$. Comme $\varphi(x) = e^{-x}$ est convexe décroissante, on voit par Jensen que

$$\mathbf{E}[Y_{n+1}|\mathscr{F}_n] = \mathbf{E}[\varphi(X_{n+1})|\mathscr{F}_n] \geq \varphi(\mathbf{E}[X_{n+1}|\mathscr{F}_n]) \geq \varphi(X_n) = Y_n.$$

Donc Y_n est une sous-martingale. Écrivons sa décomposition de Doob $Y_n = Y_0 + M_n + V_n$, où M est une martingale et où V est prévisible.

Déjà, V est p.s. croissant : $V_{n+1}-V_n = \mathbf{E}[V_{n+1}-V_n|\mathscr{F}_n] = \mathbf{E}[Y_{n+1}-Y_n|\mathscr{F}_n]-\mathbf{E}[M_{n+1}-M_n|\mathscr{F}_n] = \mathbf{E}[Y_{n+1}-Y_n|\mathscr{F}_n] \geq 0$ p.s. Donc $V_\infty = \lim_n V_n \in [0,\infty]$ existe p.s.

Montrons maintenant que $\sup_n \mathbf{E}[M_n^2] < \infty$. Comme dans la preuve précédente (voir $(*)$ et utiliser que $M_0 = 0$), on a $\mathbf{E}[M_n^2] = \sum_1^n \mathbf{E}[(\Delta M_k)^2]$ où, pour $(X_n)_{n\geq 0}$ un processus, on note $\Delta X_k = X_k - X_{k-1}$ pour $k \geq 1$. On écrit $Y_n^2 = Y_0^2 + \sum_{k=1}^n (Y_k^2 - Y_{k-1}^2)$ puis, comme $Y_k = Y_{k-1} + \Delta M_k + \Delta V_k$,

$$Y_n^2 = Y_0^2 + \sum_{k=1}^n (\Delta M_k)^2 + \sum_{k=1}^n (\Delta V_k)^2 + 2\sum_{k=1}^n Y_{k-1}\Delta M_k + 2\sum_{k=1}^n Y_{k-1}\Delta V_k + 2\sum_{k=1}^n \Delta M_k \Delta V_k.$$

On observe ensuite que $Y_0^2 + \sum_{k=1}^n (\Delta V_k)^2 \geq 0$ et $2\sum_{k=1}^n Y_{k-1}\Delta V_k \geq 0$ (car $Y \geq 0$ et V est croissant) :

$$\sum_{k=1}^n (\Delta M_k)^2 + 2\sum_{k=1}^n (Y_{k-1} + \Delta V_k)\Delta M_k \leq Y_n^2 \leq 1.$$

De plus, on a $\mathbf{E}[(Y_{k-1} + \Delta V_k)\Delta M_k] = \mathbf{E}[Y_{k-1}\mathbf{E}[\Delta M_k|\mathscr{F}_{k-1}]] = 0$ car $Y_{k-1} + \Delta V_k$ est $\mathscr{F}_{k-1}$-mesurable et car M est une martingale. Enfin, on conclut que $\sum_{k=1}^n \mathbf{E}[(\Delta M_k)^2] \leq 1$ et donc que $\mathbf{E}[M_n^2] = \sum_1^n \mathbf{E}[(\Delta M_k)^2] \leq 1$.

On déduit donc du lemme 155 que $M_\infty = \lim_n M_n$ existe p.s. Ainsi, $Y_\infty = \lim_n Y_n = Y_0 + M_\infty + V_\infty$ p.s. De plus, $Y_\infty \in [0,1]$ p.s., puisque Y_n est à valeurs dans $[0,1]$. On en déduit que $\lim_n X_n = X_\infty$ p.s., où $X_\infty = -\ln Y_\infty \in [0,\infty]$ (avec $-\ln 0 = \infty$), et il ne reste plus qu'à voir que $X_\infty < \infty$ p.s. Mais on voit facilement par récurrence que $\mathbf{E}[X_n|\mathscr{F}_0] \leq X_0$ puis, par Fatou, que $\mathbf{E}[X_\infty|\mathscr{F}_0] \leq X_0 < \infty$ p.s., ce qui implique le résultat. ■

Preuve du point (ii) du théorème 153 : Soit donc X_n une sous-martingale bornée dans L^1. Posons $K = \sup_n \mathbf{E}[|X_n|]$. Montrons qu'il existe deux sur-martingales positives Y_n, Z_n telles que $X_n = Y_n - Z_n$. Cela impliquera, par le point (i), que $X_\infty = \lim X_n$ existe p.s. (et on aura $\mathbf{E}[|X_\infty|] \leq K$ par Fatou).

Le processus $(X_n)_+$ est une sous-martingale, car $x \mapsto x_+$ est convexe croissante. Écrivons sa décomposition de Doob $(X_n)_+ = (X_0)_+ + M_n + V_n$. On vérifie comme sans la preuve du (ii) que V est croissant p.s. (et donc positif puisque $V_0 = 0$). De plus $\mathbf{E}[V_n] = E[(X_n)_+] - \mathbf{E}[(X_0)_+] \leq E[(X_n)_+] \leq K$. Donc $V_\infty = \lim_n V_n$ existe p.s. et de plus, $\mathbf{E}[V_\infty] \leq K$ par convergence monotone.

Posons $Y_n = (X_0)_+ + M_n + \mathbf{E}[V_\infty|\mathscr{F}_n]$. C'est bien sûr une martingale, et elle est positive car $Y_n \geq (X_0)_+ + M_n + V_n = (X_n)_+ \geq 0$ (comme V est p.s. croissant, $\mathbf{E}[V_\infty|\mathscr{F}_n] \geq \mathbf{E}[V_n|\mathscr{F}_n] = V_n$).

Enfin, $Z_n = Y_n - X_n$ est une sur-martingale (car c'est la différence d'une martingale et d'une sous-martingale) et on a $Z_n \geq (X_n)_+ - X_n = (X_n)_- \geq 0$. ■

Convergence dans L^1.

Théorème 156 *Soit $(X_n)_{n\geq 0}$ une $(\mathscr{F}_n)_{n\geq 0}$-martingale. Les points suivants sont équivalents.*

(i) X_n converge dans L^1.

(ii) $\sup_n \mathbf{E}[|X_n|] < \infty$ et $X_n = \mathbf{E}[X_\infty|\mathscr{F}_n]$ pour tout $n \geq 0$ (où $X_\infty = \lim_n X_n$ p.s. existe par le théorème 153).

(iii) Il existe X intégrable telle que $X_n = \mathbf{E}[X|\mathscr{F}_n]$ pour tout $n \geq 0$.

(iv) La famille $\{X_n, n \geq 0\}$ est U.I.

Preuve : Montrons que (i) implique (ii) : si X_n converge dans L^1 vers X_∞, alors $\sup_n \mathbf{E}[|X_n|] < \infty$, donc le théorème 153 nous dit que X_n converge p.s., nécessairement vers X_∞. Pour montrer que $X_n = \mathbf{E}[X_\infty|\mathscr{F}_n]$, on écrit $X_n = \mathbf{E}[X_{n+m}|\mathscr{F}_n]$ et on fait tendre $m \to \infty$ de la façon suivante :

$$\mathbf{E}[|X_n - \mathbf{E}[X_\infty|\mathscr{F}_n]|] = \mathbf{E}[|\mathbf{E}[X_{n+m} - X_\infty|\mathscr{F}_n]|] \leq \mathbf{E}[|X_{n+m} - X_\infty|] \to 0.$$

(ii) implique trivialement (iii) car $X_\infty = \lim_n X_n$ p.s. est intégrable par Fatou car $\sup_n \mathbf{E}[|X_n|] < \infty$.

(iii) implique (iv) par le lemme.

Enfin, (iv) implique (i), car $\{X_n, n \geq 0\}$ U.I. implique que $\sup_n \mathbf{E}[|X_n|] < \infty$, donc le théorème 153 nous dit que X_n converge p.s., et donc aussi dans L^1 puisque $\{X_n, n \geq 0\}$ U.I. (par le théorème 95 de convergence dominée optimal). ■

Convergence dans L^p.

Théorème 157 *Soit $(X_n)_{n\geq 0}$ une $(\mathscr{F}_n)_{n\geq 0}$-martingale. Si $\sup_n \mathbf{E}[|X_n|^p] < \infty$ pour un $p \in]1, \infty[$, alors il existe $X_\infty \in \mathrm{L}^p$ t.q. $X_\infty = \lim_n X_n$ p.s. et dans L^p. De plus, $X_n = \mathbf{E}[X_\infty|\mathscr{F}_n]$ pour tout $n \geq 0$.*

Preuve : Comme $\sup_n \mathbf{E}[|X_n|^p] < \infty$ avec $p > 1$, la famille $\{X_n, n \geq 0\}$ est U.I. et le théorème 156 implique que $X_\infty = \lim_n X_n$ p.s. et dans L^1 avec de plus $X_n = \mathbf{E}[X_\infty|\mathscr{F}_n]$ pour tout n. Il ne nous reste qu'à montrer que $X_n \to X_\infty$ dans L^p. On utilise la convergence dominée : bien sûr, $|X_n - X_\infty|^p \to 0$ p.s. D'autre part, $\sup_{n\geq 0} |X_n - X_\infty|^p \leq 2^p(\sup_{n\geq 0} |X_n|^p + |X_\infty|^p) \leq 2^{p+1} \sup_{n\geq 0} |X_n|^p$, qui est intégrable par l'inégalité de Doob (et car on a supposé que $\sup_n \mathbf{E}[|X_n|^p] < \infty$). ■

IV.3 Temps d'arrêt

Définition 158 *Soit $(\mathscr{F}_n)_{n\geq 0}$ une filtration. Une v.a. T à valeurs dans $\mathbb{N} \cup \{\infty\}$ est un $(\mathscr{F}_n)_{n\geq 0}$-temps d'arrêt si $\forall\, n \in \mathbb{N}$, $\{T = n\} \in \mathscr{F}_n$. On pose alors $\mathscr{F}_T = \{A \in \mathscr{F}_\infty \ :\ A \cap \{T = n\} \in \mathscr{F}_n,\ \forall\, n \in \mathbb{N}\}$.*

Voici quelques propriétés des temps d'arrêt dont les preuves sont élémentaires.

Proposition 159 *Soit $(\mathscr{F}_n)_{n\geq 0}$ une filtration.*

(i) Une v.a. T à valeurs dans $\mathbb{N} \cup \{\infty\}$ est un $(\mathscr{F}_n)_{n\geq 0}$-temps d'arrêt ssi $\{T \leq n\} \in \mathscr{F}_n$, pour tout $n \in \mathbb{N}$ ssi $\{T > n\} \in \mathscr{F}_n$, pour tout $n \in \mathbb{N}$.

(ii) Si T est un $(\mathscr{F}_n)_{n\geq 0}$-temps d'arrêt, alors $\mathscr{F}_T$ est une sous-tribu de $\mathscr{F}_\infty$.

(iii) Si S, T sont deux $(\mathscr{F}_n)_{n\geq 0}$-temps d'arrêt, alors $S + T$, $S \wedge T$, $S \vee T$ sont des $(\mathscr{F}_n)_{n\geq 0}$-temps d'arrêt. De plus, $\mathscr{F}_{S\wedge T} = \mathscr{F}_S \cap \mathscr{F}_T$.

(iv) Si S, T sont deux $(\mathscr{F}_n)_{n\geq 0}$-temps d'arrêt avec $S \leq T$ p.s., alors $\mathscr{F}_S \subset \mathscr{F}_T$.

Remarque 160 *Toute v.a. déterministe $T = n_0 \in \mathbb{N}$ est un temps d'arrêt, et on a $\mathscr{F}_T = \mathscr{F}_{n_0}$. Si $(X_n)_{n\geq 0}$ est une suite $(\mathscr{F}_n)_{n\geq 0}$-adaptée de v.a.r., alors pour tout $B \in \mathscr{B}(\mathbb{R})$, $T_B = \inf\{n \geq 0 \ :\ X_n \in B\}$ est un $(\mathscr{F}_n)_{n\geq 0}$-temps d'arrêt (avec la convention $\inf \emptyset = \infty$).*

Proposition 161 *Soit $(X_n)_{n\geq 0}$ une $(\mathscr{F}_n)_{n\geq 0}$-martingale (resp. sur-martingale, resp. sous-martingale) et T un $(\mathscr{F}_n)_{n\geq 0}$-temps d'arrêt. Alors $(X_{n\wedge T})_{n\geq 0}$ est une $(\mathscr{F}_n)_{n\geq 0}$-martingale (resp. sur-martingale, sous-martingale).*

Preuve : Déjà, pour chaque $n \geq 0$, $|X_{n\wedge T}| \leq \sum_0^n |X_k|$ est intégrable. Ensuite, pour chaque $n \geq 0$, $X_{n\wedge T} = \sum_{k=0}^n X_k \mathbf{1}_{\{T=k\}} + X_n \mathbf{1}_{\{T>n\}}$ est $\mathscr{F}_n$-mesurable. Enfin, pour $n \geq 0$, on a $X_{(n+1)\wedge T} = X_{n+1}\mathbf{1}_{\{T>n\}} + \sum_{k=0}^n X_k \mathbf{1}_{\{T=k\}}$. Du coup, comme $\{T > n\} \in \mathscr{F}_n$, et comme $\sum_{k=0}^n X_k \mathbf{1}_{\{T=k\}}$ est $\mathscr{F}_n$-mesurable,

$$\mathbf{E}[X_{(n+1)\wedge T}|\mathscr{F}_n] = \mathbf{E}[X_{n+1}|\mathscr{F}_n]\mathbf{1}_{\{T>n\}} + \sum_{k=0}^n X_k \mathbf{1}_{\{T=k\}}$$

Donc si par exemple on a affaire à un martingale, on trouve

$$\mathbf{E}[X_{(n+1)\wedge T}|\mathscr{F}_n] = X_n \mathbf{1}_{\{T>n\}} + \sum_{k=0}^n X_k \mathbf{1}_{\{T=k\}} = X_{n\wedge T}.$$

IV.4 Problèmes d'arrêt.

On se donne $(X_n)_{n\geq 0}$ une $(\mathscr{F}_n)_{n\geq 0}$-martingale et T un $(\mathscr{F}_n)_{n\geq 0}$-temps d'arrêt. A-t-on $\mathbf{E}[X_T] = \mathbf{E}[X_0]$? Si tout va bien, oui, car $X_{T\wedge n}$ est une martingale, donc $\mathbf{E}[X_{T\wedge n}] = \mathbf{E}[X_0]$, et il suffit de faire tendre $n \to \infty$. Mais ce n'est pas automatique.

Si par exemple $X_n = \xi_1 + \cdots + \xi_n$ est une marche aléatoire simple symétrique (les ξ_i sont i.i.d. et $\mathbf{P}(\xi_1 = 1) = \mathbf{P}(\xi_1 = -1) = 1/2$), $(X_n)_{n\geq 0}$ une $(\mathscr{F}_n)_{n\geq 0}$-martingale avec $\mathscr{F}_n = \sigma(\xi_1, \ldots, \xi_n)$, et $T_1 = \inf\{n \geq 0 \ : \ X_n = 1\}$ est un $(\mathscr{F}_n)_{n\geq 0}$-temps d'arrêt fini p.s. Donc $X_{T_1} = 1$ et du coup, $\mathbf{E}[X_{T_1}] = 1 \neq 0 = \mathbf{E}[X_0]$.

Proposition 162 *Soit $(X_n)_{n\geq 0}$ une $(\mathscr{F}_n)_{n\geq 0}$-martingale et soit T, un $(\mathscr{F}_n)_{n\geq 0}$-temps d'arrêt. On suppose que $T < \infty$ p.s. et qu'il existe une v.a.r. Z intégrable t.q. $\forall n \in \mathbb{N}$, $|X_{n\wedge T}| \leq Z$. Alors X_T est une variable intégrable et $\mathbf{E}[X_0] = \mathbf{E}[X_T]$.*

Preuve : Comme $X_{T\wedge n}$ est une martingale, $\mathbf{E}[X_{T\wedge n}] = \mathbf{E}[X_0]$. De plus, comme $T < \infty$ p.s., on a $\lim_n X_{T\wedge n} = X_T$. Il n'y a plus qu'à utiliser le théorème de convergence dominée. ■

Application à la ruine du joueur. Soit $(\xi_n)_{n\geq 1}$ i.i.d. avec $\mathbf{P}(\xi_n = 1) = p$ et $\mathbf{P}(\xi_n = -1) = q$, où $p \in]0,1[$ et $q = 1 - p$. On pose $X_0 = 0$, $\mathscr{F}_0 = \{\emptyset, \Omega\}$, et pour tout $n \geq 1$, $X_n = \xi_1 + \ldots + \xi_n$ et $\mathscr{F}_n = \sigma(\xi_n, \ldots, \xi_n)$. Pour tout $c \in \mathbb{Z}$, on pose $T_c = \inf\{n \in \mathbb{N} : S_n = c\}$, avec la convention $\inf \emptyset = \infty$. C'est un $(\mathscr{F}_n)_{n\geq 0}$ temps d'arrêt. On fixe $a, b \in \mathbb{N}^*$ et on pose $T_{-a,b} = T_{-a} \wedge T_b$, on admet qu'il est p.s. fini. On a alors

$$\mathbf{P}(S_{T_{-a,b}} = -a) = \begin{cases} \frac{1-(q/p)^b}{(q/p)^{-a}-(q/p)^b} & \text{si } p \neq q, \\ \\ \frac{b}{a+b} & \text{si } p = q = 1/2. \end{cases}$$

Cas $p = q = 1/2$: $(S_n)_{n\geq 0}$ est une $(\mathscr{F}_n)_{n\geq 0}$-martingale et $|S_{n\wedge T_{-a,b}}| \leq \max(a,b)$. La proposition 162 s'applique, donc $\mathbf{E}[S_{T_{-a,b}}] = \mathbf{E}[S_0] = 0$. Mais $\mathbf{E}[S_{T_{-a,b}}] = -a\mathbf{P}(S_{T_{-a,b}} = -a) + b\mathbf{P}(S_{T_{-a,b}} = b)$. Comme $\mathbf{P}(S_{T_{-a,b}} = b) = 1 - \mathbf{P}(S_{T_{-a,b}} = -a)$, on conclut que $0 = -(a+b)\mathbf{P}(S_{T_{-a,b}} = -a) + b$. ■

Cas $p \neq q$: soit $x = q/p$. En utilisant que $px + qx^{-1} = 1$, il n'est pas trop dur de montrer que $\mathbf{E}[x^{\xi_1}] = 1$ puis que $X_n = x^{S_n}$ est une $\mathscr{F}_n$-martingale. Clairement $X_{n\wedge T_{-a,b}}$ est bornée (par $\max_{-a\leq k\leq b} x^k$). La proposition 162 s'applique, donc $\mathbf{E}[X_{T_{-a,b}}] = \mathbf{E}[X_0] = 1$. Mais $\mathbf{E}[X_{T_{-a,b}}] = x^{-a}\mathbf{P}(S_{T_{-a,b}} = -a) + x^b\mathbf{P}(S_{T_{-a,b}} = b)$. Donc $1 = (x^{-a} - x^b)\mathbf{P}(S_{T_{-a,b}} = -a) + x^b$. ■

Théorème 163 **(Théorème d'arrêt de Doob)** *Soit* $(X_n)_{n\geq 0}$ *une* $(\mathscr{F}_n)_{n\geq 0}$*-martingale U.I. On note* X_∞ *sa limite presque sûre, qui est aussi sa limite dans* L^1*. Soient aussi* S, T*, deux* $(\mathscr{F}_n)_{n\geq 0}$*-temps d'arrêt (à valeurs dans* $\mathbb{N}\cup\{\infty\}$*). On définit* $X_T = \sum_{n\geq 0} X_n \mathbf{1}_{\{T=n\}} + X_\infty \mathbf{1}_{\{T=\infty\}}$ *et* X_S *similairement.*

(i) X_T *est* $\mathscr{F}_T$*-mesurable, intégrable,* $X_T = \mathbf{E}[X_\infty | \mathscr{F}_T]$ *(et par conséquent* $\mathbf{E}[X_T] = \mathbf{E}[X_0]$ *puisque* $X_0 = \mathbf{E}[X_\infty | \mathscr{F}_0]$*).*

(ii) $(X_{n\wedge T})_{n\geq 0}$ *est une* $(\mathscr{F}_n)_{n\geq 0}$*-martingale U.I. qui converge p.s. et dans* L^1 *vers* X_T*.*

(iii) $\mathbf{E}[X_T|\mathscr{F}_S] = X_{S\wedge T}$*.*

Preuve : Montrons (i). Pour montrer que X_T est $\mathscr{F}_T$-mesurable, considérons $A \in \mathscr{B}(\mathbb{R})$ et montrons que $\{X_T \in A\} \in \mathscr{F}_T$, i.e. que pour tout $n \in \mathbb{N}$, $\{T = n\} \cap \{X_T \in A\} \in \mathscr{F}_n$, ce qui est clair puisque $\{T = n\} \cap \{X_T \in A\} = \{T = n\} \cap \{X_n \in A\}$. Rappelons ensuite que X_∞ est intégrable et montrons que $\mathbf{E}[X_\infty|\mathscr{F}_T] = X_T$, ce qui impliquera que X_T est aussi intégrable. On doit montrer que X_T est $\mathscr{F}_T$-mesurable (déjà fait) et que pour tout $B \in \mathscr{F}_T$, on a $\mathbf{E}[X_T \mathbf{1}_B] = \mathbf{E}[X_\infty \mathbf{1}_B]$. Pour cela, on écrit

$$\mathbf{E}[X_T \mathbf{1}_B] = \sum_{n\geq 0} \mathbf{E}[X_n \mathbf{1}_B \mathbf{1}_{\{T=n\}}] + \mathbf{E}[X_\infty \mathbf{1}_B \mathbf{1}_{\{T=\infty\}}]$$

par définition de X_T. Comme $B \cap \{T = n\} \in \mathscr{F}_n$ et comme $X_n = \mathbf{E}[X_\infty|\mathscr{F}_n]$, on a $\mathbf{E}[X_n \mathbf{1}_B \mathbf{1}_{\{T=n\}}] = \mathbf{E}[X_\infty \mathbf{1}_B \mathbf{1}_{\{T=n\}}]$. Ainsi,

$$\mathbf{E}[X_T \mathbf{1}_B] = \sum_{n\geq 0} \mathbf{E}[X_\infty \mathbf{1}_B \mathbf{1}_{\{T=n\}}] + \mathbf{E}[X_\infty \mathbf{1}_B \mathbf{1}_{\{T=\infty\}}] = \mathbf{E}[X_\infty \mathbf{1}_B].$$

Montrons (ii). On sait déjà que $(X_{n\wedge T})_{n\geq 0}$ est une $(\mathscr{F}_n)_{n\geq 0}$-martingale. Comme $n \wedge T$ est un $(\mathscr{F}_n)_{n\geq 0}$-temps d'arrêt, le (i) implique que $X_{T\wedge n} = \mathbf{E}[X_\infty|\mathscr{F}_{T\wedge n}]$. Donc $(X_{T\wedge n})_{n\geq 0}$ est U.I. par le lemme 144, et la martingale $X_{T\wedge n}$ converge p.s. et dans L^1 quand $n \to \infty$. Sa limite p.s. (et donc aussi L^1) est bien sûr X_T (sur l'évènement $\{T < \infty\}$ c'est évident et sur l'évènement $\{T = \infty\}$ aussi).

Pour (iii), il suffit d'appliquer (i) avec la martingale U.I. $(X_{n\wedge T})_{n\geq 0}$ (dont la limite est X_T) et le temps d'arrêt S, ce qui donne $X_{S\wedge T} = \mathbf{E}[X_T| \mathscr{F}_S]$. ■

Il existe d'autres théorèmes d'arrêt. Une CNS pour avoir $\mathbf{E}[X_T] = \mathbf{E}[X_0]$ est que la martingale $(X_{n\wedge T})_{n\geq 0}$ soit U.I. Cette condition n'est pas très explicite. En gros, il faut que la martingale soit *sympathique*, comme dans le théorème d'arrêt de Doob, **ou** que le temps d'arrêt soit suffisamment intégrable, **ou** un peu des deux. Par exemple, on a

Remarque 164 *Pour toute* $(\mathscr{F}_n)_{n\geq 0}$*-martingale* $(X_n)_{n\geq 0}$ *et tout* $(\mathscr{F}_n)_{n\geq 0}$*-temps d'arrêt* borné, *on a* $\mathbf{E}[X_T] = \mathbf{E}[X_0]$.

Preuve : Si T est borné par n_0, il suffit d'appliquer la proposition 162 avec $Z = \sum_0^{n_0} |X_k| \in \mathrm{L}^1$. ■